Analysis of Intraseasonal and Interannual variability of the Asian summer monsoon using a Hidden Markov Model

Jin Ho Yoo[1], Andrew W. Robertson[2], and In-Sik Kang[3]

[1]The Abdus Salam International Centre for Theoretical Physics, Italy

[2]International Research Institute for Climate and Society, USA

[3]Seoul National University, Korea

May, 2010

Revised manuscript

Jin Ho Yoo

The Abdus Salam International Centre for Theoretical Physics,

Strada costiera 11, 34014 Trieste, Italy

Email: jyoo@ictp.it

Abstract

Intraseasonal and interannual variability of Asian summer monsoon rainfall in pentad precipitation data is examined using a Hidden Markov Model (HMM). The spatial patterns of discrete rainfall states derived with the HMM and the associated transition probabilities between the states are shown to represent well the principal Asian summer monsoon intraseasonal oscillation (ISO), propagating eastward and northward with a period of 40–50 days. Stochastic simulations made with the HMM reasonably reproduce the canonical ISO propagation and its observed statistics such as the frequency of ISO events.

The interannual modulation of the ISO associated with El Niño-Southern Oscillation (ENSO) is assessed by employing a nonhomogeneous HMM (NHMM) with summer-mean NINO3.4 index prescribed as an input variable. ENSO influence on the ISO is found to manifest as preferences toward particular ISO phases depending on the ENSO condition, thus adding an asymmetry to the ISO. In the presence of seasonal mean anomalies, it is shown that the El Niño seasonal mean rainfall anomaly pattern is identified by the HMM as a distinct state, in addition to the ISO states, whereas the La Niña seasonal mean rainfall anomaly pattern does not appear distinct from the ISO states.

1. Introduction

The Asian summer monsoon (ASM) exhibits pronounced variability on both intraseasonal and interannual time scales. On the intraseasoanl scale, it is characterized by active (wet) and break (dry) phases within a season (e.g. Goswami et al. 2006; Hoyos and Webster 2007), which have large agricultural impacts. The intraseasonal variation of rainfall is often associated with an irregularly propagating large-scale intraseasonal oscillation (ISO) with a period of 30–60 days (e.g. Yasunari 1979; Lau and Chan 1986) which also modulate the onset and withdrawal of monsoon season over India (Gadgil 2003). The eastward propagation of convective anomalies along the equator (e.g. Madden Jullian Oscillation; MJO) and northward propagation over the Indian Ocean and western Pacific sectors typify the structure of the summer monsoon ISO, with northeastward propagating rainfall anomalies spread out in the northwest-southeast direction. However, eastward and northward propagations are not always phase-locked to each other (e.g. Wang and Rui 1990; Klingaman et al. 2008), so that the structure and propagation of the summer monsoon ISO are more irregular than those of the winter time ISO, which is dominated by eastward propagation. Due to its complexity and crucial influence on the monsoon weather statistics, the prediction and simulation of summer monsoon ISO is one of the most challenging problems in monsoon studies.

On the interannual time scale, the ASM is known to be strongly related with El Niño-Southern Oscillation (ENSO) (e.g. Rasmusson and Carpenter 1983; Shukla and Paolina 1983; Webster and Yang 1992; Webster et al. 1998), although this relationship appears

to have weakened in recent decades (Krishna-Kumar et al., 1999; Torrence and Webster, 1999; Kucharski et al. 2007). The ENSO teleconnection is broadly characterized by suppressed convection over the Indian subcontinent and Maritime Continent during the summer of the developing phase of El Niño (e.g. Webster et al. 1998). This seasonal mean circulation anomaly associated with ENSO also results in a change of synoptic variability such as typhoon activity over the western North Pacific (Kang et al. 1995; Wang and Chan 2002) and monsoon depressions over the Bay of Bengal (Singh et al. 2002). Thus the influence of ENSO can be found at various scales of ASM variability.

While the ISO and interannual ENSO-monsoon relationships have conventionally been considered separately, recent studies have pointed to parallels in spatial structure, suggesting that interannual ASM variability may be a result of rectification of intraseasonal "active" and "break" phase events (e.g. Goswami and Mohan 2001) whose probability distribution may be affected by external forcing (Palmer 1994). On the other hand, Krishnamurthy and Shukla (2007) showed clear separation between intraseasonal and interannual modes of variability. Most of studies on the interannual variability of ISO found that there is no significant correlation between ISO activity and ENSO both in summer and winter (Salby and Hendon 1994; Hendon et al. 1999; Slingo et al. 1999; Lawrence and Webster 2001).

There are very few studies found a relationship between ENSO and ISO. Tam and Lau (2005) found that the phase speed of winter time ISO (MJO) is modulated by the ENSO. Yun et al. (2009) discussed the influence of East Asian summer ISO on the ENSO in a Quasi Biannial time scale. For the south Asian summer monsoon ISO, Teng

and Wang (2003) found that ISO activity is enhanced over western north Pacific during El Niño-developing summers through the enhancement of easterly vertical shear of mean flow; Kim et al. (2008) found similar relationships in model simulations. To date, most studies have focused on the relationship between ENSO and ISO activity (strength) but ENSO influence on its other characteristics such as phase transition properties or changes in the spatial structure have yet to be explored.

The aim of this paper is to examine ASM intraseasonal and interannual variability as expressed in historical pentad gridded rainfall analysis data with an emphasis on the influence of ENSO. In many studies, quasi-regular propagating ISO signals have been extracted by statistical methods based on lead-lag covariance analysis such as Extended Empirical Orthogonal Function (Lau and Chan 1986) and Multichannel Singular Spectrum Analysis (Krishnamurthy and Shukla 2007). In addition, pre-filtered data has often been used, tending to yield large-scale structure and smooth propagation characteristics. In reality, however, there are large variations in the ISO propagation, which often exhibits sudden jumps of convective activity and/or unexpected long stay in particular phase of ISO cycle (Singh et al. 1992; Wu et al. 1999; Chattopadhyay et al. 2008). This irregularity motivates the use of a probabilistic framework to model ASM rainfall variability, but there are very few examples to date of such approaches to the ISO.

In contrast to most previous ASM analyses, we build a probabilistic model that explicitly accounts for the stochastic aspect of atmospheric variability, and can be used to generate a large number of stochastic rainfall simulations for hypothesis testing. The

modeling framework is provided by a Hidden Markov Model (HMM), in which a latent state variable is introduced to enable a simplified factorization of the probability distribution function (PDF) of observed rainfall. The HMMs have been used previously to model daily rainfall at a set of gauge stations (Robertson et al. 2004, 2006), and to build downscaling models in which large-scale predictor variables are introduced into the HMM to from a so-called non-homogeneous HMM (NHMM) (Hughes and Guttorp 1994, Charles et al. 1999, Robertson et al. 2009). In the NHMM, the exogenous variables influence daily rainfall via their impact on the transition probabilities that control the temporal transitions between states; the transition matrix is thus no longer homogeneous. Recently, an HMM has also been applied to the identification of flow regimes in atmospheric data (Franzke et al. 2008).

In the present study, we analyze intraseasonal and interannual aspects of pentad rainfall data, in which the gridded data is firstly projected onto a basis comprising its leading empirical orthogonal functions (EOFs). The principal component (PC) time series are then modeled, via the HMM, in terms of Gaussian distributions; the resulting HMM is then equivalent to a Gaussian mixture model (e.g. Smyth et al. 1999) with an explicit Markovian time dependence. This approach allows us firstly to build and test a model for intraseasonal time scale rainfall variability alone, by subtracting out the mean seasonal cycle and interannual variability. Secondly, we examine the impact of ENSO on intraseasonal time scales, before finally adding back the interannual variations in rainfall.

Section 2 of the paper describes the data and methods. Results of the intraseasonal

HMM are described in Section 3. In Section 4, the nonhomogeneous HMM is used to examine the relationships between ENSO and ASM rainfall variability on intraseasonal and interannual timescales. The paper concludes with a summary and discussion in Section 5.

2. Data and Method

a. Data

The 5-day mean (pentad) Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997) data, given on a 2.5° latitude-longitude grid, is used in this study. We analyze the variability of precipitation in the Asian monsoon region ($60^{o}E$–$150^{o}E$, $10^{o}S$–$35^{o}N$) from the 25 boreal summers 1979–2003 (32 pentads for each summer (May–September, MJJAS). At each grid point and each pentad, the climatological seasonal cycle was removed by subtracting the 25-year mean value for that pentad. In order to obtain the purely intraseasonal components in Sect. 3, the seasonal (MJJAS) average value of each year is also subtracted, thus removing interannual variability. To reduce the dimension of the 37x19 grid point fields for each pentad, an EOF analysis of the pentad precipitation anomalies is performed, and the resulting 4 leading PC time series are used as the rainfall observations in the HMM. In the NHMM analysis, the June–August (JJA) mean sea surface temperature anomaly (SSTA) averaged over the NINO3.4 region [$90^{o}W$-$150^{o}W$, $5^{o}S$-$5^{o}N$] from the Extended Reconstruction of Sea Surface Temperature (ERSST; Smith and Reynolds 2004) is used as the exogenous data to represent ENSO. The JJA mean NINO3.4 index is chosen to represent the ENSO state during the core summer season. Daily outgoing longwave

radiation (OLR) dataset (Liebmann and Smith 1996) for same spatial domain and same time period with CMAP precipitation data is also used to verify results of precipitation data.

b. Hidden Markov Model (HMM)

Let $\mathbf{R}_t = (R_t^1, \ldots, R_t^M)$ be a vector of the leading M CMAP rainfall PC loadings on day t, and let $\mathbf{R}_{1:T}$ denote a time sequence of such vectors $\mathbf{R}_1, \ldots, \mathbf{R}_T$. The sequence of rainfall PCs $\mathbf{R}_{1:T}$ is assumed to be generated by a Markov chain of hidden (unobserved) weather states $\mathbf{S}_t = (S_1, \ldots, S_T)$, where the latent variable S_t takes values from 1 to K. Making a first-order Markov assumption, we factor the joint distribution over the hidden state sequence as:

$$p(S_1, \ldots, S_T) = p(S_1) \prod_{t=2}^{T} p(S_t \mid S_{t-1}) \tag{1}$$

where $p(S_t \mid S_{t-1})$ is modeled as a $K \times K$ stochastic matrix of state transition probabilities $\mathbf{\Gamma} = \{\gamma_{ij}\}, 1 \leq i, j \leq K$, and $p(S_1)$ is a set of initial state probabilities.

The second major assumption in the HMM is that the instantaneous rainfall PC $\mathbf{R}_t$ for a particular day (pentad) t is assumed to depend only on the state on day t, so that all rainfall autocorrelation is treated at the state level. Thus,

$$p(\mathbf{R}_{1:T} \mid \mathbf{S}_{1:T}) = \prod_{t=1}^{T} p(\mathbf{R}_t \mid S_t) \tag{2}$$

and for the joint distribution we have

$$p(\mathbf{R}_{1:T} \mid \mathbf{S}_{1:T}) = \left[p(S_1) \prod_{t=2}^{T} p(S_t \mid S_{t-1}) \right] \left[\prod_{t=1}^{T} p(\mathbf{R}_t \mid S_t) \right] \tag{3}$$

The probability models for rainfall PCs are assumed to be Gaussians, i.e.,

$$p(\mathbf{R}_t = \mathbf{r}_t \mid S_t = i) = (2\pi)^{-\frac{M}{2}} |\mathbf{\Sigma}_i|^{-1} \exp\left(-\frac{1}{2}(\mathbf{r}_t - \boldsymbol{\mu}_i)\mathbf{\Sigma}_i^{-1}(\mathbf{r}_t - \boldsymbol{\mu}_i)^T \right) \tag{4}$$

where $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ denote means and covariance matrix of PCs, respectively, for given state i. The superscript T indicates transpose of vector.

With the number of hidden states specified, the parameters of the model are estimated from the rainfall PCs in a standard manner using the Expectation-Maximization (EM) iterative algorithm (Dempster et al. 1977), run 10 times from different initial seeds, selecting the run with the highest log-likelihood. The data consist of 25 sequences of 32 pentads each, for each of the 4 PCs.

c. Nonhomogeneous Hidden Markov Model (NHMM)

In a non-homogeneous HMM the state-transition matrix $\mathbf{\Gamma}$ is no longer stationary, and the transition probabilities are defined to be a function of a multivariate "predictor" input time series $\mathbf{X}_{1:T}$, corresponding to other variables that can influence the evolution of the weather state sequence $\mathbf{S}_{1:T}$. Here we describe experiments with such a model, where the transition probabilities are defined as a logistic function f of the predictor variables, i.e.,

$$\gamma_{ij}(t) = \gamma_{ij}(t) \times f(\mathbf{X}_t) \tag{5}$$

More complete details on this model are provided in Hughes et al. (1999) and Robertson et al. (2003).

A single univariate time series is used as input, given by the JJA mean value of NINO3.4 SSTA. For each MJJAS season, the JJA value is held constant and simply repeated for each pentad during the season.

3. Hidden Markov Model of intraseasonal variability

In this section, purely intraseasonal variability is examined using the HMM of pentad precipitation data, with both the mean seasonal cycle and the seasonal average subtracted. Before applying the HMM, the dominant characteristics of intraseasonal precipitation variability can be seen in an EOF analysis. The two leading EOF modes are shown in Fig. 1, explaining 7.5% and 5.7%, respectively, of the intraseasonal variance of pentad precipitation. The first mode consists of a southwest-northeast rainfall dipole structure. In the second mode, this dipolar precipitation anomaly pattern is shifted northeastward and an additional negative anomaly appears over the equatorial Indian Ocean. These two EOF patterns are very similar to those seen in the previous studies (Lau and Chan 1986; Lawrence and Webster 2001; Jiang and Li 2005), based on OLR data, bandpass filtered to retain intraseasonal time scales.

These two EOF modes can be considered to be different phases of the northeastward propagating summertime ISO, as indicated by the time series of two modes (Fig. 1c). The time variation of first mode appears to lead that of second mode by one or two time pentads. The maximum lag-correlation between the two PCs is 0.42 at 10-day lag (PC1 leads PC2), yielding a period of around 40 days. Although the spatial patterns of two EOF modes exhibit a clear 90-degree phase quadrature, Fig. 1c shows that the lag relationship between the two PCs is often quite irregular, motivating the use of the

HMM.

The HMM is now trained on the leading 4 PCs of intraseasonal pentad MJJAS precipitation, 1979–2003, yielding a 4-dimensional mean vector and covariance for each state in PC space, accounting for 20.1% of the total variance. In terms of explained variance, the first two modes are relatively distinctive from other modes. The next four modes explain 3.8, 3.1, 2.9, and 2.7 % of total variance. Spatial patterns of the mean for each state are then obtained by transforming back to physical space via the EOF patterns. These mean spatial patterns for each HMM state (i.e. from Eq. 4) are the analogs of the cluster means in EOF-based cluster analyses (Kimoto and Ghil 1993, Michelangeli et al. 1995). The Gaussianity assumption was verified at the 95% significance level for each of the 4 PCs using a Kolmogorov-Smirnov test. A scatter plot of the two leading PCs did not exhibit any noticeable skewness (not shown). The result of this study is not sensitive to the number of PCs due to the relatively small contribution of higher modes.

An important question in any HMM or cluster analysis concerns the appropriate number of states to be specified. This choice depends on the goal of the analysis. Objective measures can often provide guidance, such as cross-validated likelihood, or penalized likelihood measures such as the Bayes information criterion. Figure 2 shows cross-validated likelihood of HMM with different number of states. As number of states increases, cross-validated likelihood increases and become saturated when the number of states is more than 5. However, when, as here, a major goal of the analysis is to construct a model for diagnostic purposes, the number of states suggested by these

measures may be too large to be useful. In this case, the principal of parsimony is more appropriate. We thus select the minimal model (least number of states) that is capable of representing the main features of the target phenomena, here the ISO and ENSO's impact on the ASM. In addition, if a particular state of HMM represents a physically meaningful rainfall pattern, it should be picked up by the NHMM as well, since the impact of the input variable in the NHMM is through the state-transition probabilities. For our purposes, it is desirable that introducing external input data for transition probability change should not affect the selection of the hidden states significantly. We compared the spatial patterns of the states derived with an HMM and an NHMM with different number of states, using JJA NINO3.4 SSTA as the input. The spatial patterns of the states from HMM and NHMM were found to be consistent up to 5-state model, while models with more states produce some of states are not shared between the models. Because the spatial patterns from 4-state HMM and 5-state HMM appear to approximate the same physical process, we thus chose the 4-state model (HMM4) for parsimony.

The state-mean precipitation anomaly patterns of HMM4 are shown in Fig. 3. State-1 (S1) is characterized by a positive precipitation anomaly over the equatorial Indian Ocean and a negative anomaly stretching from India to the tropical western North Pacific. In the S2, the positive rainfall anomaly is enhanced and displaced northward and eastward; the negative anomaly in the Indian sector is reduced but that in the western north Pacific is shifted further north. S3 and S4 are almost mirror images of S1 and S2. The four HMM states clearly contain the main features seen in EOFs 1 and 2 (Fig. 1), and are ordered in Fig. 3 such that their sequence from left to right clearly

shows the eastward and northward propagation of precipitation anomalies, characteristic

of the ISO. Note that four states are required to represent the two possible polarities of

the two EOFs in Fig. 1.

If HMM4 does represent the ISO, the transitions among the states should be cyclic. The

transition probabilities between the states are shown in Table 1 and do indeed imply a

clear cyclic evolution of states (S1 → S2 → S3 → S4 → S1). The self-transition

(persistence) probabilities of four states have similar magnitudes (~0.5), signifying that

the cyclic evolution between the four states is approximately equally spaced in time. In

order to deduce the period of oscillation from the transition probabilities, we performed

simple tests of the HMM with synthetic data. We generated a pair of univariate time

series data (x_1, x_2) representing a propagating oscillation superposed on a white-noise

background, defined as,

$$
\begin{aligned}
x_1(t) &= \cos(\frac{t}{\tau}) + \varepsilon_1 \\
x_2(t) &= \sin(\frac{t}{\tau}) + \varepsilon_2
\end{aligned}
\tag{6}
$$

where τ and ε denote the period of oscillation and amplitude of the Gaussian white

noise, respectively. HMMs are then trained to fit the pair of time series, for different

choices of τ and ε, and the number of states. Fig. 4 is an example of (a) time series

and (b) data points in the phase space. The probability distributions of four states

identified by an HMM are shown as contours. Although the absolute positions of states

in the 2-D phase space are arbitrary and vary with different subsets of data, their relative

positions are robust and evenly spaced in a circle with center at the origin of the phase

space, and the transition probabilities show clear cyclic transitions. Since the period of

13

oscillation is fixed in this data, the transition probabilities from one state to the next state of cycle are almost identical for all the states. In this test, we found the relationship between the transition probability and the period of oscillation as follows.

$$p_{TRANS} = \frac{N_S}{N_\tau} \tag{7}$$

where p_{TRANS} is the transition probability to the next state, N_S the number of states, and N_τ the number of time steps for one cycle of oscillation. In case of HMM4, the transition probability in the Table 1 is approximately 0.4~0.5, Eq. (7) yields N_τ = 8~10, which corresponds to a period of 40~50 days for pentad data.

For the real pentad data, there are, however, some noticeable deviations in transition probability (Table 1) from what would be expected in the regular ISO cycle. The relatively large persistence probability of S4 and small transition probability from S4 to S1 indicate that the re-initiation of convection over the equatorial Indian Ocean (S1) is often delayed. On the other hand, there is also a chance of transition from S3 to S1 without passing through S4, suggesting an accelerated re-initiation of the ISO. These results suggest that the re-initiation of convective activity in the equatorial Indian Ocean is more variable than other phases of the ISO and therefore, less predictable. The complexity of the re-initiation of convection has been reported in previous studies and it is associated with the thermodynamic state of background, such as the variation of western Indian Ocean SST and slow annual cycle of Asian monsoon system (Kembell-Cook and Wang 2001; Jiang and Li 2005; Wang et al. 2005; Goswami et al. 2006).

One of the strengths of the HMM lies in its ability to generate stochastic simulations

which can be used to validate the model. Here we use HMM4 to generate a 25-year MJJAS synthetic precipitation data set to assess the ability of the model to simulate the ISO. Since the direct outputs of the simulations are PCs, we first examined periodicity of simulated PCs in comparison with observed PCs. The power spectra of leading 2 PCs of observation and simulation show clear peaks in the 30-60 days range (not shown), indicating that the 30–60-day oscillations represented by the two leading PCs are reasonably simulated by HMM4. The notable difference is, however, that the simulated PCs show relatively higher power in the high frequency range (10-30days) than observation. Since the HMM is a purely stochastic model governed by a set of state transition probabilities,unrealistically fast or slow cycles can be simulated (cf. Jones 2009). Moreover, the amplitude of PCs for given HMM state is randomly determined by Eq. 4 without any memory of past state. Because of the random amplitude change, there can be sudden activations or terminations of the simulated ISO. The overestimation of the high frequency activity may be related with this abrupt change of magnitude and some unregulated fast phase transitions in the HMM simulation.

Figure 5 shows the evolution of rainfall anomalies in the CMAP observations (left two columns: Fig. 5a to 5i) compared to those simulated by HMM (right two columns: Fig. 5j to 5t). These maps are lead-lag regression coefficients of precipitation anomalies of individual grid points onto the precipitation time series at the reference point (87.5^{o}E and 15°N) for 800 time steps; the maps are made in exactly the same way for the observations and simulations. Only regression coefficients significant at the 90% significance according to a two-sided t-test are plotted, with the number of degrees of freedom 30. The regression maps of the observational data show the evolution of ISO

with northward and eastward propagation of precipitation anomalies. From the -15days to the +15days lag, the regression patterns of HMM4 have significant resemblance with those of the observations. Moreover, the magnitudes of regression coefficients of HMM simulation are similar to those of observation. This suggests that HMM4 reproduces the ISO and its irregularity relatively well. It is noted that the regressed precipitation anomaly of simulation is weaker than observation but the variance of simulated rainfall of each grid point is very similar to that of observed rainfall reconstructed with 4 leading EOF modes. At ±20-days lag, however, the HMM simulations do tend to underestimate the ISO signal, suggesting that a higher-order Markov model might be preferable.

Further analysis on the statistics of active ISO events was performed. A couple of studies (Matthews 2008; Jones 2009) have classified MJOs into primary (isolated or the first event in the series of consecutive MJOs) and successive (events following previous active MJO) events. In a similar way to the study of Matthews (2008), discrete active ISO events were identified using two leading PCs of pentad precipitation anomaly. To reduce the impact of noisy fluctuations, a three-pentad moving average is applied to each PC before identifying ISO events. To identify phase transitions, the phase space of the two leading PCs (PC1 as x-axis and PC2 as y-axis) is divided into 4 phase categories depending on the signs of the two PCs. These quadrants represent 4 ISO phases, which are similar to the 4 states of HMM4 (Fig. 3). Based on the values of PC1 and PC2, each pentad can be categorized into one of four phases of ISO and then a sequence of ISO phases can be obtained. The proper sequence of phases (e.g. 1-1-2-2-3-4-4-1) represents ISO propagation. The ISO event is identified when there is a

continuous phase transition across at least 3 phases (e.g. 1-2-2-3 or 4-4-1-1-1-2-3-3). When the sequence of phases reverses (e.g. 1-2-3-2) or jumps (e.g. 1-3-4), it is not considered as a continuous transition and therefore not a proper ISO event. Another criterion to identify active ISO episodes is that the amplitude $a = \sqrt{PC1^2 + PC2^2}$ should be bigger than a threshold value. Here one standard deviation (sd) of a is used for observation. If there is one inactive pentad (a<1sd) between two active pentads in the sequence of proper phase transition, it is counted as an active pentad.

Based on the above criteria, 38 observed ISO events were identified. This is a little less than the result of Jones and Carvalho (2006), who identified 46 events for the same period (1979-2003). This may be related with the slightly different definition of ISO event and longer summer season (MJJASO) in their study. Among the 38 events, 27 (70%) events were classified as primary and 11 events were successive events. This contrasts with the statistics of previous studies showing that only around 40% of all season MJO events are primary, and is due to our focus on the boreal summer season when the ISO is relatively weak. When the ISO amplitude is small, the successive events of ISO may be classified as isolated as primary events. For the simulation, due to aforementioned randomness and stronger high frequency activity in HMM simulation, the threshold of amplitude is reduced to 0.7 and only the events longer than 5 pentads are counted. A total of 40 events were identified in the simulation, of which 32 (80%) events were classified as primary events. Although HMM simulates slightly more isolated events than in observations, the general statistics of simulated ISO events are in good agreement with the observations.

So far, we have provided evidence that a 4-state HMM can well represent general ISO variability, when no interannual variability is present. We next investigate the impact of ENSO on ISO activity, using a four-state NHMM (NHMM4), with the NINO3.4 SSTA index as the input. The same procedure is followed as in the HMM4 case, except that the transition probabilities in the NHMM are also made a function of the NINO3.4 SSTA index.

The precipitation anomaly patterns of four states of NHMM4 are shown in Fig. 6. Despite some small shifts in position, the NHMM patterns are quite similar to those of HMM4. In addition, simulations generated with NHMM4 exhibit almost the same propagating ISO signal to that in Fig. 5 (not shown). Thus, the ISO captured by NHMM4 is very similar to that of HMM4.

The interesting aspect of NHMM4 is that it associates the transition probabilities with ENSO. The transition probabilities between the states, which are the function of input variable (NINO3.4 SSTA), are shown in Fig. 7. Generally, the cyclic transitions between states (S1 → S2 → S3 → S4 → S1) are still apparent regardless of the value of NINO3.4 SSTA. However, the relationship between transition probability and NINO3.4 shows a notable dependence on ENSO phase. During El Niño (La Niña), the transition probabilities to S2 (S1) and S4 (S3) increase. Therefore, S2 (S1) and S4 (S3) will be more persistent during El Niño (La Niña). Since the seasonal mean anomaly was removed from the data at the outset, this preference toward certain states is not related to the well-known interannual variability of summer mean precipitation associated with ENSO. Rather it shows that particular phases of the ISO can be more prevalent

according to ENSO phase.

The prevalence of the ISO phases will be reflected in the dominant standing mode because the pairs of preferred states have similar patterns with opposite signs. The states preferred in the El Niño years show stronger signature in the northern equatorial Indian Ocean sector, whereas ISO activity is stronger over the western north Pacific and Bay of Bengal during La Niña years. Furthermore, the nodal line of the ISO dipole structure during El Niño years will be located further north than in La Niña years. To corroborate this ENSO dependence of ISO phases, Fig. 8 shows the precipitation anomaly pattern of leading EOF mode, calculated separately for the 4 years of positive (1982, 1987, 1991, 1997) and negative (1985, 1988, 1999, 2000) summer mean NINO3.4 index. The spatial structure of leading mode indeed resembles that of the preferred ISO states in each ENSO phase: the variability over the Indian Ocean is stronger in El Niño years while that over the western North Pacific is stronger in La Niña years. Also, the change of position of nodal lines of dipole in EOF modes is consistent with preferred ISO phases.

Typical cases of ISO temporal evolution during an El Niño and La Niña year are shown in Fig. 9. The left panels are sequence of pentad precipitation anomalies reconstructed with four leading EOF modes from 23rd of May (bottom) to 2nd of July (top) in 1997 (El Niño). The right panels are those from 13th of June to 20th of July in 1988 (La Niña). The pentad number is shown above each panel. At the right side of each panel, the pattern correlation between precipitation anomaly of each pentad and four states of NHMM4 are listed in an order of S1 (top) to S4 (bottom). The highest value among four correlation values are shown in red, indicating which state is most close to each pentad

precipitation anomaly.

The first precipitation anomaly patterns of both cases (Fig. 9a and Fig.9k) show the highest pattern correlation with S3 and they are very similar to each other. In general, both cases (1997, 1988) exhibit a typical ISO evolution: The northward and eastward propagation of the negative anomaly over the equatorial Indian ocean forms a slantwise negative precipitation anomaly band from India to the Maritime Continent. In the meantime, a positive anomaly appears over the equatorial Indian ocean and then propagates mostly northward.

Detailed temporal evolutions of El Niño and La Niña year show some notable differences. During the El Niño year, the negative anomaly over the Indian ocean tends to persist with its center around $5^{o}N$-$10^{o}N$ (Fig. 9b-d), while the positive anomaly over the equatorial Indian ocean first appears four pentads later (Fig. 9e). This positive anomaly remains weak (Fig. 9f) and then slowly propagate northward similar to previous negative anomaly (Fig. 9g-i). In contrast, during the La Niña year, the northward propagation of the negative anomaly over the Indian ocean occurs rather rapidly (Fig. 9l and m) and a positive anomaly appears three pentad after the initial state (Fig. 9n). This positive anomaly stays and develops near equator (Fig. 9n-p) and then rapidly propagates northward to the central Indian region (Fig. 9q-s). The sequences of states with the highest pattern correlation for each pentad are $3\rightarrow3\rightarrow4\rightarrow4\rightarrow4\rightarrow4\rightarrow4\rightarrow2\rightarrow2$ for El Niño year and $3\rightarrow4\rightarrow4\rightarrow1\rightarrow1\rightarrow1\rightarrow3\rightarrow3\rightarrow3$ for the La Niña year. It clearly shows how two preferred states (S2 and S4 for El Niño, S1 and S3 for La Niña) appears in the evolution of ISO. Another notable difference is the

20

appearance of stronger anomalies in the western North Pacific during El Niño year. It shows weak signature of westward propagation of positive anomaly (Fig. 9e-g). However, the relationship between this westward propagating mode over WNP and preference of ISO phase with respect to ENSO is not clear.

The preference of ISO phases as a function of ENSO is also examined in terms of persistence time scale by comparing the autocorrelation function of the projection of each state of NHMM4 (Fig. 10). In the figure, the colored curves show the autocorrelation functions of projection coefficient of patterns of S1(Fig. 10a) and S2(Fig. 10b) onto the daily precipitation anomaly data, for El Niño (red) and La Niña (blue) years. The daily precipitation data is constructed by linearly interpolating pentad mean precipitation reconstructed using the four leading EOFs. Thin colored curves are autocorrelation of four individual El Niño (red) and La Niña (blue) years. The thick colored curves are mean of the four years. The black curves show the all-years mean values.   The upper and lower bounds of insignificant (with 90% of significance level) autocorrelation values are shown as solid green lines, based on white noise with 32 samples (i.e. the number of pentads per season). A higher positive autocorrelation value at a particular lead time indicates that the pattern persists longer. In general, the autocorrelations of projection coefficients of both patterns are significantly positive up to around 6-10 days lag.

Since S3 and S4 are the mirror images of S1 and S2, the results of S1 and S2 are only shown here. For S1, the autocorrelations during all four El Niño years are lower than those of all four La Niña years, indicating that S1 persists longer during La Niña years.

21

The situation is opposite for the S2 up to 7 days lag with less contrast between El Niño and La Niña compared to S1. It is also consistent with the preference of ISO phases found in NHMM4. The solid black curves with open circles show the correlation coefficients between JJA mean NINO3.4 index and autocorrelation value for each lag, which is calculated every year. As expected by comparing individual autocorrelations of El Niño and La Niña years, the autocorrelation of S1, namely persistence time of S1, shows significant negative correlation with NINO3.4 SSTA for all lags with maximum around 8-12 days lag. whereas the autocorrelation of S2 shows significant positive correlation with NINO3.4 SSTA up to 5-6 days lag. The threshold value of significant correlation with 90% significance level are shown as green dashed line. The result suggests that transition from S1 to S2 is faster in El Niño years and slower in La Niña years. The situation will be same for S3 and S4 as for S1 and S2. In a same way to precipitation data, correlations are calculated by projecting S1 and S2 onto daily OLR anomaly data (black dashed curves with closed dot). For the OLR data, the correlations between the autocorrelations and the JJA NINO3.4 SSTA are qualitatively similar to those of the precipitation data but for S2, the correlation is not significant at the 90% significance level. Nonetheless, even assuming that the persistence time of S1 only is associated with the ENSO, the preference toward S1(S3) in La Niña years and S2(S4) in El Niño years can be still expected, provided there is no significant change in the period of the ISO (Lawrence and Webster 2001). This is the case because if the ISO persists in a particular phase S1(S3) for longer, the next phase S2(S4) has to be shorter for a fixed oscillation period.

4. Nonhomogeneous Hidden Markov Model of intraseasonal variability and

interannual variability

In this section, the interannual variability of CMAP pentad precipitation is included in the analysis, subtracting out only the climatological seasonal cycle. The objective here is to examine whether interannual variability and intraseasonal variability are associated with different spatial structures, expressed as different HMM states. This may help resolve the issue of whether interannual variability of ASM is primarily generated by nonlinear rectification of high frequency variability (Ferranti et al. 1997) such as the ISO, or whether it has its own distinct structure independent of the ISO (Krishnamurthy and Shukla 2007). The 4-state NHMM with NINO3.4 SSTA input (NHMM4-IA) is trained with the 4 leading PCs, as in the previous section.

The leading EOF modes of CMAP pentad precipitation anomaly do not change significantly when interannual variability is retained. Fig. 11 shows mean precipitation anomaly patterns of four states identified from the NHMM4-IA. The spatial patterns of the states show general resemblance with those of NHMM4 as well as HMM4 indicating that the ISO still dominates the variability in the pentad precipitation when interannual variability is retained. However, there are some notable differences. Considering the canonical evolution of the ISO seen in HMM4 and NHMM4 (Figs. 3 and 6), the S3 and S4 are expected to be mirror images of S1 and S2, respectively. However, the positive anomaly over the western North Pacific in S3 is expanded and shifted to the north compared to the counterpart of S1. The negative anomalies stretching from India to Maritime Continent in S4 are also slightly shifted northward, and the positive anomaly over western north Pacific is much weaker than expected from

S2. Thus, the ISO phases represented by S3 and S4 appear somewhat ahead in phase compared to the canonical ISO cycle, rather than being opposite phases of S1 and S2, as in HMM4 and NHMM4 when interannual variation was excluded from the data. Furthermore, in S4, the weak positive anomaly over western north Pacific and negative anomaly covering rest of domain except for western equatorial Indian Ocean cannot be explained by any phases of ISO evolution shown in previous section. Since the only difference between NHMM4-IA and NHMM4 is the presence of interannual variability in the former: year-to-year differences in the seasonal mean field must be responsible for this difference, particularly the structure of S4.

To examine the interannual variability of state occurrence frequency, the most-probable temporal (pentad) sequence of states during the 25 summers 1979–2003 is estimated using the *Viterbi* algorithm (see Robertson et al. 2006). From this sequence of states, the number of pentads allocated to each state is counted for each May–September season. The resulting time series of state occurrence frequency are shown in the lower panel of Fig. 12. The upper panel shows the time series of JJA mean NINO3.4 SSTA. The colored lines in the lower panel indicate the percentage of time spent in each state. For example, the S4 (green line) is the most probable state of almost 90% of total pentads during MJJAS season of 1982, whereas in 1985, about 60% of pentads are allocated to the S3. The frequency of S4 shows a strong relationship with El Niño events. The five distinctive years with large occurrence frequency of S4 are 1982, 1987, 1991, 1997 and 2002, which have higher values of NINO3.4 SSTA shown in upper panel.. Thus, the distinctive spatial pattern and strong interannual variability in the state occurrence suggest that S4 represents not only a particular phase of ISO, but the interannual

24

variability associated with El Niño as well. The interannual variability of state occurrence associated with La Niña is less clear in this case. State S3 tends to be most frequent during non-El Niño years and the frequency of S2 also increases for developing summer of some La Niña years such as 1988, and 1998.

In the previous section, it was shown that the ISO was reasonably modeled by a 4-state HMM (HMM4 and NHMM4) in the absence of interannual variability. If S4 of NHMM4-IA is not unambiguously a part of the ISO cycle, it might be necessary to increase the number of states to resolve the ISO in the presence of interannual variability. Indeed, a 6-state model (NHMM6-IA) is found to exhibit spatial patterns characteristic of both the ISO and interannual variability, as shown in Fig. 13; a 5-state model also shows very similar states but without S1. Although the patterns do not exactly match those of NHMM4, the four states S2, S3, S4, and S5 largely represent the four phases of the ISO. The S6 pattern is very similar to that of S4 in NHMM4-IA and, therefore, can be associated with El Niño. The pattern of S1 shows a positive anomaly extending from the eastern equatorial Indian Ocean to the Maritime Continent, and negative anomalies over the rest of domain (south of 20°N), particularly over the western Indian Ocean and western north Pacific. In some regions such as near the equator and western north Pacific, the S1 pattern is similar to the reverse polarity of S6, and thus might be expected to be related to La Niña. However, the relationship between the state and La Niña is unclear and appears more complex than for El Niño.

The transition probabilities of NHMM6-IA are shown in Fig. 14. The transition probability in the normal condition of ENSO (NINO3.4 SSTA = 0) show that most

probable transition is cyclic showing ISO evolution (S2→S3→S4→S5→S6→S2). In the case of El Niño, transition probabilities to S6 and its persistence probability increase rapidly as NINO3.4 index increases. Although transition probability from S3 to S4 is also increased (Fig. 14c), transition from S4 to S6 will occur rapidly (Fig. 14d). Therefore, S6 becomes persistent. Although it seems to lie between S5 and S2 regarding the ISO phase transition, and thus it could be interpreted as a phase of ISO, the strong dependency on the NINO3.4 SSTA and the pronounced association with El Niño (Figs. 12 and 15) suggest that it's more representing a seasonal mean anomaly during El Niño rather than one phase of normal ISO. The increased transition probability from S3 to S4 (Fig. 14c) during El Niño may reflect a preferred phase of the "pure" ISO (S2 of NHMM4, Fig. 6b) based on its resemblance with S4 of NHMM6-IA (the pattern correlation is 0.54), or a manifestation of the fast transition from S3 towards S6.

On the other hand, transition probabilities to both of S3 and S5 increase in the La Niña situation. Those two states are also similar to two preferred phases of the pure ISO (Fig. 6a, c). The preference of ISO phase with respect to the ENSO still holds here but is not apparent in the El Niño situation due to the strong persistent seasonal mean signal (S6). It is noted that S1, which was expected to be related with La Niña by its spatial pattern shown in Fig. 13a, indeed shows increased persistence probability in the La Niña. However, it is shown that probabilities of transition to S1 from other states are very small. It might be that S1 represents seasonal mean residual signal between two dominant states (S3 and S5) or just a weak seasonal mean signal.

Fig. 15 shows the interannual variation of state occurrence frequency of NHMM6-IA

computed from the *Viterbi* sequence. Similar to Fig. 12, the preference for S6 (orange color) during El Niño is clear, while the associations with La Niña are less clear as expected from transition probabilities; S3 is dominant during 1988 and 1998 and S5 is dominant in 1985 and 2000. These two states appear to compete during La Niña years. Note that occurrence of S1 is rare, suggesting that S1 is neither an ISO phase nor detectable seasonal mean signal in the pentad precipitation anomaly.

Simple MJJAS composites of precipitation anomalies during strong El Niño and La Niña years are plotted in Fig. 16; the two panels show composites of precipitation anomalies for four years with strong positive JJA mean NINO3.4 index (1982, 1987, 1991, and 1997) and negative NINO3.4 index (1985, 1988, 1999, and 2000). The seasonal mean precipitation anomaly of El Niño years resembles the pattern of S6 of NHMM6-IA (Fig. 13f) showing overall negative anomaly at the south of 20°N with the strongest anomaly extending from the equatorial western Indian ocean to the Maritime Continent and a weak positive anomaly over the eastern equatorial Indian ocean as well as east of the Philippines. The seasonal mean anomaly of La Niña years is similar to S1 showing a positive anomaly from the equatorial western Indian ocean to the Maritime Continent and negative anomaly from Bay of Bengal to east of the Philippines. The pattern correlation between S6 and El Niño composite is 0.67 and S1 and La Niña composite is 0.64, indicating reasonable resemblance between them.

In El Niño years, there is a strong persistent seasonal mean signal apart from ISO. The signal is detectable in the pentad mean rainfall anomaly. On the other hand, the seasonal mean anomaly of La Niña years is detected as a separate state but it is not strong enough

to be observed in the pentad mean rainfall data. According to the well-known ENSO-Monsoon relationship, however, the Indian summer monsoon rainfall should be increased in La Niña years. But the composite map shows weak signal over the India and pattern of S1 doesn't show the positive rainfall anomaly over India either. The S6 pattern and El Niño composite seasonal mean anomaly also shows some discrepancies in the location of the negative anomaly near the Indian peninsula. This may be related with the weakening of ENSO-Monsoon relationship in recent decades (Krishna-Kumar et al., 1999; Torrence and Webster, 1999; Kucharski et al. 2006) and also indicates that complexity of Indian summer monsoon variability and its connection with ENSO (Sperber et al. 2000).

5. Summary and discussion

In this study, the intraseasonal and interannual variability of Asian summer monsoon precipitation is investigated in the framework of Hidden Markov Model (HMM) with pentad precipitation analysis data for the 1979–2003 period. In some recent studies, it is shown that HMM is a useful tool to detect hidden attractors (regimes) in noisy weather data (Franzke et al. 2008; Horenko 2008). Indeed, the HMM is shown to be able to detect prominent intraseasonal modes of variability in the south Asian monsoon precipitation data; this is especially clear when interannual variability was subtracted prior to running the model. A four-state HMM shows eastward and northward propagation of precipitation anomalies (Fig. 3) and their transition probabilities support clear cyclic transitions with a period of about 40 days (Table 1). It is also found that re-initiation of convective activity over equatorial Indian Ocean is more variable than other phases of the ISO. Even with the Markovian assumption that no long-term memory

exists, the HMM is able to generate stochastic simulations that exhibit the main features of ISO propagation (Fig. 5) due to the cyclical transition probabilities. Also, it is found that the numbers of active ISO events and statistics of primary and successive ISO events in the simulations are in good agreement with those of observation. However, due to the lack of an explicit constraint to control the time scale of ISO (Jones, 2009), the HMM tends to simulate more high frequency variability than the observation. In future work, an autoregressive HMM, in which an additional autoregressive dependence between $\mathbf{R}_t$ and $\mathbf{R}_{t-1}$ (Sect. 2b) is encoded into the model (Kirshner 2005), could be considered to deal with this limitation. Along with this study, the realistic stochastic simulation of MJO by homogenous and non-homogenous Markov model (Jones 2009; Jones and Carvalho 2009) suggests that short-term ISO forecast based on the Markov process may be possible.

The interannual modulation of the ISO by ENSO is identified using the nonhomogeneous HMM (NHMM) together with the summer mean NINO3.4 index as an external input . In contrast to previous studies that mainly focused on the relationship between ISO intensity and ENSO, by associating transition probabilities between states with the NINO3.4 SSTA index, it is found that there are preferred pairs of phases of ISO with respect to ENSO phases (S2 and S4 for El Niño, S1 and S3 for La Niña; Figs. 6 and 7). These preferred phases are characterized by a shift of the nodal line of ISO dipole structure and centers of precipitation anomalies. The spatial patterns of these preferred phases show significant similarity with those of leading EOF modes during El Niño and La Niña years, respectively (Fig. 8). The preference of phases during El Niño and La Niña year is able to be verified from typical cases of ISO evolution (Fig. 9). To

further assess the modulation of the ISO by ENSO, the persistence of each ISO phase is estimated by calculating its autocorrelation function, calculated by projecting each phase's pattern onto the daily precipitation data anomaly time series. A significant correlation between NINO3.4 SSTA and the autocorrelation function is found for S1 (and to some extent S2) that supports the preference of the ISO phases with respect to ENSO state (Fig. 10). Thus, during El Niño years, the residence time of S2 (S1) tends to be longer (shorter) compared to La Niña years. This preference of particular phases may be related with the background state modified by ENSO. Several studies have pointed out that ISO convection tends to follow the region of climatologically favored convection (Salby and Hendon 1994; Li and Wang 1994; Kembell-Cook and Wang 2001). Similarly the changes in the seasonal mean flow and the SST associated with ENSO can cause shift of active convection region and it is reflected by different persistence time of ISO phases.

Previous studies have identified a correlation between ENSO and northward and eastward propagating ISO activity during early summer (Lawrence and Webster 2002; Teng and Wang 2003). In our NHMM, the impact of ENSO on the ISO is solely through changes in the transition probabilities between states, it is not possible to verify any relationship between the intensity of ISO and ENSO. In addition, the autocorrelation of projection coefficients of two ISO phases (S1 and S2) of NHMM4 shown in Fig. 10 shows only small seasonality when the autocorrelations are calculated for the first and second half of season separately (not shown). In the early summer of El Niño years, the residence time S2 tends to be even longer while that of S1 becomes less short, indicating that period of ISO may be increased. If the ISO with longer life cycle

tends to be stronger (Seo and Kumar 2008), the enhanced positive relationship between ENSO and ISO activity found by Lawrence and Webster (2002) and Teng and Wang (2003) may be explainable, although further analysis will be required to confirm this hypothesis.

When the interannual variability of the CMAP pentad precipitation is retained in the NHMM analysis, the seasonal anomalies associated with ENSO are represented by the NHMM as distinctly different states from the purely ISO ones (Fig. 13). Although the state associated with the seasonal mean anomaly of La Niña is less significantly detected in a pentad data, it generally supports the result of Krishnamurthy and Shukla (2000, 2007), who argued that the ASM mode of interannual (seasonal mean) variability has a distinct spatial structure that is independent of the ISO mode. Sperber et al. (2000), on the other hand have emphasized a common spatial structure on both subseasonal and interannual time scales, with the impact of ENSO taking place by shifting the mean of the sub-seasonal PC's probability distribution function (PDF). This ENSO dependency, however, was found by Sperber et al. (2000) to disappear when the low-frequency variability (e.g. interannual variability) is subtracted from the data. In the present study, the influence of ENSO on the ISO exists as a preference of certain phases even in the data without interannual variability. When the seasonal mean anomaly is retained in the data, the ENSO related seasonal mean will projects on preferred ISO phases with similar pattern and it may give the impression of a shift in PDF.

Like most of climate phenomena, the Asian summer monsoon ISO is a nonlinear convective process so that the linear analysis may have a limitation to describe whole

characteristics of the phenomena (Chattopadhyay et al. 2008). The result of this study shows that there are nonlinear influences of ENSO in the Asian summer monsoon ISO and the HMM effectively detects those nonlinear features, which was not clearly reported in the previous studies. However, the Asian summer monsoon ISO consists of various components with different spatial structures and different time scales (e.g. 10-20 day mode and 30-60day mode). Because it assigns one discrete state at one time step, the HMM is not able to explain when multiple ISO modes are active at the same time.

We have tested the sensitivity of the HMM results when the analysis domain is reduced to cover only the Indian ocean region. Because of short zonal extent of domain, leading EOFs only explain northward propagating mode. Krishnamurthy and Shukla (2008) have argued that the northward propagating mode may be either associated with the MJO or independentof it, with these two types of ISO having different periods of oscillation.. Perhaps as a result, the transition probabilities of the HMM trained over the reduced domain were not found to be as strongly cyclic as those in the present study. Thus, the ISO depicted by HMM in the present study is mainly MJO in the boreal summer, which has northward and eastward propagating components.

It is still un open question that whether the propagation of convective activity with certain periods ("oscillation view") is a sufficient explanation of intraseasonal rainfall variability of ASM. There are other explanations such as the intermittent developments of tropical convective zone in certain preferred locations ("regime view", e.g. Chao 2000). Ghil and Robertson (2002) discussed that a complementary approach of the "regime" and "oscillation" views in the analysis of low frequency variability will

provide deeper understanding of it. In contrast to the conventional intraseasonal oscillation analysis, the regime-like description (e.g. HMM in the present study) of the intraseasonal rainfall variability may provide a wider scope leading to in-depth understanding.

References

Chao, W.C., 2000: Multiple Quasi Equilibria of the ITCZ and the Origin of Monsoon Onset. *J. Atmos. Sci.*, 57, 641–652.

Charles, S. P., B. C. Bates, and J. P. Hughes, 1999: A spatiotemporal model for downscaling precipitation occurrence and amounts. *J. Geophys. Res.*, **104**(D24), 31657–31669.

Chattopadhyay, R., A. K. Sahai, and B. N. Goswami, 2008: Objective Identification of Nonlinear Convectively Coupled Phases of Monsoon Intraseasonal Oscillation: Implications for Prediction. *J. Atmos. Sci.*, **65**, 1549–1569.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.*, **39B**, 1–38.

Ferranti, L., J. M. Slingo, T. N. Palmer, and B. J. Hoskins, 1997: Relations between interannual and intraseasonal monsoon variability as diagnosed from AMIP integrations. *Quat. J. Roy. Meteor. Soc.*, **123**, 1323-1357

Franzke, C., D. Crommelin, A. Fischer, and A. J. Majda, 2008: A Hidden Markov Model perspective on regimes and metastability in atmospheric flows. *J. Climate*, **21**, 1740-1757.

Gadgil S., 2003: The Indian monsoon and its variability, *Ann. Rev. Planet Sci.*, **31**, 429-467.

Ghil, M., and A. W. Robertson, 2002: "Waves" vs. "particles" in the atmosphere's phase space: A pathway to long-range forecasting?, *Proc. Natl. Acad. Sci.*, **99** (Suppl. 1), 2493-2500

Goswami, B. N., and R. S. A. Mohan,2001: Intraseasonal oscillations and interannual variability of the Indian summer monsoon. *J. Climate*, **14**, 1180-1198.

Goswami, B. N., G. Wu, and T. Yasunari, 2006: The annual cycle, intraseasonal oscillations, and roadblock to seasonal predictability of the Asian summer monsoon. *J. Climate*, **19**, 5078-5098.

Hendon, H. H., C. Zhang, C., and J. Glick, 1999: Interannual fluctuations of the Madden-Julian oscillation during austral summer. *J. Climate*, 12, 2538-2550.

Horenko, I, 2008: On Simultaneous Data-Based Dimension Reduction and Hidden Phase Identification. *J. Atmos. Sci.*, **65**, 1941–1954.

Hoyos, C. D., and P. J. Webster, 2007: The role of intraseasonal variability in the nature of Asian monsoon precipitation. *J. Climate*, **20**, 4402-4424.

Hughes, J. P. and P. Guttorp 1994: Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Applied Meteorology*, **33**, 1503–1515.

Hughes, J. P., P. Guttorp, and S. P. Charles, 1999: A non-homogeneous hidden Markov model for precipitation occurrence. *J. Royal Stat. Soc.* Series C Applied Statistics, **48(1)**, 15–30.

Jiang, X.-A., and T. Li, 2005: Reinitiation of the boreal summer intraseasonal oscillation in the tropical Indian Ocean. *J. Climate*, **18**, 3777-3795.

Jones, C., 2009: A Homogeneous Stochastic Model of the Madden–Julian Oscillation. *J. Climate*, 22, 3270–3288.

Jones, C. and L. M. V. Carvalho, 2006: Changes in the activity of the Madden-Julian Oscillation during 1958-2004. *J. Climate*, **19**, 6353-6370

Jones, C. and L. M. V. Carvalho, 2009: Stochastic simulations of the Madden-Julian oscillation activity. *Climate dyn.*, DOI 10.1007/s00382-009-0660-2

Kang, I.-S., Y.-M. Lee, and S.-I. An, 1995: Interannual variability of Typhoon activity over the western Pacific and El Niño, *Asia-Pacific J. Atmos. Sci.*, **31**, 15-26.

Kembell-Cook, S., and B. Wang, 2001: Equatorial waves and air-sea interaction in the boreal summer intraseasonal oscillation. *J. Climate*, **14**, 2923-2942.

Kim, H.-M., I.-S. Kang, B. Wang, and J.-Y. Lee, 2008: Interannual variations of the boreal summer intraseasonal variability predicted by ten atmosphere-ocean coupled models, *Climate Dyn.*, **30**, 485-496.

Kimoto, M., and M. Ghil, 1993: Multiple flow regimes in the northern hemisphere winter. Part I: Methodology and Hemispheric Regimes. *J. Atmos. Sci.* **50**, 2625-2644.

Kirshner, S., 2005: Modeling of multivariate time series using hidden Markov models. Ph.D. thesis, University of California, Irvine.

Klingaman, N.P., H. Weller, J. M. Slingo, and P. M. Inness, 2008: The Intraseasonal Variability of the Indian Summer Monsoon Using TMI Sea Surface Temperatures and ECMWF Reanalysis. *J. Climate*, **21,** 2519-2539.

Krishnamurthy, V., and J. Shukla, 2000: Intraseasonal and interannual variability of rainfall over India. *J. Climate*, **13**, 4366-4377.

Krishnamurthy, V., and J. Shukla, 2007: Intraseasonal and Seasonally Persisting Patterns of Indian Monsoon Rainfall. *J. Climate, 20*, 3-20.

Krishnamurthy, V., and J. Shukla, 2008: Seasonal persistence and propagation of intraseasonal patterns over the Indian monsoon. *Climate Dyn., 30*, 353-369.

Krishna Kumar, K., B. Rajagopalan, and M. A. Cane, 1999: On the weakening relationship between the Indian monsoon and ENSO. *Science*, 284, 2156-2159.

Kucharski, F., A. Bracco, J. H. Yoo, and F. Molteni, 2006: Low-frequency variability of Indian monsoon-ENSO relationship and the tropical Atlantic: The 'weakening' of the '90s and '90s. *J. Climate*, **20**, 4255–4266.

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

Lau K. M., and D. E. Waliser, 2005: Intraseasonal Variability in the Atmosphere–Ocean Climate System. *Springer,* 436 pp.

Lawrence, D. M., and P. J. Webster, 2001: Interannual variations of the intraseasonal oscillation in the south Asian summer monsoon region. *J. Climate*, **14**, 2910-2922.

Li, T., and B. Wang, 1994: The influence of sea surface temperature on the tropical intraseasonal oscillation: A numerical experiment. *Mon. Wea. Rev.*, **122**, 2349-2362.

Liebmann B. and C.A. Smith, 1996: Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset. *Bull. Amer. Meteor. Soc*., 77, 1275-1277.

Matthews, A. J. 2008: Primary and successive events in the Madden-Julian Oscillation. . *Quart. J. Roy. Meteor. Soc.,* **134**, 439-453.

Michelangeli P. A., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi-stationarity. *J. Atmos. Sci.*, 52, 1237–1256.

Palmer, T.N., 1999: A Nonlinear Dynamical Perspective on Climate Prediction. *J. Climate*, **12**, 575–591.

Robertson, A. W., S. Kirshner, and P. J. Smyth, 2003 Hidden Markov models for modeling daily rainfall occurrence over Brazil. *Tech. Rep. ICS-TR* 03-27, Information and Computer Science, University of California, Irvine, CA, 38pp.

Robertson, A. W., S. Kirshner, and P. Smyth, 2004: Downscaling of daily rainfall occurrence over Northeast Brazil using a Hidden Markov Model. *J. Climate*, 17,

4407–4424.

Robertson, A. W., S. Kirshner, P. Smyth, S. P. Charles, and B. C. Bates, 2006: Subseasonal-to-Interdecadal variability of the Australian monsoon over North Queensland. *Quat. J. Roy. Meteor. Soc.*, **132**, 519-542.

Robertson, A. W. , V. Moron, and Y. Swarinoto, 2009: Seasonal predictability of daily rainfall statistics over Indramayu district, Indonesia. *Int. J. Climatology*, **29**, 1449-1462

Salby, M. L., and H. H. Hendon, 1994: Intraseasonal behavior of clouds, temperature, and motion in the Tropics. *J. Atmos. Sci.*, 51, 2207-2224.

Seo, K.-H. and A. Kumar, 2008: The onset and life span of the Madden–Julian oscillation. *Theor. Appl. Climatol.,* **94** 13-24

Shukla J., and D. A. Paolino, 1983: The Southern Oscillation and long-range forecasting of the summer monsoon rainfall over India. *Mon. Wea. Rev.*, 111, 1830–1837.

Singh, S. V., R. H. Kripalani, and D. R. Sikka, 1992: Interannual variability of the Madden-Julian oscillations in Indian summer monsoon rainfall. *J. Climate*, **5**, 973-978.

Slingo, J. M., D. P. Rodwell, K. R. Sperber, and F. Nortley, 1999: On the predictability of the interannual behavior of the Madden-Julian oscillation and its relationship with El Niño. *Quat. J. Roy. Meteor. Soc.*, **125**, 583-609.

Smith, T.M., and R.W. Reynolds, 2004: Improved Extended Reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466-2477.

Smyth, P., K. Ide and M. Ghil, 1999: Multiple Regimes in Northern Hemisphere Height Fields via Mixture Model Clustering. *J. Atmos. Sci.*, **56**, 3704–3723.

Sperber, K.R., J.M. Slingo, and H. Annamalai, 2000: Predictability and the relationship

between subseasonal and interannual variability during the Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.,* **126**, 2545–2574.

Tam, C-Y. and N.-C. Lau, 2005: Modulation of the Madden-Julian Oscillation by ENSO: Inferences from Observations and GCM Simulations". *J. Meteor. Soc. Japan*, **83**, 727-743. (2005) .

Teng, H., and B. Wang, 2003: Interannual variations of the boreal summer intraseasonal oscillation in the Asian-Pacific region. *J. Climate*, **16**, 3572-3584.

Torrence, C., and P. J. Webster, 1999: Interdecadal changes in the ENSO-monsoon system. *J. Climate*, **12**, 2679-2690.

Wang, B., and J.C.L. Chan, 2002: How Strong ENSO Events Affect Tropical Storm Activity over the Western North Pacific. *J. Climate*, 15, 1643–1658.

Wang, B., and H. L. Rui, 1990: Synoptic climatology of transient tropical intraseasonal convection anomalies: 1975-1985. *Meteor. Atmos. Phys.*, **44**, 43-61.

Wang, B., P. J. Webster, H. Teng, 2005: Antecedents and self-induction of the active-break Indian summer monsoon. *Geophys. Res. Lett.* 32, L04704, doi:10.1029/ 2004GL020996.

Webster, P. J., V. O. Magana, T. N. Palmer, J. Shukla, R. A. Tomas, M. Yanai, T. Yasunari, 1998: The monsoon: processes, predictability, and the prospects for prediction. *J. Geophys. Res.*, **103**, 14451-14510.

Webster, P. J. and S. Yang, 1992: Monsoon and ENSO: Selectively Interactive Systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877-926.

Wu, M. L. C., S. Schubert, and N. E. Huang, 1999: The development of the south Asian summer monsoon and the intraseasonal oscillation. *J. Climate*, **12**, 2054-2075.

Xie, P. and P. A. Arkin, 1997: Global Precipitation: A 17-year monthly analysis based on

gauge observations, satellite estimates, and numerical model outputs, *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558.

Yasunari, T., 1979: Cloudiness fluctuations associated with the Northern Hemisphere summer monsoon. *J. Meteor. Soc. Japan*, **57**, 227–242.

Yun, K.-S., B. Ren, K.-J. Ha, J. C.L. Chan, and J.-G. Jhun, 2009: The 30-60 day Oscillation in the East Asian Summer Monsoon and its Time-dependent Association with the ENSO, *TellusA*, **56**, 565-578

Figure captions

Figure 1. (a) The first and (b) the second EOF patterns of CMAP pentad precipitation without seasonal cycle and seasonal mean anomaly. (c) Sample of PC time series of the first (black) and the second (red) EOF modes for 4 years (MJJAS only).

Figure 2. Cross-validated log-likelihood of HMM with different number of states. LL is total log-likelihood and N is number of variables used in HMM.

Figure 3. (a) Pentad precipitation anomaly[mm day^{-1}] corresponding State 1(S1) determined in 4 states HMM (HMM4). (b)-(d) Same as in (a) except for State 2-4.

Figure 4. (a) Time series of two synthetic data (x_1, x_2). (b) Scatter plot of synthetic data in the phase space of (x_1, x_2). Each dot corresponds to each time step. Contours are emission probability density function (p.d.f.) determined by HMM (arbitrary interval). Four states are identified as four centers of p.d.f.

Figure 5. Left panels: (a)-(i) one point lead-lag regression coefficients maps of CMAP pentad precipitation anomaly. Right panels: (j)-(t) same as in left panels except for HMM simulations. The precipitation anomaly time series at individual grid points are regressed onto that of the point of 87.5^{o}E and 15°N, which is located in Bay of Bengal. The values with 90% significance level are plotted.

Figure 6. Same as in Figure 3 except for NHMM with JJA mean NINO3.4 SSTA as an input data (NHMM4)

Figure 7. (a) Probabilities of transition from State 1 to State 1 (black), State 2 (red), State 3 (blue) and State 4 (green) as a function of JJA mean NINO3.4 SSTA value (x-axis). (b)-(d) Same as (a) but from State 2-4.

Figure 8. (a) Regressed pattern of CMAP pentad precipitation anomaly[mm day^{-1}] to the first EOF PC calculated only in the year with the largest four positive values of summer NINO3.4 SSTA (1982, 1987, 1991, 1997). (b) Same as (a) except

41

for years with negative NINO3.4 SSTA (1985, 1988, 1998, 1999)

Figure 9. Sequence of pentad precipitation anomaly [mm day^{-1}] reconstructed by using first 4 leading EOFs during early summer of (a)-(j) 1997 (left panels, bottom to top) and (k)-(s) 1988 (right panels, bottom to top). The pattern correlations between each precipitation anomaly and patterns of each NHMM4 state in Fig. 6 are shown at the right of each panel (top: S1, bottom: S4) The highest pattern correlations among 4 values are shown in red color. The pentad number is shown above each panel (e.g. 30 is 30th pentad of year; 26th of May to 30th of May)

Figure 10. Autocorrelations (AC) of projection coefficients of (a) S1 of NHMM4 and (b) S2 onto daily precipitation anomaly for El Niño years (82, 87, 91, 97; red) and La Niña years (85, 88, 98, 99; blue). Thin and thick colored lines correspond to AC of individual years and mean value of four El Niño(red) and La Niña(blue) years, respectively. Black solid lines are mean AC of all years. The upper and lower bound of insignificant values are shown in green solid lines. Autocorrelation between upper and lower bounds are not significant with 90% significance. Black curves with open circles are correlations between JJA mean NINO3.4 SSTA and autocorrelation values at each lag day. Black dashed curves with closed circles are correlations between JJA mean NINO3.4 SSTA and the autocorrelation of projection coefficient of (a) S1 and (b) S2 onto daily OLR anomaly. The upper and lower bound of insignificant values (with 90% significance) are shown in green dashed lines.

Figure 11. Same as in Figure 6 except for the NHMM4-IA.

Figure 12. Lower panel: Interannual variation of the frequency of occurrence of states in the NHMM4-IA. The frequency of states is represented by percentage. Colors of states are marked on the right side of panel. Upper panel: JJA mean NINO 3.4 SST anomaly [deg]

Figure 13. Same as in Figure 11 but for 6 states model (NHMM6-IA).

Figure 14. Same as in Figure 7 except for NHMM6-IA. The lines with black, red, blue, green, violet and orange colors indicate transition probabilities *to* S1, S2, S3, S4, S5, and S6, respectively.

Figure 15. Same as in Figure 12 except for NHMM6-IA

Figure 16. Composite of MJJAS mean precipitation anomalies [mm day^{-1}] for (a) El Niño years (82, 87, 91, 97) and (b) La Niña years (85, 88, 98, 99)

Table 1. Probabilities of transition from the state at one time step to the state at next
time step for HMM4. Probability values shown are multiplied by 100.

From \ To	S1	S2	S3	S4
S1	55	45	0	0
S2	0	58	42	0
S3	4	0	42	54
S4	37	0	0	63

Figure 1. (a) The first and (b) the second EOF patterns of CMAP pentad precipitation without seasonal cycle and seasonal mean anomaly. (c) Sample of PC time series of the first (black) and the second (red) EOF modes for 4 years (MJJAS only).

Figure 2. Cross-validated log-likelihood of HMM with different number of states. LL is total log-likelihood and N is number of variables used in HMM.

Figure 3. (a) Pentad precipitation anomaly[mm day^{-1}] corresponding State 1(S1) determined in 4 states HMM (HMM4). (b)-(d) Same as in (a) except for State 2-4.

Figure 4. (a) Time series of two synthetic data (x_1, x_2). (b) Scatter plot of synthetic data

    in the phase space of (x_1, x_2). Each dot corresponds to each time step.

    Contours are emission probability density function (p.d.f.) determined by

    HMM (arbitrary interval). Four states are identified as four centers of p.d.f.

Figure 5. Left panels: (a)-(i) one point lead-lag regression coefficients maps of CMAP pentad precipitation anomaly. Right panels: (j)-(t) same as in left panels except for HMM simulations. The precipitation anomaly time series at individual grid points are regressed onto that of the point of $87.5^{o}E$ and $15^{o}N$, which is located in Bay of Bengal. The values with 90% significance level are plotted.

Figure 6. Same as in Figure 3 except for NHMM with JJA mean NINO3.4 SSTA as an input data (NHMM4)

Figure 7. (a) Probabilities of transition from State 1 to State 1 (black), State 2 (red), State 3 (blue) and State 4 (green) as a function of JJA mean NINO3.4 SSTA value (x-axis). (b)-(d) Same as (a) but from State 2-4.

51

Figure 8. (a) Regressed pattern of CMAP pentad precipitation anomaly[mm day^{-1}] to the first EOF PC calculated only in the year with the largest four positive values of summer NINO3.4 SSTA (1982, 1987, 1991, 1997). (b) Same as (a) except for years with negative NINO3.4 SSTA (1985, 1988, 1998, 1999)

Figure 9. Sequence of pentad precipitation anomaly [mm day^{-1}] reconstructed by using first 4 leading EOFs during early summer of (a)-(j) 1997 (left panels, bottom to top) and (k)-(s) 1988 (right panels, bottom to top). The pattern correlations between each precipitation anomaly and patterns of each NHMM4 state in Fig. 6 are shown at the right of each panel (top: S1, bottom: S4) The highest pattern correlations among 4 values are shown in red color. The pentad number is shown above each panel (e.g. 30 is 30th pentad of year; 5days from 26th of May to 30th of May)

53

Figure 10. Autocorrelations (AC) of projection coefficients of (a) S1 of NHMM4 and (b) S2 onto daily precipitation anomaly for El Niño years (82, 87, 91, 97; red) and La Niña years (85, 88, 98, 99; blue). Thin and thick colored lines correspond to AC of individual years and mean value of four El Niño(red) and La Niña(blue) years, respectively. Black solid lines are mean AC of all years. The upper and lower bound of insignificant values are shown in green solid lines. Autocorrelation between upper and lower bounds are not significant with 90% significance. Black curves with open circles are correlations between JJA mean NINO3.4 SSTA and autocorrelation values at each lag day. Black dashed curves with closed circles are correlations between JJA mean NINO3.4 SSTA and the autocorrelation of projection coefficient of (a) S1 and (b) S2 onto daily OLR anomaly. The upper and lower bound of insignificant values (with 90% significance) are shown in green dashed lines.

54

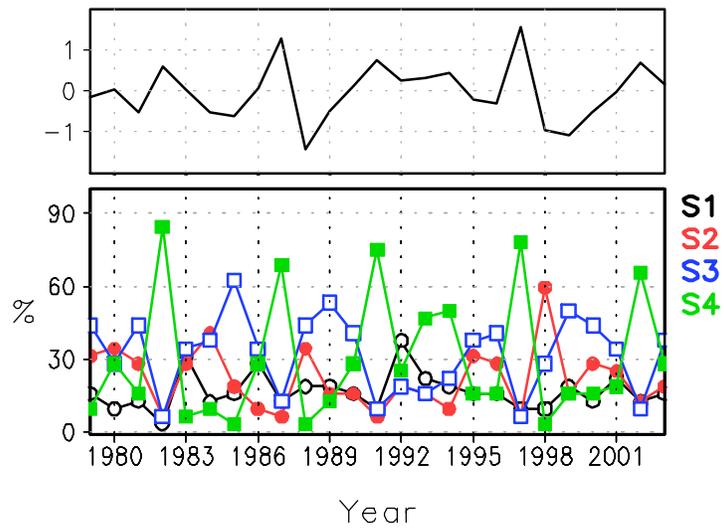Figure 11. Same as in Figure 6 except for the NHMM4-IA.

Figure 12. Lower panel: Interannual variation of the frequency of occurrence of states in
the NHMM4-IA. The frequency of states is represented by percentage. Colors
of states are marked on the right side of panel. Upper panel: JJA mean Nino
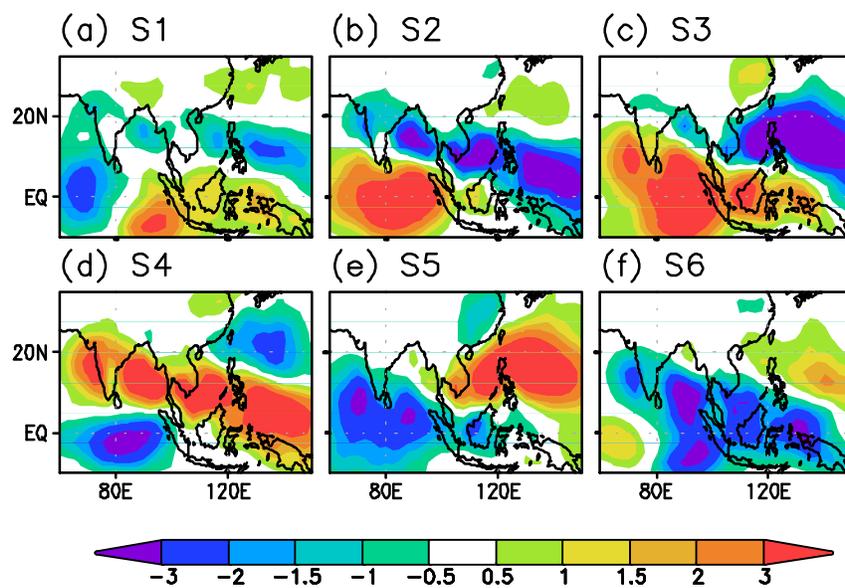3.4 SST anomaly [deg]

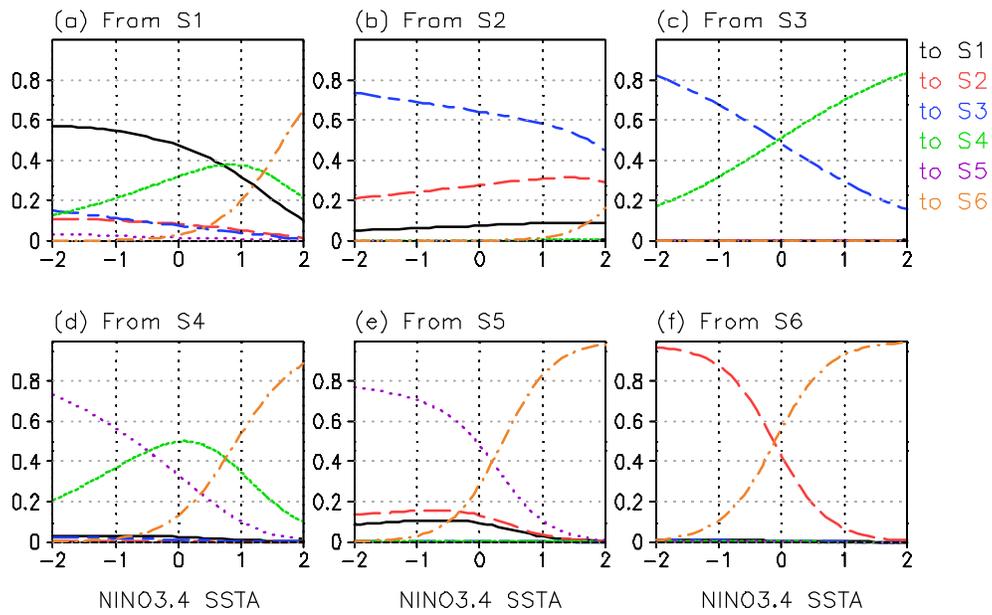Figure 13. Same as in Figure 11 but for 6 states model (NHMM6-IA).

Figure 14. Same as in Figure 7 except for NHMM6-IA. The lines with black, red, blue, green, violet and orange colors indicate transition probabilities *to* S1, S2, S3, S4, S5, and S6, respectively.
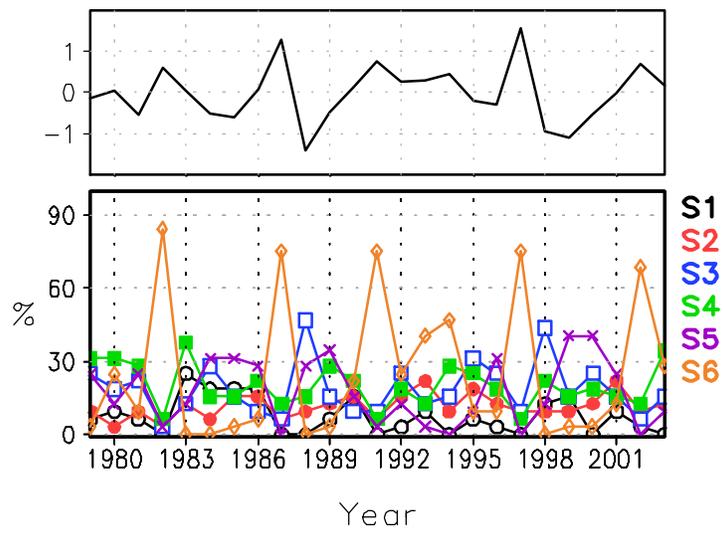
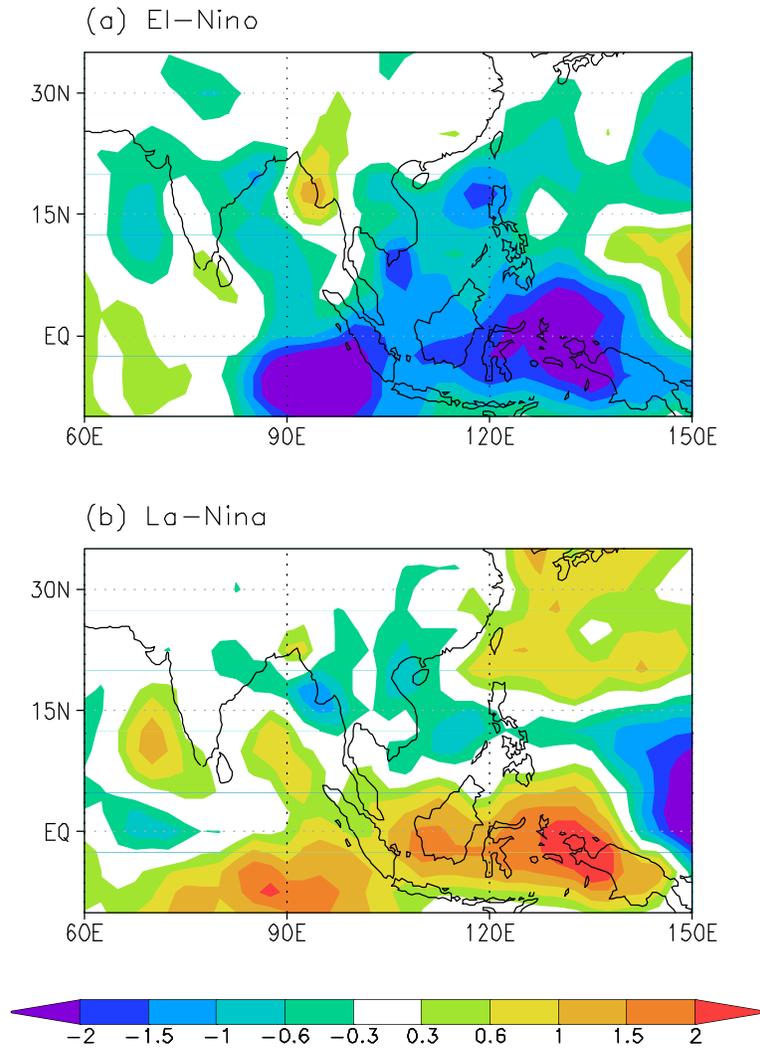Figure 15. Same as in Figure 12 except for NHMM6-IA

Figure 16. Composite of MJJAS mean precipitation anomalies [mm day^{-1}] for (a) El Niño years (82, 87, 91, 97) and (b) La Niña years (85, 88, 98, 99)