Climate Dynamics manuscript No.
(will be inserted by the editor)

Extraction and prediction of monsoon intraseasonal oscillations: An approach based on nonlinear Laplacian spectral analysis

C. T. Sabeerali · R. S. Ajayamohan · Dimitrios Giannakis · Andrew J. Majda

Received: date / Accepted: date

C. T. Sabeerali
Center for Prototype Climate Modeling, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE
E-mail: sabeer@nyu.edu

R. S. Ajayamohan
Center for Prototype Climate Modeling, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE
E-mail: Ajaya.Mohan@nyu.edu

Dimitrios Giannakis
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
E-mail: dimitris@nyu.edu

Andrew J. Majda
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
Center for Prototype Climate Modeling, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE
E-mail: jonjon@cims.nyu.edu

Abstract An improved index for real-time monitoring and forecast verification of monsoon intraseasonal oscillations (MISOs) is introduced using the recently developed nonlinear Laplacian spectral analysis (NLSA) technique. Using NLSA, a hierarchy of Laplace-Beltrami (LB) eigenfunctions are extracted from unfiltered daily rainfall data from the Global Precipitation Climatology Project over the south Asian monsoon region. Two modes representing the full life cycle of the northeastward-propagating boreal summer MISO are identified from the hierarchy of LB eigenfunctions. These modes have a number of advantages over MISO modes extracted via Extended Empirical Orthogonal Function (EEOF) analysis, including higher memory and predictability, stronger amplitude and higher fractional explained variance over the western Pacific, Western Ghats, and adjoining Arabian Sea regions, and more realistic representation of the regional heat sources over the Indian and Pacific Oceans. The skill of the NLSA-based indices in real-time prediction of MISO is demonstrated using extended-range hindcasts of the NCEP version 2 Coupled Forecast System (CFSv2) model. It is shown that these indices yield a significantly higher prediction skill than conventional indices supporting the use of NLSA in real-time prediction of MISO.

Keywords Monsoon Intraseasonal Oscillations · Nonlinear Laplacian Spectral Analysis · CFSv2

1 Introduction

The boreal summer monsoon rainfall over south Asia shows a strong intraseasonal variability with two dominant modes: a northeastward propagating mode with 30–60 day periodicity (Sikka and Gadgil, 1980; Goswami and Ajayamohan, 2001) and a westward propagating biweekly mode with 10–20 day periodicity (Krishnamurti and Bhalme, 1976; Chatterjee and Goswami, 2004). The low-frequency northeastward-propagating mode is generally known as the Monsoon Intraseasonal Oscillation (MISO; Kikuchi et al, 2012; Lee et al, 2012). The propagating characteristics of the MISO are more complex compared to the eastward-propagating

Madden Julian Oscillation (MJO) due to its interaction with the mean monsoon circulation and other modes of tropical variability. The phase of MISO occurring during the early and late monsoon season influences the timing of the onset and withdrawal of the Indian summer monsoon, respectively, and thereby the length of the rainy season (Sabeerali et al, 2012). MISO also affects rainfall over the Indian subcontinent, playing a fundamental role in the strength of the seasonal mean Indian summer monsoon and its predictability (Goswami and Ajayamohan, 2001; Ajayamohan and Goswami, 2003; Gadgil, 2003). Hence, an accurate prediction of various characteristics MISO phases and extreme events associated with the Indian summer monsoon is highly significant. In particular, the extended range prediction of MISO phases and real-time monitoring of the MISO is vital for agricultural planning like sowing, harvesting and water management (Sahai et al, 2013; Abhilash et al, 2014).

Several indices have been proposed in recent years for real-time monitoring and forecast verification of the MJO and MISO (Wheeler and Hendon, 2004; Lee et al, 2012; Kikuchi et al, 2012; Suhas et al, 2013). Among these, the multivariate RMM index (Wheeler and Hendon, 2004), constructed through multivariate Empirical Orthogonal Function (EOF) analysis of Outgoing Longwave Radiation (OLR) and zonal wind data, is primarily designed to monitor the MJO, which peaks in boreal winter. For that reason, the RMM index fails to capture the northeastward propagation of the MISO (Lee et al, 2012; Kikuchi et al, 2012; Suhas et al, 2013). By applying Extended EOF (EEOF) analysis on bandpass-filtered OLR data, a bimodal MJO-BSISO index was introduced by Kikuchi et al (2012) to represent the state of the intraseasonal variability during all seasons. Other indices (Lee et al, 2012; Suhas et al, 2013) are based on similar multivariate EOF and EEOF techniques. In particular, the MISO index proposed by Suhas et al (2013, hereafter EEOF MISO index) has been used since its introduction by the Indian Institute of Tropical Meteorology (IITM) for real-time MISO prediction (Sahai et al, 2013; Abhilash et al, 2013). This index is based on EEOF analysis of longitudinally averaged

JJAS rainfall data over the Indian Monsoon region, and captures the spatial and temporal MISO patterns reasonably well, isolating the northeastward-propagating 30–60 day periodicity band from the high-frequency westward propagating band (Suhas et al, 2013; Abhilash et al, 2013, 2014). Yet, the seasonal extraction and longitudinal averaging required to compute these indices can potentially lead to loss of predictive information or mixing with other other modes. More broadly, it is evident that discrepancies among these indices are caused by factors such as the physical variables, geographical domain, data preprocessing, and statistical analysis technique used in their definition. Indeed, an accurate and objective identification of tropical intraseasonal oscillations, including the MJO and MISO, remains a challenging open problem (Kiladis et al, 2014).

In this work, we introduce a new MISO index based on the Nonlinear Laplacian Spectral Analysis (NLSA) technique (Giannakis and Majda, 2012b,a), and use that index to explore the possibilities of improving the real-time monitoring and prediction of MISO. NLSA is a nonlinear data analysis technique that combines ideas from delay embeddings of dynamical systems (Packard et al, 1980; Sauer et al, 1991) and kernel methods for harmonic analysis and machine learning (Belkin and Niyogi, 2003; Coifman and Lafon, 2006a) to extract spatiotemporal modes of variability from high-dimensional timeseries. These modes are computed using the eigenfunctions of a discrete Laplace-Beltrami operator—an operator which can be thought of as a local analog of the temporal covariance matrix employed in EOF and EEOF techniques, but adapted to the nonlinear geometry of data generated by complex dynamical systems. A key advantage of NLSA over classical covariance-based approaches is that it is able to extract modes spanning multiple timescales without requiring ad hoc preprocessing (e.g., seasonal partitioning or bandpass filtering) of the input data. Thus, the method is well-suited for objectively identifying MISO patterns in noisy precipitation data.

NLSA has previously been employed to extract families of modes of variability from equatorially averaged (Giannakis et al, 2012; Tung et al, 2014) and two-dimensional (2D) (Székely et al, 2016a,b) brightness temperature (T_b) data spanning interannual to diurnal timescales without prefiltering the input data (hereafter, we collectively refer to these references as GMST). These mode families include representations of the MJO and BSISO with higher temporal coherence (Székely et al, 2016b) and stronger discriminating power between eastward and poleward propagation (Székely et al, 2016a) than patterns extracted through EOF and EEOF approaches. The MJO and BSISO modes from NLSA have also been used in low-order forecast models based on nonlinear stochastic oscillators (Chen et al, 2014; Chen and Majda, 2015) and ensembles of analogs (Alexander et al, 2016) with useful predictive skill extending out to 40–50 day leads.

Here, we demonstrate that NLSA yields physically meaningful and highly predictable MISO modes when applied to unprocessed daily precipitation data from Global Precipitation Climatology Project (GPCP; Huffman et al, 2001) over the south Asian monsoon region. We find that compared to the conventional EEOF MISO indices, the NLSA-based MISO indices have higher memory and predictability. Further, we demonstrate the skill of the NLSA based MISO modes in real time prediction of the MISO using the NCEP Climate Forecast System version 2 (CFSv2; Saha et al, 2014) model hindcast data.

The plan of this paper is as follows. An overview of the datasets and NLSA methodologies used in this study are presented in sections 2 and 3, respectively. Section 4 presents the hierarchy of modes extracted by NLSA applied on spatiotemporal data, focusing on the temporal and spatial properties of the MISO modes. A comparison of the NLSA modes with the conventional EEOF-based MISO modes is presented in section 5, and section 6 discusses real-time MISO forecasting with the NLSA modes. The paper ends in section 7 with a summary discussion and concluding remarks.

2 Dataset description

We apply NLSA on daily GPCP rainfall data (Huffman et al, 2001) over the Asian summer monsoon region (20°S–40°N, 30°E–160°E) for the period 1997–2014. The spatial resolution of this dataset is 1°X 1°, amounting to $n = 5500$ gridpoints for the Asian summer monsoon region. The number of temporal samples is $s = 6574$. Note that we analyze the raw GPCP data for the full year period without performing any pre-filtering. To create the MISO phase composites, we use daily averaged outgoing longwave radiation (OLR) data from the NOAA advanced very high resolution radiometer (Liebmann, 1996) and lower level (850 hPa) wind anomalies obtained from the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay et al, 1996) for the period 1998–2013. The horizontal resolution of these two datasets are $2.5^\circ \times 2.5^\circ$.

As hindcast data, we use precipitation fields from 45 day operational integrations of NCEP CFSv2. The CFSv2 is a fully coupled ocean-atmosphere-land model, with modified physics and higher resolution compared to its earlier version (CFSv1; Saha et al (2014)). In addition, this model has been identified as the base model for the Monsoon Mission project of the Government of India. Earlier studies have reported that the CFSv2 is able to adequately simulates the mean Indian summer monsoon features (George et al, 2016; Chattopadhyay et al, 2015; Ramu et al, 2016) and the subseasonal variability associated with it (Sabeerali et al, 2013; Goswami et al, 2014). For extended range MISO forecasts, 45 day lead time model integrations were performed at IITM using the CFSv2 coupled model (Sahai et al, 2013; Abhilash et al, 2014). In each monsoon season, 25 simulations with different initial conditions were performed starting from May 31 to September 28 at 5 day intervals and each initial condition runs involve 40 ensemble members (a total of 25×40 runs for each year). For verifying the NLSA MISO forecasts, we use the ensemble mean of each initial condition run.

3 NLSA methodology

In what follows, we first summarize the NLSA methodology to compute the Laplace-Beltrami eigenfunctions and associated spatiotemporal patterns from the training (GPCP) data (section 3.1), and then describe the procedure to compute the eigenfunctions from previously unseen forecast data using out-of-sample extension techniques (section 3.2). More detailed discussions on NLSA and the out-of-sample extension procedure can be found in GM, and in Zhao and Giannakis (2014) and Comeau et al (2016), respectively.

3.1 Overview of NLSA algorithms

Let $x(t_i)$ be an n-dimensional vector of gridded precipitation values over the South Asia monsoon region at time $t_i = (i-1)\,\delta t$. Here, δt represents the 21 day sampling interval of the data, and i is an integer ranging from 1 to s so that the start date of the training dataset (January 1, 1997) is assigned the reference time $t_1 = 0$. Using the data $\{x(t_1), \dots, x(t_s)\}$, NLSA computes a hierarchy Laplace-Beltrami eigenfunctions $\phi_0(t_i), \phi_1(t_i), \dots, \phi_l(t_i)$ (which are temporal patterns that can be thought of as nonlinear analogs of the principal components (PCs) in EEOF analysis), and a corresponding collection of reconstructed spatiotemporal patterns $\{x^{(0)}(t_i), x^{(1)}(t_i), \dots, x^{(l)}(t_i)\}$ such that $\sum_{k=0}^{l} x^{(k)}(t_i)$ approximates the input signal $x(t_i)$. The NLSA pipeline consists of three main steps, as follows.

The first step, which is in common with EEOF analysis, is to construct a higher-dimensional, time-lag embedded dataset using Takens' method of delays. Fixing a positive integer parameter q (the number of lags), each snapshot $x(t_i)$ with $i \geq q$ is mapped to the lagged sequence $X(t_i) = (x(t_i), x(t_{i-1}), \dots, x(t_{i-q+1}))$. Note that the dimension of the vectors $X(t_i)$ is $N = nq$, and that after time-lagged embedding $n - q + 1$ samples are available for analysis. Following GMST, we set $q = 64$; this choice corresponds to an intraseasonal embedding window of length $q\,\delta = 64$ days. We verified our results with different embedding windows by computing eigenfunctions for $q = 34$, 48, and 90. Eigenfunctions computed using $q = 34$ and 40 exhibit mixing of different timescales, whereas those computed using $q = 90$ are in good agreement with our nominal choice, $q = 64$.

The next step in NLSA is to compute the kernel matrix K with entries $K_{ij} = K(X(t_i), X(t_j))$ given by

$$K(X(t_i), X(t_j)) = \exp\left(-\frac{\|X(t_i) - X(t_j)\|^2}{\epsilon\xi(t_i)\xi(t_j)}\right).$$

In the above, ϵ is a positive kernel bandwidth parameter, and the quantities $\xi(t_i)$ are "phase space velocities" measuring the local time-tendency of the data through $\xi(t_i) = \|X(t_i) - X(t_{i-1})\|$. The kernel values $K(X(t_i), X(t_j))$ provide a nonlinear measure of similarity between samples $X(t_i)$ and $X(t_j)$ with $K(X(t_i), X(t_j))$ close to 1 or 0 meaning that $X(t_i)$ and $X(t_j)$ are highly similar or highly dissimilar, respectively. Due to the exponential decay of the kernel, this measure of similarity is local in the sense that for a fixed reference point $X(t_i)$ sufficiently small ϵ, $K(X(t_i), X(t_j))$ is appreciable only in a small neighborhood of $X(t_i)$ where the local geometry of the data (viewed as a cloud of points in $\mathbb{R}^N$) is approximately linear. Intuitively, operators constructed from $K(X(t_i), X(t_j))$ smoothly interpolate between such local linear patches that together make up the global nonlinear geometry of the data. This approach has been widely used in machine learning algorithms (e.g., Belkin and Niyogi, 2003; Coifman and Lafon, 2006a), but the novelty of the NLSA kernel lies in the fact that $K(X(t_i), X(t_j))$ depends on the dynamical system generating the data due to both time lagged embedding (since changing the dynamics would change the snapshot sequences present in the time-lagged vectors) and the local phase space velocities $\xi(t_i)$. Time-lagged embedding is crucial for obtaining timescale separation in the eigenfunctions ϕ_i, and the phase space velocities enhances the ability of the algorithm to capture intermittent rapid transitions. Since ther calculation of $\xi(t_i)$ "uses up" the initial lagged-embedded sample $X(t_q)$, the kernel matrix K has size $S \times S$ where $S = s - q$. Due to the exponential decay of the kernel, the entries of K below a given threshold can be set to zero leading to a sparse matrix. Here, following GMST, we work with the bandwidth parameter value $\epsilon = 2$, and retain the largest 650 nonzero entries in each row of K (which corresponds to $\simeq 10\%$ of the total number of samples). To verify the sensitivity of our results to the value of ϵ, we repeated our analysis with different ϵ values. We found that choosing ϵ in the interval 2–5 does not make qualitative changes in the results.

Having computed the sparse kernel matrix K, NLSA proceeds by normalizing it to obtain a Markov (row-stochastic) matrix P using the normalization procedure introduced in the diffusion maps algorithm Coifman and Lafon (2006a). Specifically, the matrix elements P_{ij} are computed through the sequence of operations

$$q_i = \sum_{j=1}^{S} K_{ij}, \quad K'_{ij} = \frac{K_{ij}}{q_i q_j}, \quad d_i = \sum_{j=1}^{S} K'_{ij}, \quad P_{ij} = \frac{K_{ij}}{d_i}, \tag{1}$$

and it follows immediately that $\sum_{j=1}^{S} P_{ij} = 1$. The NLSA temporal patterns $\phi_k(t_i)$ are then determined by the eigenvectors of the Laplacian matrix $L = I - P$. That is, we solve the sparse eigenvalue problem

$$L\boldsymbol{\phi}_k = \lambda_k \boldsymbol{\phi}_k, \quad \boldsymbol{\phi}_k = (\phi_{1k}, \phi_{2k}, \ldots, \phi_{Sk})^\top,$$

and set $\phi_k(t_i) = \phi_{ik}$. It follows from standard properties of ergodic Markov chains that the eigenvalues λ_i admit the ordering $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_S$. Moreover, the eigenfunctions can be chosen to be orthonormal with respect to the weighted inner product $\langle \phi_j, \phi_k \rangle := \sum_{i=1}^{S} \mu_i \phi_{ij} \phi_{ik} = \delta_{ik}$, where μ_i are positive weights with $\sum_{i=1}^{S} \mu_i = 1$, given by the entries of the (unique) left eigenvector of P with corresponding eigenvalue 1. Conceptually, these Laplace-Beltrami eigenfunctions can be treated as nonlinear analogs of the principle components (PCs) in (E)EOF analysis, and can be used e.g., to create spatiotemporal reconstructions and phase composites. In particular, an exact recovery of the input signal is possible using all S eigenfunctions, although of course in practice one works with the leading few eigenfunctions.

In a suitable limit of large data ($S \to \infty$ and $\epsilon \to 0$) L converges to the Laplace-Beltrami operator on the manifold sampled by the lag-embedded data $X(t_i)$ for a Riemannian geometry that depends on the kernel K (Coifman and Lafon, 2006a). That is, L generates a diffusion process (random walk) on the nonlinear data manifold sampled by the data, which is statistically isotropic (i.e., the random walker takes steps with equal probability in every direction), but the notion of isotropy is with respect to a modified geometry that depends on the choice of kernel. The eigenfunctions ϕ_k correspond to preferred classes of functions that remain statistically invariant (up to an eigenvalue-dependent scaling) under that diffusion process. Moreover, the corresponding eigenvalues λ_k can be interpreted as a measure of roughness (called Dirichlet energy) of the ϕ_k viewed as functions on the data manifold, much like the Laplacian eigenvalue k^2 corresponding to a Fourier function $e^{ik\theta}$ on a periodic domain measures roughness associated with the wavenumber k.

It is well known that for appropriate choices of kernel, eigenfunctions of diffusion operators on manifolds can reveal important relationships in complex data (Belkin and Niyogi, 2003; Coifman and Lafon, 2006a). In particular, a popular approach in harmonic analysis and machine learning is to use the ϕ_i as nonlinear dimension reduction maps, sending the n-dimensional snapshots $x(t_i)$ to the l-dimensional vectors $(\phi_1(t_i), \phi_2(t_j), \ldots, \phi_l(t_j))$ where $l \ll n$. Ordering the eigenfunctions in order of increasing corresponding eigenvalues, leads to the least rough l-dimensional dimension reduction map in the kernel dependent geometry. For the class of kernels in time-lagged embedding space used in NLSA it can be shown that as the number of lags q increases, the leading eigenfunctions become increasingly sensitive towards the subset of dynamical degrees of freedom with large Lyapunov stability, filtering out the unstable degrees of freedom. Quasi-periodic patterns, such as intraseasonal oscillations, are likely to be well represented by stable degrees of freedom, making NLSA a suitable technique for their detection in high-dimensional complex data (Berry et al, 2013). Indeed in section 4 ahead, we will see that NLSA recovers MISO from precipitation data through a doubly-degenerate pair of eigenfunctions with more realistic corresponding spatial features and higher predictability than the corresponding EEOF modes.

3.2 Out-of-sample extension

In real-time monitoring and forecasting applications it is important to be able to compute the values of NLSA eigenfunctions for previously unseen samples. Specifically, suppose that we are given a lagged sequence $Y = (y(t'_i), y(t'_{i-1}), \dots, y(t'_{i-q+1}))$ of precipitation snapshots, where t'_i represents time at forecast verification and the $y(t'_j)$ are n-dimensional vectors storing precipitation data over the South Asian monsoon region in the same manner as the training data $x(t_i)$. In the application of interest here, Y will be constructed from CFSv2 output, or a concatenated sequence to CFSv2 output and GPCP data (to provide precipitation snapshots at times prior to CFSv2 initialization). To that end, we employ so-called Nyström out-of-sample extension techniques, originally introduced in the 1930s for interpolation of solutions of integral eigenvalue problems and adopted to the setting of kernel methods on manifolds by Coifman and Lafon (2006b).

Consider now the eigenfunction time series $\phi_k(t_i)$ with corresponding eigenvalue λ_k. Each value $\phi_k(t_i)$ of that time series can be naturally associated with the training sample $X(t_i)$ in lagged embedding space $\mathbb{R}^N$; i.e., we have the mapping $X(t_i) \mapsto \phi_k(t_i)$. In the Nyström method, that mapping is extended to arbitrary points $Y \in \mathbb{R}^N$ subject to a consistency requirement on the training data. That is, given $Y \in \mathbb{R}^N$, we compute a quantity $\hat{\phi}_k(Y)$ such that if Y happens to be equal to some $X(t_i)$ in the training dataset, then $\hat{\phi}_k(Y) = \phi_k(t_i)$.

The procedure to compute $\hat{\phi}_k(Y)$ has its foundations in the theory for function interpolation in reproducing kernel Hilbert spaces, and follows closely the diffusion maps construction described in section 3.1. Specifically, we first compute the pairwise kernel values between Y and the samples in the training dataset, $\hat{K}_j(Y) = K(Y, X(t_j))$, and then perform the diffusion maps normalization procedure,

$$\hat{K}_j(Y) = \frac{\hat{K}_j(Y)}{q_j}, \quad \hat{d}(Y) = \sum_{j=1}^{S} \hat{K}'_j(Y), \quad \hat{P}_j(Y) = \frac{K'_j(Y)}{\hat{d}(Y)},$$

where q_j is determined from (1). Note that $\sum_{j=1}^{S} P'_j(Y) = 1$, and if $Y = X(t_i)$ then $\hat{P}_j(Y) = P_{ij}$. Introducing the row vector $\hat{P}(Y) = (\hat{P}_1(Y), \ldots, \hat{P}_S(Y))$, the out-of-sample extension of ϕ_k is then given by

$$\hat{\phi}_k(Y) = \frac{1}{1-\lambda_k} \hat{P}(Y) \boldsymbol{\phi}_k. \tag{2}$$

The consistency condition on the training data follows from the facts that $\hat{P}(Y)$ is equal to the i-th row of the matrix P from (1) when $Y = X(t_i)$, and that $\boldsymbol{\phi}_k$ is an eigenvector of P corresponding to the eigenvalue $1 - \lambda_k$.

It is evident from (2) that Nyström extension becomes ill-conditioned when $1 - \lambda_k \approx 0$, and this is consistent with our interpretation of the eigenvalues as measures of eigenfunction roughness (see section 3.1). That is, eigenfunctions with low roughness have $\lambda_k \ll 1$, and intuitively such eigenfunctions should be robustly extendable to previously unseen points Y, but eigenfunctions with large roughness have $\lambda_k \approx 1$ and cannot be robustly extended.

4 Hierarchy of spatiotemporal modes revealed by NLSA

Applying the NLSA algorithm to the raw GPCP rainfall data as described in section 3.1, yields a hierarchy of Laplace-Beltrami eigenfunctions capturing coherent patterns of rainfall variability. In order to identify the modes northward propagating boreal summer MISO, we examine the frequency spectra of the the eigenfunction time series, as well as spatial reconstructions and composites. Following the convention of section 3.1, we order the eigenfunctions in order of increasing eigenvalue; the latter are displayed in Figure 1. In what follows, we focus on the leading six eigenfunctions, whose time series and power spectral densities are displayed in Figure 2.

4.1 Periodic modes

As is evident by their strong spectral peak at the frequency 1/yr, the first two eigenfunctions, ϕ_1 and ϕ_2 (Figure 2a,b) represent the annual cycle. The timeseries of these eigenfunctions have the structure of a periodic wave (which is nearly sinusoidal in the case of ϕ_1, whereas ϕ_2 also exhibits higher-frequency overtones). Eigenfunctions ϕ_1 and ϕ_2 also exhibit discernible semiannual and triennial spectral peaks, respectively. Modes ϕ_3 and ϕ_4 (Figure 2c,d) have strong spectral peaks at the frequency 2/yr representing semiannual variability.

In spatiotemporal reconstructions (not shown here for brevity), mode ϕ_1 shows a seasonal (winter to summer) shift of precipitation anomalies between the two hemispheres with strong precipitation anomalies in winter and summer months and relatively weak precipitation anomalies in other months. Moreover, the precipitation anomalies associated with this mode are stronger over land than over the ocean. On the other hand, the annual mode ϕ_2 shows significant precipitation anomalies over oceanic region compared to land region and it shows strong anomalies during spring and autumn season. The semiannual modes ϕ_3 and ϕ_4 show significant precipitation anomalies over the equatorial Indian Ocean, and these anomalies appear twice a year in association with the ITCZ movement. Precipitation anomalies are initially seen over the the equatorial Indian Ocean, and then propagates poleward towards the Indian subcontinent.

4.2 MISO modes

Eigenfunctions ϕ_5 and ϕ_6 represent the dominant MISO activity over the south Asian monsoon region. These eigenfunctions form a doubly-degenerate pair (Figure 1) of 90° out-of-phase amplitude-modulated waves with a spectral peak in the 1/(30 day)–1/(60 day) frequency band (Figure 2e,f). Moreover, they exhibit strong seasonality with the bulk of their activity taking place during the boreal summer months. The temporal evolution of eigenfunctions ϕ_5 and ϕ_6 is shown in more detail in Figure 3 for a two-year reference period, where the 90° phase difference and seasonality are clearly evident. The detailed view in Figure 3 also illustrates the absence of high-frequency noise from the ϕ_5 and ϕ_6 time series. Another important feature of eigenfunctions ϕ_5 and ϕ_6 is their non-Gaussian statistics. As shown in Figure 4, the probability density functions (PDFs) of the ϕ_5 and ϕ_6 timeseries have fat tails when computed from the year-round data, and their kurtosis values ($\kappa = 7.6$ and 3.8, respectively) are significantly higher than the $\kappa = 3$ kurtosis of the Gaussian distribution. Computed over JJAS, the PDFs of ϕ_5 and ϕ_6 become platykurtic (i.e., have lighter tails than a Gaussian distribution) with $\kappa = 1.5$ and 1.4, respectively. The non-Gaussianity of the NLSA eigenfunction PDFs contribute to their higher discriminating power compared to classical linear approaches (Székely et al, 2016b).

In the spatial domain, NLSA MISO modes display the characteristic pattern of northeastward propagating anomalies associated with the MISO. This pattern is illustrated in Figure 5 with a spatiotemporal reconstruction of the 2004 monsoon season. The wet phase of MISO seen at the third week of June 2004 (Figure 5c) over the western/central tropical Indian Ocean propagates in the northeastward direction in the following days and reaches the foothills of Himalayas by the third week of July 2004 (Figure 5f). Following this event, a new wet phase of MISO initiates over the western/central tropical Indian Ocean in the last week of July 2004, and reaches the Himalayan foothills by the end of August 2004. The cycle continues with the initiation of convection over the central equatorial Indian Ocean in first week of September 2004 and propagates northeastward.

Together, eigenfunctions ϕ_5 and ϕ_6 delineate the full life cycle of the northward propagating boreal summer convection band, and can be used to determine the phase and amplitude of the poleward-propagating rainfall anomalies associated with the MISO. Hereafter, we refer to eigenfunctions ϕ_5 and ϕ_6 as MISO1 and MISO2, respectively. Following previous works (Kikuchi et al, 2012; Székely et al, 2016a,b), we also define the NLSA MISO amplitude at time t via

$$r(t) = \sqrt{\frac{\mathrm{MISO1}(t)^2}{\sigma_1^2} + \frac{\mathrm{MISO2}(t)^2}{\sigma_2^2}}, \tag{3}$$

where $\sigma_i = 1.03$ are the standard deviations of the $\mathrm{MISO}_i(t)$ time series.

4.3 Real-time monitoring via NLSA MISO indices

The daily evolution of the MISO can be monitored from the two-dimensional (2D) phase space diagram constructed from the NLSA MISO indices, shown in Figure 6 for three drought years. Note that flood years are not present in the 1998–2003 analysis period. In Figure 6, the 2D phase space diagram is plotted for the extreme rainfall years where the All India summer monsoon rainfall (AISMR) index exceeds ± 1 of its standard deviation (this corresponds to a $\pm 10\%$ fractional rainfall anomalies). In the period 1998–2013, there are only three years where AISMR is less than -1 (the drought years 2002, 2004, and 2009); the rest are normal rainfall years with $|\mathrm{AISMR}| < 1$. A list of all drought and flood years for the period 1871–2015 can be found in the IITM website (http://www.tropmet.res.in/ kolli/mol/Monsoon/Historical/air.html).

Figure 6 shows the strong MISO activity during June and July months of 2002 and the subdued MISO activity during the ensuing August and September months. In contrast, in spite of it being a drought year, MISO activity during 2004 is persistently strong throughout the boreal summer. In 2009, MISO activity is weak during the late monsoon season. Such day to day evolution of MISO can be used for real-time monitoring of monsoon intraseasonal rainfall variability (Abhilash et al, 2014). It is evident from 6 that MISO activity does not always begin in phase 1 and end in phase 8; a behavior which has also been observed in the case of the MJO (Straub, 2013; Stachnik et al, 2015; Székely et al, 2016b). To illustrate the relationship between the NLSA MISO indices plotted in Figure 6 with actual rainfall data, in Figure 7 we compare the MISO2 time series against the corresponding bandpass-filtered (25–90 d) and unfiltered JJAS rainfall anomalies over the central Indian domain. Evidently, in all the three drought years the NLSA index is able to capture the active and break phases associated with the Indian summer monsoon. NLSA mode MISO1 also correlates well with the active and break phases in those years, but because this mode has a 90deg phase difference with MISO2, the correlation exhibits a time lag (not shown). Following the familiar approach from RMM (Wheeler and Hendon, 2004) and EEOF (Suhas et al, 2013; Abhilash et al, 2013) indices, we divide the 2D phase space into eight phases, and compute phase composites by conditional averaging in each phase subject to the requirement that the instantaneous MISO amplitude $r(t)$ from (3) is greater than 1. In what follows, we use this threshold to identify significant MISO events.

The resulting composites for bandpass-filtered OLR and 850 hPa wind anomalies are shown in Figure 8. The composites indicate that an anticlockwise rotation from the phase 1 through phase 8 in the 2D phase space represents the poleward propagation of the MISO. In particular, phase 1 represents the formation of enhanced convection anomalies (negative OLR anomalies) over the Indian Ocean, phases 2 and 3 (Figure 8b,c) the subsequent movement of convection towards the Indian subcontinent, phases 4–6 (Figure 8d,e,f) the propagation of enhanced convection over the subcontinent and Bay of Bengal, and phases 7 and 8 (Figure 8g,h) the breaking over the subcontinent. The composites for bandpass-filtered rainfall (Figure 8i–p) also exhibit consistent propagating MISO patterns. The realistic northward and eastward propagation characteristics of the NLSA MISO modes can also be seen in phase-latitude and phase-longitude plots in Figure 9. There, the phase-latitude diagrams of both OLR and precipitation field show a clear northward propagation of the convective anomalies from the equatorial Indian Ocean (5°S) into the northern latitudes (around 25°N) and a southward propagation from 5°S into the southern ocean (Figure 9a,b). Moreover, the longitude-phase diagram of OLR and precipitation anomalies averaged over the equatorial belt shows a clear eastward propagation of convective anomalies from the western equatorial Indian Ocean to the tropical western Pacific (Figure 9c,d).

A number of studies argue that Rossby wave emanation from eastward-propagating convective anomalies is responsible for the poleward propagation of the MISO (Wang and Xie, 1997; Kemball-Cook and Wang, 2001; Annamalai and Sperber, 2005; Ajayamohan et al, 2010). Therefore, realistic simulation of this eastward propagating convective anomalies in a model is thought to be essential for the realistic northward propagation of the MISO (Sabeerali et al, 2013). The phase relationship between convection and circulation in Figure 8 shows evidence of the Rossby wave emanation. In particular, the wind pattern in phases 3 and 4 displays a classical Matsuno-Gill Kelvin-Rossby wave response (Matsuno, 1966; Gill, 1980) with easterly anomalies along the equatorial western Pacific and two cyclonic gyres on either side of the equatorial Indian Ocean (Figure 8). This wind pattern exhibits an asymmetry about the equator, indicating the role of Rossby wave propagation in modulating MISO's poleward propagation.

This Rossby wave propagation brings out the importance of the western Pacific and maritime continents in determining the structure of MISO rainfall. Another important feature of the MISO is the quadrupole-like convection pattern over the Asian monsoon region in which positive (negative) anomalies persist as a tilted band extending from the Indian subcontinent to the western Pacific and negative (positive) anomalies exist to the south of this pattern over the Indian Ocean and western Pacific (Annamalai and Sperber, 2005; Pillai and Sahai, 2015). This structure is clearly captured in the OLR composites in Figure 8, especially in phases 5 and 6 where the amplitude of convection over the western Pacific is strong and extends beyond the date line.

5 Comparison with EEOF-based MISO indices

To place our results in context, we compare the NLSA-based MISO modes with the EEOF-based modes of Suhas et al (2013). As stated in section 1, the EEOF-based MISO indices are currently used for real-time monitoring of MISO at IITM, and one of the objectives of our study is to explore ways to improve the skill of these real-time forecasts.

We have computed the EEOF MISO modes as described in Suhas et al (2013) using same daily GPCP rainfall dataset described in section 2. Specifically, we perform EEOF analysis on longitudinally averaged (over 60.5°E–95.5°E) GPCP rainfall data and the latitudes 12.5°S–30.5°N, after removing the climatological mean and first three harmonics of the seasonal cycle. We use 15 EEOF lags, sampled once per day. At a given time t, we define the MISO indices $\mathrm{MISO1_E}(t)$ and $\mathrm{MISO2_E}(t)$ from EEOF PCs 1 and 2 (ordered in order of decreasing explained variance), and also define the EEOF-based MISO amplitude index (cf. (3))

$$r_{\mathrm{E}}(t) = \sqrt{\frac{\mathrm{MISO1_E}(t)^2}{\sigma_{1\mathrm{E}}^2} + \frac{\mathrm{MISO2_E}(t)^2}{\sigma_{2\mathrm{E}}^2}}.$$

where $\sigma_{1\mathrm{E}} = 39.2$ and $\sigma_{2\mathrm{E}} = 33.5$ are the standard deviations of the $\mathrm{MISO1_E}(t)$ and $\mathrm{MISO2_E}(t)$ time series, respectively. Similarly to section 4.3, we use $r_{\mathrm{E}}(t) \geq 1$ as a threshold for significant MISO events based on EEOFs.

Figure 10 displays the joint temporal evolution of the MISO1 and MISO2 indices and the corresponding amplitudes obtained via NLSA and EEOF analysis for the 1998–2013 JJAS period. There, it can be seen that the NLSA and EEOF time series are in moderately good qualitative agreement, although the temporal evolution of the NLSA modes is markedly more coherent. Moreover, as shown in the amplitude plots in Figure 10(c), the significant MISO events detected via NLSA tend to be more persistent. Examined in terms of their statistics (Figure 4), the EEOF-based MISO indices are more Gaussian than their NLSA counterparts.

Next, we compare the NLSA and EEOF MISO indices in terms of their power spectral densities (Figure 11) and temporal correlation structure (Figure 12). As shown in Figure 11, the indices obtained via either of the two methods capture the central peak between 1/30 and 1/60 d^{-1} observed in the raw rainfall anomalies, and are also effective in removing the high-frequency content present in rainfall data. In general, the spectra of the NLSA indices have smaller high-frequency power than the EEOF spectra, which is consistent with the remark made earlier that the time evolution of the former is more coherent than the latter. In Figure 12a, the autocorrelation functions of the of NLSA and EEOF MISO modes are compared with that of the observed bandpass filtered (25–90 d) rainfall anomalies. In general, the autocorrelation functions of the NLSA modes are closer to observations than the EEOF modes, especially at longer (± 20 d) lags. In Figure 12b, the cross correlation function between the two NLSA MISO modes, which are uncorrelated at lag zero by orthogonality of the eigenfunctions, exhibits a near-sinusoidal behavior with a reemergence of correlations ($\simeq 0.95$ values) at ± 11 day lags. This behavior is indicative of a coherent, and hence predictable, harmonic oscillator. In the case of the EEOF modes, the cross-correlation function is characterized by a marked amplitude decay, with the minima/maxima occurring earlier (at ± 7 d) and attaining smaller absolute values ($\simeq 0.7$). Overall, these results indicate that the NLSA indices retain their memory for a longer period (Figure 12), while capturing the dominant spectral peak of MISO efficiently (Figure 11).

We now turn attention to spatial composites. Figure 13 shows similar OLR and wind composites to the NLSA-based composites in Figure 8, constructed via the EEOF MISO indices. These composites clearly exhibit the typical lifecycle of the MISO, including its northeastward propagation and zonal and meridional structure, but certain features are not as well represented as in NLSA. In particular, the EEOF-based composites have weaker loadings of convection anomalies over the Maritime continent, a less coherent quadrupole structure, and a less developed tilted zonal convection band. These features are also evident in rainfall composites (Figure 13). To further assess the skill of NLSA and EEOF analysis in capturing the regional heat sources we examine spatial maps (Figure 14) showing the percentage of fractional variance of bandpass-filtered rainfall anomalies explained by the spatial composites from the two methods. Consistent with the spatial composites in Figure 8, NLSA yields a realistic variance pattern and captures the regional centers of MISO activity. Compared to the EEOF-based variance maps, NLSA explains larger fractional variance over important MISO regions including the western Pacific, Western Ghats, the adjoining Arabian Sea. Note that capturing the variability over Indo-West Pacific region is particularly important in determining the propagation characteristics of MISO (e.g. Pillai and Sahai, 2015). In summary, the results in Figures 8, 13, and 14 indicate that NLSA outperforms EEOF analysis in capturing variability over the regional heat sources associated with the MISO.

6 Application to extended-range MISO prediction

In this section, we demonstrate the skill of the NLSA MISO modes identified in section 4 in extended-range MISO prediction. In particular, we use the CFSv2 operational data described in section 2 to create hindcasts of the NLSA MISO1 and MISO2 indices, and assess the skill of these hindcasts by comparing the predicted values of the indices against the true values computed from GPCP data.

Recall from section 3.2 that the Laplace-Beltrami eigenfunctions (including the NLSA MISO indices) can be evaluated for an arbitrary lagged sequence Y using out-of-sample extension techniques. In the scenario of interest here, Y has the structure $Y_{\text{pred}}(t'_i) = (y(t'_i), y(t'_{i-1}), \ldots, y(t'_{i-q+1})$, where t'_i is the forecast verification time for the i-th hindcast experiment under study, and $y(t'_{i-j})$ is the vector predicted rainfall values over the Asian summer monsoon region at time t'_{i-j}, $j \in \{0, 1, \ldots, q-1\}$. When t'_{i-j} is smaller than the forecast initialization time, τ_i, we set $y(t'_{i-j})$ equal to the historically observed GPCP rainfall $x(t'_{i-j})$. This takes into account that the fact that evaluation of the NLSA MISO indices requires information from a time interval containing q rainfall snapshots, and if $t'_{i-j} \leq \tau_i$, this interval includes times prior to CFSv2 initialization time. The predicted value $\hat{\phi}_k(Y_{\text{pred}})$ for the MISO indices is then determined via Nyström extension using (2). We also use (2) to compute the true values for the monsoon indices, replacing $Y_{\text{pred}}(t'_i)$ with the lagged vector $T_{\text{true}}(t'_i) = (x(t'_i), x(t'_{i-1}), \ldots, x(t'_{i-q+1}))$ constructed from the GPCP data.

We have performed such hindcast experiments using CFSv2 runs for the period 2009-2010, initialized at five-day intervals from May 31 to September 28 of each year. Figure 15 shows the corresponding pattern correlation (PC) and root mean square error (RMSE) scores computed for lead times ranging from 0 to 45 days. The PC scores for both MISO1 and MISO2 (Figure 15a) exhibit an initial period of persistence to $\gtrsim 0.9$ values for up to $\simeq 16$ day leads. The PC scores then begin a more rapid decay, but MISO1 (MISO2) remain greater than 0.8 for $\simeq 22$ ($\simeq 25$) days. The RMSE scores (Figure 15b) show a near- linear increase with lead time, and remain less than one for up to $\simeq 22$ day. These results indicate that in CFSv2, an accurate prediction of MISO for a minimum of 22 days can be achieved using NLSA based MISO indices.

To further assess the skill of NLSA-CFSv2 for real-time MISO forecasts, we examine in Figure 16 phase space trajectories of the MISO1 and MISO2 indices for four representative hindcast experiments. The cases shown in Figure 16a,b,e,f are examples of successful forecasts. In Figure 16a, the truth signal shows a MISO event that starts at phase 4 in May 31, 2009 and subsequently moves northward, decaying at phase 8 in July 2, 2009. The predicted trajectory successfully tracks the truth for up to 32 days, and then slightly deviate from the truth (Figure 16e). Similarly, in Figure 16b, the observed MISO becomes significant in September 2, 2010 in phase 2 and then follows its northward propagation until it reaches phase 7 in the end of September. The predicted trajectory realistically captures the truth until the middle of September 2010 and then a moderately small deviation can be seen from the truth (Figure 16f). On the other hand, the examples in Figure 16c,d,g,h are unsuccessful forecasts. In these two cases, the forecasted MISO trajectory is reasonably good for up to 10 day leads, and then fails to track the truth trajectory. It is found that out of the 50 test cases analyzed, 78% are comparably successful to the cases in Figure 16a,b,e,f and 22% are comparably unsuccessful as the cases in Figure 16c,d,g,h. Overall, the results in Figure 16 illustrate that the forecast skill can have large spread depending on the initial data, though on average the NLSA MISO modes generated using CFSv2 runs are useful for at least 22 day leads.

As a comparison with EEOF-based indices, we note that Suhas et al (2013) have estimated the MISO prediction skill using CFSv1 (an earlier version of CFSv2), and found that MISO1 (MISO2) forecasts have skill for up to 13 (9) days. In their study, they used a lag of 15 days to resolve the northward propagating MISO. Using the same EEOF-based indices, (Abhilash et al, 2014) have reported that the MISO1 (MISO2) prediction skill of CFSV2 is 17 (14) days. A difference between these approaches and our NLSA-based approach is that we use a longer, 64 day, embedding window in conjunction with kernel eigenfunctions to resolve a coherent MISO evolution. As a result, our forecasts depend more strongly on past observations of nature as opposed to CFSv2 output, especially for short leads.

In general, a direct comparison between data-driven indices, including EEOFs and NLSA, is not very meaningful since all such indices have a degree of subjectivity (though NLSA attempts to minimize that subjectivity by avoiding pre-processing of the input data). Instead, a more appropriate comparison would involve using these indices to predict physical observable (e.g., average rainfall over a given region) of interest to forecasters and stakeholders. While such a comparison is beyond the scope of this work, the fact that the NLSA MISO modes realistically capture the structure of a number of key physical variables associated with the MISO (in particular, rainfall, convection (OLR), and circulation; see Figures 8, 9, and 14) is encouraging for future applications of NLSA in real-time monitoring and forecasting of aspects of MISO beyond indices.

7 Summary and conclusion

In this paper, we have developed improved indices for real-time monitoring and forecast verification of the MISO using NLSA; an objective data analysis technique for decomposition of high-dimensional time series. A key advantage of NLSA over classical eigen decomposition techniques is improved timescale separation and ability to detect intermittent patterns through the use of kernel methods in conjunction with Takens delay embeddings. Applied to GPCP rainfall data over the Asian summer monsoon region, NLSA yields a hierarchy of spatiotemporal modes spanning annual to subseasonal timescales. This hierarchy includes an in-quadrature pair of modes representing the full life cycle of MISO with improved temporal and spatial characteristics compared to the conventional EEOF-based MISO indices (Suhas et al, 2013). These features include improved temporal phase coherence while maintaining the ability to isolate the northeastward-propagation and 30–60-day MISO periodicity from the broad band rainfall data, as well as strong seasonal activity in the boreal summer (emerging without having to partition the input data). Moreover, the NLSA modes seems to better-resolve the tilted structure of MISO convention and its associated quadrupole circulation structure through phase composites, and also explain more fractional variance over the western Pacific and Western Ghats and adjoining Arabian Sea regions. This is a value added feature of MISO as the regional heat sources and Pacific variability has a significant influence over the monsoon variability.

Using NLSA based MISO indices, we also demonstrated the skill of NLSA in real-time prediction of MISO. The forecast skill of MISO is verified using hindcasts of CFSv2 extended range prediction runs. It is found that NLSA yields a significantly higher prediction skill than conventional MISO indices. The better skill of NLSA may be due to the ability of NLSA algorithm to capture the non linear features of MISO. These above mentioned merits of the NLSA over EEOF gives a scope for using this technique for the real-time monitoring and forecast verification of the MISO and can supplement to the existing EEOF based index used at Indian Institute of Tropical Meteorology, Pune, India. Real-time monitoring of the monsoon intraseasonal oscillation using a global coupled model assume significance in light of its applications in agriculture, construction and hydro-electric power sectors.

Acknowledgments

Authors gratefully acknowledge the financial support given by the Earth System Science Organization, Ministry of Earth Sciences (MoES), Government of India (Grant no.MM/SERP/NYU/2014/SSC-01/002) to conduct this research under Monsoon Mission. The Center for Prototype Climate Modelling (CPCM) is fully funded by the Abu Dhabi Government through New York University Abu Dhabi (NYUAD) Research Institute grant. The computations were carried out on the High Performance Computing resources at NYUAD. We thank IITM, Pune, India for providing NCEP CFSv2 model simulations for verification. We thank Dr. M. Rajeevan, secretary, MoES for constant encouragement. We acknowledge the support from Dr. S. A. Rao, Dr. A. K. Sahai and Dr. S. Abhilash for providing model data.

References

Abhilash S, Sahai AK, Pattnaik S, Goswami BN, Kumar A (2013) Extended range prediction of active-break spells of Indian summer monsoon rainfall using an ensemble prediction system in NCEP Climate Forecast System. Int J Climatol DOI 10.1002/joc.3668

Abhilash S, Sahai A, Borah N, Chattopadhyay R, Joseph S, Sharmila S, De S, Goswami B, Kumar A (2014) Prediction and monitoring of monsoon intraseasonal oscillations over indian monsoon region in an ensemble prediction system using cfsv2. Clim Dynam 42(9-10):2801–2815, DOI 10.1007/s00382-013-2045-9,2801-2815

Ajayamohan RS, Goswami BN (2003) Potential predictability of the Asian summer monsoon on monthly and seasonal time scales. Meteorol Atmos Phys 84:83–100, DOI 10.1007/s00703-002-0576-4

Ajayamohan RS, Annamalai H, Luo JJ, Hafner J, Yamagata T (2010) Poleward propagation of boreal summer intraseasonal oscillations in a coupled model: Role of internal processes. Clim Dynam DOI 10.1007/s00382-010-0839-6

Alexander RP, Zhao Z, Székely E, Giannakis D (2016) Kernel analog forecasting of tropical intraseasonal oscillations. J Atmos Sci In review

Annamalai H, Sperber KR (2005) Regional heat sources and the active and break phases of boreal summer intraseasonal (30-50 day) variability. J Atmos Sci 62:2726–2748

Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15:1373–1396, DOI 10.1162/089976603321780317

Berry T, Cressman R, Greguric Ferencek Z, Sauer T (2013) Time-scale separation from diffusion-mapped delay coordinates. SIAM J Appl Dyn Sys 12:618–649, DOI 10.1137/12088183X

Chatterjee P, Goswami BN (2004) Structure, genesis and scale selection of the tropical quasi-biweekly mode. Quart J Roy Meteorol Soc 130(599):1171–1194, DOI 10.1256/qj.03.133

Chattopadhyay R, Rao SA, Sabeerali C, George G, Rao DN, Dhakate A, Salunke K (2015) Large-scale teleconnection patterns of Indian summer monsoon as revealed by cfsv2 retrospective seasonal forecast runs. Int J Climatol DOI 10.1002/joc.4556

Chen N, Majda AJ (2015) Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. Math Climate Wea Forecast 1(1):1–20, DOI 10.1515/mcwf-2015-0001

Chen N, Majda AJ, Giannakis D (2014) Predicting the cloud patterns of the Madden-Julian Oscillation through a low-order nonlinear stochastic model. Geophys Res Lett 41(15):5612–5619, DOI 10.1002/2014GL060876

Coifman RR, Lafon S (2006a) Diffusion maps. Appl Comput Harmon Anal 21:5–30, DOI 10.1016/j.acha.2006.04.006

Coifman RR, Lafon S (2006b) Geometric harmonics: A novel tool for multiscale out-of-sample of empirical functions. Appl Comput Harmon Anal 21:31–52, DOI j.acha.2005.07.005

Comeau D, Zhao Z, Giannakis D, Majda AJ (2016) Data-driven prediction strategies for low-frequency patterns of north pacific climate variability. Climate Dyn DOI 10.1007/s00382-016-3177-5, in press

Gadgil S (2003) The Indian monsoon and it's variability. Annu Rev Earth Planet Sci 31:429–467

George G, Rao DN, Sabeerali C, Srivastava A, Rao SA (2016) Indian summer monsoon prediction and simulation in CFSv2 coupled model. Atmospheric Science Letters 17(1):57–64, DOI 10.1002/asl.599

Giannakis D, Majda AJ (2012a) Comparing low-frequency and intermittent variability in comprehensive climate model through nonlinear Laplacian spectral analysis. Geophys Res Lett 39:L10710, DOI 10.1029/2012GL051575

Giannakis D, Majda AJ (2012b) Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. Proc Nat Acad Sci USA

109:2222–2227, DOI 10.1073/pnas.1118984109

Giannakis D, Tung Ww, Majda AJ (2012) Hierarchical structure of the Madden-Julian oscillation in infrared brightness temperature revealed through nonlinear Laplacian spectral analysis. In: 2012 Conference on Intelligent Data Understanding (CIDU), Boulder, Colorado, pp 55–62, DOI 10.1109/CIDU.2012.6382201

Gill AE (1980) Some simple solutions for heat-induced tropical circulation. Quart J Roy Meteorol Soc 106:447–462

Goswami BB, Deshpande M, Mukhopadhyay P, Saha SK, Rao SA, Murthugudde R, Goswami BN (2014) Simulation of monsoon intraseasonal variability in ncep cfsv2 and its role on systematic bias. Clim Dynam 43(9):2725–2745, DOI 10.1007/s00382-014-2089-5

Goswami BN, Ajayamohan RS (2001) Intraseasonal oscillations and interannual variability of the Indian summer monsoon. J Climate 14:1180–1198, DOI 10.1175/1520-0442(2001)014⟨1180:IOAIVO⟩2.0.CO;2

Huffman GJ, Adler RF, Morrissey MM, Curtis S, Joyce R, McGavock B, Sisskind J (2001) Global precipitation at one degree daily resolution from multisatellite observations. J Hydrometeor 2:35–50

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, et al (1996) The ncep/ncar 40-year reanalysis project. Bulletin of the American meteorological Society 77(3):437–471

Kemball-Cook SR, Wang B (2001) Equatorial waves and air-sea interaction in the boreal summer intraseasonal oscillation. J Climate 14:2923–2942

Kikuchi K, Wang B, Kajikawa Y (2012) Bimodal representation of the tropical intraseasonal oscillation. Clim Dynam 38(9–10):1989–2000, DOI 10.1007/s00382-011-1159-1

Kiladis GN, Dias J, Straub KH, Wheeler MC, Tulich SN, Kikuchi K, Weickmann KM, Ventrice MJ (2014) A comparison of OLR and circulation-based indices for tracking the MJO. Mon Wea Rev 142:1697–1715, DOI 10.1175/mwr-d-13-00301.1

Krishnamurti TN, Bhalme HN (1976) Oscillations of monsoon system. Part I: Observational aspects. J Atmos Sci 45:1937–1954

Lee JY, Wang B, Wheeler MC, Fu X, Waliser DE, Kang IS (2012) Real-time multivariate indices for the boreal summer intraseasonal oscillation over the asian summer monsoon region. Climate Dynamics 40(1):493–509, DOI 10.1007/s00382-012-1544-4

Liebmann B (1996) Description of a complete (interpolated) outgoing longwave radiation dataset. Bull Amer Meteor Soc 77:1275–1277

Matsuno T (1966) Quasi-geostrophic motions in the equatorial area. J Meteorol Soc Japan 44(1):25–43

Packard NH, et al (1980) Geometry from a time series. Phys Rev Lett 45:712–716, DOI 10.1103/physrevlett.45.712

Pillai PA, Sahai AK (2015) Moisture dynamics of the northward and eastward propagating boreal summer intraseasonal oscillations: possible role of tropical indo-west pacific sst and circulation. Clim Dyn pp 1–16, DOI 10.1007/s00382-015-2904-7

Ramu D, Sabeerali C, Chattopadhyay R, Rao DN, George G, Dhakate A, Salunke K, Srivastava A, Rao SA (2016) Indian summer monsoon rainfall simulation and prediction skill in the CFSv2 coupled model: Impact of atmospheric horizontal resolution. Journal of Geophysical Research: Atmospheres DOI 10.1002/2015JD024629

Sabeerali C, Ramu Dandi A, Dhakate A, Salunke K, Mahapatra S, Rao SA (2013) Simulation of boreal summer intraseasonal oscillations in the latest CMIP5 coupled GCMs. J Geophys Res 118(10):4401–4420, DOI 10.1002/jgrd.50403

Sabeerali CT, Rao SA, Ajayamohan RS, Murtugudde R (2012) On the relationship between indian summer monsoon withdrawal and indo-pacific sst anomalies before and after 1976/1977 climate shift. Climate dynamics 39(3-4):841–859

Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou YT, Chuang Hy, Iredell M, et al (2014) The NCEP climate forecast system version

2. J Climate 27(6):2185–2208, DOI 10.1175/JCLI-D-12-00823.1

Sahai AK, Sharmila S, Abhilash S, Chattopadhyay R, Krishna NBRPM, Joseph S, Roxy M, De S, Pattnaik S, Pillai PA (2013) Simulation and extended range prediction of monsoon intraseasonal oscillations in NCEP CFS/GFS version 2 framework. Curr Sci 104(10):1394–1408

Sauer T, Yorke JA, Casdagli M (1991) Embedology. J Stat Phys 65(3–4):579–616, DOI 10.1007/bf01053745

Sikka DR, Gadgil S (1980) On the maximum cloud zone and the ITCZ over Indian longitude during southwest monsoon. Mon Weather Rev 108:1840–1853

Stachnik J, Waliser D, Majda AJ (2015) Precursor environmental conditions associated with the termination of MaddenJulian oscillation events. J Atmos Sci 72:19081931, DOI 10.1175/JAS-D-14-0254.1

Straub KH (2013) MJO initiation in the real-time multivariate MJO index. J Climate 26:1130–1151, DOI 10.1175/jcli-d-12-00074.1

Suhas E, Neena J, Goswami B (2013) An indian monsoon intraseasonal oscillations (miso) index for real time monitoring and forecast verification. Clim Dynam 40(11-12):2605–2616, DOI 10.1007/s00382-012-1462-5

Székely E, Giannakis D, Majda AJ (2016a) Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data. Climate Dyn 46(5):1473–1502

Székely E, Giannakis D, Majda AJ (2016b) Initiation and termination of intraseasonal oscillations in nonlinear Laplacian spectral analysis indices. Math Climate Wea Forecast In minor revision

Tung Ww, Giannakis D, Majda AJ (2014) Symmetric and antisymmetric signals in MJO deep convection. Part I: Basic modes in infrared brightness temperature. J Atmos Sci 71:3302–3326, DOI 10.1175/jas-d-13-0122.1

Wang B, Xie X (1997) A model for the boreal summer intraseasonal oscillations. J Atmos Sci 54:72–86

Wheeler MC, Hendon HH (2004) An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. Mon Weather Rev 132(8):1917–1932

Zhao Z, Giannakis D (2014) Analog forecasting with dynamics-adapted kernels. Nonlinearity In minor revision, arXiv preprint 1412.3831

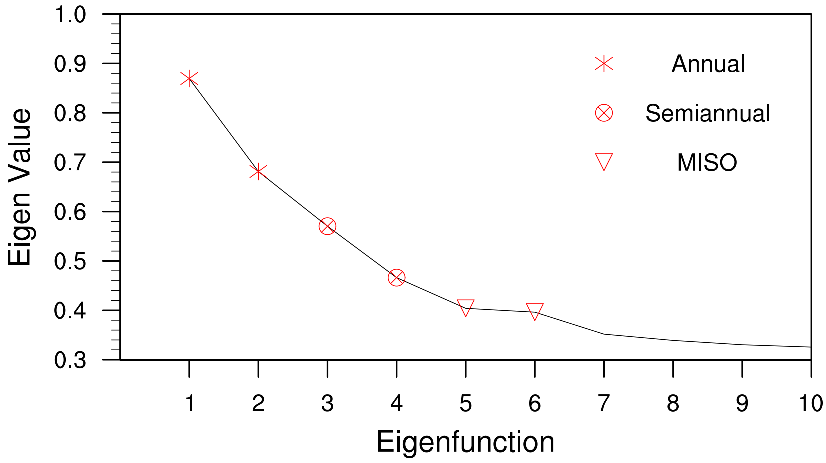


Fig. 1 Eigenvalues corresponding to the leading 10 Laplace-Beltrami eigenfunctions. Asterisks represent annual modes, crossed circles represent semiannual modes, and inverted triangles represent monsoon intraseasonal oscillation (MISO) modes.

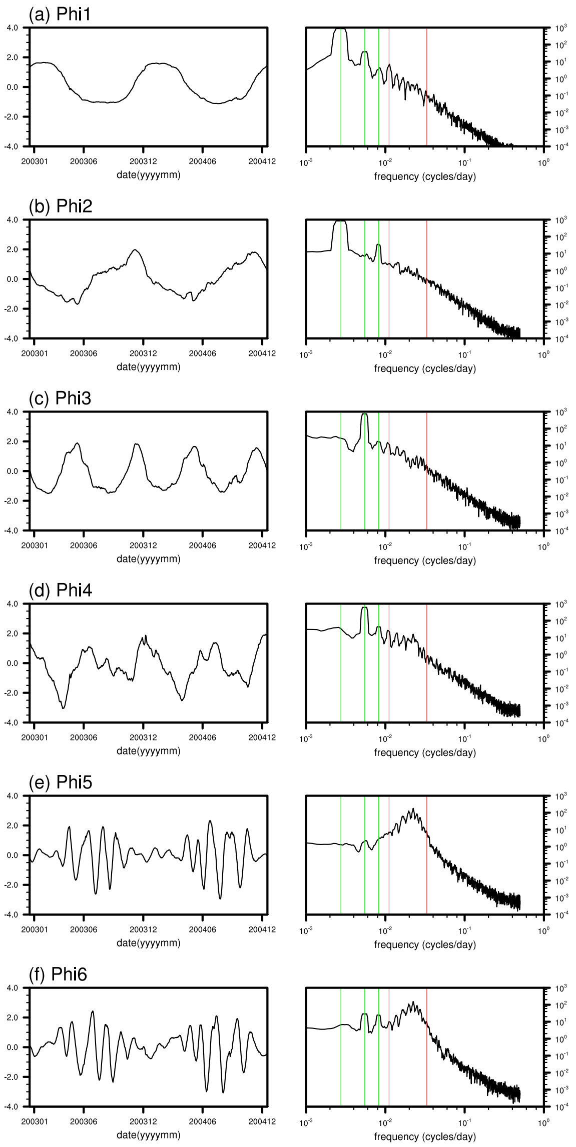


Fig. 2 Leading six Laplace-Beltrami eigenfunctions for the period January 2003 to December 2004 (left panels) and the corresponding power spectra (right panels). The power spectra are computed for the period January 1998 to December 2013. The red lines represent the 1/(90 days) and 1/(30 days) frequencies, and the green lines represent the 1/year, 2/year and 3/year frequencies.

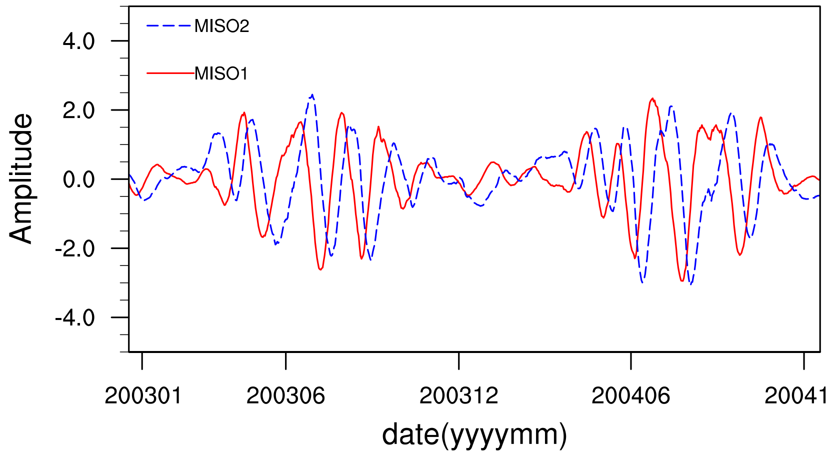


Fig. 3 Laplace-Beltrami eigenfunctions corresponding to the monsoon intraseasonal oscillation (NLSA MISO1 and NLSA MISO2) plotted together for the period January 2003 to December 2004.

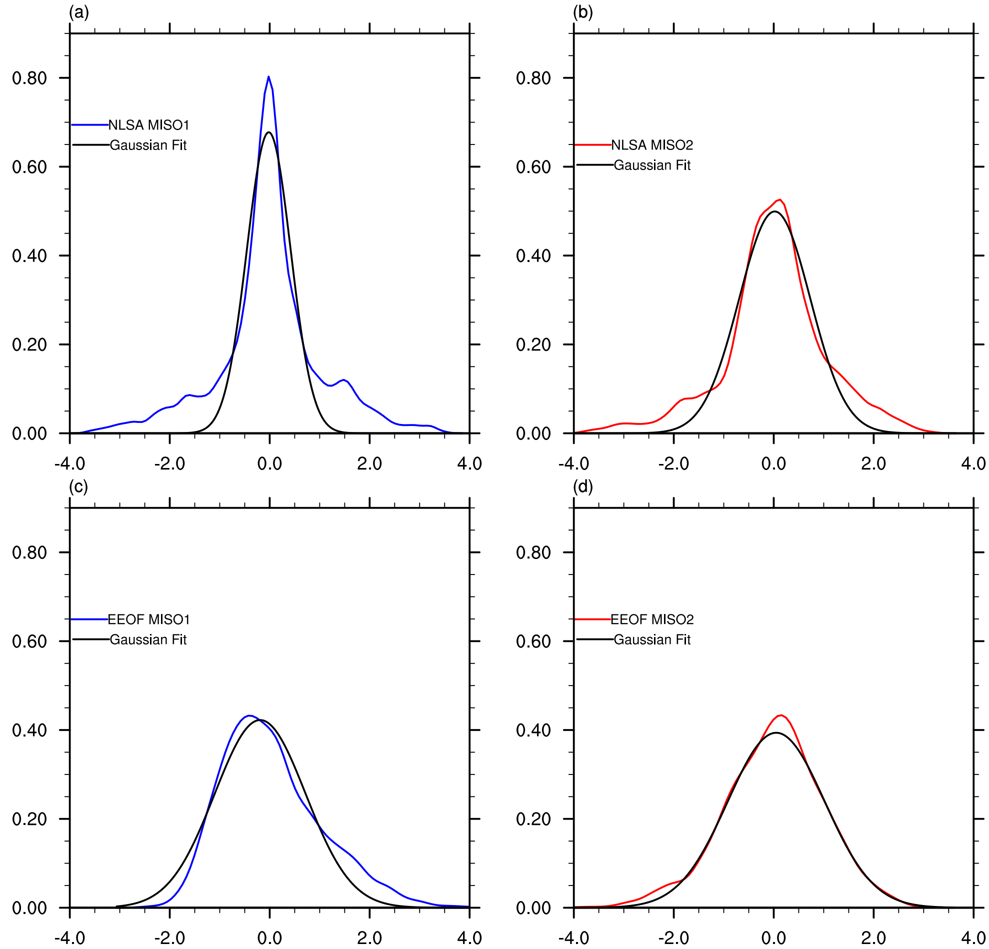


Fig. 4 PDFs of MISO indices from NLSA (a,b) and EEOF analysis (c,d). The black curves show Gaussian fits estimated via nonlinear least squares.

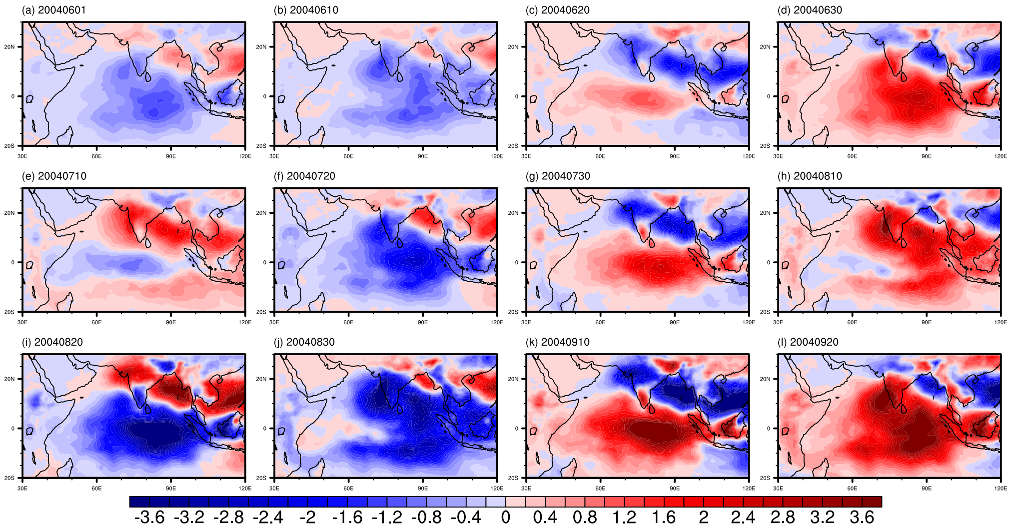


Fig. 5 Reconstruction of the MISO evolution for the period June 2004 to September 2004. The spatiotemporal map represent the GPCP rainfall anomalies (mm/day) obtained from the NLSA MISO indices for the period June 2004-September 2004

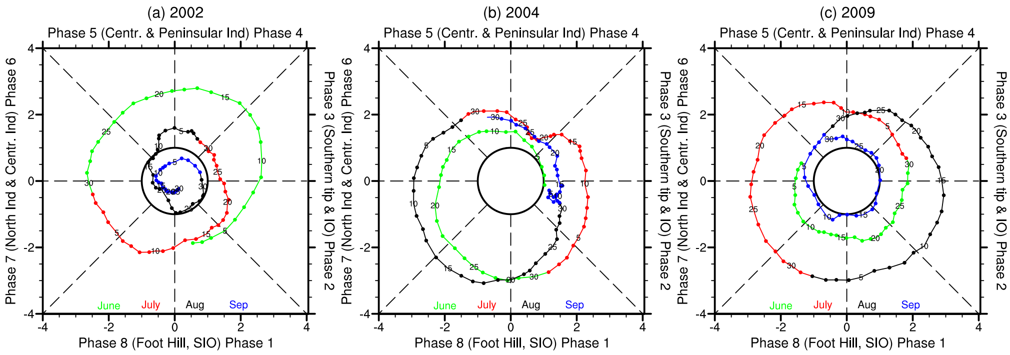


Fig. 6 2D phase space diagrams for the NLSA MISO indices, showing the significant MISO events in three typical drought years: (a) 2002, (b) 2004, and (c) 2009. An anticlockwise propagation from the phase 1 represents MISO's northward propagation. The circle centered at the origin has radius 1 standard deviation 0.89 of the MISO amplitude index $r(t)$ from (3).

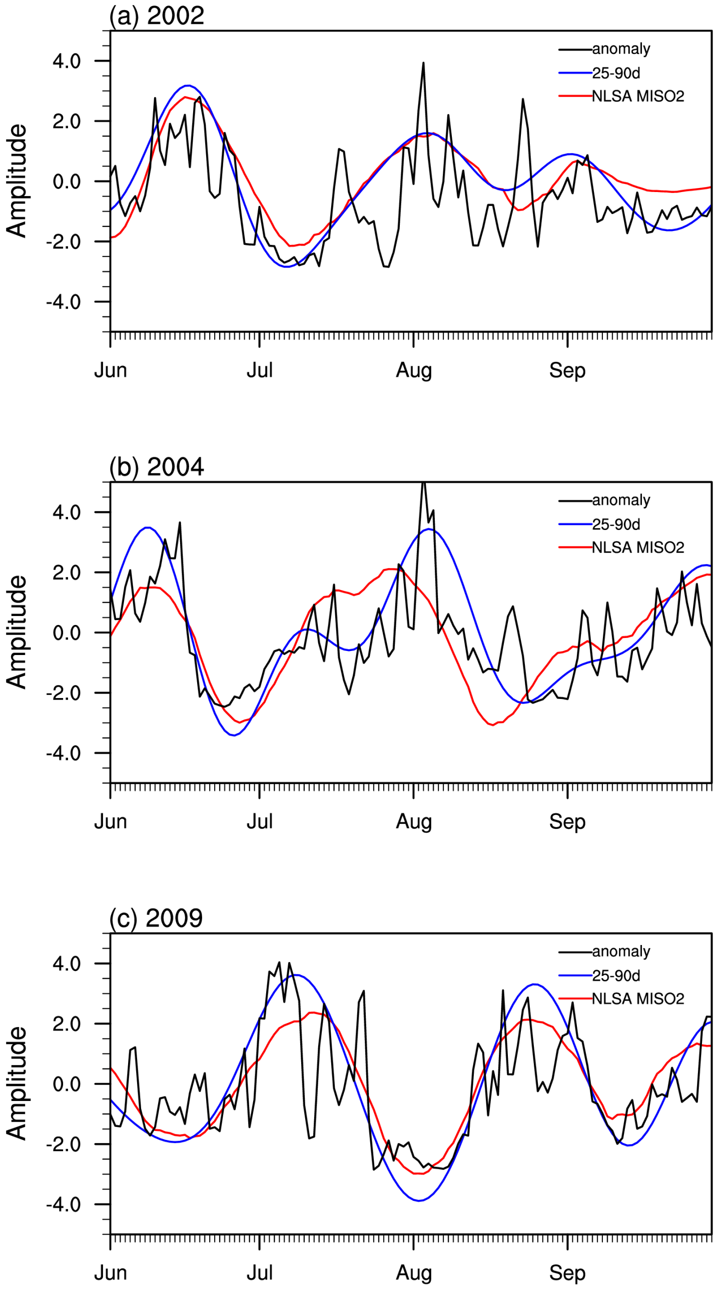


Fig. 7 Time series of the MISO2 index from NLSA and bandpass-filtered (25–90d) and unfiltered rainfall anomalies averaged over the central Indian domain (10.5°N–25.5°N, 70.5°E–85.5°E) for the JJAS seasons of the three drought years depicted in Figure 6.

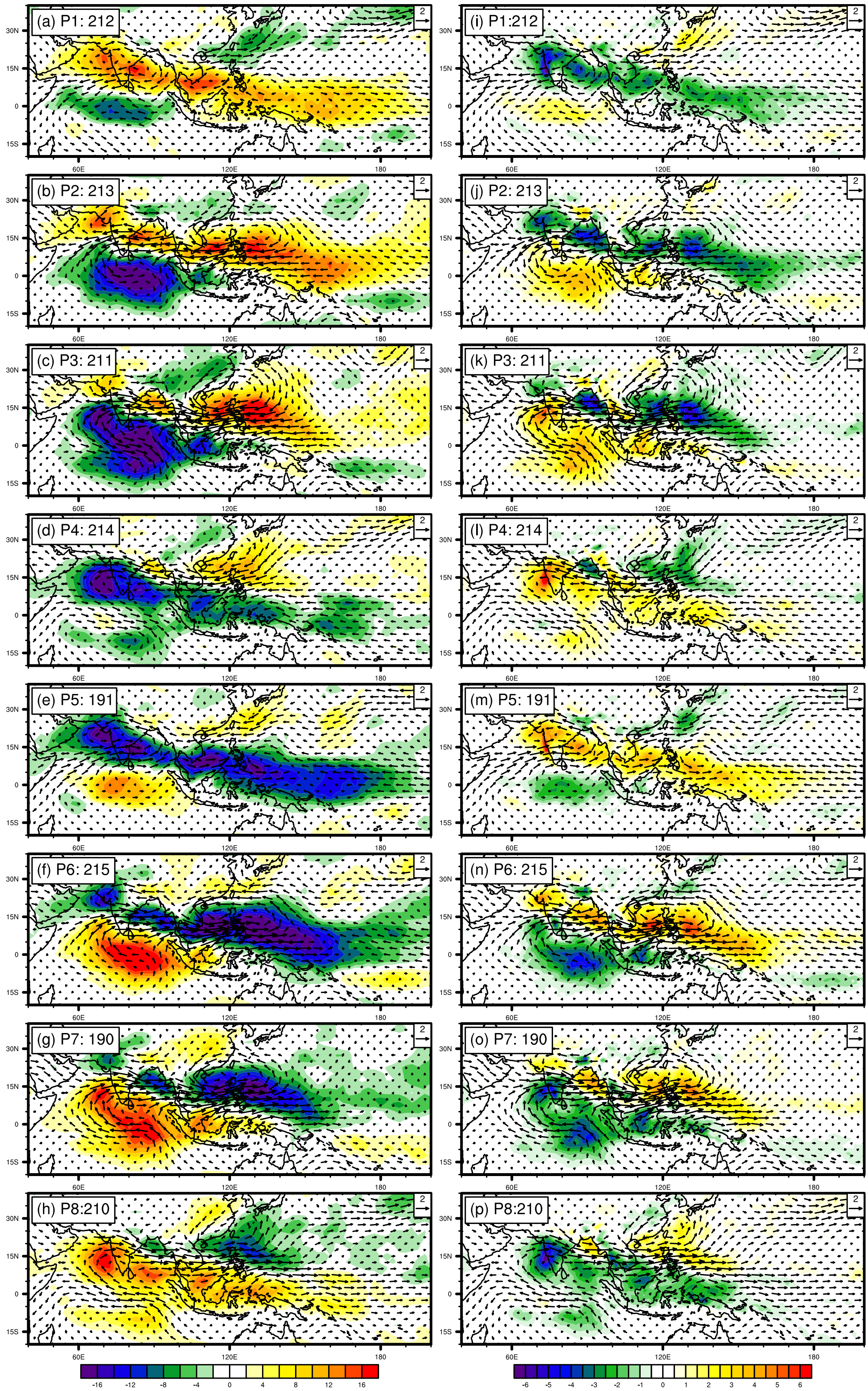


Fig. 8 (a-h) Phase composites of bandpass-filtered (25–90d) OLR (colors) and 850 hPa winds (vector) anomalies obtained from NLSA MISO modes.(i-p) same as (a-h) but for the bandpass-filtered (25–90d) rainfall (colors) and 850 hPa winds (vector) anomalies. The number of days used to create each composite is shown at the top left of each panel.

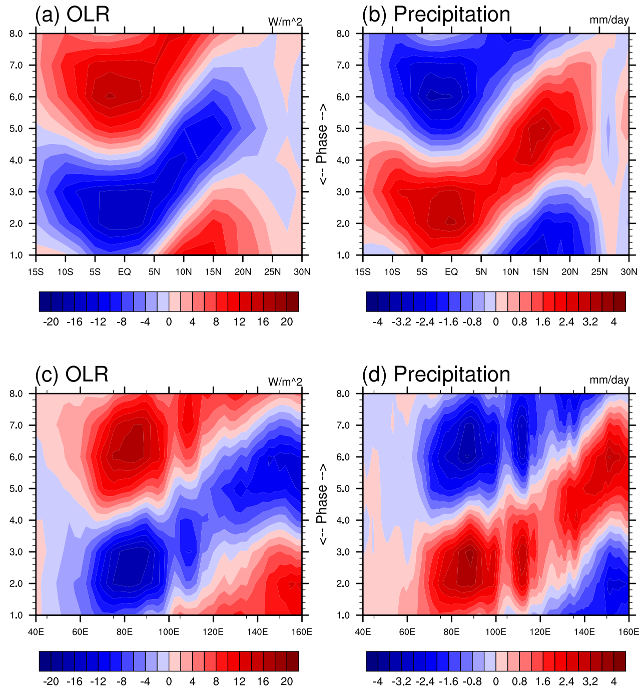


Fig. 9 (a,b) Latitude-phase diagrams for the phase composites of (a) OLR anomalies (b) rainfall anomalies from Figure 8, averaged over 70°E–100°E. (c,d) The corresponding longitude-phase diagrams for anomalies averaged over 5°S–5°N. For non integer phase values, the values are computed by interpolating between the 8 phases.

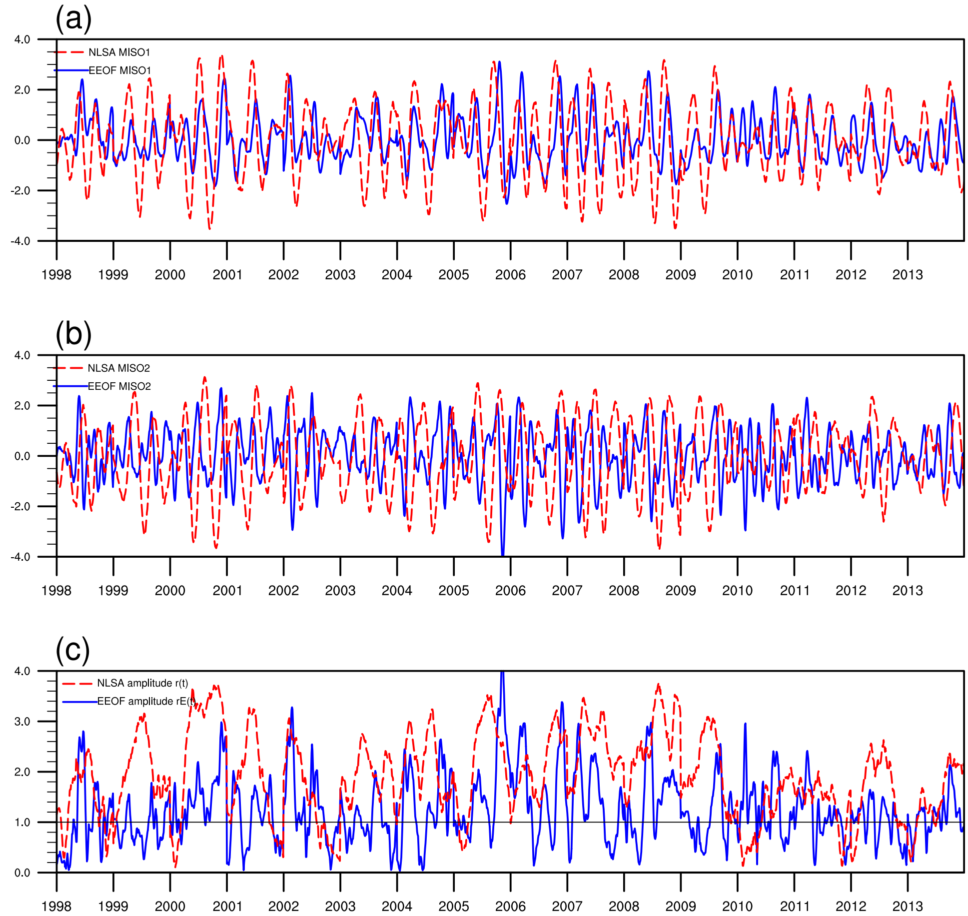


Fig. 10 (a) MISO1 indices for the 1998–2013 JJAS period obtained from NLSA (red line) and EEOF analysis (blue line). (b) same as (a) but for the MISO2 indices. Each indices are normalized by its own standard deviation (c) MISO amplitude index for the 1998-2013 JJAS period obtained from NLSA (r(t); red line) and EEOF analysis ((r$_E$)(t); blue line). Horizontal black line indicate the threshold for significant MISO events.

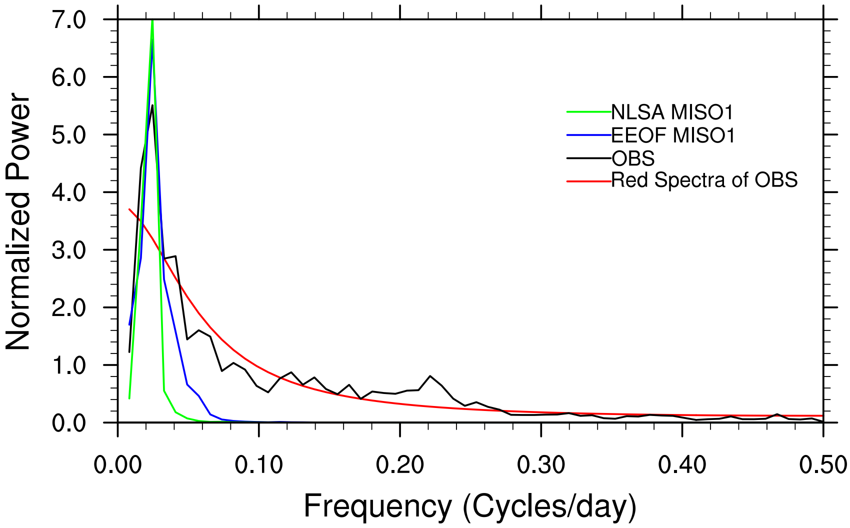
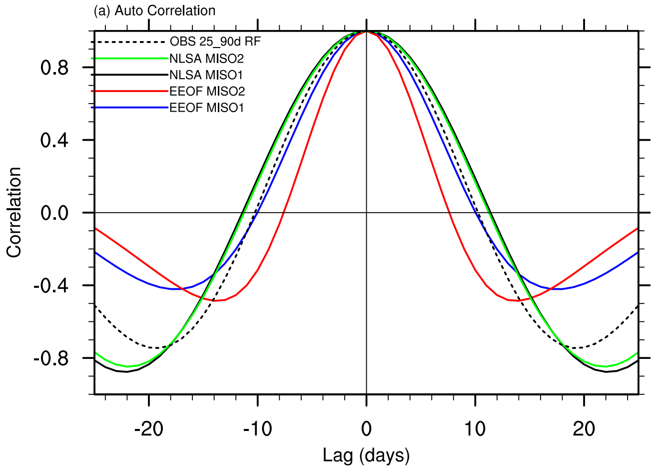


Fig. 11 Composites of the power spectra of rainfall anomalies over the monsoon core region (10.5°N–25.5°N, 70.5°E–085.5°E). Green lines represent NLSA MISO1, blue lines represent EEOF MISO1 and red lines represent Markov Red noise spectrum. Sixteen boreal summer season (1998-2013, JJAS) rainfall data is used for this calculation.



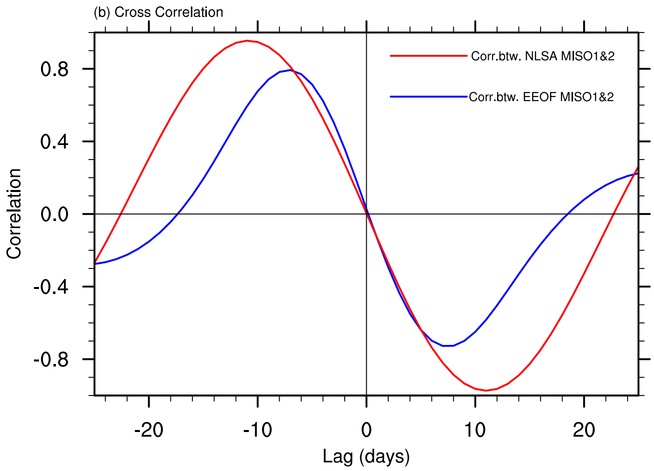


Fig. 12 (a) Autocorrelation function of the NLSA and EEOF MISO modes compared with the autocorrelation function of bandpass filtered (25–90 d) rainfall anomalies over the monsoon core region.(b) Cross-correlation functions of the NLSA and EEOF MISO modes.

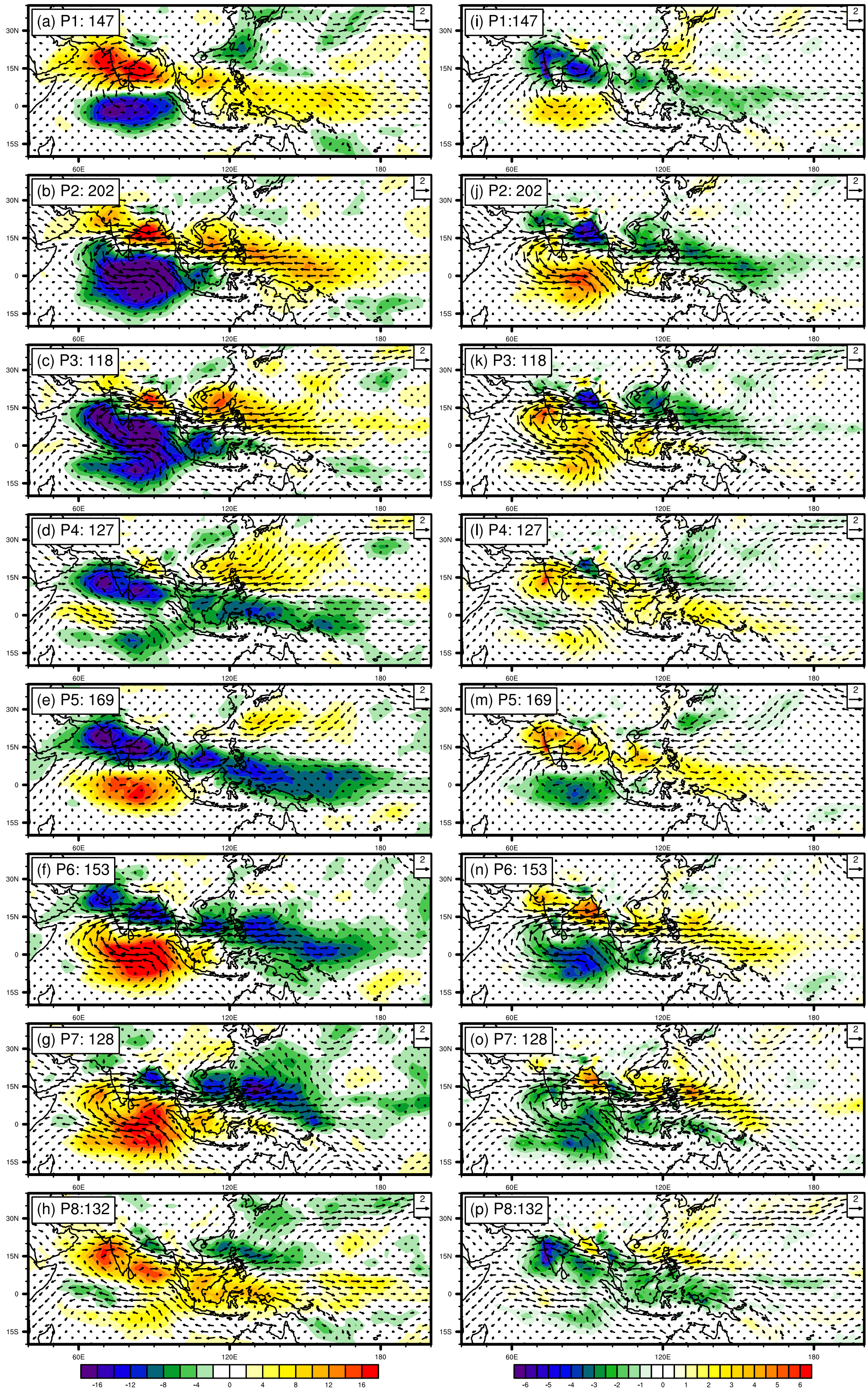


Fig. 13 (a-h) Phase composites of bandpass-filtered (25–90d) OLR (colors) and 850 hPa winds (vector) anomalies obtained from EEOF MISO modes. (i-p) same as (a-h) but for the bandpass-filtered (25–90d) rainfall (colors) and 850 hPa winds (vector) anomalies. The number of days used to create each composite is shown at the top left of each panel.

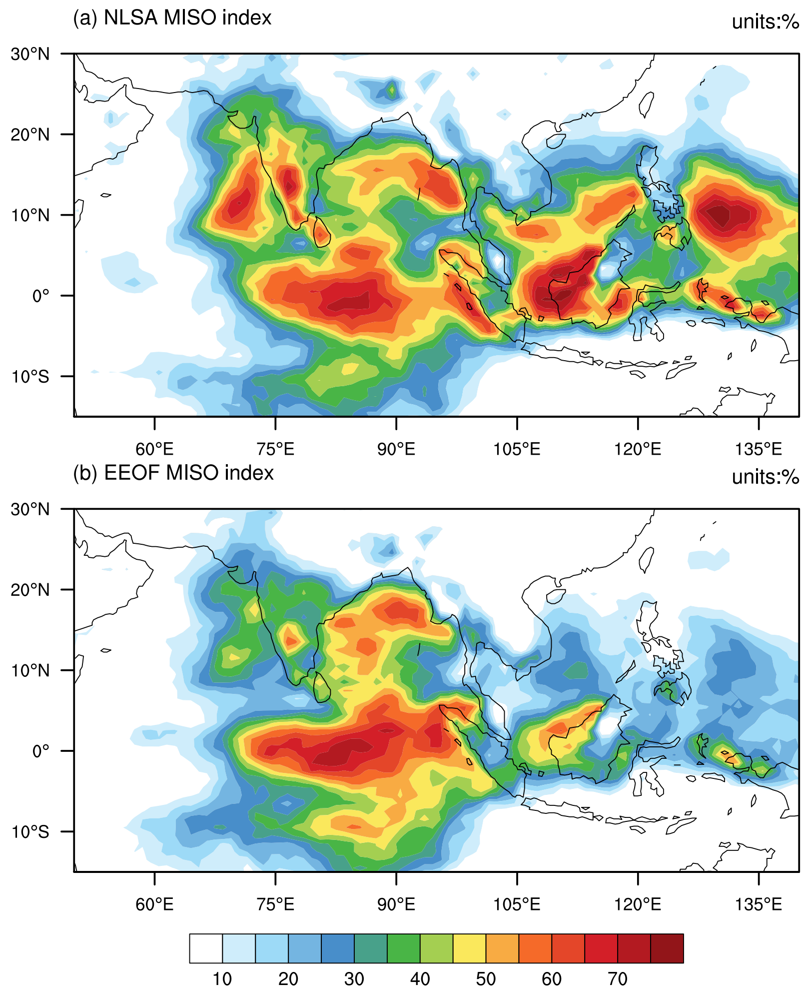


Fig. 14 Aggregate fractional variance associated with the (a) NLSA and (b) EEOF phase composites of bandpass-filtered rainfall anomalies. The aggregate fractional variance at each gridpoint is estimated as the ratio between the variance of phase composites and the total bandpass-filtered rainfall anomalies. The variance of phase composites is estimated from the eight life cycle composites (from Fig 8i-p and Fig 13i-p). The total bandpass-filtered rainfall anomalies is calculated for the period 1998-2013.

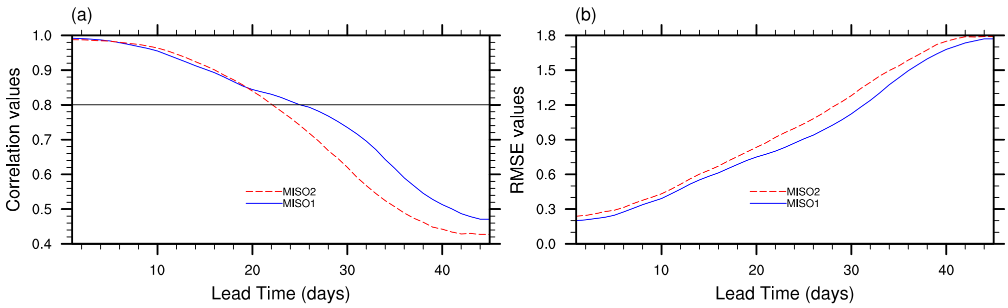


Fig. 15 (a) Extended range prediction skill of MISO modes and (b) root mean square error (RMSE) of the predicted MISO modes at each lead time estimated via out-of-sample extension of the NLSA modes using the CFSv2 hindcast data.

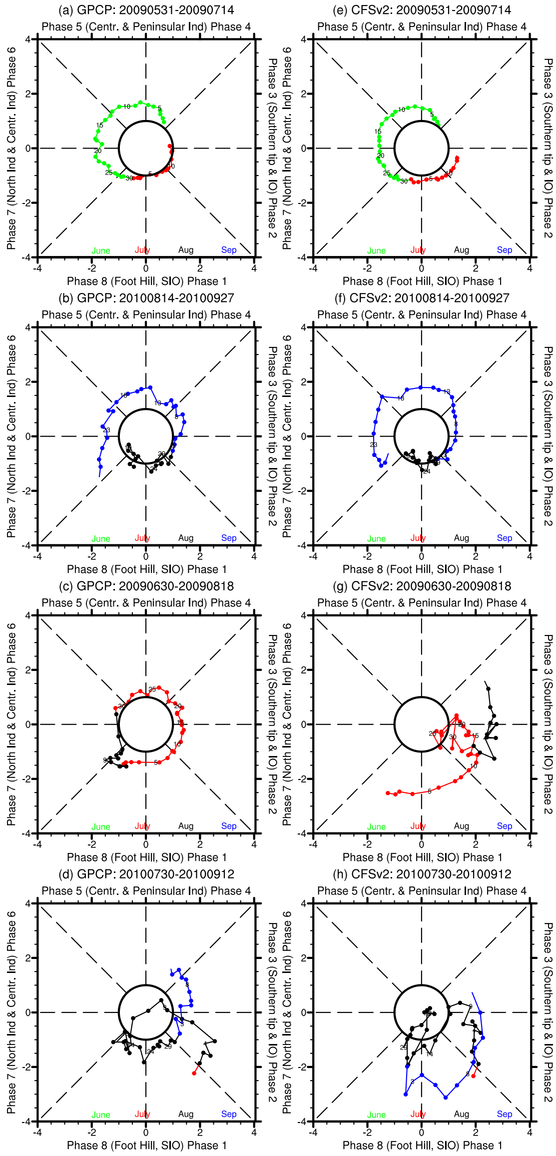


Fig. 16 Forecasts of the NLSA MISO indices for four initial condition runs of CFSv2 (right panels, e–h). Forecasts shown in lower panels are verified with the GPCP rainfall observations (left panels, a–d). Colors denote month.