

# *20th century intraseasonal Asian monsoon dynamics viewed from Isomap*

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access

Hannachi, A. and Turner, A. G. ORCID: <https://orcid.org/0000-0002-0642-6876> (2013) 20th century intraseasonal Asian monsoon dynamics viewed from Isomap. *Nonlinear Processes in Geophysics*, 20 (5). pp. 725-741. ISSN 1607-7946 doi: <https://doi.org/10.5194/npg-20-725-2013> Available at <https://centaur.reading.ac.uk/33824/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.5194/npg-20-725-2013>

Publisher: European Geosciences Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online





# 20th century intraseasonal Asian monsoon dynamics viewed from Isomap

A. Hannachi<sup>1</sup> and A. G. Turner<sup>2</sup>

<sup>1</sup>Department of Meteorology, Stockholm University, Stockholm, Sweden

<sup>2</sup>NCAS-Climate, University of Reading, Reading, UK

Correspondence to: A. Hannachi (a.hannachi@misu.su.se)

Received: 31 May 2013 – Revised: 12 August 2013 – Accepted: 23 August 2013 – Published: 8 October 2013

**Abstract.** The Asian summer monsoon is a high-dimensional and highly nonlinear phenomenon involving considerable moisture transport towards land from the ocean, and is critical for the whole region. We have used daily ECMWF reanalysis (ERA-40) sea-level pressure (SLP) anomalies on the seasonal cycle, over the region 50–145° E, 20° S–35° N, to study the nonlinearity of the Asian monsoon using Isomap. We have focused on the two-dimensional embedding of the SLP anomalies for ease of interpretation. Unlike the unimodality obtained from tests performed in empirical orthogonal function space, the probability density function, within the two-dimensional Isomap space, turns out to be bimodal. But a clustering procedure applied to the SLP data reveals support for three clusters, which are identified using a three-component bivariate Gaussian mixture model. The modes are found to appear similar to active and break phases of the monsoon over South Asia in addition to a third phase, which shows active conditions over the western North Pacific. Using the low-level wind field anomalies, the active phase over South Asia is found to be characterised by a strengthening and an eastward extension of the Somali jet. However during the break phase, the Somali jet is weakened near southern India, while the monsoon trough in northern India also weakens. Interpretation is aided using the APHRODITE gridded land precipitation product for monsoon Asia. The effect of large-scale seasonal mean monsoon and lower boundary forcing, in the form of ENSO, is also investigated and discussed. The outcome here is that ENSO is shown to perturb the intraseasonal regimes, in agreement with conceptual ideas.

## 1 Introduction

The South Asian and East Asian monsoon region constitutes one of the most populated areas on earth with around one-third of the world's population. The lives and well-being of society in this region depend critically on the Asian summer monsoon rainfall, which represents around 80 % of annual precipitation. As such, a lot of effort has been expended to understand and predict monsoon variability, an exercise that started from the days of H. F. Blanford (Blanford, 1884) and Sir Gilbert Walker (Walker, 1910, 1922), who attempted to predict the Indian monsoon but ended up discovering the Southern Oscillation (Walker, 1910). The fundamental driving mechanisms of the Asian monsoon can be linked to the differential heating between land and ocean and the associated moisture transport (Webster et al., 1998).

Potential predictability of seasonal mean monsoon rainfall implied by large-scale processes and slowly varying lower boundary conditions is reasonably well understood and is used routinely for rainfall forecasting (Charney and Shukla, 1981), particularly with empirical models (most recently, Rajeevan et al., 2007). However, the skill, measured by spatial correlation, at which current dynamical models can predict monsoon rainfall lies at around only 0.45 (Rajeevan et al., 2012), leading to the suggestion that other avenues for predictability should be explored. Subseasonal variations of monsoon activity, which are of particular importance to the local population for agriculture, are poorly understood (Turner and Slingo, 2009) and this remains a major source of uncertainty limiting the dynamical predictability of the Asian summer monsoon using general circulation models (GCMs) (Brankovich and Palmer, 2000). Furthermore, the relationship between this intraseasonal component of the monsoon

and seasonal mean forcing conditions is unclear (Turner and Hannachi, 2010; TH10 hereafter). Among the hypotheses put forward regarding monsoon intraseasonal variability (MISV) is that of a chaotic model (Palmer, 1994) with fluctuations between active and break monsoon phases (Meehl, 1994; Webster et al., 1998), analogous to the wings in the classical Lorenz system (Lorenz, 1963), in some way determined by large-scale boundary forcing. Some modelling studies using simplified GCMs (Carl, 1994; Tschentscher et al., 1994) as well as also comprehensive climate models (Goswami, 1997) have even suggested a chaotic monsoon system (see also Rangarajan and Sant, 2004). Large-scale forcing acts to change the stability and hence the population of one phase at the expense of the other as in Palmer (1999). Sperber et al. (2000) investigated the relationships between patterns of subseasonal and interannual variability of the Asian summer monsoon. They used the 40 yr National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR) reanalyses of low-level wind fields and precipitation data to test the hypothesis that the characteristics of intraseasonal variability of the monsoon are modulated on interannual timescales in a systematic manner. In particular, they showed that the probability distribution function (PDF) of intraseasonal monsoon variability was Gaussian, suggesting that anomalous monsoons are not associated with changes in regime behaviour. However, by using EOF analysis to break the intraseasonal variability down into its principal components, Sperber et al. (2000) noted that only a small subset of MISV (PC3) can be perturbed by the large-scale forcing. In another work, Straus and Krishnamurthy (2007) analysed pooled 5 day means of rotational 850 and 250 hPa wind and showed that (hints of) bimodality only exist under certain conditions: namely during weak monsoon years. Clear bimodality of the South Asian and East Asian summer monsoon activity, however, has not been established with certainty (Hannachi and Turner, 2013). Hence the relationship between long and short timescale modes of monsoon variability, and the inherent predictability it may bring, offers great opportunities for further research (Turner and Annamalai, 2012).

The Asian summer monsoon (ASM) is a complex and high-dimensional nonlinear dynamic system interacting with other components of the climate system, including large-scale processes and lower boundary conditions such as the El Niño Southern Oscillation (ENSO). The so-called *active* and *break* phases of intraseasonal monsoon activity are well recognised through their impacts, which could hint at a bimodal behaviour. This in turn would suggest the possible existence of a low-dimensional manifold that captures important dynamical features of the system. This low-dimensional space should also allow us to reveal the relationship to large-scale flow and lower boundary forcing. One would therefore hope that any analysis method aiming to characterize Asian monsoon dynamics in reanalyses or centennial-length model simulations would be able to identify those main features.

Looking for a low-dimensional embedding of ASM dynamics belongs to the class of dimensionality reduction in pattern recognition. One of the most commonly used methods of dimensionality reduction is the empirical orthogonal function (EOF) method (Jolliffe, 2002; von Storch and Zwiers, 1999; Hannachi et al., 2007). The EOF method is based on an eigen-analysis of the data covariance matrix and projects the data onto the leading modes of variability or EOFs to yield the principal component (PC) time series. The EOF method is known to have several severe geometric constraints, such as orthogonality in time and space, often preventing the analyst from obtaining a meaningful physical interpretation of the data in the low-dimensional space of EOFs/PCs (see e.g. Hannachi, 2006). The main feature of this low-dimensional space is that it explains maximum variance compared to other linear spaces with similar dimensions, but no dynamical feature is necessarily involved. Another serious weakness of the method is that it only enables projection onto linear subspaces of the original high-dimensional data. In the context of nonlinear processes with nonlinear structures as the ASM may be, this means that the method may not enable us to capture the intrinsic structure present in the data. For example, if the system's attractor lies on a nonlinear manifold then the EOF method will simply yield the linear tangent space and therefore can fail to detect the full structure of the nonlinear data manifold.

Another classical method for dimensionality reduction is the so-called multidimensional scaling (MDS), also known as principal coordinates (Young and Householder, 1938; Torgerson, 1952; Borg and Groenen, 1997). It has been used in applied sciences but not much in atmospheric science. MDS finds a linear low-dimensional embedding that preserves interpoint distances. A direct consequence of MDS is that data points that are close together in a high-dimensional space remain close in the low-dimensional space. Therefore MDS is a proximity-preserving dimensionality reduction. Note that the EOF method is similar, but not identical, to MDS in that it attempts to preserve the data variance as measured in the original high-dimensional space. But both the methods become equivalent when the distance used in MDS becomes Euclidean. MDS, like EOF, is unable to detect nonlinear structures in the data, but both share the advantage of being simple and efficiently computable.

The limitation of these linear projective methods has led to the development of nonlinear dimensionality reduction methods that can handle nonlinear structures present in the data. Here we simply mention examples like local linear embedding (Roweis and Saul, 2000), and neural networks methods such as nonlinear principal component analysis (Monahan, 2001; Hsieh, 2004), but in this paper we focus on the isometric feature mapping (Isomap) method (Tenenbaum et al., 2000). In climate, only a few studies have applied Isomap to ENSO (Gamez et al., 2004; Ross et al., 2008) and the Asian monsoon (Hannachi and Turner, 2013). In Hannachi and Turner (2013), for example, we simply explored the

bimodality of ASM without a deep and systematic analysis of the PDF structure of ASM. In this paper, a deeper analysis of intraseasonal monsoon variability is performed by genuinely investigating the structure of the system's PDF. We use sophisticated clustering methodology combined with the multivariate Gaussian mixture model to estimate the probability density function of ASM within the obtained Isomap low-dimensional space. The data and methodology are presented in Sect. 2. In Sect. 3, we analyse the climatology and variability of ASM. The application of Isomap and the mixture model is then performed in Sect. 4. The relationship to seasonal averages of the large-scale monsoon and of the large-scale forcing is analysed in Sect. 5, and a summary and discussion are presented in the last section.

## 2 Data and methodology

### 2.1 Data

The data used here come from the European Re-analysis (ERA-40) project (Uppala et al., 2005) of the European Centre for Medium-Range Weather Forecasts (ECMWF) and consist of sea-level pressure (SLP) and lower tropospheric (850 hPa) zonal and meridional wind fields over the Asian summer monsoon region (50–145° E, 20° S–35° N) for June to September of the ERA-40 period 1958–2001. The data are supplied on a Gaussian grid of approximately 1.125° spacing. Daily anomalies, with total sample size of 5368 (44 × 122 days of JJAS), are obtained by removing the seasonally varying mean field based on daily averages over the years, i.e. daily anomalies to the seasonal cycle. In addition, the data are detrended by removing a linear trend using least-squares fit.

We also use observed rainfall data for *monsoon Asia* from the APHRODITE project<sup>1</sup> (Yatagai et al., 2012). This is a daily gridded dataset at 0.5° resolution based on rain gauge information. As such, it covers the land domain only, and we curtail the data to match the ERA-40 period as above.

To characterise the large-scale seasonal mean monsoon, we have used the WY dynamical index proposed by Webster and Yang (1992) and area-averaged Indian rainfall (IR). The WY index is a proxy for the (diabatic) heating in the atmospheric column and is defined as the JJAS average of anomalous zonal wind shear between the lower (850 hPa) and upper (200 hPa) troposphere averaged over the band 40–110° E, 5–20° N. The IR index is derived from an area-weighted average of 2140 rain gauge stations from the India Meteorological Department's 1 degree gridded product (Rajeevan et al., 2006), averaged over the summer (JJAS) months and curtailed to match the ERA-40 period. To characterise ENSO, the major driver of interannual variability in the monsoon (see, e.g., Straus and Krishnamurthy, 2007, among many others) we use the JJAS average SST in the Niño-3 region

(210–270° E, 5° S–5° N) from HadISST data (Rayner et al., 2003).

## 2.2 Methodology

### 2.2.1 Metric scaling

Multidimensional scaling is a geometric method for reconstructing a configuration from its interpoint distances and enables visualising proximities in low-dimensional space. The EOF method, for example, can find a low-dimensional embedding of the data that preserves variances, as measured in a high-dimensional input space. MDS, on the other hand, finds embedding that preserves interpoint distances and consequently attempts to preserve agglomerations. To start, we let  $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kp})^T$  designate the  $k$ th point in the original  $p$  dimensional space  $E_p$ , where  $k = 1, \dots, n$ , and  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  is the associated centered  $n \times p$  data matrix. In the classical metric problem (Young and Householder, 1938; Torgerson, 1952; Borg and Groenen, 1997), we are given a matrix of distances

$$\mathbf{D} = (d_{ij}) = (\|\mathbf{x}_i - \mathbf{x}_j\|) \quad (1)$$

between different pairs of the  $n$  given points within a high-dimensional space and the objective is to find a low-dimensional embedding space of the coordinates or configuration of the data points. In order for the problem to be well posed, the elements of  $\mathbf{D}$  have to satisfy the criteria of distances or dissimilarities, namely: positivity ( $d_{ij} \geq 0$ ); non-degeneracy ( $d_{ii} = 0$ ); and triangular inequality ( $d_{ij} \leq d_{ik} + d_{kj}$ ). We also denote the  $n \times n$  matrix of scalar product by

$$\mathbf{A} = \mathbf{X}\mathbf{X}^T. \quad (2)$$

When the distances defined in Eq. (1) are Euclidean a simple algebra based on the identity  $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 - 2 \langle \mathbf{x}, \mathbf{y} \rangle$ , applied to the elements of  $\mathbf{A}$  yields the identity:

$$\mathbf{A} = -\frac{1}{2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{I}\mathbf{I}^T \right] \mathbf{D}_2 \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{I}\mathbf{I}^T \right]^T, \quad (3)$$

where  $\mathbf{D}_2 = (d_{ij}^2)$ ,  $\mathbf{I}_n$  is the identity matrix of order  $n$  and  $\mathbf{I} = (1, \dots, 1)^T$  is the vector of length  $n$  containing ones. Note that since  $(\mathbf{I}_n - \frac{1}{n} \mathbf{I}\mathbf{I}^T)$  is a centering operator, Eq. (3) represents a double centering of the matrix  $\mathbf{D}_2$ . The classical MDS problem seeks a low-dimensional space  $E_r$ , where the distances  $d_{ij}^*$  are as close as possible to their original analogues, achieved by minimizing  $\sum (d_{ij} - d_{ij}^*)^2$ . Since  $\mathbf{A}$  is symmetric semi-definite positive it can be decomposed using the singular value decomposition (SVD) procedure as  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$  where  $\mathbf{\Lambda} = (\text{diag}(\lambda_i))$  is a positive diagonal matrix. Using Eqs. (2) and (3) the principal coordinate matrix  $\mathbf{X}$  can be retrieved and yields:

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}. \quad (4)$$

<sup>1</sup>available for free from [www.chikyu.ac.jp/precip](http://www.chikyu.ac.jp/precip)

When the dissimilarities are not Euclidean, the matrix  $\mathbf{D}_2$ , and hence the matrix  $\mathbf{A}$  in Eq. (3), may cease to be semi-definite positive. The classical solution to this problem is to simply choose the first  $r$  largest positive eigenvalues and associated eigenvectors of the matrix  $\mathbf{A}$ . The usual hope then is that  $\lambda_r \gg \lambda_{r+1}$ , for some  $r < p$ , as usually applied in EOF analysis.

### 2.2.2 Isomap

In Isomap the underlying assumption is that the data belong to a nonlinear manifold. The metric therefore is not Euclidean and has to be computed carefully. The basic idea in Isomap (Tenenbaum et al., 2000) is to use a geodesic<sup>2</sup> distance, i.e. a dissimilarity measured along the (generally unknown) nonlinear manifold of the data. This is achieved by using local paths connecting the data points and then distances are computed along those (geodesic) paths. The Isomap algorithm has three main steps. The first step is to use the available interpoint (usually Euclidean) distances,  $d_{ij}$ , for all  $i$  and  $j$ , to construct neighbouring points. This can be done either by selecting the  $K$  nearest (for some value of  $K$ ) points to, or by choosing points falling inside a ball of radius  $\varepsilon$  centred at a target point. The neighbourhood is then defined as a weighted graph where the weight of the edges is represented by the distances  $d_{ij}$ , for all  $i$  and  $j$ . The second step consists of defining the geodesic distance  $\delta_{ij}$  between any two points using the shortest path following the graph constructed in step 2. Once the dissimilarity matrix  $\mathbf{\Delta} = (\delta_{ij})$  is obtained the last step of Isomap consists of applying the classical MDS procedure to find the embedding space and the associated principal coordinates.

### 2.2.3 The mixture model

To estimate the PDF of the data within the low-dimensional space we use, in addition to the non-parametric kernel estimate (Silverman, 1981), the parametric multivariate Gaussian mixture model (Hannachi, 2007, 2010; Woollings et al., 2010a; Rust et al., 2010). Within this framework any PDF  $f(\mathbf{x})$  can be expressed as a convex combination of  $M$  multivariate Gaussian distributions:

$$f(\mathbf{x}) = \sum_{k=1}^M \alpha_k g_k(\mathbf{x}, \mathbf{\Sigma}_k, \boldsymbol{\mu}_k), \quad (5)$$

where  $\alpha_1, \dots, \alpha_M$  are the  $M$  mixing proportions or weights of the mixture model and they satisfy:

$$0 < \alpha_k < 1, \text{ for } k = 1, \dots, M, \text{ and } \sum_{k=1}^M \alpha_k = 1, \quad (6)$$

and  $\boldsymbol{\mu}_k$  and  $\mathbf{\Sigma}_k$  are, respectively, the mean and the covariance matrix of the  $k$ th,  $k = 1, \dots, M$ , multivariate normal density

<sup>2</sup>geodesics are the shortest path curves joining points on a given manifold. For example, great circles are geodesics on the sphere.

function  $g_k$ :

$$g_k(\mathbf{x}, \mathbf{\Sigma}_k, \boldsymbol{\mu}_k) = (2\pi)^{-q/2} |\mathbf{\Sigma}_k|^{-1/2} \exp[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)], \quad (7)$$

where  $q$  is the state space dimension, i.e. the dimension of state vector  $\mathbf{x}$ . The  $\frac{M(q+1)(q+2)-2}{2}$  unknown parameters  $\boldsymbol{\mu}_k$ ,  $\mathbf{\Sigma}_k$ ,  $k = 1, \dots, M$ , and  $\alpha_k$ ,  $k = 1, \dots, M - 1$  are obtained using the expectation-maximization (EM) algorithm (Everitt and Hand, 1981; McLachlan and Basford, 1988; Hannachi and O'Neill, 2001).

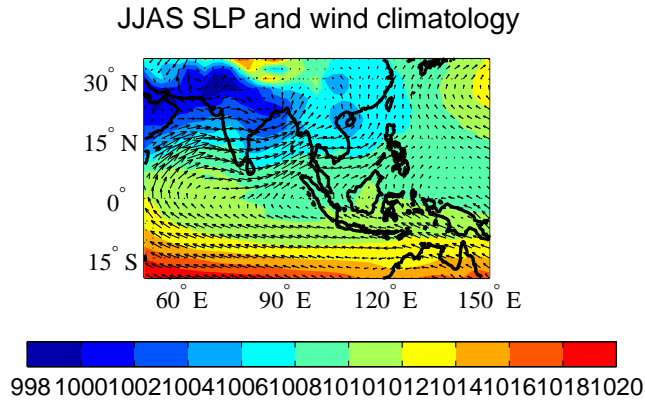
Estimating the number of components in the mixture model (5) proves to be the most difficult part and can be merged with the problem of finding the number of clusters in a dataset. In this manuscript we focus on a particular and robust method, namely the gap statistic (Tibshirani et al., 2001), which can be applied to any clustering method (e.g. Hannachi et al., 2011). It is based on a comparison between the within-dispersion index of the data and that expected from a null distribution, normally taken to be a homogeneous random point process. The computation of the gap statistic is deferred until Sect. 4.1 in connection with clustering.

## 3 ASM climatology and variability

### 3.1 ASM climatology

The South Asian monsoon is a seasonal feature in the summer (June to September) and peaks around July/August. During boreal spring, differential heating of the Eurasian land-mass versus the Indian Ocean drives a pressure gradient and development of a heat low over the northern Indian subcontinent. Aided by the extension of the meridional temperature gradient to significant depth in the troposphere by the Tibetan Plateau (Li and Yanai, 1996) and isolation from drier mid-latitude air by the Himalayas (Boos and Kuang, 2010), the pressure gradient leads to a reversal of the seasonal winds and advection of moist air to the South Asian subcontinent. For further details see the recent review in Turner and Annamalai (2012).

Figure 1 shows the summer (JJAS) climatology of SLP over the ASM region overlaid with the 850 hPa wind field. First and foremost we can see the low pressure trough sitting over the Asian continent with the lowest values, reaching 998 hPa, located over most of the Indian sub-continent, western Arabia, the Bay of Bengal and the Bay of Vietnam in the South China Sea. The high pressure system is pushed south over the southern Indian Ocean (not shown). Next, the low-level wind field shows clearly the Somali jet over the Arabian Sea, reaching speeds of about  $18 \text{ m s}^{-1}$ , crossing India and the Bay of Bengal to Southeast Asia. Similarly, one can also see cyclonic circulation of moist air around the monsoon trough west from the head of the Bay of Bengal, in Southeast Asia and the western North Pacific, which can give rise to monsoon precipitation in those regions. The southerlies



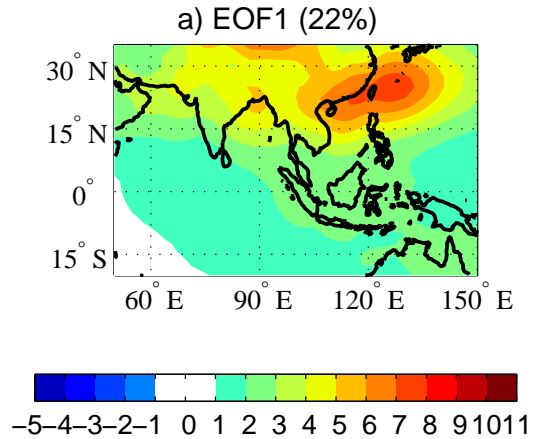
**Fig. 1.** JJAS climatology of the ERA-40 SLP and 850 hPa wind field over the Asian monsoon region. SLP units: 2 hPa, and the maximum wind speed is  $18 \text{ m s}^{-1}$ .

incident on East Asia can also be perturbed by the western North Pacific high further to the east.

### 3.2 ASM variability

Figure 2 shows the leading two EOFs of JJAS detrended SLP anomalies over the ASM region. The leading mode of variability (Fig. 2a) explains about 22% of the total JJAS SLP variance, and is well separated from the rest of the covariance matrix spectrum (not shown). This EOF shows a monopole over most of the domain with most of the loading located over land masses and with the centre of action situated over the coast of East Asia and the western North Pacific centred near  $25^\circ \text{ N}$ ,  $120^\circ \text{ E}$ . Since the sign of EOFs is irrelevant, EOF1 reflects a strengthening and weakening of the main centre of action of the Asian monsoon in terms of anomalous surface pressure, over the western North Pacific. The second EOF, with about 15% explained variance, shows two centres located respectively in the west Pacific near the East China Sea at  $30^\circ \text{ N}$ ,  $130^\circ \text{ E}$  and most of the Indian Ocean with the strongest loadings situated in the Arabian Sea and the Indian subcontinent. The eigenvalue associated with EOF2 is also well separated from the remaining covariance spectrum. The second mode of ASM variability (Fig. 2b) reflects a see-saw of monsoon variability showing opposite behaviour between western Indian Ocean and north-west Pacific. The third EOF (not shown) indicates mainly a north-east/south-west dipole with centres of action located respectively over the north-west Pacific ( $18^\circ \text{ N}$ ,  $130^\circ \text{ E}$ ) and east China, and explains only 9% of the total variance.

To investigate the behaviour of the PDF of the leading two modes of variability, we show in Fig. 3 the Gaussian kernel estimate of the ASM PDF using the leading two PCs of the SLP anomaly over the ASM region. Given a sample of data  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , this PDF estimate is given by

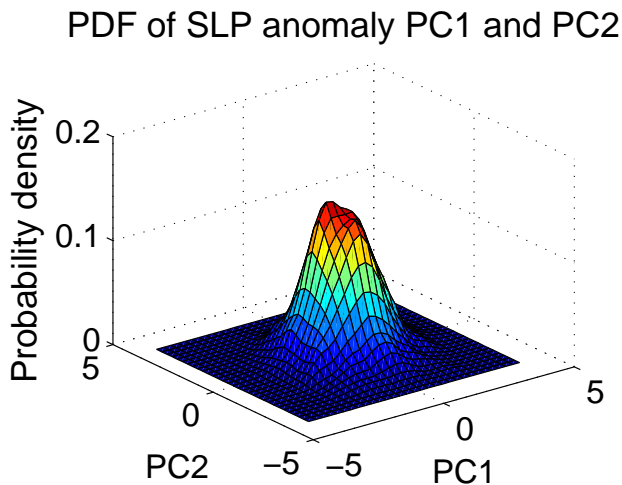


**Fig. 2.** The leading two EOFs of JJAS detrended SLP anomalies over the Asian summer monsoon region. The percentage of explained variance of the EOFs are also shown. Units are arbitrary.

$$\hat{f}(\mathbf{x}) = \frac{1}{2\pi n \sqrt{h_1 h_2}} \sum_{k=1}^n \exp \left[ -\frac{1}{2\sqrt{h_1 h_2}} (\mathbf{x} - \mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) \right]. \quad (8)$$

In Eq. (8), we have used the optimal kernel width,  $h_k = \sigma_k n^{-1/6}$ , where  $n$  is the sample size and  $\sigma_k$  is the standard deviation of the  $k$ th ( $k = 1, 2$ ) PC. The PDF in Fig. 3 is clearly unimodal, although not Gaussian. There does seem to be minor evidence of two peaks in the data, although these are not significant when decomposing the data using the linear EOF analysis. The skewness estimates of PC1 and PC2 are, respectively,  $-0.05$  and  $-0.23$ . Given the standard deviation of the sample skewness,  $\sqrt{6/n}$ , with  $n$  being the number of independent samples in the data, and taking heuristically every 5 to 10 data points of the daily detrended SLP anomalies as being approximately independent, only PC2 is significantly skewed (at the 5% significance level). The PC1 component, however, is not significantly skewed nor kurtotic, and hence is not significantly different from a Gaussian as suggested by Sperber et al. (2000). A similar skewness behaviour was obtained using the leading PC of outgoing long-wave radiation (OLR) by TH10. The OLR PC1 was also unimodal, but TH10 interpreted the skewness in terms of a two-component





**Fig. 3.** Gaussian kernel PDF estimate of the leading two JJAS SLP PCs.

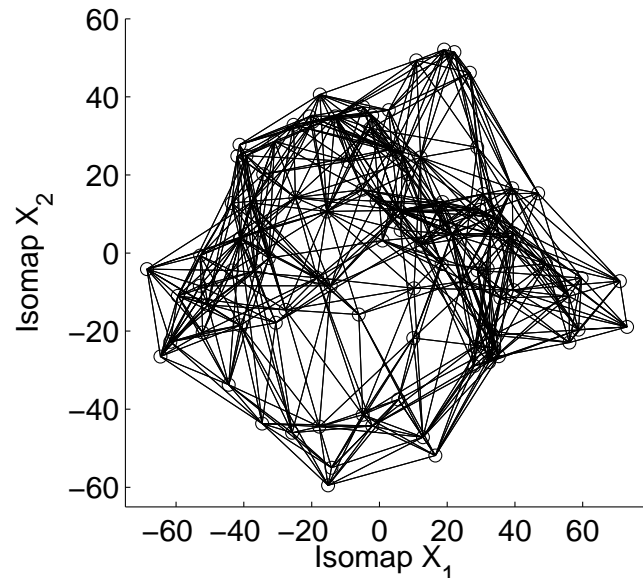
Gaussian mixture model, the components analogous to active and break phases of ASM. The unimodality of the PCs is not surprising since EOFs/PCs aim only at maximising variance but do not take into account the topology of the data manifold, which is investigated next.

## 4 Nonlinear dynamics of intraseasonal ASM

### 4.1 Isomap low-dimensional structure of ASM

#### 4.1.1 Variance in the leading modes

Given that the PCs do not generally reflect the inherent structure of the manifold present in the data, in this section we set out to apply Isomap to ASM. Note that Isomap is customarily illustrated using the swiss-roll data as in Tenenbaum et al. (2000) and Ross et al. (2008), and so we do not repeat this exercise here but refer the reader to those references. Since Isomap takes into account the neighbourhood structure of the data it is hoped that the low-dimensional projection would reveal the nonlinear feature of ASM dynamics. The interpoint distance matrix  $\mathbf{D}$  between all JJAS daily observation pairs of the SLP anomalies is first computed based on the Euclidean distance. To construct the neighbourhood graph we connect the data using the  $K$  nearest neighbour method. Accordingly, two points in state space that are not  $K$  nearest neighbours are not connected directly but through other intermediate points. For example for  $K = 2$ , every point is connected to the nearest two points. Of course a very small value of  $K$  may lead to a poor graph that may not connect all the data points whereas a quite large value may lead to a metric that is not too different from the Euclidean metric. We followed some guidelines given in Ross et al. (2008) for the neighborhood size, and we found a value of the order  $K = 12$ , which

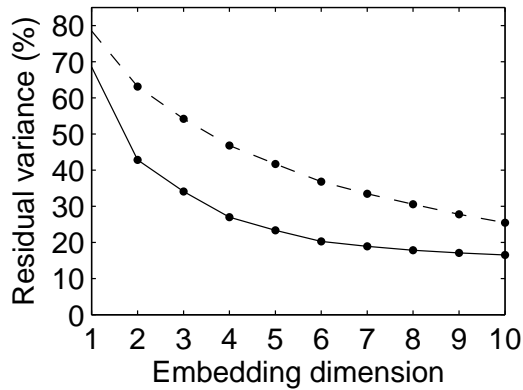


**Fig. 4.** Illustration showing the neighborhood graph for the first 100 anomalous SLP observations within the two-dimensional Isomap embedding space. The data have been scaled to zero-mean and unit variance. Axis units arbitrary.

we use for the discussion to follow. Sensitivity to this parameter and to the choice of the domain have been investigated and are discussed below (see Sect. 4.1.3). The obtained graph allows us to construct the geodesic distance  $\delta_{ij}$  between any two points  $i$  and  $j$  by choosing the shortest path between them following the graph. Note that this graph cannot be visualized as is because it is in a high-dimensional space, but we show below an illustrative example from the data when embedded in two dimensions. Once computed the geodesic dissimilarity matrix,  $\mathbf{\Delta} = (\delta_{ij})$ , is submitted to an MDS decomposition. As an illustrative example we show in Fig. 4 the two-dimensional embedding of the neighborhood graph based only on the first 100 data points, using  $K = 10$ , for visualization purposes. The data are shown in small circles and the graph is shown, which connects the data points using the 10 nearest points.

The Isomap goodness of fit is normally measured using the residual variance explained by the retained Isomap components (Tenenbaum et al., 2000). Figure 5 shows the Isomap residual variance of the SLP anomalies as a function of the embedding dimension. The residual variance associated with the leading Isomap component is 69%, and is 44% for the first two components taken together. The same residual variance is also plotted for the leading 10 EOFs. Since the variances in EOFs are additive, residual variances are computed from the explained variances of these same EOFs. For example, the residual variance based on the leading EOF is 78% and is 63% for the leading two EOFs taken together. These values are larger than their Isomap analogues because





**Fig. 5.** Residual variance obtained from the Isomap application using  $K = 12$  (solid) and from the EOFs (dashed) of the Asian monsoon SLP anomalies versus the embedding dimension.

Isomap attempts to follow the nonlinear manifold of the data, unlike the EOFs which are obtained by looking for optimal linear subspaces. Figure 5 also shows an elbow, at the embedding dimension  $d = 2$ , associated with a break in the slope of the residual variance curve. Note the absence of this elbow in the residual variance curve associated with the EOFs, which decreases, somehow, in a more smooth fashion. The existence of such an elbow is normally taken as the dimension of the nonlinear manifold of the data (Roweis and Saul, 2008; Ross et al., 2008). We note, however, that the main objective here is not to identify the dimensionality of the ASM, an important subject in its own right and which is deferred to another study, but to look for possible nonlinearity using a simple 2-D embedding of the ASM via Isomap.

#### 4.1.2 Manifestation of the leading modes

Figure 6a shows the leading two time series obtained from the 2-D embedding. As for the PCs, these two time series are uncorrelated, but they do not explain maximum variance. To learn about the spatial features associated with these time patterns, we show in Fig. 7a and b the spatial patterns obtained by regressing the (detrended) SLP anomalies onto the time series (Fig. 6a). A similar regression is performed between the Isomap time series and the 850 hPa wind anomalies. Further, regressions of APHRODITE precipitation are shown for Isomap time series 1 and 2 in Fig. 7c and d, respectively.

The low-pressure anomaly over the East China Sea (Fig. 7a) is associated with a low-level anomalous cyclonic circulation with an eastward extension of the Somali jet. This pattern appears to represent a strengthening and weakening of the large-scale Asian monsoon flow: strengthening in the west including the Somali jet as it crosses the equator in the Indian Ocean before extending on to India and Southeast Asia. As it crosses into the western Pacific it counteracts the mean flow into southern China. However, looking

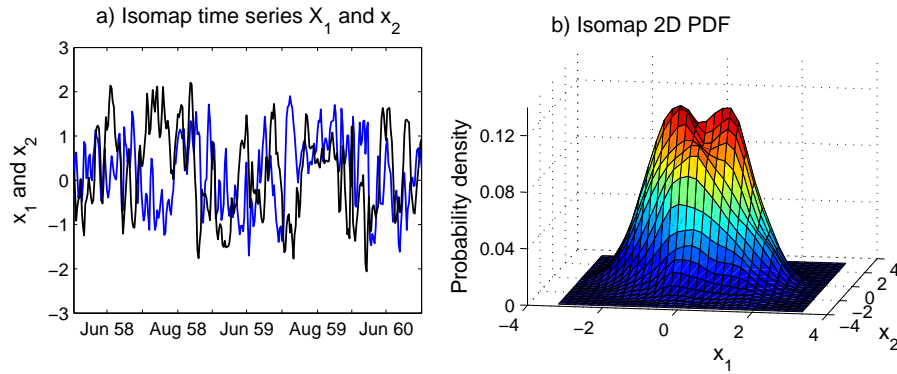
more closely and aided in our interpretation by the rainfall (Fig. 7c), we see the flow to deviate southwards as it reaches the Bay of Bengal. One can see much of the peninsular of India covered by negative rainfall anomalies, while positive rainfall anomalies push further north associated with cyclonic rainfall at the head of the Bay of Bengal. The pattern of anomalies suggests somewhere in the transition between an active and break phase for Indian monsoon rainfall (e.g., see Krishnamurthy and Shukla, 2007, Fig. 6). Over China, reductions in rainfall are seen, consistent with the northerly flow anomalies preventing the advection of moisture inland.

The dipole pattern in the SLP field (Fig. 7b) is associated with a cyclonic circulation over the north-western Indian Ocean/Arabian sea (showing strengthening of the monsoon trough) as well as intensification of the Somali jet near southern India. The rainfall pattern shows enhancement over much of the country, as is common during an active phase. Indeed the flow pattern over India is reminiscent of an active phase as in Krishnan et al. (2000). The anomalous anticyclonic circulation off the coast of China can also bring additional moisture across its southern coast, reflected in the strengthening of precipitation particularly in the north of China (Fig. 7d). One of the most important features seen in Fig. 7a and b is thus the large-scale nature of the patterns, associated with contributions from low-frequency variability (Fig. 6a). The other main feature is that these patterns are not very different from the leading EOFs (Fig. 2).

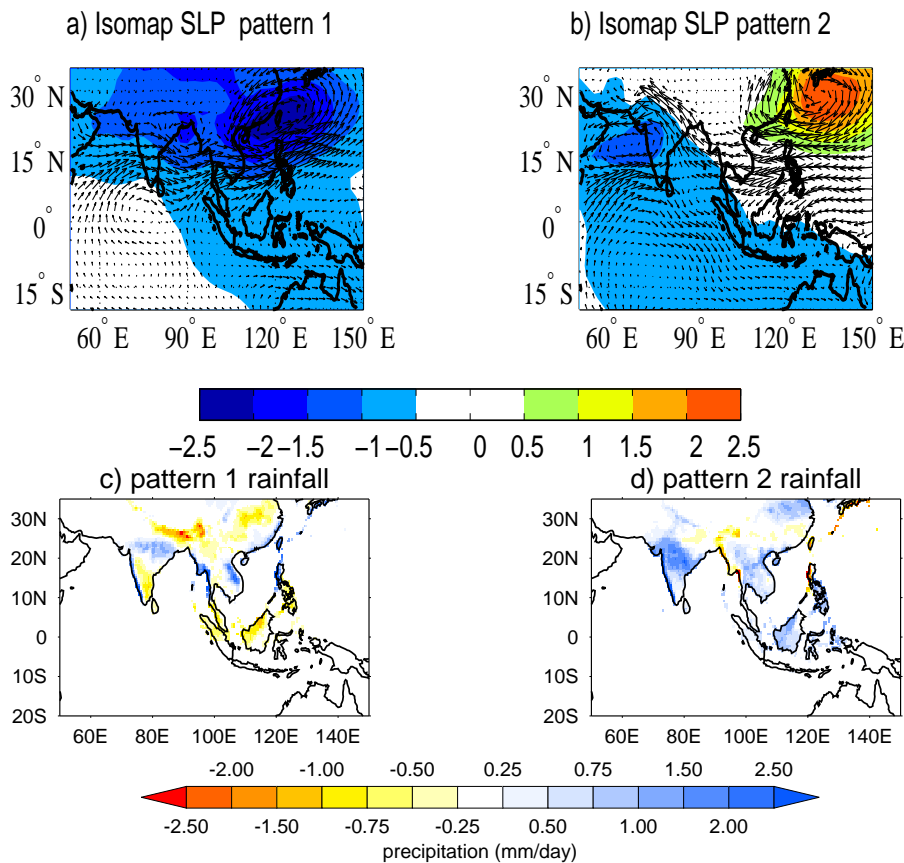
The correlation coefficients between the two leading PCs and the corresponding leading Isomap time series are a little larger than 0.9. In fact, the few leading PCs are highly correlated with the associated Isomap time series (Table 1). The difference, of course, between Fig. 2 and Fig. 7a and b is that the patterns in the latter figure are not exactly orthogonal. This is an example where one of the main properties of EOFs/PCs, i.e. orthogonality in space and non-correlation in time, is compromised as in orthogonal EOF rotation (Hannachi et al., 2007) except that rotation is linear but Isomap is not. This closeness between the leading modes of variability and the associated Isomap components shows that the nonlinear manifold is only slightly different from the linear EOF space. As is shown below, however, this subtle distinction is sufficient to reveal the nonlinear behaviour of ASM.

#### 4.1.3 Clustering of the leading modes

As for the PCs, we now investigate the PDF of ASM using the 2-D embedding space. Figure 6b shows the Gaussian kernel estimate of this PDF using the leading two Isomap time patterns shown in Fig. 6a. Unlike the PCs (Fig. 3), the 2-D PDF (Fig. 6b) is much more clearly bimodal as reported in Hannachi and Turner (2013). As was mentioned in that paper, the obtained bimodality is quite robust to reasonable changes of the number  $K$  around the value  $K = 12$  in the Isomap as well as changes in the ASM domain. We have also tested the Isomap method with OLR, and the



**Fig. 6.** The leading two Isomap time series from the SLP anomalies (a) and the associated kernel PDF estimate (b).



**Fig. 7.** Regression onto the first (a) and the second (b) Isomap time series of the SLP and 850 hPa wind anomalies. Units: hPa for SLP and maximum wind speed is  $2.5 \text{ m s}^{-1}$ . Regressions for APHRODITE daily precipitation are also shown in (c) and (d).

bimodality was particularly strong, but we choose not to use this field because of uncertainties especially prior to the satellite era when it was wholly modelled (see Hannachi and Turner, 2013). Here, however, we would like to go deeper in analysing the PDF rather than being content with the obtained feature. In fact, this bimodality suggests that the data within the low-dimensional Isomap space could well be clustered, a clear indication of nonlinear behaviour, for which

an appropriate test would be required. One way to test this bimodal behaviour is to use a Monte-Carlo-based bootstrap test as suggested by Silverman (1981), see e.g. Woollings et al. (2010b). A more elegant and robust way is to use a clustering test by applying concepts from homogeneous random point processes combined with copula (Stephenson et al., 2004; Hannachi, 2010; Hannachi et al., 2012). Clustering is a property that does not depend in general on the marginal

**Table 1.** Correlation coefficients between the leading 10 PCs and the Isomap time series of the SLP anomalies.

PCs and Isomap time series										
Number	1	2	3	4	5	6	7	8	9	10
Correlation	0.97	0.97	0.93	0.91	0.85	0.84	0.80	0.69	0.52	0.23

distributions of the data, and the method of copula attempts to factor out precisely the (undesired) effect of the marginal distributions. This is achieved by mapping each variable  $X_k$ ,  $k = 1, 2$ , onto the unit, or probability, interval  $[0, 1]$  using the cumulative distribution function  $F_k()$ , of the same variable, i.e.

$$U_k = F_k(X_k), \tag{9}$$

as a result of which  $U_k$ ,  $k = 1, 2$ , becomes uniform. Within this probability plane  $(U_1, U_2)$  an assessment of clustering is easily achieved using concepts from point processes (e.g., Ripley, 1976). Now, given any target point within this plane the mean number of points,  $n(d)$ , within a distance  $d$  from this point can be expressed as

$$n(d) = K(d)p(u_1, u_2), \tag{10}$$

where  $p(u_1, u_2)$  is the mean density of points and  $K(d)$  is a function of  $d$  known as *Ripley's K-function*. It has been shown (Ripley, 1976) that for a randomly uniform distribution  $K(d)$  reduces to the area of the circle of radius  $d$ . Hence the departure of the L-function:

$$L(d) = \sqrt{\frac{K(d)}{\pi d^2}}, \tag{11}$$

from  $L = 1$  provides evidence of departure from homogeneity.

Figure 8a shows a scatter plot of the leading two Isomap time patterns mapped onto the probability plane. The figure shows a region of low density around the diagonal separating two regions of higher density situated near the lower left and upper right parts of the scatter. We highlight these features more clearly later in Fig. 9a. This inhomogeneity is further quantified using the L-function shown in Fig. 8b. The shading represents the region delimited by the upper and lower envelopes of similar L-functions computed from 100 random samples of a homogeneous Poisson point process (uniform distribution on the unit square) with the same sample size as the data. The L-function is clearly outside the shading for distances  $d \leq 0.25$ , and provides evidence for the clustering and the presence of nonlinearity in the ASM.

In order to characterize the clustering, we show in Fig. 9a the kernel estimate of the data PDF within the same probability plane of Fig. 8a, where the two regions discussed above are singled out and now appear much more obvious. In addition, the PDF shows a further (third) peak, suggesting that

there may be three preferred modes<sup>3</sup>. To compute the most likely number of clusters we use the gap statistic (Tibshirani et al., 2001; Hannachi et al., 2011). For a given number of clusters,  $k$ , the sum of pairwise distances for all points in the  $m$ th cluster,  $D_m$ , with size  $n_m$ , for  $m = 1, \dots, k$  is computed and yields the so-called *within dispersion*

$$W_k = \sum_{m=1}^k \frac{D_m}{2n_m}. \tag{12}$$

The gap statistic is then given by

$$G(k) = E(\text{Log}(W_k^*)) - \text{Log}(W_k), \tag{13}$$

where  $W_k^*$  is the within dispersion computed from the null distribution, and  $E()$  is the expectation operator. A large sample of  $W_k^*$  is obtained, then its average  $\text{Log}(W_k^*)$  and standard deviation  $s_k$  are computed together. The optimum number of clusters,  $k_{\text{opt}}$ , is then given by the smallest  $k$  satisfying:

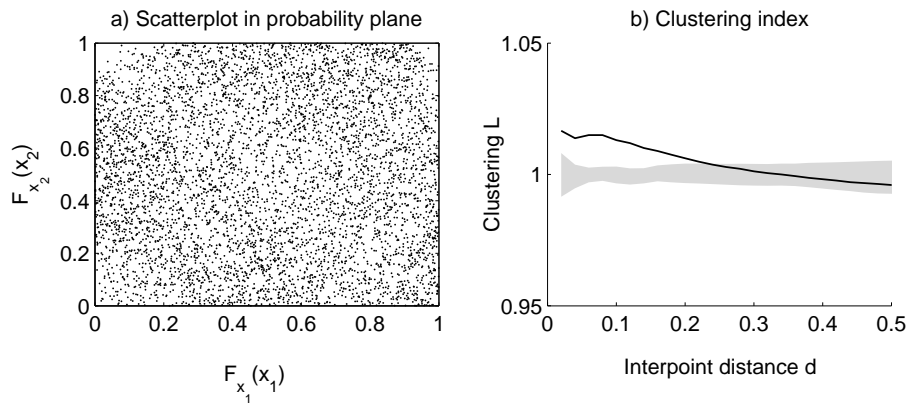
$$G(k) \geq G(k + 1) - s_{k+1}. \tag{14}$$

We have computed  $G()$  using both the 2-D Isomap time series with a (autocorrelated) Gaussian null, and the data mapped onto the probability plane with a homogeneous random point process for the null. Both consistently yield three clusters. Figure 9b shows the gap statistic computed using the probability plane using the data shown in Fig. 9a. The figure clearly shows that the gap at  $k = 3$  is greater than the upper bound of the gap at  $k = 4$  (see the condition in Eq. 14), and this does not occur for smaller  $k$ , i.e.  $k = 1, 2$ , supporting three clusters, which are investigated next.

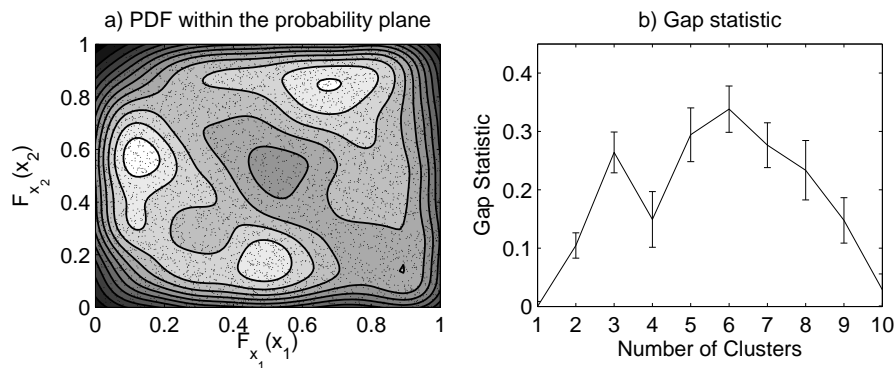
#### 4.2 Nonlinearity and ASM phases

In Hannachi and Turner (2013) the bimodality of the intraseasonal Asian monsoon PDF was discussed in terms of active versus break phases. The above result, however, lends strong evidence for three clusters. To consolidate these results further, we show in Fig. 10 the kernel estimate of the PDF within the Isomap plane along with a two- (Fig. 10a) and a three- (Fig. 10b) component mixture model (dashed line). There is clear evidence that the two-component model (Fig. 10a), suggested by Hannachi and Turner (2013), fails to match the kernel modes, whereas the three-component model

<sup>3</sup>We note here that the skewed PDF in the EOF space has a small shoulder but no signature of bimodality even within the probability space (not shown).



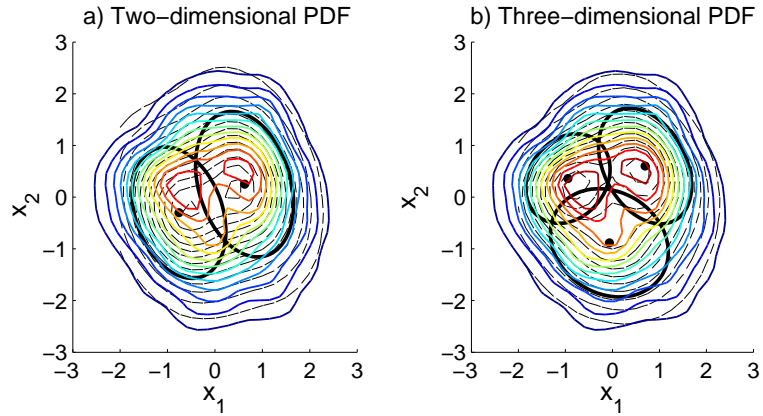
**Fig. 8.** Scatter plot of the leading two Isomap time series within the probability plane (a) and the associated clustering L-curve (b). The shading in (b) represents the upper and lower envelopes of clustering indexes obtained from 100 simulated homogeneous Poisson point process.



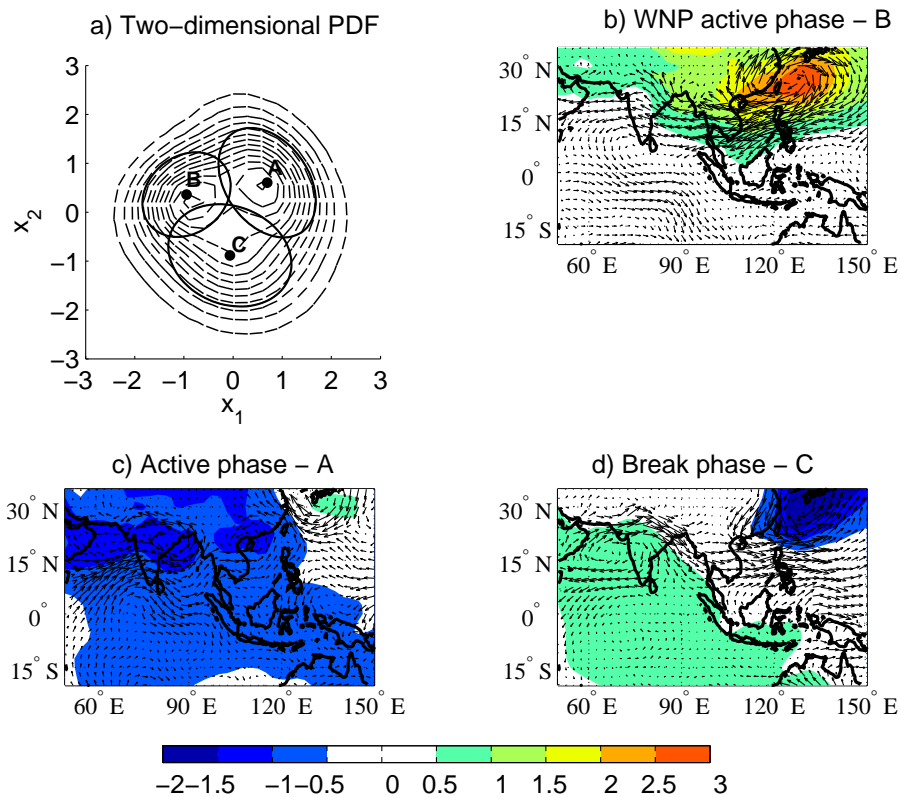
**Fig. 9.** Kernel PDF estimate of the leading two Isomap time series within the probability plane (a) and the gap statistic (b). The vertical bars in (b) refer to one standard deviation above and below the gap statistic based on the homogeneous Poisson point process null hypothesis (see text for details).

(Fig. 10b) matches very well the kernel modes. For example, the modes of the two-component mixture model PDF (Fig. 10a, dashed) do not match the modes of the kernel PDF (Fig. 10a, continuous), whereas the three-component PDF modes do (Fig. 10b). In addition, the PDF shoulder is well captured by the three-component model but not the two-component model. In order to identify the ASM states associated with the PDF modes we can use the three-component mixture model and analyse the individual components. Alternatively, we can use a simple composite analysis of the ASM states located near these modes. Both methods yield similar conclusions and we focus here on the mixture model. The black dots in Fig. 10b refer to the centres of the individual bivariate Gaussians of the mixture model and the ellipses represent the associated covariance matrices. Each covariance ellipse delimits around 84% of the total mass of the corresponding component, and is used to show the shape and orientation of the individual bivariate Gaussians.

The mixture model PDF with three Gaussian components is shown again in Fig. 11a, but without the kernel PDF, and each component is now labelled to enable clear discussion. The component centres are labelled A, B and C for what we describe here as the South Asian monsoon active phase, a western North Pacific (WNP) active phase and a South Asian monsoon break phase, respectively. The ASM phases associated with the centres of the bivariate Gaussian components are shown in Fig. 11b, c and d, respectively. They are obtained by averaging the SLP anomalies associated with the closest 300 states to the individual centres of the three-component mixture model (Fig. 11a). Superimposed on these maps are the associated 850 hPa wind averages. The first ASM phase (Fig. 11b) corresponds to the left-hand side centre of Fig. 11a, labelled B. It clearly shows a high pressure anomaly over most of the domain with enhancement over the north-eastern part (the WNP) accompanied by an anticyclonic wind circulation. This phase we describe here as the WNP active phase of the Asian monsoon.



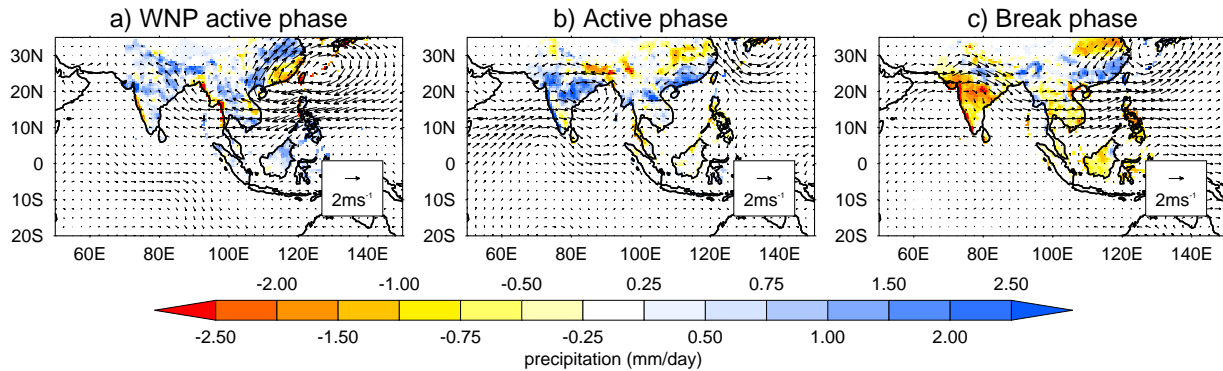
**Fig. 10.** Kernel PDF estimate of the leading two Isomap time series along with the two- (a) and three- (b) component gaussian mixture model. The mixture PDF is shown by the dashed lines. The covariance ellipses of each component along with the respective centres (black dots) are also shown.



**Fig. 11.** Three-component Gaussian mixture PDF of the leading two Isomap time series (a), composites of SLP and 850 hPa wind anomalies based on the 300 closest data points to the centres of the mixture PDF for the western North Pacific (b), the active (c) and the break (d) phases of the ASM. The labels A, B and C in (a) represent, respectively, the active, the WNP and the break phases. Contour interval 0.5hPa and the maximum wind speed for the break, active and China Sea active phases are, respectively,  $3 \text{ m s}^{-1}$ ,  $2 \text{ m s}^{-1}$  and  $2.3 \text{ m s}^{-1}$ .

The strong anticyclonic low level wind circulation becomes nearly easterly around  $20^\circ \text{ N}$ ,  $130^\circ \text{ E}$  and makes its way westward across Thailand and Cambodia through to India and the Arabian sea. Here it opposes and hence weakens the westerly and south-westerly jet responsible for monsoon activity.

There is also an indication of a cyclonic circulation slightly west of the southern tip of the Indian peninsula, but it is very weak and cannot generate a noticeable low pressure anomaly, a fact due most probably to the lack of geostrophy near the equator, compared to further poleward. We note that



**Fig. 12.** Composites of APHRODITE precipitation and 850 hPa wind anomalies based on the 300 closest data points to the centres of the mixture PDF for the WNP phase (a), the active phase (b) and the break (c) of the ASM. The unit wind vector is  $2 \text{ m s}^{-1}$ .

the anomalously strong western North Pacific high suggests enhanced convection over southern China. This zonal pattern of west-to-east reduced then enhanced convection is part of a wider quadrupole pattern of intraseasonal monsoon variability (see, for example, the composite observations in Fig. 9a of Sperber and Annamalai, 2008).

The second ASM phase (Fig. 11c), which corresponds to the right-hand side centre in Fig. 11a and labelled A, shows a low pressure anomaly over most of the domain with high pressure anomaly over a smaller region in the north-eastern part near to Japan. Note in particular the low pressure anomalies and associated cyclonic circulation over the Arabian Sea, the Bay of Bengal/north-west India and north-western part of Vietnam. This flow regime corresponds to the active phase of the ASM. The anticyclonic wind circulation in this case is localized quite close to the high pressure anomaly centre, and the easterly wind in the eastern part of the domain is blocked by the Indonesian islands and is deviated southward.

The last ASM phase (Fig. 11d) shows a high pressure anomaly over most of the domain with an anomalous low pressure system localized mostly over the north-eastern part. We note that the circulation anomalies in the Indian region are approximately opposite to those in the active phase (Fig. 11c) and so we term this the break phase. Similarly, circulation anomalies over the western North Pacific are cyclonic and roughly opposite to those in the WNP phase (Fig. 11b). Figure 10 (see also Fig. 11a) shows that the modes of the kernel estimate and the mixture models are quite close to the active and WNP phases, which means that these two phases dominate the ASM dynamics. The mixing proportions (Eq. 5) of the WNP phase (Fig. 11b), the active phase (Fig. 11c) and the break phase (Fig. 11d) of the ASM are respectively 26 %, 38 % and 36 %. The analysis presented here shows that the monsoon over South Asia is closely connected to that over Southeast and East Asia in a complex manner on intraseasonal timescales (see, e.g., Annamalai and Sperber, 2005). This points to the fact that the Asian monsoon system should be looked at in a wider context, bringing in the

possible influence of the large-scale on the intraseasonal variability and regimes of ASM convection, which is discussed next.

For comparison, we also plot the composite phases using the 300 closest data point method for the APHRODITE rainfall, as shown in Fig. 12. For ease of interpretation, we show again the wind field as in Fig. 11. In Fig. 12a, the WNP phase shows a large-scale reduction in the strength of the westerly flow across the Bay of Bengal, Southeast Asia and western North Pacific, driven by strengthening of the western North Pacific high. For India and Burma, reduction in the mean westerlies on the west coast causes negative anomalies in the orographic rainfall there. Meanwhile, inflow from the Bay of Bengal enhances rainfall over northern and eastern India. In China, the enhanced high strengthens the convergence inland, moving rainfall further from the coast and intensifying it. The flow pattern in Fig. 12b, which we noted indicates an active phase for India, shows enhanced rainfall on the west coast of India, and in north-east India associated with the cyclonic anomaly. Meanwhile in China, rainfall shifts southwards associated with the anomalous divergence at around  $30^\circ \text{ N}$ . In the break phase (Fig. 12c), the large-scale weakening of the monsoon trough over northern India and weakening of the Somali jet near southern India lead to a large negative rainfall anomaly there, consistent with a break phase in the Indian monsoon. Over China, the cyclonic circulation anomaly leads to added convergence and rainfall in southern China, opposite to that shown in Fig. 12a.

Composites of the three-component mixture of the leading two Isomap time series are thus consistent between SLP, lower tropospheric winds and precipitation.

## 5 ASM and large-scale forcing

It has been suggested that the primary reason behind the lack of dynamical predictability of the monsoon is poor simulation of teleconnections between the monsoon and large-scale boundary forcing in GCMs due to the dominance of model

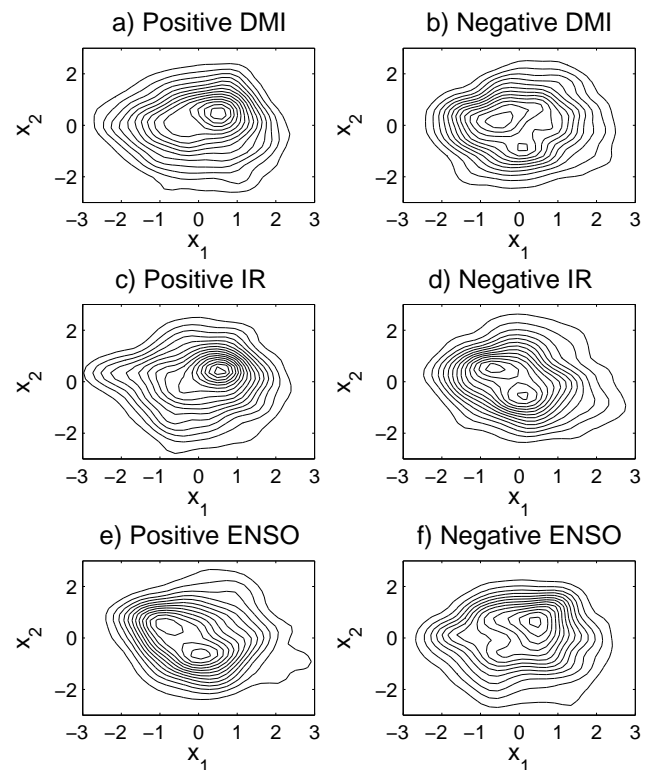
errors (Sperber et al., 2000; Sperber and Palmer, 1996). However, we now understand more on the role of systematic mean state biases acting detrimentally against monsoon-ENSO teleconnections and inhibiting seasonal prediction (Turner et al., 2005; Annamalai et al., 2007). In addition, more recent comparisons of coupled models used in seasonal prediction mode in the EU-ENSEMBLES project reveal skill levels of around 0.45 in the multi-model mean (Rajeevan et al., 2012). Nonetheless, predictability may still be limited by the lack of understanding of the connections between monsoon variability on shorter and longer timescales.

Although Sperber et al. (2000) did not find bimodality in the monsoon intraseasonal variability (MISV), they found that the PDF mean of EOF3/PC3 was systematically perturbed during weak and strong monsoon years categorized in terms of seasonal mean all-India rainfall (AIR). Here we investigate the relationship between regimes of MISV and the large-scale seasonal mean monsoon based on measures of both the dynamical monsoon (using the WY dynamical monsoon index; DMI) and our own area-averaged Indian rainfall (IR). We also relate MISV to the most effective slowly varying lower boundary condition acting to perturb the seasonal mean monsoon, ENSO, using the summer (JJAS) Niño-3 index.

To characterise these relationships, we simply classify the daily SLP anomalies within the Isomap space, according to the seasonal (JJAS) means of large-scale measures being larger (smaller) than one (minus one) standard deviation  $\pm\sigma$  (as in TH10). The large-scale measures used are IR, DMI or ENSO. For example, if IR is larger than one standard deviation in a given season then all SLP data during that season are classified as IR+.

### 5.1 Relationship with seasonal mean dynamical monsoon

Figure 13 shows the two-dimensional PDF of the ASM using the leading two Isomap time series during the positive (Fig. 13a) and negative (Fig. 13b) phases of the DMI. Although the PDF during both phases combined (not shown) is not bimodal but skewed towards the right-hand side mode (Figs. 10b and 11a), i.e. the active phase, it is clear that during the positive DMI phase (Fig. 13a) we have a positive skewness associated with a substantial increase of probability of the active monsoon phase consistent with the results in Hannachi and Turner (2013). During the negative DMI phase (Fig. 13b), on the other hand, we have a tendency for an increase in frequency of the WNP active monsoon phase, with its associated large-scale easterly anomalies over India. This is consistent with the definition of the WY index, based on large-scale zonal wind shear in the vertical. The change of frequency, however, is not symmetrical between the positive and negative DMI phases, with much higher probability of the active phase when the DMI is positive. This behaviour is reminiscent of the paradigm put forward by



**Fig. 13.** Kernel PDF estimates of the leading two Isomap time series during positive (a, c, e) and negative (b, d, f) phases of the dynamical monsoon (DMI/WY) index (a, b), area-averaged India rainfall, IR, (c, d) and ENSO (e, f).

Palmer (1999) regarding the change of regime frequency under forcing changes in a chaotic system. Palmer (1994) also suggests that the seasonal mean condition relates to preference for a particular monsoon phase. Here we see that seasons with strong broad-scale monsoon heating are related to a much higher likelihood for the active monsoon phase. Note also that during the negative DMI phase (Fig. 13b) there is a slight enhancement of the probability of the third monsoon phase (Fig. 11d), with break circulation and precipitation (Fig. 12c) over India. We also note here that TH10 found about equal probabilities of both the break/active monsoon phases during positive DMI phase, but with a greater probability of break events in the DMI negative phase (see Fig. 3b of TH10).

### 5.2 Relationship with seasonal mean Indian rainfall

A similar analysis is applied using our IR index. The index is scaled to zero-mean and unit standard deviation, and periods when it is greater (smaller) than 1 ( $-1$ ) are selected and referred to as positive (negative) IR. Figure 13 shows the kernel PDF estimate of ASM during positive (Fig. 13c) and negative (Fig. 13d) IR phases. During positive IR (Fig. 13c) the PDF is positively skewed with large frequency increase of



the active phase at the expense of the remaining two phases. The SLP anomaly associated with this mode (not shown) is similar to the active phase of Fig. 11c, but is much stronger.

During the negative phase of IR (Fig. 13d) the PDF is clearly bimodal. The two modes represent, respectively, the second and the third monsoon regimes, that is the WNP active phase and break phase of the ASM (compare the position to Figs. 10b and 11a). The emergence of the break phase during the negative IR phase is not surprising, but it is also interesting to see an increase of probability of the western North Pacific active phase. Note, in addition, there is equal probability of the break phase and the western North Pacific active phase of ASM during negative IR conditions. The mode associated with the WNP active phase (top left mode of Fig. 13d) is quite similar to the mode shown in Fig. 11b, but slightly weaker in amplitude. The second (bottom right) mode of Fig. 13b, associated with negative IR (not shown) is quite similar to the break phase shown in Fig. 11d.

### 5.3 Influence of ENSO

It is well known that ENSO is one of the fundamental predictors of seasonal mean monsoon rainfall in the tropics (see, e.g., Webster and Yang, 1992, among many others), and that this can be simulated to some degree in models (Annamalai et al., 2007). While ENSO is known to lend predictability to seasonal mean, countrywide rainfall anomalies for India (Krishnamurthy and Shukla, 2000, 2007), its role in perturbing MISV is uncertain and offers the prospect of enhanced predictability of monsoon rainfall.

Here we investigate this issue by applying the same methodology as with seasonal mean indices IR and DMI. The PDF of the ASM using the two Isomap time series during the positive ENSO phase is shown in Fig. 13e and that during the negative ENSO phase is shown in Fig. 13f. During negative ENSO the PDF has a positive skewness associated with an increase of probability of the active monsoon phase, suggesting that the large-scale slowly varying boundary conditions are also lending predictability to intraseasonal modes. The SLP anomaly associated with this mode (not shown) is similar to the active phase (Fig. 11c) with a particularly low pressure anomaly system of about  $-2$  hPa around northern Australia, the Indonesian archipelago, India and the Arabian Sea. The positive SLP anomaly is located in the upper right corner of the domain.

During the positive ENSO phase (Fig. 13e) the PDF is bimodal with modes coinciding with the modes associated with the break phase and the western North Pacific active phase. There is nearly equal likelihood for both these phases during positive ENSO phases. The SLP anomalies associated with these modes (not shown) compare well with the break and the western North Pacific active phases, and also with the modes obtained during the negative AIR phase. The cases associated with positive and negative phases of ENSO are

equivalent respectively to those associated with negative and positive IR phases, and those in DMI.

## 6 Summary and discussion

The Asian summer monsoon is one of the most large-scale atmospheric phenomena involving one of the largest moisture volume transports from the ocean to land. This moisture transport is what makes the monsoon special as it is the main source of rainfall for the most populated region on earth. Rather than the summer seasonal mean rainfall, which is relatively constant, it is the variability in the monsoon on intraseasonal timescales that has most impact on society and leads to the importance of studying the dynamics and predictability of the Asian monsoon. The ASM is a highly nonlinear and high dimensional phenomenon. One way to understand the dynamics of ASM is to find ways to reduce the dimensionality of the system in a way that could help capture the main features of its nonlinear behaviour.

We have investigated the ASM variability using nonlinear dimensionality reduction based on Isomap of ERA-40 SLP anomalies over the Asian monsoon region ( $50$ – $145^\circ$  E,  $20^\circ$  S– $35^\circ$  N) for the summer season June–September (JJAS) 1958–2001. In this study, we have revisited and systematically extended Hannachi and Turner (2013). The Isomap projection technique is based on computing local geodesic distances between the atmospheric states. The data points are first connected by a graph based on the  $K$  nearest points, then distances between any two states are computed based on the previous graph. Multidimensional scaling (MDS) is then used to get an Isomap embedding. We have used  $K = 12$ , and considered a two-dimensional embedding space. A kernel PDF estimate is fitted to the data and reveals bimodality, compared to the unimodal PDF obtained within the EOF space. Furthermore, the data are shown to be clustered and support three clusters. The ASM phases associated with these clusters are determined using a three-component bivariate Gaussian mixture model.

The first mode corresponds to an active phase in the WNP. It is associated with a high pressure anomaly over most of the domain with a particular increase of SLP in the western North Pacific. The associated low-level wind field has a westward component that extends through to the Indian Ocean and Arabian Sea, which contributes also to a large-scale weakening of the Somali jet. The second mode is associated with a typical monsoon active phase for India. This mode has a dipole with high pressure anomaly around the East China Sea and a low pressure anomaly elsewhere, with a particular enhancement over the Arabian Sea and India. The low-level flow is characterised by an anti-cyclonic circulation over the high-pressure anomaly, an easterly flow west of  $120^\circ$  E, and an extension of the Somali jet eastward. The last ASM phase found here represents a classic Indian monsoon break phase. It is also associated with an anomalous low

pressure centre over the China and Philippine Seas extending north-eastward into China.

We have also investigated the effect of the large scale seasonal monsoon and large-scale boundary forcing using seasonal mean area-averaged Indian rainfall (IR), large-scale monsoon heating represented by the dynamical monsoon index (DMI) and an ENSO index. During the positive phase of the DMI, we find substantial increased probability of the active phase whereas during the opposite phase, we find a significantly increased probability of the break phase and active WNP phases of the ASM.

During the positive phase of IR we have, as expected, an increase of probability of the active phase. During the negative IR phase, however, we get equal likelihood of the break phase and the western North Pacific active phase. A similar behaviour, with exchanged roles, is obtained with the ENSO phases, i.e., the positive phase of ENSO is equivalent to the negative phase of IR and vice-versa, as expected. This supports the conceptual idea that large-scale lower boundary forcing (such as ENSO) exerts influence on modes of MISV, thereby extending its predictive influence beyond seasonal mean rainfall anomalies alone (Charney and Shukla, 1981). By further understanding these relationships between the large-scale and regimes of intraseasonal monsoon convection we hope to increase the prospects of dynamical prediction of the monsoon. For example, the analysis of the intraseasonal ASM predictability in relation to large-scale forcing permits us, via conditional probabilities, to compute probabilistic forecast of the state of the monsoon given the state of the large-scale flow. This is certainly of benefit to society and agriculture.

In future work we plan to investigate these relationships in long integrations of state-of-the-art coupled GCMs, judged capable of simulating the full spectrum of monsoon variability as suggested in Turner and Annamalai (2012), and also study the behaviour of ASM phases in a warmer climate (Schewe and Levermann, 2012).

*Acknowledgements.* The authors would like to thank Jacob Schewe and an anonymous reviewer for their constructive comments. A. G. Turner was funded by a NERC Fellowship (NE/H015655/1).

Edited by: J. Kurths

Reviewed by: J. Schewe and one anonymous referee

## References

- Annamalai, H., Hamilton, K., and Sperber, K. R.: The South Asian monsoon and its relationship with ENSO in the IPCC AR4 Simulations, *J. Climate*, 20, 1071–1092, 2007.
- Blanford, H. F.: On the connection of the Himalaya snowfall with dry winds and seasons of drought in India, *Proc. Roy. Soc. London*, 37, 1–23, 1884.
- Boos, W. R. and Kuang, Z.: Dominant control of the south Asian monsoon by orographic insulation versus plateau heating, *Nature*, 436, 218–222, 2010.
- Borg, I. and Groenen, P.: *Modern Multidimensional Scaling, Theory and Applications*, Springer Verlag, New York, 1997.
- Brankovich, C. and Palmer, T.: Seasonal skill and predictability of ECMWF PROVOST ensembles, *Quart. J. Roy. Meteorol. Soc.*, 126, 2035–2067, 2000.
- Carl, P.: Monsoon dynamics in a low-dimensional GCM, WCRP-84, WMO/TD-No 619, Volume II, 773–780, 1994.
- Charney, J. G. and Shukla, J.: *Monsoon Dynamics: Predictability of Monsoons*, edited by: Lighthill, J. and Pearce, R., Cambridge University Press, Cambridge, UK, 1981.
- Everitt, B. S. and Hand, D. J.: *Finite Mixture Distributions*, Chapman and Hall, London, 143 pp., 1981.
- Gómez, A. J., Zhou, C. S., Timmermann, A., and Kurths, J.: Nonlinear dimensionality reduction in climate data, *Nonlin. Processes Geophys.*, 11, 393–398, doi:10.5194/npg-11-393-2004, 2004.
- Goswami, B. N.: Chaos and predictability of the Indian summer monsoon, *Pramana J. Phys.*, 48, 719–736, 1997.
- Hannachi, A.: Pattern hunting in climate: a new method for finding trends in gridded climate data, *Int. J. Climatol.*, 27, 1–15, 2006.
- Hannachi, A.: Tropospheric planetary wave dynamics and mixture modeling: Two preferred regimes and a regime shift, *J. Atmos. Sci.*, 64, 3521–3541, 2007.
- Hannachi, A.: On the origin of planetary-scale extratropical winter circulation regimes, *J. Atmos. Sci.*, 67, 1382–1401, 2010.
- Hannachi, A. and O'Neill, A.: Atmospheric multiple equilibria and non-Gaussian behaviour in model simulations, *Quart. J. Roy. Meteor. Soc.*, 127, 939–958, 2001.
- Hannachi, A. and Turner, A. G.: Isomap nonlinear dimensionality reduction and bimodality of Asian monsoon convection, *Geophys. Res. Lett.*, 40, 1653–1658, 2013.
- Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol.*, 27, 1119–1152, 2007.
- Hannachi, A., Mitchell D., Gray L., and Charlton-Perez A.: On the use of geometric moments to examine the continuum of sudden stratospheric warmings, *J. Atmos. Sci.*, 68, 657–674, 2011.
- Hannachi, A., Woollings T., and Fraedrich, K.: The North Atlantic jet stream: Preferred positions, paths and transitions, *Quart. J. Roy. Meteor. Soc.*, 138, 862–877, 2012.
- Hsieh, W. W.: Nonlinear principal component analysis by neural networks, *Tellus*, 53A, 599–615, 2001.
- Jolliffe, I. T.: *Principal Component Analysis*, Springer, second Edition, New York, 2002.
- Krishnamurthy, V. and Shukla, J.: Intraseasonal and interannual variability of rainfall over India, *J. Climate*, 13, 4366–4377, 2000.
- Krishnamurthy, V. and Shukla, J.: Intraseasonal and seasonally persisting patterns of Indian monsoon rainfall, *J. Climate*, 20, 3–20, 2007.

- Krishnan, R. and Zhang, C., and Sugi, M.: Dynamics of Breaks in the Indian Summer Monsoon, *J. Atmos. Sci.*, 57, 1354–1372, 2000.
- Li, C. and Yanai, M.: The onset and interannual variability of the Asian summer monsoon in relation to land–sea thermal contrast, *J. Climate*, 9, 58–375, 1996.
- Lorenz, E. N.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- McLachlan, G. J. and Basford, K. E.: *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, Inc. New York, 1988.
- Meehl, G. A.: Coupled ocean-atmosphere-land processes and south Asian monsoon variability, *Science*, 265, 263–267, 1994.
- Monahan, A. H.: Nonlinear principal component analysis: Tropical Indo-Pacific sea surface temperature and sea level pressure, *J. Climate*, 14, 219–233, 2001.
- Palmer, T. N.: Chaos and predictability in forecasting the monsoons, *Proc. Indian Natl. Sci. Acad.*, 60, 57–66, 1994.
- Palmer, T. N.: A nonlinear dynamical perspective on climate prediction, *J. Climate*, 12, 575–591, 1999.
- Rajeevan, M., Bhate, J., Kale, J. D., and Lal, B.: High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells, *Current Sci.*, 91, 296–306, 2006.
- Rajeevan, M., Pai, D. S., Anil Kumar, R., and Lal, B.: New statistical models for long-range forecasting of southwest monsoon rainfall over India, *Clim. Dynam.*, 28, 813–828, 2007.
- Rajeevan, M., Unnikrishnan, C. K., and Preethi, B.: Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability, *Clim. Dynam.*, 38, 2257–2274, 2012.
- Rangarajan, G. and Sant, D. A.: Fractal dimensional analysis of Indian climatic dynamics, *Chaos Soliton Fractals*, 19, 285–291, 2004.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108, 4407, doi:10.1029/2002JD002670, 2003.
- Ripley, B. D.: The second-order analysis of stationary point processes. *J. Appl. Probability*, 13, 255–266, 1976.
- Ross, I., Valdes, P. J., and Wiggins, S.: ENSO dynamics in current climate models: an investigation using nonlinear dimensionality reduction, *Nonlin. Processes Geophys.*, 15, 339–363, doi:10.5194/npg-15-339-2008, 2008.
- Roweis, S. T. and Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290, 2323–2326, 2000.
- Rust, W. H., Vrac M., Lengaigne M., and Sultan, B.: Quantifying differences in circulation patterns based on probabilistic models: IPCC AR4 multimodel comparison for the North Atlantic, *J. Climate*, 23, 6573–6589, 2010.
- Schewe J., and Levermann, A.: A statistically predictive model for future monsoon failure in India, *Environ. Res. Lett.*, 7, 044023, doi:10.1088/1748-9326/7/4/044023, 2012.
- Silverman, B. W.: Using kernel density estimates to investigate multimodality. *J. R. Statist. Soc.*, 43, 97–99, 1981.
- Sperber, K. R. and Annamalai, H.: Coupled model simulations of boreal summer intraseasonal (30–50 day) variability, Part 1: Systematic errors and caution on use of metrics, *Clim. Dynam.*, 31, 345–372, 2008.
- Sperber, K. R. and Palmer, T. N.: Interannual tropical rainfall variability in general circulation model simulations associated with the Atmospheric Model Intercomparison Project, *J. Climate*, 9, 2727–2750, 1996.
- Sperber, K. R., Slingo, J. M., and Annamalai, H.: Predictability and the relationship between subseasonal and interannual variability during the Asian summer monsoon, *Q. J. R. Meteorol. Soc.*, 126, 2545–2574, 2000.
- Stephenson, D. B., Hannachi, A., and O'Neill, A.: On the existence of multiple climate regimes, *Q. J. Roy. Meteor. Soc.*, 130, 583–605, 2004.
- Straus, D. and Krishnamurthy, V.: The preferred structure of the interannual Indian monsoon variability, *Pure Appl. Geophys.*, 164, 1717–1732, 2007.
- Tenenbaum, J. B., de Silva, V., and Langford, J. C.: A global geometric framework for nonlinear dimensionality reduction, *Science*, 290, 2319–2323, 2000.
- Tibshirani, R., Walther, G., and Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, *J. Roy. Statist. Soc. B*, 63, 411–423, 2001.
- Torgerson, W. S.: *Multidimensional scaling I: Theory and method*, *Psychometrika*, 17, 401–419, 1952.
- Tschantcher, I., Worbs, K. D., and Carl, P.: Frequency drift and retreat variability of a GCM's monsoon oscillator, WCRP-84, WMO/TD-No 619, Volume II, 781–788, 1994.
- Turner, A. G. and Annamalai, H.: Climate Change and the South Asian summer monsoon, *Nature Climate Change*, 2, 587–595, 2012.
- Turner, A. G. and Hannachi, A.: Is there regime behavior in monsoon convection in the late 20th century? *Geophys. Res. Lett.*, 37, L16706, doi:10.1029/2010GL044159, 2010.
- Turner, A. G. and Slingo, J. M.: Subseasonal extremes of precipitation and active-break cycles of the Indian summer monsoon in a climate change scenario, *Q. J. R. Meteorol. Soc.*, 135, 549–567, 2009.
- Turner, A. G., Inness, P. M., and Slingo, J. M.: The role of the basic state in the ENSO-monsoon relationship and implications for predictability, *Q. J. R. Meteorol. Soc.*, 133, 1143–1157, 2005.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 Re-Analysis, *Quart. J. Roy. Meteorol. Soc.*, 131, 2961–3012, 2005.
- von Storch, H. and Zwiers, F. W.: *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 1999.
- Walker, G. T.: Correlation in seasonal variations of weather II, *Mem. Indian Meteor. Dept.*, 21, 445–452, 1910.
- Walker, G. T.: Correlation in seasonal variations of weather, VII: The local distribution of monsoon rainfall, *Mem. Indian Meteor. Dept.*, 21, 12–15, 1922.

- Webster, P. J., and Yang, S.: Monsoon and ENSO—Selectivity interactive systems, *Quart. J. Roy. Meteorol. Soc.*, 118, 877–926, 1992.
- Webster, P. J., Magana, V. O., Palmer, T. N., Shukla, J., Tomas, R. A., Yanai, M., and Yasunari, T.: Monsoons: Processes, predictability, and the prospects for prediction, *J. Geophys. Res.*, 103, 14451–14510, 1998.
- Woollings, T., Hannachi, A., Hoskins, B. J., and Turner, A. G.: A regime view of the North Atlantic Oscillation and its response to anthropogenic forcing, *J. Climate*, 23, 1291–1307, 2010a.
- Woollings, T., Hannachi, A., and Hoskins, B. J.: Variability of the North Atlantic eddy-driven jet stream, *Q. J. Roy. Meteor. Soc.*, 136, 856–868, 2010b.
- Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a Long-term Daily Gridded Precipitation Dataset for Asia based on a Dense Network of Rain Gauges, *Bull. Amer. Meteor. Soc.*, 93, 1401–1415, doi:10.1175/BAMS-D-11-00122.1, 2012.
- Young, G. and Householder, A. S.: Discussion of a set of points in terms of their mutual distances, *Psychometrika*, 3, 19–22, 1938.