

Spatial structures and directionalities in Monsoonal precipitation over South Asia

N. Malik^{1,2}, N. Marwan¹, and J. Kurths^{1,3}

¹Potsdam Institute for Climate Impact Research, P.O. Box 601203, 14412, Potsdam, Germany

²Institute of Physics, University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Germany

³Department of Physics, Humboldt University, Newtonstr. 15, 12489 Berlin, Germany

Received: 20 January 2010 – Revised: 11 August 2010 – Accepted: 17 August 2010 – Published: 1 September 2010

Abstract. Precipitation during the monsoon season over the Indian subcontinent occurs in form of enormously complex spatiotemporal patterns due to the underlying dynamics of atmospheric circulation and varying topography. Employing methods from nonlinear time series analysis, we study spatial structures of the rainfall field during the summer monsoon and identify principle regions where the dynamics of monsoonal rainfall is more coherent or homogenous. Moreover, we estimate the time delay patterns of rain events. Here we present an analysis of two separate high resolution gridded data sets of daily rainfall covering the Indian subcontinent. Using the method of event synchronization (ES), we estimate regions where heavy rain events during monsoon happen in some lag synchronised form. Further using the delay behaviour of rainfall events, we estimate the directionalities related to the progress of such type of rainfall events. The Active (break) phase of a monsoon is characterised by an increase(decrease) of rainfall over certain regions of the Indian subcontinent. We show that our method is able to identify regions of such coherent rainfall activity.

1 Introduction

Monsoons are among the most prominent and dynamic phenomena of the climate system, manifesting over large parts of the tropics. The origin of monsoons is in the seasonal reversal of wind directions due to the strong differential heating of land and the surrounding tropical Oceans. This in turn creates a very strong seasonality in rainfall at the annual scale. In particular, the Indian subcontinent receives over 75% of its annual rainfall over the

year during the four summer months of June, July, August, and September(JJAS). As a consequence, the monsoon tends to have a large socioeconomic impact on the inhabitants of the region. Hence, understanding monsoon and its variability on different time scales is of considerable importance in climate sciences and in society (Webster et al., 1998; Gadgil, 2003).

The summer monsoon over South Asia does not only show very strong temporal variations on multiple time scales but also an enormous spatial variability. During its active phase monsoon could evolve on large spatial scales within a short period of time and hence large parts of the land mass could receive rainfall simultaneously or within short time delay of a few days. Active periods of monsoon seem to be formed as a cumulative effect of several events scattered over large spatial scales (100–1000 km), which may also include massive rain events resulting in 400 mm of rainfall at a single day (May, 2004; Stephenson et al., 1999). Such massive events take place due to the formation of monsoon depressions (Sikka, 1977) or midtropospheric cyclones (Keshavamurthy, 1973).

The spatial variability and patterns of monsoon over South Asia is governed by both dynamics of monsoonal circulations and orography. The influence of orography on monsoon is most visible in regions of Western Ghats and Himalayas. Also, monsoonal rainfall is not localized at a particular point in space rather it covers a certain region in form of many events spread over large regions. So, it is very important to identify such regions. We here employ the method of event synchronization (ES) to find out spatial regions where the specific type of rain events occur either simultaneously or within certain delay. In this way we are trying to regionalize the precipitation field. This kind of regionalization has been endeavoured on precipitation with different methods (Iyengar and P.Basak, 1994; Gadgil et al.,



Correspondence to: N. Malik
(malik@pik-potsdam.de)

1993; Gutierrez et al., 2006). The main idea behind our approach is not to find regions of similar variability but regions of similar dynamics of rainfall.

Another aspect of monsoonal rainfall are inter-seasonal oscillations (ISO). They are composed of two distinct phases, known as the “Active Phase” and “Break Phase”. During an active period heavy rainfall appears over most parts of India; in a break period the phase reverses and rainfall is recorded only along the Himalayan foothills, south-east peninsular India (Singh et al., 1992; Krishnamurthy and Shukla, 2000). We try to understand these phenomena using ES and to find out those regions where monsoonal activity has higher order of synchronization than others.

In the next section we introduce the data sets used in the analysis. In Sect. 3 we discuss the climatology of the region. In Sect. 4 we present the concept of event synchronization, algorithm used for clustering and the method used to derive directionalities from the data. Then we proceed with results, discussion and summary in Sects. 5 and 6, respectively.

2 Data

The main data set we have used is a daily precipitation gridded data set for 1961–2004 developed as part of the project – Asian Precipitation Highly Resolved Observational Data Integration Towards the Evaluation of Water Resources (APHRODITE) (Yatagai et al., 2009). It is freely downloadable from the website – <http://www.chikyu.ac.jp/precip/>. We have extracted the data for the South Asian region from the 0.5 degree resolution data set (APHRO-V0902) for monsoon Asia. We will refer this data set as APHRO-V0902 in the paper. For the purpose of comparing our results we have also used another high resolution precipitation data set collected and developed by the Indian Meteorological Department (IMD, Pune) from 1951–2004. (Rajeevan et al., 2006). It has a daily, one degree latitude, and one degree longitude resolution. But its spatial coverage includes only the political boundary of India. In the following we will refer the data as IMD-D. In both these data sets the density of rain gauge stations varies over the whole subcontinent. There are certain regions which do not seem to be well represented, for example one of such regions is the state of Jammu and Kashmir in IMD-D. Where the number of stations are 2 to 4 but the area interpolated is 23 grid points which is roughly equivalent to 2500 km². Hence we have tried to exclude this region in our analysis in IMD-D. For the description of the interpolation method, the exact locations of rain gauge stations, and the quality of raw data used in IMD-D see the references Rajeevan et al. (2006), and Rajeevan et al. (2005). For further comparison with wind directions during certain type of rainfall events, we have used the NCEP/NCAR reanalysis data from 1951–2004 with 2.5 degree resolution, provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, available at <http://www.esrl.noaa.gov/psd/>.

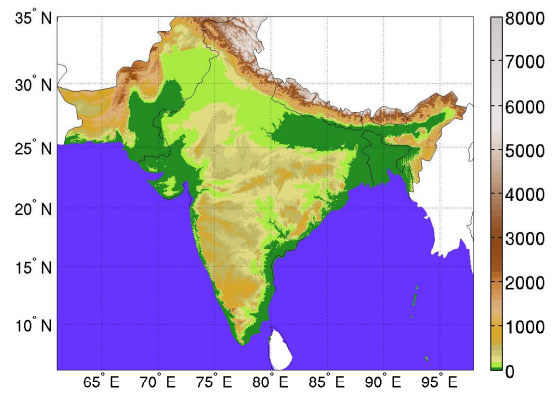


Fig. 1. Topographic map of the region concerned in the data set and the analysis. The colour represents the height above sea level in meters. This map is obtained from ETOPO1, the 1 arc min gridded global relief data provided by NOAA.

3 Patterns of mean and variability of monsoonal rainfall

Before describing further details of methods and results, it is essential to understand the climatology of the region analysed and especially to know the critical regions of monsoonal rainfall over South Asia.

The Indian monsoonal system covers some of the wettest to the driest regions on Earth. It shows an immense amount of spatial variability over the South-Asian subcontinent in terms of net amount of rainfall over a season. The north east and western coast along the Western Ghats are the wettest regions, whereas north west is the driest. We observe that rainfall regions on annual scale ranges from below 150 mm per season to above 2100 mm per season during summer monsoon (JJAS months). Apart from the monsoonal circulation themselves, the orographic forcing drives this spatial variability too. The strongest topographic barriers exist along the western coast in form of the Western Ghats and in the north and north east in form of Himalayas (Fig. 1).

In the first step we coarse grain the mean seasonal rainfall during JJAS at each grid point into 15 equally distributed partitions with each having a range of 300 mm. Then we merge the top 6 partitions into one partition; i.e. we club together the regions having rainfall above 2400 mm into a single partition. So, now we have 9 partitions of mean seasonal rainfall, each plotted with different colour in Fig. 2. One aspect that is obvious from this plot is a very strong gradient in mean rainfall when we cross the Western Ghats from west to east. The regions on the eastern side of the Western Ghats are quite dry, whereas on the western side we have some of the wettest regions. So, western ghats does act as a very strong orographic barrier to the moisture coming from the Arabian Sea branch of the Indian Monsoon. A similar barrier exists in form the Himalayas all along the eastern boundary of the region (Bookhagen and Burbank,

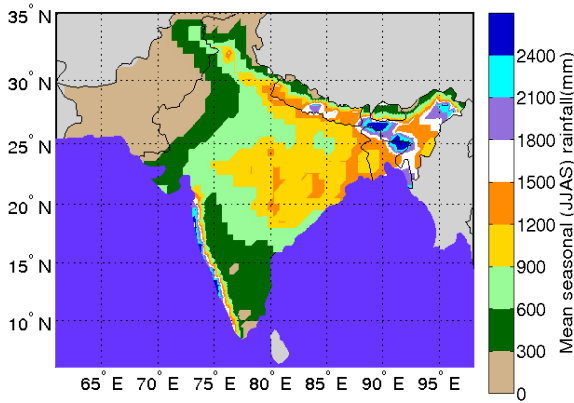


Fig. 2. Partitions of mean seasonal rainfall based on the data analysed from 1961 to 2004 (data used APHRO-V0902).

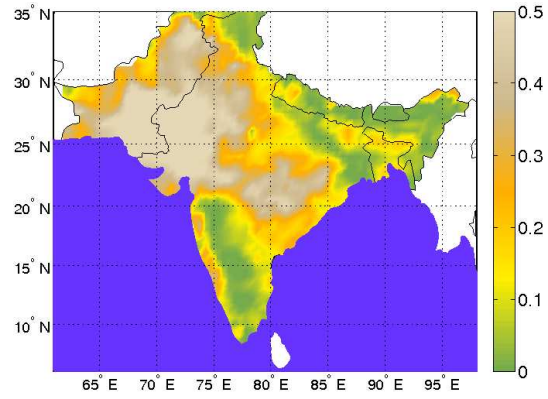


Fig. 3. Variance of the mean seasonal rainfall at each grid point. The colour scale is relative hence each grid point could be compared with the any other grid point (data used APHRO-V0902).

2006). Also, the north-east of India is another very wet region. Central India is made of many different regions in terms of mean annual rainfall. The north west again is very dry which includes the Thar Desert.

To evaluate the patterns of variability of mean rainfall during the summer monsoon season, we next determine the variance at k -th grid as

$$\sigma^2(k) = \sum_{i=1}^m \lambda_i |e_i^k|^2,$$

where e^i is the i -th empirical orthogonal function (EOF), λ_i is the corresponding Eigen value, and m is the total number of EOF, which in this case will be equal to the total number of grid points (Trauth, 2006). The result is shown in Fig. 3. We examine this variability using a relative scale ensuring that each grid point could be compared with any other one (Fig. 3). We infer that the regions of maximum rainfall are not the most variable ones in terms of mean seasonal rainfall. We find that the central Indian region and the north west parts of the subcontinent are the most variable regions in terms of fluctuation of mean seasonal rainfall during summer monsoon.

After having a detailed introduction to the data and climatology of the region, we next introduce the special methods used in our data analysis for identifying those regions of rainfall which show higher level of coherence in terms of rainfall activity.

4 Methods

4.1 Event synchronization

In this section we describe the concept of ES (Quiroga et al., 2002). Let R_i^j be the net amount of rainfall received at grid point j on the i -th day during the JJAS months of any year in the data. To classify events into a “heavy rainfall

event” or a rainfall event which is a result of increased monsoonal activity, the threshold T_j on a rain event at grid point j is calculated by taking the α_j percentile of rain events occurring at grid j of all wet days ($R_i^j > 0$). Next, we determine the time indeces of events $R_i^j \geq T_j$ in our data. Let such an event occur at time t_l^j at grid point j and t_m^k at grid point k , l and m are the time indices assigned to the events $l = 1, 2, \dots, s_j, m = 1, 2, \dots, s_k$. Where, s_j and s_k are total number of such events occurring at grid point j and k , respectively and within a minimal separation time $\pm \tau_{lm}^{jk}$ for each pair of events (l, m), where it is defined as

$$\theta_{lm}^{jk} = \min \left\{ t_{l+1}^j - t_l^j, t_l^j - t_{l-1}^j, t_{m+1}^k - t_m^k, t_m^k - t_{m-1}^k \right\} / 2 \quad (1)$$

τ_{lm}^{jk} is the minimum time between two succeeding rainfall events. We need to count the number of times an event occurs at j after it appears at k and vice versa. This is achieved by defining the quantities $c(j|k)$ and $c(k|j)$:

$$c(j|k) = \sum_{l=1}^{s_j} \sum_{m=1}^{s_k} J_{lm}^{jk} \quad (2)$$

with

$$J_{lm}^{jk} = \begin{cases} 1 & \text{if } \tau'^{jk} < t_l^j - t_m^k < \tau^{jk} \\ 1/2 & \text{if } t_l^j - t_m^k = \tau'^{jk} \\ 0 & \text{else,} \end{cases} \quad (3)$$

where τ^{jk} is the upper limit of the range of time delays and τ'^{jk} is the lower limit. These time delay limits will be used to explicitly take into account the different time scales on which monsoonal precipitation can evolve (as in Sect. 5.3). If not explicitly mentioned then we have estimated the delays from minimum separation time, Eq. (1), i.e. $\tau^{jk} = \theta_{lm}^{jk}$ and $\tau'^{jk} = 0$ (Sects. 5.1 and 5.2).

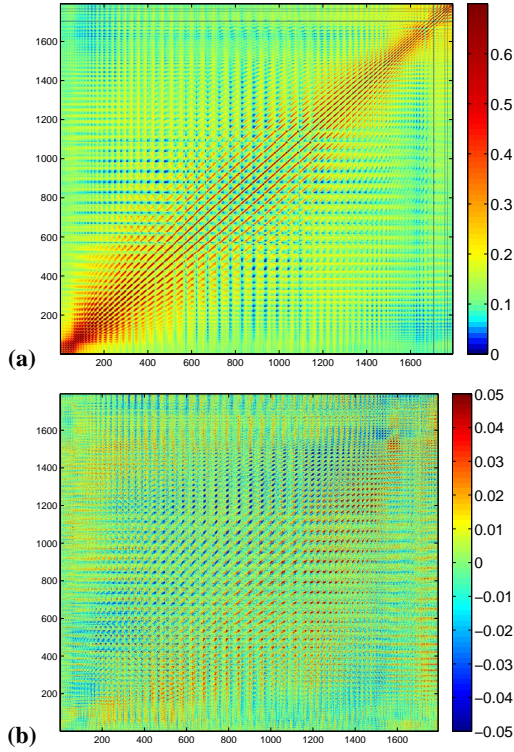


Fig. 4. (a) It is the matrix of \mathbf{Q} for $\alpha = 94\%$, (b) matrix for \mathbf{q} for $\alpha = 94\%$.

Similarly we can also define $c(k|j)$ and from these quantities we obtain

$$\mathbf{Q}_{jk} = \frac{c(j|k) + c(k|j)}{\sqrt{(s_j s_k)}} \quad (4)$$

$$\mathbf{q}_{jk} = \frac{c(j|k) - c(k|j)}{\sqrt{(s_j s_k)}} \quad (5)$$

\mathbf{Q}_{jk} is the measure of event synchronization between grid points j and k . The \mathbf{q}_{jk} measures the delay behaviour. These measures are normalized to $0 \leq \mathbf{Q}_{jk} \leq 1$ and $-1 \leq \mathbf{q}_{jk} \leq 1$. $\mathbf{Q}_{jk} = 1$ implies complete synchronization and $\mathbf{q}_{jk} = 1$ means that events at k precedes events in j . Along the diagonals we find higher values of \mathbf{Q}_{jk} as the elements along the diagonals are closer in space (Fig. 4).

4.2 Clustering

To achieve the task of regionalization, we have used the hierarchical clustering algorithm for forming clusters out of the matrix \mathbf{Q}_{jk} (Jain and Dubes, 1988). Hierarchical clustering can be considered as building a hierarchical tree (or dendrogram) where at each step we merge the leaf nodes sequentially based on a linkage criterion. Each of these leaf nodes is a cluster. All the clusters are fused together at the root of the hierarchical tree. In the sense of graph theory a hierarchical tree is an acyclic connected graph where each

node has zero or more children nodes and at most one parent node. As we have large data to cluster (around 1800 grid points in APHRO-V0902 and 350 in IMD-D) we have used the ‘‘average linkage’’ as the linkage criterion for merging the nodes of the tree. In the average linkage method the distance L between two clusters, say U and V , is defined as

$$L = \frac{1}{|U| \cdot |V|} \sum_{j \in U} \sum_{k \in V} d(j, k) \quad (6)$$

Where j and k are the indices for the grid points, $|U|$ and $|V|$ stands for number of elements in set U and V , respectively, and $d(j, k) = 1 - \mathbf{Q}_{jk}$ is the dissimilarity matrix. The clusters which have least L are merged together first. The hierarchical tree so formed is cut at the second highest value of the inconsistency coefficient. At the highest value all the leaf nodes of the hierarchical tree are merged together. The inconsistency coefficient is given by the relation

$$I_l = \frac{z_l - \mu_l}{\sigma_l} \quad (7)$$

where μ_l is the mean of the heights (height of a node is the length of the longest downward path to a leaf from that node) of all the links included in the calculation and σ_l is the standard deviation. z_l is the distance between the links joined at the level l (level of a node is the length of the path to its root). After obtaining the tree, we recognize only those clusters which contain more than or equal to 3 grid points. The smaller clusters are discarded.

4.3 Time-delay patterns of events

From Eq. (5) we extract the time delay patterns of events. Remember that $q_{jk} > 0$ means that events in k precede events in j and for $q_{jk} < 0$ vice-versa. So it is easy to determine the delay direction of events occurring at location j and k . We determine the time delay patterns as follows. In accordance with the convention that negative q_{jk} means time delay is from j to k and vice-versa for positive q_{jk} , we try to calculate the average direction to the nearest neighbours. Say a vector $\mathbf{Y}_{jk} = (x_{jk}, y_{jk})$ gives the direction between two nearest neighbor grids (see Fig. 5) and $x_{jk} = \cos \theta_{jk}$ and $y_{jk} = \sin \theta_{jk}$, where θ_{jk} could have discrete values $\{0, \pi/4, \pi/2, 3\pi/4, 5\pi/4, 3\pi/2, 7\pi/4, 2\pi\}$. Then the average direction will be

$$\mathbf{Y}'_{jk} = \frac{\sum_{\substack{\text{nearest} \\ \text{neighbours} \\ \text{of } j}} q_{jk} \mathbf{Y}_{jk}}{\sum_{\substack{\text{nearest} \\ \text{neighbours} \\ \text{of } j}} q_{jk}} \quad (8)$$

We plot the direction at the corresponding grid points and compare with wind data from NCEP/NCAR reanalysis.

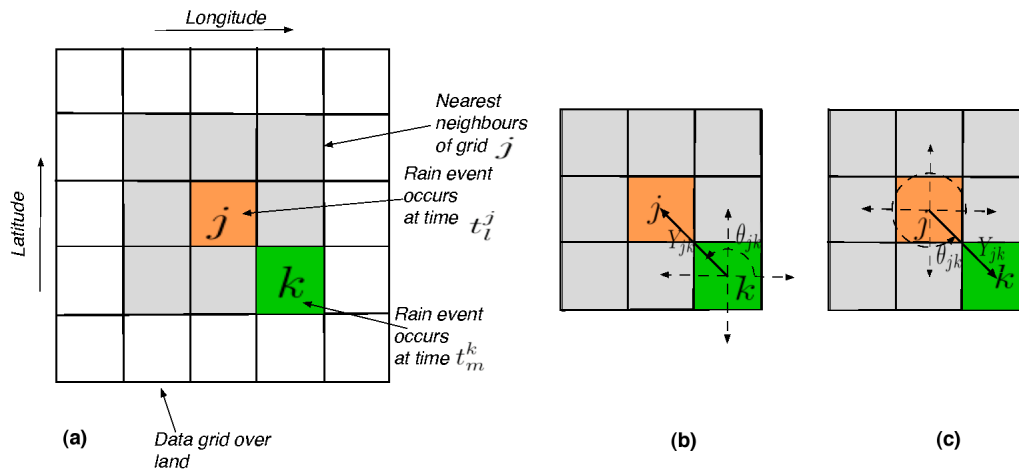


Fig. 5. (a) Representation of the grid for estimating Q_{jk} and q_{jk} , (b)–(c) diagrammatic representation for calculating time delay patterns.

Table 1. Summary of results of Cluster analysis.

Data Used	Threshold α percentile	I_l	Figure no.	Total Clusters	Clusters containing greater than 3 grid points
APHRO-V0902	94%	1.547	6a	67	38
APHRO-V0902	90%	1.547	6b	45	27
IMD-D	94%	1.538	6c	14	10
IMD-D	90%	1.538	6d	19	14

5 Results and discussion

5.1 Spatially coherent zones of heavy rain events

For calculating the ES we have used two different thresholds of $\alpha = 94$ percentile and $\alpha = 90$ percentile. These thresholds could be interpreted as very heavy to heavy rainfall events. Our motivation for choosing these thresholds is based on the fact that such rainfall events at a given location can only occur when monsoon is in its active phase over a large spatial region containing this particular location. Such events can be scattered over many locations within a large spatial region occurring either at the same time or with some delay. To get these regions where monsoonal activity is more coherent compared to any other region, we applied a clustering analysis to the ES measure of this coherent activity. The result of clustering are shown in Fig. 6 where clusters containing at least 3 or more grid points are shown in colour and the remaining grid points are marked in black. We have marked some of the larger clusters by numbers for easy reference. A summary of the analysis is given in Table 1. We do not expect to generate exactly equal structures for both data sets used in our analysis, as spatial resolution, spatial extent and time period are different (see

Sect. 2). Nevertheless, we find some striking similarities in the structures in all the figures (Fig. 6a–d). In all cases studied in clustering analysis we see a cluster forming along the strong orographic barrier of Western Ghats on the west coast of India (cluster number 4 in Fig. 6a–d) pointing the important role played by orographic barriers in generation of rainfall. Interestingly, the whole of north east (cluster number 10 in Fig. 6) emerge as a single cluster even though the region comprises of a very variable topography, including many high mountain ranges. We also observe many other similarities in the results like formation of separate clusters in south east peninsular India (cluster number 5 in Fig. 6a–d).

The basic motivation for the above analysis was to identify regions where monsoonal rainfall dynamics is more coherent or connected than others. So, in a way we have been able to identify the following major dynamical regions of monsoonal activity over the Indian subcontinent, these are: i) along the west coast (cluster number 4), ii) north east of India and parts of Bangladesh (cluster number 10), iii) north west of the subcontinent (cluster number 1), iv) western and adjoining parts of central India (cluster number 2), v) south east peninsular India (cluster number 5). Some of the clusters are large in size covering thousands of kilometers. A possible mechanism for the existence of such long range spatial

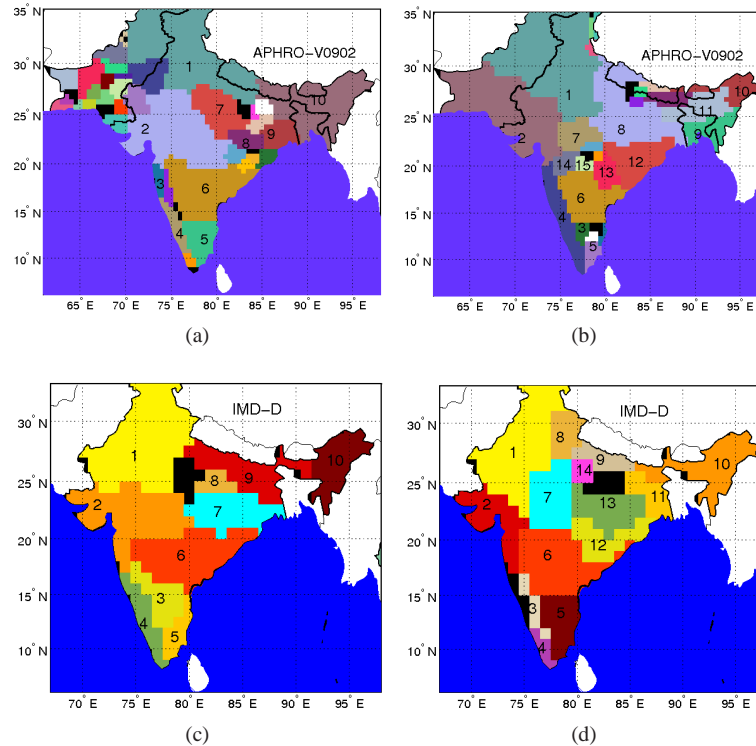


Fig. 6. Different colours here gives the different clusters obtained using Hierarchical clustering algorithm using matrix for Q as dissimilarity matrix. (a) Data used is APHRO-V0902, threshold $\alpha = 94\%$, (b) data used is APHRO-V0902, threshold $\alpha = 90\%$, (c) data used is IMD-D, threshold $\alpha = 94\%$, (d) data used is IMD-D, threshold $\alpha = 90\%$.

correlations could be due to clustering of synoptic activity during the active phase of monsoon (Goswami et al., 2003). Such type of monsoonal rainfall activity seems to be a result of large spatial scale atmospheric activity, like the formation and clustering of large low pressure systems. In one monsoon season there could be 3–4 active (break) periods and slowly evolving between each other (Lawrence and Webster, 2002). Very heavy rain events are associated with the active phase of monsoon. So from our method we could obtain the regions where monsoon gets active simultaneously or within some delay. During the active phase of monsoon, large parts of the country experience heavy rainfall (above normal rainfall). But some parts like the south east peninsular India are out of phase with it (Krishnamurthy and Shukla, 2000; Singh et al., 1992), and receive rain during the break phase. We also see in our clusters that south west peninsular India is always a small separate cluster. Different rain producing monsoonal systems can evolve at many different time scales, in Sect. 5.3 we make an explicit distinction between different modes of rainfall activity present in monsoon and estimate coherent zones based on time scales involved in monsoonal precipitation.

Coherent rainfall zones over India have been previously recognised by Gadgil et al. (1993). The data used in their work was monthly rainfall from 459 stations spread over India for the period 1907–1981. The delineation of the

rainfall zones was based on the variation of monthly rainfall. They have estimated 31 zones (limited to political boundary of India) in number and quite smaller in size than we have estimated above. In present work we had maximum of 14 to 19 clusters when we considered data only for India (see Fig. 6c and d and Table 1). Most of the basic structures we have found and mention above were also present in Gadgil et al. (1993) but few of them defragmented into smaller zones. This difference seems to emerge due to two main reasons – (1) dissimilarities in the spatial and temporal resolution and spatial extent of the data analysed in the present study. We used spatially interpolated daily rainfall data compared to the monthly rainfall station data in Gadgil et al. (1993). (2) In our study we have aimed at understanding the dynamical coherence in the daily rainfall activity during monsoon season. For measuring the dynamical coherence we have used a nonlinear measure of correlations rather than coherence in variation of monthly rainfall as was done in Gadgil et al. (1993).

5.2 Time delay directions

The method of ES as defined before, allows for obtaining time delay patterns of events. The time delay patterns could be further used in studying the paths of rain events and their moisture sources or paths of synoptic scale atmospheric

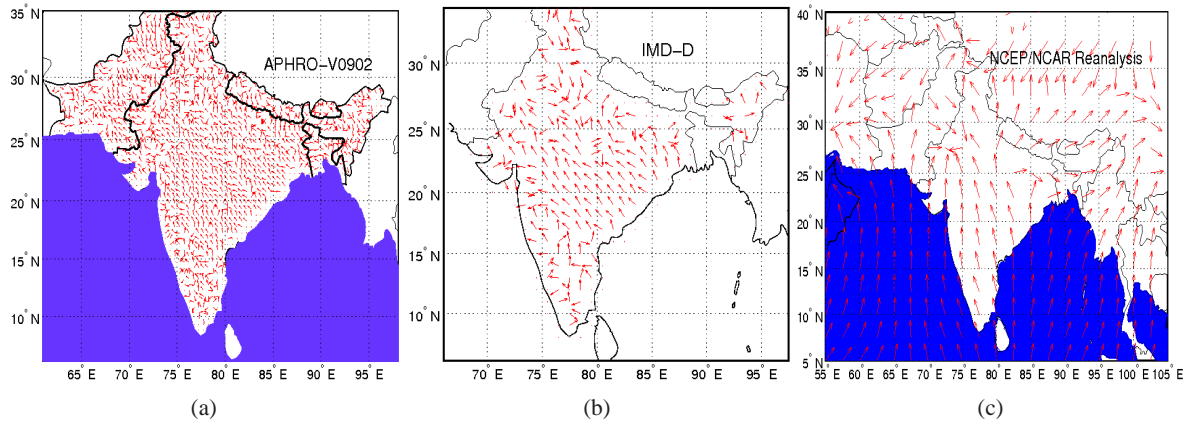


Fig. 7. (a) Time delay patterns as obtained from APHRO-V0902 for $\alpha = 94\%$. (b) Time delay patterns as obtained from IMD-D for $\alpha = 94\%$. (c) Mean wind direction at the height of 850 hPa during rain events above the threshold of $\alpha = 94\%$, derived from NCEP/NCAR data set with time indices for the events calculated from NCEP/NCAR precipitation data.

disturbances and their evolution. This is significant in the light that Indian summer monsoon has two major branches or sources of moisture. The first one is the Bay of Bengal on the east and the second one is Arabian sea on the west. In Fig. 7 we show the time delay patterns obtained employing the approach discussed in Sect. 4.3. We observe similar patterns for both data sets. We compare these results with wind directions as calculated from the reanalysis data at 850 hPa height for the same threshold Fig. 7c. To obtain wind directions we have taken the mean of the wind directions over the days when there were events above an α percentile in the precipitation data set of NCEP/NCAR reanalysis data sets, i.e., the time indices are also derived from the same NCEP/NCAR reanalysis data set. The directions estimated by our proposed method are almost identical to the wind directions (Fig. 7) derived from reanalysis data. We have obtained similar results for the other thresholds too. The similarity between time delay patterns and wind directions indicate that our method allows to estimate the paths of moisture movement over the land mass or the directions involved in atmospheric circulations during moisture transport in the monsoon season. It is interesting to note that the moisture source of heavy rainfall events over central Indian region is the Bay of Bengal. It is known that rainfall over the central part of India is received by north ward movement of low pressure areas from the Bay of Bengal (Lal et al., 1995). Therefore we find strong directionality of arrows from the Bay of Bengal to north over central India (Fig. 7).

5.3 Separating time scales

The Monsoon’s inter seasonal oscillations (ISO) are another intriguing feature of monsoon dynamics. These oscillations are part of an internal dynamics of monsoon and also govern its inter seasonal variability (ISV). The ISO is composed of a hierarchy of quasi-periods like 3 to 7 days,

10 to 20 days and 30 to 60 days (Waliser, 2006; Ding and Sikka, 2006). The 3 to 7 days mode is caused by the oscillation of monsoon trough over the Indo-Gangetic plains. Both 10–20 days (Krishnamurti and Arduay, 1980) (quasi-biweekly oscillation – QBM) and 30–60 days (Sikka and Gadgil, 1980) (Madden-Julian oscillations – MJO) are associated with planetary scale waves.

As a first exercise we try to identify the important time delays in our data at a particular threshold on the rain events. For this purpose we use the IMD-D data set and we calculate the frequency distribution of the delays $p(n_i^d) = p(|t_l^j - t_m^k|)$ for a given distance d between the grid points j and k on the lattice where $n = |l - m|$, $n = 1, 2 \dots 122$ (days) and i is an index for a specific comparison between the grid points j and k satisfying the distance criterion ($maxnorm(j, k) = d$). To calculate the relative significance (not in a strict statistical sense) of delays, we calculate Shannon’s entropy for each delay at $S_n^d = \sum_{i=1}^r p(n_i^d) \ln p(n_i^d)$. Clearly S_n^d is a measure of how uniform the distribution is for a particular delay over all r comparisons. The more uniform the particular delay the more it has relative importance in the system. Further we normalize S_n^d by, $S = \frac{|S_n^d - S_n^{d,max}|}{S_n^{d,max}}$ where $S_n^{d,max}$ is the maximum of S_n^d over all the delays. In Fig. 8 the lowest values (white colour) of S gives the most significant delays. We observe that the most important delay lies between 10 to 40 days with a maximum spatial extent for 10 to 20 days. We may conclude that the 10 to 20 days mode is the most influential one and has largest spatial scales.

For the purpose of our study on the spatial organization of the rainfall field and the formation of coherent zones, we use four different delay ranges: 3–7 days, 10–20 days (QBM), 30–60 days (MJO), and 0–7 days to account for the fluctuations in synoptic systems such as lows and depressions. Applying these time slices in Eq. (3) we

Table 2. Summary of results of the Cluster analysis.

Data Used	Threshold α percentile	time delay range	I_l	Figure no.	Total Clusters	Clusters plotted
IMD-D	94%	10–20 days	1.1496	9a	119	14
APPHRO-V0902	94%	10–20 days	1.1546	9b	239	10
IMD-D	94%	30–60 days	1.1467	9c	159	6
APPHRO-V0902	94%	30–60 days	1.1546	9d	28	6
IMD-D	94%	3–7 days	1.1545	9c	69	15
APPHRO-V0902	94%	3–7 days	1.1546	9d	213	19
IMD-D	94%	0–7 days	1.148	9e	29	18
APPHRO-V0902	94%	0–7 days	1.15469	9f	32	19

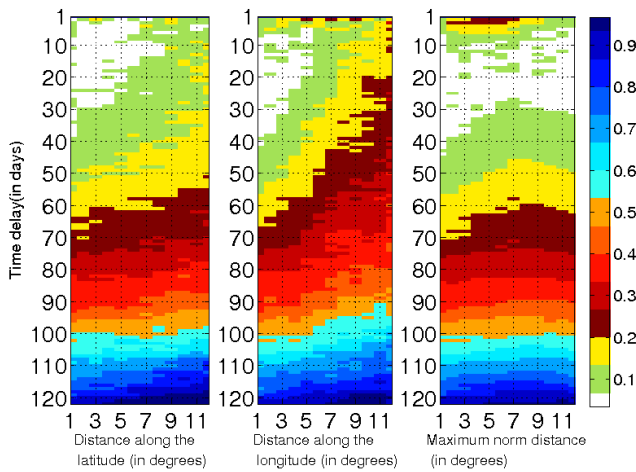


Fig. 8. Shannon entropy for the distribution of each delay for all comparisons between grid with a given distance. The data used to generate above image is IMD-D.

calculate the ES matrix \mathbf{Q} and perform a clustering analysis on it. This allows us to determine the spatial scales under explicit consideration of the aforementioned time scales.

In Fig. 9a and b we consider $\tau^{'jk} = 10$ and $\tau^{jk} = 20$, corresponding to the QBM. In this case we get geographically discontinuous clusters. We find a strong coherence between the west coast of peninsular India, parts of central India and the northeast of India (see green colour along the west coast and north east of India in Fig. 9a and yellow colour in Fig. 9b marked as 1a, 1b and 1c). The black colour means the formation of very small clusters or defragmentation. Another important aspect to note is the formation of a huge cluster formed over parts of central India, north India and Pakistan Fig. 9a. These findings suggest that two zones of rainfall activity exist on these times scales, one along the west coast, parts of central India and north east and another one consists of the Indo-Gangetic plains. The detected spatial scales are now rather large, as also expected because this particular mode is associated with the planetary scale phenomena of westward moving waves originating in the Pacific.

For the case of the MJO scale we take $\tau^{'jk} = 30$ and $\tau^{jk} = 60$. Again we find a formation of geographically discontinuous clusters, with largest coherence emerging between the west coast of peninsular India, parts of central India, the northeast of India and the foothills of the Himalayas (see yellow colour along the west coast, along Himalayas, central India and parts of north east of India in Fig. 9c and yellow colour in Fig. 9d marked as 1a, 1b and 1c). The high coherence in rain intensities over the west coast and central India on MJO timescales has also been reported by Singh et al. (1992). However the coherence observed between the Himalayan foothills and the north east has not yet been known. One striking dissimilarity between the results for the two data set is the formation of a single cluster covering the east peninsular India and Indo-Gangetic plains (green colour in Fig. 9c) in APHRO-V090 as compared to IMD-D Fig. 9d where the region is defragmented into very small clusters. The clustering algorithm is not able to fully resolve the structures in this particular case of the IMD-D data set.

In the next case we consider the time scale $\tau^{'jk} = 3$ and $\tau^{jk} = 7$. Now we do not find too much geographically discontinuity in clusters (Fig. 9e and f). There does not exist very long range correlations as observed before when considering larger time scales. We observe clear and strong clusters along both east and west peninsular India, the former being the larger one (Fig. 9e and f; clusters 4 and 5). We find a strong separate clustering over central India and north east too. Finally, we take a temporal thresholds of $\tau^{'jk} = 0$ and $\tau^{jk} = 7$, putting together all the synoptic scale activities and fluctuations into one single time scale. Obviously, very long range correlations and geographically discontinuous spatial correlations as observed above do not exist, we also have similar observation for these time scales in the Fig. 8. The last two time scales are indicative for spatially local variations in monsoonal rainfall due to synoptic scale activities like monsoonal troughs, depressions and lows. Even with all the dissimilarities of IMD-D and APHRO-V090 we see striking similarities in the clusters in Fig. 9g and h. The clustering analysis is summarized in Table 2.

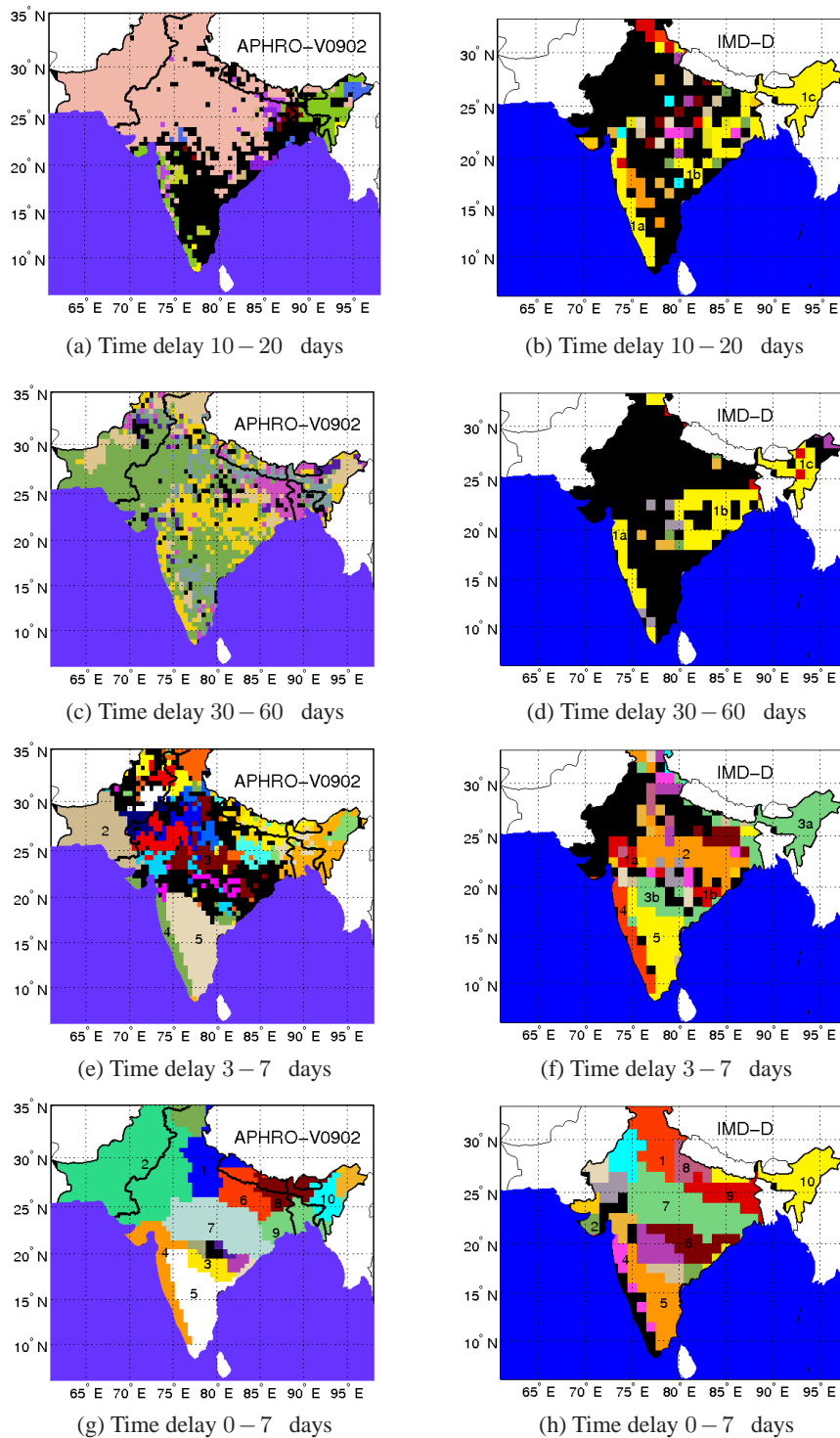


Fig. 9. Results of clustering at different delays.

From the above results we can conclude that during monsoon the rainfall field exhibits an obvious spatial organisation. Our study provides a qualitative view of this spatial organization. It has a strong dependence on the time scales of the different rain producing processes and systems. At QBM

and MJO time scales, we have long range spatial correlations transacting even couples of 1000 km, where at shorter time scales of 3 to 7 days and 0 to 7 days the spatial scales are smaller. The interaction of topography and monsoonal dynamics is also apparent in the shape of the clusters.

5.4 Remarks

By using event synchronisation we have estimated different dynamical zones of monsoon and how they re-organize themselves on different time scales. Our results strongly point towards the existence of spatial organization in the rainfall field due to the interaction of physico-geographical factors with the dynamics of the monsoon. The spatial correlations studied above could only emerge if there is a recurring tendency of heavy rain events to occur over the same zone in lag synchronised form. This in turn could lead to recurring floods over the same zone. An example in case could be the north east (cluster number 10; Fig. 6) of the subcontinent and parts of Bangladesh. These regions suffer from regular floods during monsoons. In accordance with the observation these regions do emerge as separate clusters in our study and even at different time scales of monsoonal evolution. Hence, such spatial organization also opens the possibility for predictions of a probable spatial extent of monsoonal rainfall activity within a certain time delay, if some sporadic heavy rain events are reported in a part of these coherent rainfall zones. Another important application could also be in paleoclimate studies where data are usually collected from few sparse locations. From such locations of the data, we can then infer the dynamical region of past monsoonal rainfall it represents. Furthermore we could spatially extrapolate such data to the dynamical region it belongs to. In the light of the above facts, a further study of the spatial organization of monsoonal rainfall is necessary to both understand climatological causes and its application in prediction of extreme monsoonal events like floods and droughts.

6 Summary

The Monsoon over South Asia is manifested in a form of a very complex spatio-temporal behaviour. It shows an enormous amount of spatial variability. Here we have presented a method based on event synchronization of rain events to identify regions where the dynamics of monsoonal rainfall is more coherent or homogenous. We have applied this method on two separate data sets of different spatial resolution and spatial extent and also different time periods. The regions obtained in both data sets are strikingly similar emphasizing the underlying structure in the physical processes responsible for the generation of rainfall activity, whether it is the atmosphere circulation itself or the topography of the region. Furthermore, we have developed a method to construct patterns of time delay of rainfall events. We showed that these time delay patterns seem to follow the prevalent wind directions during the period. Hence, we can establish the path ways of moisture movement over land during monsoon. At distinct time scales, we observe the emergence of spatial correlations on very different spatial

scales. At the time scales of QBM and MJO, which are related to the planetary scale waves, we have seen very long range spatial correlations crossing over even the geographically discontinuous regions. Our results also seem to be consistent with the observation of recurring extreme events (floods) over these regions. We suggest that the possibility of employing such approach in the prediction of extreme events at different time scales should be explored.

Acknowledgements. The authors are thankful to Krishna Kumar, IITM, Pune for helping in getting data from IMD, Pune. We are also thankful to Bodo Bookhagen for helpful discussions. We also thank the anonymous referees for their valuable comments especially on the need of separation of time scales in monsoonal rainfall. This work is supported by doctoral scholarship from DFG Graduate School 1364 and DFG (SFB 555) project C4.

Edited by: O. Talagrand

Reviewed by: four anonymous referees

References

- Bookhagen, B. and Burbank, D. W.: Topography, relief, and TRMM-derived rainfall variations along the Himalaya, *Geophys. Res. Lett.*, 33, L08405, doi:10.1029/2006GL026037, 2006.
- Ding, Y. and Sikka, D. R.: Synoptic Systems and Weather, in: *The Asian Monsoon*, edited by: Wang, B., Springer, 2006.
- Gadgil, S.: The Indian Monsoon and its variability, *Annu. Rev. Earth Pl. Sc.*, 31, 429–67, 2003.
- Gadgil, S., Yadumani, and Joshi, N.: Coherent rainfall zones of the Indian region, *Int. J. Climatol.*, 13, 547–566, 1993.
- Goswami, B. N., Ajayamohan, R. S., Xavier, P. K., and Sengupta, D.: Clustering of synoptic activity by Indian summer monsoon intraseasonal oscillations, *Geophys. Res. Lett.*, 30, 1431, doi:10.1029/2002GL016734, 2003.
- Gutiérrez, J., Galván, A., Cofinño, A., and Primo, C.: Chaos game characterization of temporal precipitation variability: Application to Regionalization, *Fractals*, 14, 87–99, 2006.
- Iyengar, R. and Basak, P.: Regionalization of Indian monsoon rainfall and long term variability signals, *Int. J. Climatol.*, 14, 1095–1114, 1994.
- Jain, A. and Dubes, R.: *Algorithms for Clustering Data*, Prentice-Hall, NJ, 1988.
- Keshavamurthy, R. N.: Power-spectra of large scale disturbances of the Indian south-west monsoon, *Indian J. Meteorol. Geophys.*, 24, 117–124, 1973.
- Krishnamurti, T. N. and Ardunay, P.: The 10 to 20 day westward propagating mode and “breaks in the monsoons”, *Tellus*, 32, 15–26, 1980.
- Krishnamurthy, V. and Shukla, J.: Interseasonal and inter annual variability of rainfall over India, *J. Climate*, 13, 4366–4377, 2000.
- Lal, M., Bengtsson, L., Cubash, U., Esch, M., and Schlese, U.: Synoptic scale disturbances of the Indian summer monsoon as simulated in a high resolution climate model, *Clim. Res.*, 5, 243–258, 1995.
- Lawrence, D. M. and Webster, P. J.: The Boreal summer inter-seasonal Oscillation: Relation between Northward and Eastward movement of convection, *J. Atmos. Sci.*, 59, 1593–1606, 2002.

- May, W.: Simulation of the variability and extremes of daily rainfall during the Indian summer monsoon for present and future times in a global time-slice experiment, *Clim. Dynam.*, 22, 183–204, 2004.
- Quiroga, R. Q., Kreuz, T., and Grassberger, P.: Event Synchronization: A simple and fast method to measure synchrony and time delay patterns, *Phys. Rev. E*, 66, 041904–1, 2002.
- Rajeevan, M., Bhate, J., Kale, J., and Lal, B.: Development of a high resolution daily gridded rainfall data for the Indian region, *Met. Monograph Climatology No. 22/2005*, Indian Meteorological Department, Pune, India, 2005.
- Rajeevan, M., Bhate, J., Kale, J., and Lal, B.: High Resolution daily gridded data for Indian region: Analysis of break and active monsoon spells, *Curr. Sci. India*, 9, 296–306, 2006.
- Sikka, D.: Some aspects of the life history, structure and movement of monsoon depressions, *Pure Appl. Geophys.*, 115, 1501–1529, 1977.
- Sikka, D. R. and Gadgil, S.: On the maximum cloud zone and ITCZ over Indian longitudes during the southwest monsoon, *Mon. Weather Rev.*, 108, 1840–1853, 1980.
- Singh, S. V., Kirpalani, R. H., and Sikka, D. R.: Interannual variability of the Madden-Julian Oscillations in Indian Summer monsoon rainfall, *J. Climate*, 5, 973–979, 1992.
- Stephenson, D., Kumar, K. R., Doblas-Reyes, F., Royer, J.-F., and Chavin, F.: Extreme rainfall events and their impact on ensemble forecasts of Indian Monsoon, *Mon. Weather Rev.*, 127, 1954–1966, 1999.
- Trauth, M. H.: *MATLAB[®] recipes for Earth Sciences*, Springer, 2006.
- Waliser, D. E.: Interseasonal variability, in: *The Asian Monsoon*, edited by: Wang, B., Springer, 2006.
- Webster, P. J., Magana, V. O., T. Palmer, Shukla, J., Tomas, R. A., Yani, M., and Yasunari, T.: Monsoons: Processes, predictability, and the prospects for prediction, *J. Geophys. Res.*, 103, 14451–14510, 1998.
- Yatagai, A., Arakawa, O., Kamiguchi, K., and Kawamoto, H.: A 44 Year Daily Gridded Precipitation Dataset for Asia Based on Dense Network of Rain Gauges, *SOLA*, 5, 137–140, 2009.