The statistical face of a region under monsoon rainfall in Eastern India

Technical Report No. ASU/2017/10
Dated: 17 July, 2017

Kaushik Jana
Applied Statistics Unit,
Indian Statistical Institute,
Kolkata 700108
kaushikjana11@gmail.com

Debasis Sengupta
Applied Statistics Unit,
Indian Statistical Institute,
Kolkata 700108
sdebasis@isical.ac.in

The Statistical Face of a Region under Monsoon Rainfall in Eastern India

Kaushik Jana[1], Debasis Sengupta[1], Subrata Kundu[2],

Arindam Chakraborty[3] and Purnima Shaw[4]

[1]Indian Statistical Institute, Kolkata, India

[2]George Washington University, Washington, D.C., USA

[3]Indian Institute of Science, Bangalore, Inidia

[4]Reserve Bank of India, Mumbai, India

Abstract

A region under rainfall is a contiguous spatial area receiving positive precipitation at a particular time. The probabilistic behavior of such a region is an issue of interest in meteorological studies. A region under rainfall can be viewed as a shape object of a special kind, where scale and rotational invariance are not necessarily desirable attributes of a mathematical representation. For modeling objects of this type, we propose an approximation of the boundary that can be represented as a single real valued function, and arrive at further approximation through functional principal component analysis. We also propose a new method of handling censored data objects, or occluded shapes, in this setting. The analysis of an open access TRMM satellite data set on monsoon precipitation over Eastern India leads to explanation

of most of the variation in shapes of the regions under rainfall through a handful of interpretable functions. The most important aspect of shape is found to be elongation, which occurs predominantly along the East-West axis. The different modes of variation are remarkably stable across calendar years and across different thresholds for minimum size of the region.

Keywords: Functional Principal Component Analysis; Shape Analysis; Mesoscale Convective System; Precipitation Area; Censored Data

1 Introduction

How large is the contiguous area under a particular spell of rainfall at a particular time? What is the typical shape of a raining cloud system? How does the shape vary from one spell to another?

While answers to these questions should be interesting by their own right, there could be other motivations for obtaining them also. Rainfall systems with different sizes and shapes have received attention from meteorologists over the past few decades (see for example von Hardenberg et al. (2003) and references cited there). In particular, precipitation areas coming from Mesoscale Convective Systems (MCS) are categorized according to their horizontal extent (Austin & Houze, 1972; Houze, 2004). These systems are major contributors to severe weather and precipitation in most parts of the earth (Zipser, 1982). Although individual convective clouds have scales from 1 to 10 km, MCS organize at 200 to 2000 km spatial scales (Houze, 1977; Johnson, 1984; Jirak et al., 2003). Despite several studies on the size of MCS, systematic accounts of their typical shape and size are not available. The shape along with size of a convective system could be an indicator of its degree of organization and rainfall rate. The organization could, in turn, depend on the state of the climate. Thus, a thorough characterization of shape and size of observed con-

vective systems could be important for understanding the properties of the raining system and its spatio-temporal evolution, which could further be compared with numerical model simulated systems.

Modeling of the precipitation area of a convective system depends crucially on the representation of the area objects. Approximate representations can be potentially simpler than exact representations. A formal assessment of the distribution of these objects is crucial to fulfilling this potential through appropriate approximation. Efficient representation should also lead to better compression of data, forecasting verification, nowcasting and so on (Micheas et al., 2007).

A statistical description of precipitation areas has to rely on simultaneous measurements over a large area. Terrestrial rainfall gauges are not particularly suitable for this purpose because of the requirement of a dense grid network. Terrestrial radar based data can be an improvement, but spatial coverage of the existing radar infrastructure in many countries is inadequate for this purpose. Direct estimates of rainfall rate from space-based instruments such as those aboard a satellite, which have been around for only a couple of decades, have the requisite coverage. Such measurements are more spatially coherent and can span large regions including oceans, which make it possible to identify and study large raining systems, such as MCS.

The present work is motivated by satellite data on rainfall rate over the region spanning from latitudes 50°N and 50°S, captured since 1998 by the Tropical Rainfall Measurement Mission (TRMM). The data are collected through imaging over large swathes of area during the flight of a satellite. We deal with images representing rainrate profile over the area covered at the time of recording. The gray level of each pixel represents the average precipitation rate over an approximately 5 km by 5 km grid. The TRMM images show clusters of contiguous pixels with positive rainfall that represent contiguous patches of areas under rainfall at the time of imaging. Very small patches are missed by these images,

because of their finite resolution. On the other hand, very large patches are not fully captured within the frames of the images. Therefore, this type of data can only be used to describe the size and shape of rainfall patches that are neither too big nor too small.

We look for a statistical description of a precipitation area. The problem of describing a spatial region falls in the domain of shape analysis, which is a well-known problem in, e.g., computer vision and medical imaging. Many shape modeling techniques have been developed over the years with different advantages and drawbacks (Rabiner, 1989; Loncaric, 1998). Shape representation and description techniques can be generally classified into contour-based and region-based methods. A typical region based method involves representation through a binary matrix of pixels and use of features for dimension reduction. A contour based method treats the description of the shape objects in terms of their boundaries, either through a finite number of landmarks or through a functional representation. We prefer a contour based approach for the TRMM rainfall data, which is rather voluminous, since a pixel based representation would involve more computation. Further, we shun the landmark based methods (Dryden & Mardia, 1998), as the extraction of landmarks/features generally require labeling of contour points with domain-specific knowledge.

In a functional representation of the boundary, a contour is typically described through a pair of continuous coordinate functions $x(t)$ and $y(t)$, where the parameter t takes value over the periphery of the unit circle. These functions are not unique, as different monotonic reparametrizations of t correspond to the same contour. A particular parametrization, called constant arc length parametrization, has been considered by Klassen et al. (2004). Srivastava et al. (2011) argued that this representation is computationally heavy, and proposed a 'square-root velocity function' (SRVF) representation together with a suitable metric. Kurtek et al. (2012) provided a general framework for representing and analysing size and orientation together with shape, by building on the method provided by Srivastava

et al. (2011). A crucial aspect of the the SRVF approach is that objects are projected to the tangent space of the original shape space at a certain point (e.g., at the Karcher mean). This projection may not be appropriate when one is dealing with heterogeneous shapes (Bhattacharya & Bhattacharya, 2012). The problem is analogous to using a unimodal approximation to a multimodal distribution.

There is another reason for us to look for an alternative representation. Unlike conventional shape objects, the size and the orientation of a particular region of rainfall are important for analysis. Thus, methods that are invariant on size and orientation would not be appropriate for these data. While the SRVF representation is not necessarily invariant to size, the distance between objects change only by the square root of any change in scale of the objects. In the sequel, we use the term 'shape' to mean the combination of shape, size and orientation of the object.

In this paper, we propose an envelope approximation of the region under precipitation, which permits simple representation of the shape objects through a real valued function defined over the unit circle. The space of all such functions is an L_2 space. This reduction enables us to analyse these infinite dimensional objects using standard statistical tools designed to deal with functional data in linear space (Ramsay & Silverman, 2005). It is known (Delaigle & Hall, 2010) that such objects cannot have a density. Any notion of distribution is typically based on finite dimensional representations. We choose to use the functional principal component (FPCA) method to analyse the shape objects. However, occluded regions of rainfall (censored shapes), pose a challenge to this analysis. Based on preliminary findings of the analysis of the motivating data set, we propose an adjustment to the FPCA for incorporating occluded regions. This adjusted FPCA leads us to a set of non-parametrically estimated basis functions that explain most of the variation present in the TRMM data. We identify certain interpretable modes of variation present in the data. The analysis culminates in a parametric representation of the shape objects with only a few

parameters that are readily interpretable. The distributions of the estimated parameters also reveal interesting patterns.

This article is organized as follows. The next section describes the TRMM rainfall data in detail. The problem formulation and methodology is described in Section 3. Analysis of the TRMM data set over a geographical region covering the state of West Bengal and neighbouring regions of Eastern India is presented in Section 4. Some concluding remarks are provided in Section 5.

2 TRMM Data

The problem considered in this paper is motivated by the rainfall rate data collected by the Tropical Rainfall Measuring Mission (TRMM) of the US National Aeronautics and Space Administration (NASA) and the Japanese Aerospace Exploration Agency (JAXA). The data (TRMM 2B31, version 7) are derived by processing images recorded by the combined Precipitation Radar (PR) and TRMM Microwave Imager (TMI), through an algorithm (2B31). Data from several locations lying on a straight line are acquiesced simultaneously, and another set of locations lying on a parallel straight line are covered in the next round of acquisition, as the satellite moves perpendicularly to these straight lines. The satellite has about sixteen revolutions around the Earth every day. However, the inclined orbit precesses (wavers) in such a way that the satellite overflies a given location at different times of the day.

The data can be freely downloaded from the website http://trmm.gsfc.nasa.gov/. The TRMM data are available for the period November 1997 to April 2015. From the beginning of the TRMM satellite operation upto the 7 August 2001, the horizontal resolution of the satellite had been 4.3 km and swathe width had been 215 km. After 24 August 2001, resolution was adjusted to 5 km and the swathe width to 247 km.

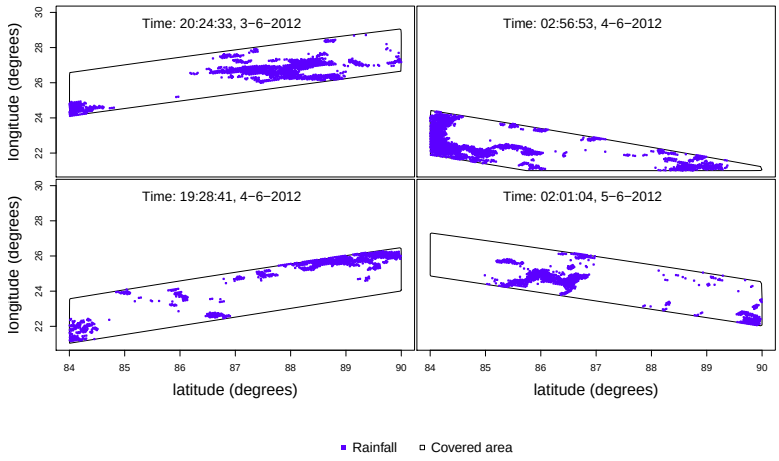

Figure 1: Covered region and areas of rainfall for four consecutive passes of TRMM satellite in June 2012

We consider data from 2002 to 2012, over the latitude range 21°N-30°N and the longitude range 84°E-90°E covering Eastern India. We exclude the earlier data to ensure uniformity of the the pixel size.

The four panels of Figure 1 show locations of rainfall as recorded by four consecutive passes of the TRMM satellite in the month of June 2012, together with the areas (within the target region) covered in each pass. The measured rate of rainfall at these locations are not shown in these figures. It is seen that these locations form spatial clusters. Each cluster corresponds to a particular contiguous region of rainfall. Since each pass of the satellite covers the entire range of observations in a matter of seconds, the clusters obtained from a particular pass may be approximated as a single snapshot of the covered region at a particular instant. Thus, the various clusters of each panel may be regarded as cross-sections of multiple spells of rainfall at a particular time.

Even though the actual set of acquired data corresponds to locations lying on a grid, the location and spatial orientation of the grid depend on the varying flight path of the satellite during data acquisition. For the image data acquired through a single pass of the satellite, the contiguous region under rainfall at a particular instant may be described as

the boundary contour of the set S consisting of a collection of grid cells (i, j), such that

1. for each $(i, j) \in S$, the rainfall rate is strictly positive;
2. one can reach any element $(i, j) \in S$ from any other element $(i', j') \in S$ through a series of neighboring cells (including diagonal neighbors) which are also in S;
3. for any $(i, j) \notin S$ that has a neighboring element in S, the observed rainfall rate is zero.

Note that each of the points in S is actually the center (intersection of the principal diagonals) of an approximately 5 km by 5 km grid cell around that location.

Once clusters of grid points are identified, their locations can be mapped onto a flat surface through any of the standard methods of projection (Keith, 2016). After that, the outer boundary of the cluster of grid points (S) gives the contour of a particular rainfall area. Thus, the objects that may eventually be used for the requisite analysis are polygons on a planar surface.

3 Methodology

3.1 Star-hull representation

We consider the radius-vector function Kindratenko (2003) to describe a contour X in the following way. We select a reference point O in the interior of a contour X that would be regarded as origin for describing X. It could be the centroid of the region lying inside the contour, the centroid of its convex hull, or any other uniquely defined point of physical importance. For $0 \leq \theta \leq 2\pi$, the set of intersections of the θ-rays with a given contour contains all the information about that contour. The region bounded by the contour is called star-shaped with respect to O, when every θ-ray has a unique point of intersection

with the contour. Thus, a star-shaped contour can be represented simply through the set of distances of the contour from O at different values of θ. This simple representation is not possible for contours that are not star-shaped. In order to circumvent this problem, we propose to use a star-shaped hull of any contour and model this hull. For a given contour and a reference point O, this hull would be defined as the locus of the furthest (from O) intersection of a θ-ray with the contour for $\theta \in [0, 2\pi]$. It may be noted that points on the star-hull over a coarse angular grid, where O is the centroid of the contour, have been used in the past as 'landmarks' to represent a contiguous region of precipitation (Micheas et al., 2007).

One might also shun any heuristic choice of the reference point, and opt for an optimal choice. Arkin et al. (1998) proposed that one should use the hull with the minimum area, and termed it the star-shaped hull. They also found that the problem of identifying the star-shaped hull of a polygon with n vertices, i.e., the search over the optimum reference point, takes $O(n^2)$ computational time. It is not clear though that this computation is a worthwhile exercise, as the optimal location of the reference is not guaranteed to be unique or stable.

Therefore, using a uniquely defined reference point, such as the centroid of the area or the centroid of the convex hull, might work better. We opt for the latter as it produces a star-hull that is necessarily contained in the convex hull. We refer to the star-shaped hull corresponding to this heuristically chosen reference point as the star-hull of the original contour. The star-hull contains the original contour. Different contours can have a common star-hull. Representation of contours through star-hull would be inaccurate in case there is a substantial difference between a contour and its hull. We show in Section 4 that this gap is not substantial in the case of the regions under rainfall identified from the TRMM data.

Once the star-hull is used to represent a region under rainfall, its contour can be repre-

sented by the radius vectors for different angles. This would be a real valued function over $[0, 2\pi]$, with identical values at 0 and 2π.

3.2 Functional Principal Component Analysis and Karhunen-Leove Expansion

We use functional principal component analysis (Ramsay & Silverman, 2005; Ferraty & Vieu, 2006) and model each contour (closed curve) as the realization of a stochastic process indexed by the closed interval $[0, 2\pi]$ and taking values in $\mathbb{R}$. Let X be a random function supported on a compact interval $[0, 2\pi]$, with mean function $\mu : [0, 2\pi] \to \mathbb{R}$ and covariance function $K : [0, 2\pi]^2 \to \mathbb{R}$. Assuming that K is positive definite, it admits the spectral decomposition

$$K(\theta, \theta') = \sum_{j=1}^{\infty} \lambda_j \psi_j(\theta) \psi_j(\theta'), \tag{1}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ are the eigenvalues, with respective orthonormal eigenfunctions $\psi_1, \psi_2, \ldots$. This decomposition leads us to the Karhunen-Loeve expansion of the random curve X (see Bosq (2000)),

$$X(\theta) = \mu(\theta) + \sum_{j=1}^{\infty} X_{[j]} \varphi_j(\theta), \quad \theta \in [0, 2\pi], \tag{2}$$

where $X_{[j]} = \int_0^{2\pi} (X(\theta) - \mu(\theta)) \varphi_j(\theta) d\theta$ are the j-th principal component(PCs) of X.

Given a sample of n contours $\{X_1(\theta), \ldots, X_n(\theta); \theta \in [0, 2\pi]\}$, estimates of μ and K may be obtained as

$$\hat{\mu}(\theta) = \frac{1}{n} \sum_{i=1}^{n} X_i(\theta) \;\; and \;\; \hat{K}(\theta, \theta') = \frac{1}{n} \sum_{i=1}^{n} \left(X_i(\theta) - \hat{\mu}(\theta)\right) \left(X_i(\theta') - \hat{\mu}(\theta')\right), \quad \theta, \theta' \in [0, 2\pi]. \tag{3}$$

The eigen elements of $\hat{K}$ provide estimators of λ_j and φ_j as $\hat{\lambda}_j$ and $\hat{\varphi}_j$, respectively, for $j = 1, 2, \ldots$, which lead to a sample version of the Karhunen-Leove expansion. The finite number of principal components to be retained is determined from the usual considerations based on explanation of most of the variability in the data.

The representation of randomness through the covariance operator and the associated Karhunen-Loeve expansion may be problematic because of the severe asymmetry in the TRMM rainfall data, as well as the fact that $X_i(\theta)$ are non-negative valued. The asymmetry arises because the smaller contours are generally more abundant than larger ones. This problem may be alleviated by using a transformation of original functional data (see Section 4).

3.3 Handling of censored data

During a particular pass of the satellite, a part of a contiguous region under rainfall may fall outside the swathe covering it. One does not have adequate information about the shape or size of such region. This is a form of spatial censoring. Treating the censored regions as complete would lead to biased estimates. On the other hand, ignoring them would lead to inefficient estimates.

We would observe in Section 4.1 that the largest component of the variation present in the data is the variation in size of the region under rainfall. We use this crucial fact for handling censored data.

We consider the area within each of the contours (star-hull of complete and censored contours) as realizations of a positive valued random variable. Some of these observations, where contours are fully observed are complete observations, and some others, where contours are partially observed, are censored. If the censored contours had been fully observed, the area enclosed by them would have been larger. Therefore, the areas (sizes) of the contours may be regarded as right-censored data. The size distribution can therefore be

estimated by using the Kaplan-Meier estimator of the survival function. The corresponding estimator of the distribution has (possibly unequal) probabilities allocated to the complete cases only. Suppose, p_i be the probability mass allocated to the complete contour X_{iu} having the ith smallest size, with i running from 1 to n_u, the number of complete contours.

The mean and the covariance functions defined in (3) are estimated as

$$\hat{\mu}(\theta) = \sum_{i=1}^{n_u} p_i X_i(\theta), \text{ and } \hat{K}(\theta, \theta') = \sum_{i=1}^{n_u} p_i \left(X_{iu}(\theta) - \hat{\mu}(\theta)\right) \left(X_{iu}(\theta') - \hat{\mu}(\theta')\right), \theta, \theta' \in [0, 2\pi]. \quad (4)$$

The eigenvalues and eigenfunctions (λ_j, φ_j), $j = 1, 2, \ldots$, are then estimated by replacing (3) with (4).

4 Data Analysis

4.1 Preliminary analysis

The TRMM product (2B31, version 7) consists of latitude and longitude of a location, average precipitation rate over an approximately 5 km x 5 km grid cell around that location and the time of recording. To find the boundary contour of a contiguous region under rainfall, we need to construct the grid cells around the given locations of measurement. In general, for a particular cell on the grid, there are four diagonal neighbors. The mid points of the lines joining the center of the given cell and its diagonal neighbors are regarded as four vertices of the cell with size approximately 5 km.

The distances between the centers of the cells that receive positive rainfall are used to find clusters of grid points with positive rainfall. Each cluster represents a geographically contiguous area that receives rainfall simultaneously. The polygon obtained by joining the outer sides of the of cells corresponding to a cluster represents the contour of a contiguous region under rainfall.

The available locational data are in terms of latitude and longitude, which describe location on the curved surface of the earth. We would like to use a planar approximation of the area under a particular spell of rainfall. For this purpose, we project all coordinates to the Euclidean space as follows. For any set S as described in Section 2, we define its bounding region on the surface of the earth as

$$B_S = [i_{min}, i_{max}] \times [j_{min}, j_{max}],$$

where

$$i_{min} = \min\{i : \ (i,j) \in C_S\}; \quad i_{max} = \max\{i : \ (i,j) \in C_S\};$$
$$j_{min} = \min\{j : \ (i,j) \in C_S\}; \quad j_{max} = \max\{j : \ (i,j) \in C_S\}.$$

and C_S is the set of latitude-longitude pairs of the boundary contour of S, described in Section **??**.

The center of the above bounding region is at $((i_{min} + i_{max})/2, (j_{min} + j_{max})/2)$. A projection of a general point within the bounding region, having latitude *lat* and longitude *long* expressed in degrees, on a flat surface with origin (0,0) is given by the x and y coordinates (in kilometers) (Keith, 2016):

$$\begin{aligned} x &= R\cos\left(lat\frac{\pi}{180}\right)\left(long - \frac{j_{min}+j_{max}}{2}\right)\frac{\pi}{180} \\ y &= R\left(lat - \frac{i_{min}+i_{max}}{2}\right)\frac{\pi}{180}, \end{aligned} \tag{5}$$

where R is the radius of the earth in kilometers.

In order to examine the accuracy of the proposed approximation of the contour of region under rainfall through their star-hull for the TRMM rainfall data, we compute the ratio

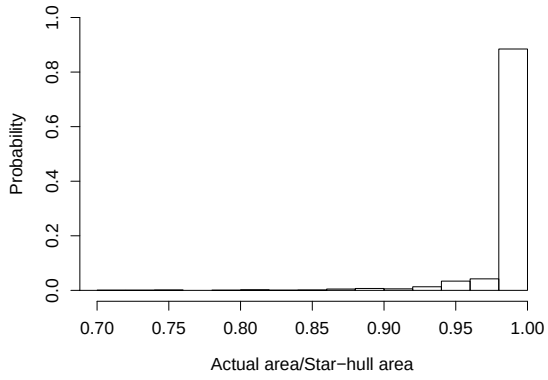

Figure 2: Histogram of ratio of actual area and star-hull approximated area

of the areas of the actual region under rainfall and its star-hull. In Figure 2 we plot the histogram of this ratio. It is observed from the Figure 2 that the most of the ratios are more than 98 per cent. The minimum value of the ratio is 70 per cent. This finding justifies the use of the star-hull approximation as a proxy of the original contour of a cluster.

A major challenge for the analysis of the star-hull contour functions $X_i(\theta)$, $\theta \in [0, 2\pi]$, $i = 1, \ldots, n$ is the asymmetric distribution of the values of the function. Let us consider a set of 1000 uniformly spaced values of θ over the interval $[0, 2\pi]$. The set of $1000n$ values of $X_i(\theta)$ for different choices of i and θ have a distribution, with estimated density shown in Figure 3. The density reveals a positively skewed distribution of the radial distances, which is rather unsuitable for the analysis of the variation through Karhunen-Loeve expansion (see Section 3.2 and 3.3). Therefore, we transform the data through a monotone function in such a way that the distribution of $X(\theta)$ (with all θ pooled together) is standard normal. Specifically, the transformation is $\Phi^{-1} \circ F_n$, where F_n is empirical distribution function of $X(\theta)$ and Φ is the standard normal distribution function. All computations in the sequel are based on this transformed data.

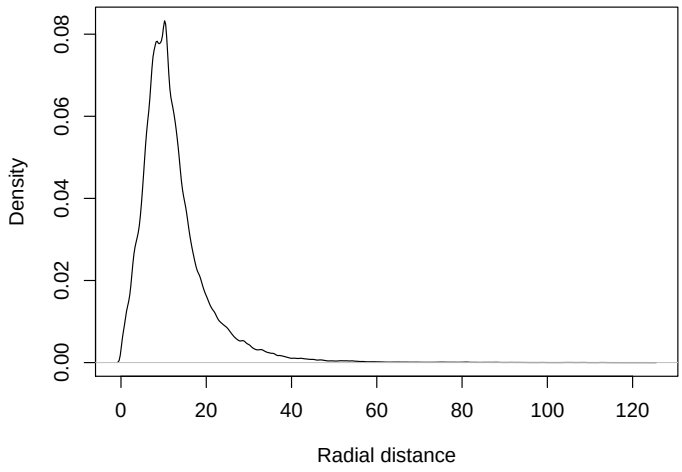

Figure 3: Estimated probability density of radial distance of star-hull

4.2 Functional Principal Component Analysis

In order to ensure homogeneity of the population, we consider region under rainfall data for only the monsoon season (June to September) of a particular year at a time. Further, to focus on mesoscale convective systems, we consider contiguous regions of size 200 square kilometers (about eight pixels) or more, having positive rainfall. It may also be noted that directionality is an important feature in the study of rainfall regions and directionality of small regions of rainfall cannot be studied meaningfully because of the limitation in resolution. On the other hand, overly large regions of rainfall are not fully captured by the images, as noted in Section 1. In view of this limitation, regions with area larger than 13500 square kilometers were also disregarded. In summary, the objects of study were limited to contiguous regions of rainfall having area 200 to 13,500 square kilometers, occurring during the monsoon months of the calendar years 2002 to 2012, over Eastern India (bound by the latitudes and longitudes indicated in Section 2).

Once the regions of rainfall are short-listed as above, we use their star-hull contours for FPCA, after transforming the radial distances (at different angles) as mentioned in the

Table 1: Percentage of variance explained by the first j principal components, for $j = 1, \ldots, 12$ in the contour of region under rainfall, with censored regions regarded as complete, for the year 2011.

j	1	2	3	4	5	6	7	8	9	10	11	12
λ_j	46.54	60.62	73.29	79.96	84.67	87.25	89.48	91.09	92.69	93.55	94.36	94.94

previous section. Initially we treat the censored regions as complete ones.

In order to choose the number of eigenfunctions that provide a reasonable approximation to the contour functions, we use the cross-validation score based on the integrated mean square of the leave-one-curve-out prediction error (Rice & Silverman, 1991), Yao et al. (2005). Minimization of the pooled integrated mean squared error across years (with year-wise FPCA) provides the optimal selection of 11 eigenfunctions for the approximation.

The estimated eigenvalues ($\hat{\lambda}_j$) corresponding to the first few principal components are reported in Table 1 in the decreasing order. The values decrease fairly quickly, with the first ten principal components accounting for 94 per cent of the total variation. In Figure 4, we plot the first ten principal component basis functions. Observe that the first principal component function (solid curve) is almost a constant. Since there is no directionality, it can be said that the first principal component, which alone accounts for about half of the total variation, captures the variation in size of the contours. The other components, which are orthogonal to the first component, capture different aspects of the variation in shape of the contours.

The same pattern is observed for data from each of the years. We skip the detailed findings for brevity. It is clear that the size of the region of rainfall is the dominant aspect of its variation. This fact justifies the adjustment for the censored regions on the basis of size, as described in Section 4.1. Note that, nearly 25 percent of the considered set of regions are censored.

We now turn to FPCA with adjustment for censoring. In this analysis, the estimates (3) of the mean and the covariance functions are replaced by the estimates given in (4).

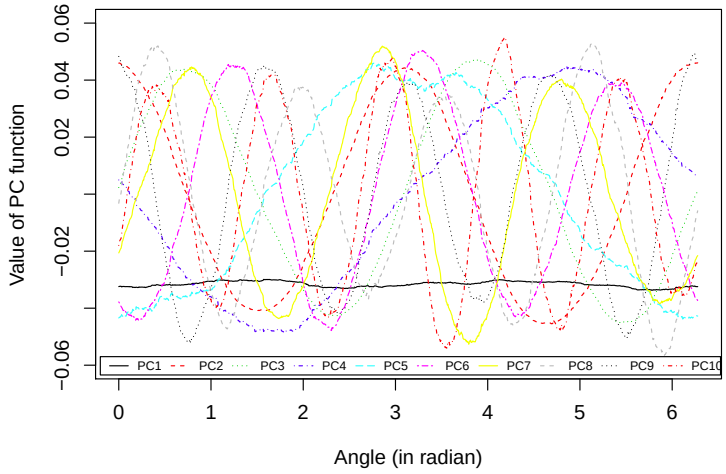

Figure 4: Plots of the first ten principal component basis functions, with censored regions regarded as complete, for the year 2011.

We present in Figure 5 the eigenvalues $(\hat{\lambda}_1, \ldots, \hat{\lambda}_{20})$ obtained by using the modified FPCA. As seen in the case of unadjusted FPCA, the eigenvalues decrease rapidly in this case also, with the first ten principal components accounting for 95 per cent of the total variation. It seems reasonable to use only the first ten principal components to describe the variation.

For visualization of the principal component functions, we use a specially designed plot. Consider the function

$$\eta_{k,\alpha}(\theta) = \mu(\theta) + \alpha\sqrt{\hat{\lambda}_k}\hat{\varphi}_k(\theta), \quad \theta \in [0, 2\pi],$$

where α is a fixed number chosen from the interval $[-2, 2]$. This function is generally referred to as the kth mode of variation (MV) (Wang et al., 2016). The function $g\left(\eta_{k,\alpha}(\theta)\right)$, where g is the re-transformation function $F_n^{-1} \circ \Phi$ and $F_n^{-1}(u) = \inf_x\{F_n(x) \geq u\}$, represents the kth MV in the original (re-transformed) scale. The plot of $g\left(\eta_{k,\alpha}(\theta)\right)$ against θ in polar coordinates would show a typical departure of the kth MV from the mean function. In order to visualize the kth MV, we overlay this graph for $\alpha = -1, 0$ and 1. The overlaid

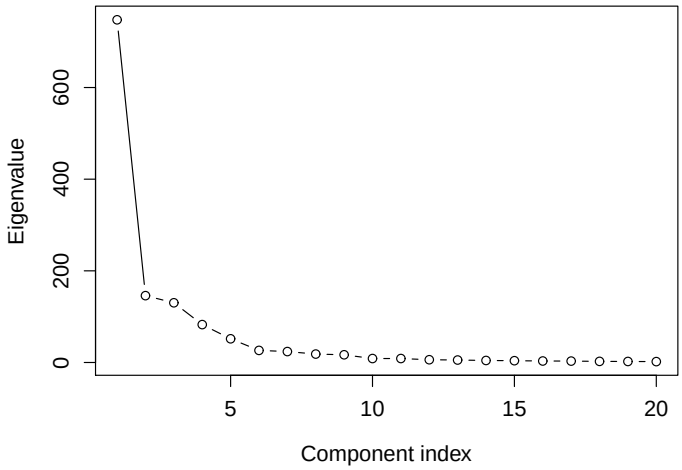

Figure 5: Scree plot of the eigenvalues obtained from the adjusted FPCA of contours of regions under rainfall for year the 2011.

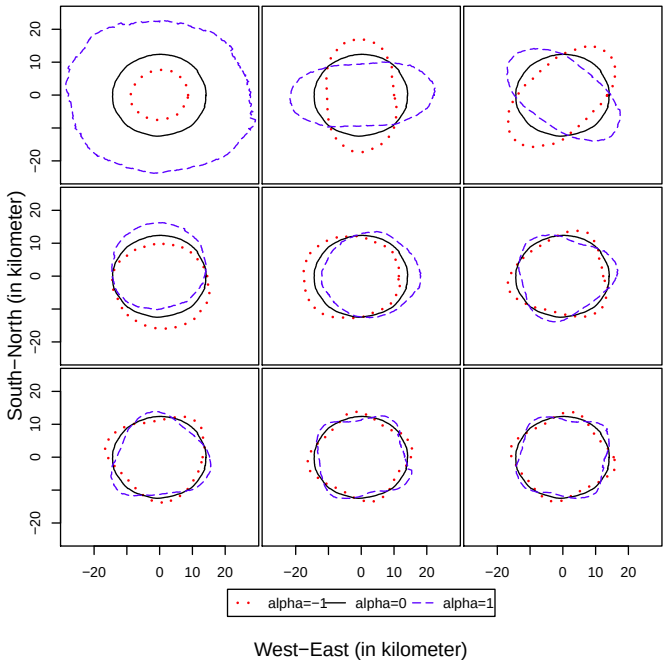

Figure 6: Plots of the first nine modes of variations (MV) for the year 2011.

graphs of the first nine MV are displayed in Figure 6.

In each of the nine panels of Figure 6, the case $\alpha = 0$ corresponds to the mean function in the re-transformed scale. Note that the mean and the median coincide for the normal distribution. Since the transformed data are approximately normal, the mean function may be thought of as the median function also. Further, as the re-transformation is also a monotone function, the MV for $\alpha = 0$ may be identified approximately as the 'median' shape of the regions under rainfall. This 'median' region happens to be approximately circular with a radius of about 10 km.

The first mode of variation, shown in the top left panel, represents the variation in size of the regions under rainfall. One standard deviation of departure from the median function in the inward direction leads to about 40% reduction in size, while the same amount of departure in the outward direction leads to about 150% increase in size.

The second MV, shown in the middle panel of the top row, represents contrasting departures from the median function in the North-South and the East-West axes. When $\alpha = 1$, we have about 40% elongation in the East-West axis and about 20% contraction in the North-South axis. When $\alpha = -1$, there is reversal in the directions of contraction and elongation. Therefore, the second MV (the second most important aspect of the total variation) relates to whether the region of rainfall is elongated in the North-South direction or in the East-West direction, and how sharp that elongation is. This pattern of departure from the median function is less pronounced than the first mode of variation, shown in the top left panel.

The third MV, shown in the top right panel, is similar to the second MV, except that the axes of contrasting elongations and contractions are rotated by about 45°. The amount of departure from the median function is comparable to the typical departure observed in the case of the second MV.

The fourth MV, shown in the left panel of the middle row, shows departures from the

median function that are confined mostly to the North and South directions. When $\alpha = 1$, there is about 25% elongation in the North and about 15% contraction in the South. The pattern reverses when $\alpha = -1$. For values of α in between 1 and -1, the departure from the median function is less pronounced. Overall, the departure is rather subdued in comparison to the cases of the first three MVs.

The fifth MV, shown in the second panel in middle row, is similar to the fourth MV, except that the contrasting elongations and contractions are in the East and West directions. The amount of departure from the median function is also comparable to the departure observed in the case of the fourth MV.

The sixth MV, shown in the right panel of the middle row, features elongations in three directions that are about 120° apart, and contractions in between. When $\alpha = 1$, one of the directions of elongation is East. The amount of departure from the median function is slightly less than that observed in the corresponding cases of the fourth and the fifth MVs. There is less departure when α is in between -1 and 1.

The seventh MV, shown in the bottom left panel, is similar to the sixth MV, except that the contrasting elongations and contractions are in different directions. When $\alpha = 1$, one of the directions of elongation is North. The amount of departure from the median function is similar to the typical departure observed in the case of the sixth MV.

The eighth MV, shown in the bottom middle panel, features elongation in four directions that are about 90° apart and contraction in between. The departure from the median function is comparable to the sixth and seventh MVs.

The ninth MV, shown in the bottom right panel, is similar to the eighth MV, except that the contrasting elongations and contractions are in different directions. The departure from the median function is also similar to the typical departure observed in the cases of sixth, seventh and eight MVs.

In Figure 7, we plot of the principal component functions corresponding to the years

2002-2012 estimated by using proposed modified FPCA based on data of different years. Each panel corresponds to a single principal component function, while each curve in a single plot represents a particular year (2002 to 2012). It is found that

- the first PC function is approximately a constant,
- the fourth and the fifth PC functions constitute an orthogonal pair of approximately sinusoidal functions with period 2π (fundamental or lowest frequency),
- the second and the third PC functions constitute an orthogonal pair of approximately sinusoidal functions with period π (first harmonic),
- the sixth and the seventh PC functions constitute an orthogonal pair of approximately sinusoidal functions with period $2\pi/3$ (second harmonic),
- the eighth and the ninth PC functions constitute an orthogonal pair of approximately sinusoidal functions with period $2\pi/4$ (third harmonic).

This pattern is remarkably stable across the years. There is occasional change in phase from one year to another, but the shapes of the functions are consistent across the years. The variances of principal components associated with these functions, reported in Table 2, also show consistent proportions of the total variation being explained by the different PC functions. The pairs of sinusoids of a particular frequency/period have about the same value of the associated variance of principal component.

The threshold for minimum size of a region under rainfall, used in the above analysis is 200 sq km. The modes of variation are found to be remarkably stable when this threshold is as small as 50 square km or as large as 1000 square km.

Table 2: Percentage of variance explained by each of the first j principal components, for $j = 1, \ldots, 12$. Each row represents eigenvalues corresponding to a particular year (2002 to 2012).

	Eigenvalue of principal component number											
Year	1	2	3	4	5	6	7	8	9	10	11	12
2002	54.36	11.47	10.70	5.54	4.31	2.47	2.14	1.36	1.23	0.81	0.77	0.48
2003	52.26	11.78	10.44	6.64	4.81	2.33	2.00	1.60	1.31	0.82	0.81	0.57
2004	52.77	11.34	10.37	6.87	4.73	2.25	2.14	1.43	1.39	0.82	0.74	0.56
2005	52.96	11.86	10.49	7.06	4.47	2.07	1.94	1.38	1.33	0.80	0.65	0.54
2006	52.13	11.94	10.61	7.44	4.20	2.21	1.99	1.39	1.31	0.82	0.78	0.54
2007	56.51	10.58	9.66	6.33	4.28	2.19	1.94	1.39	1.18	0.71	0.56	0.46
2008	57.37	11.14	9.52	5.31	4.33	2.13	1.92	1.20	1.17	0.69	0.68	0.48
2009	58.70	10.08	8.86	5.70	4.33	1.94	1.84	1.30	1.23	0.71	0.66	0.48
2010	55.47	11.64	9.81	5.70	4.22	2.27	1.97	1.26	1.21	0.76	0.73	0.53
2011	56.60	11.04	9.86	6.25	3.91	1.99	1.80	1.39	1.27	0.65	0.65	0.45
2012	53.57	11.43	11.05	6.55	4.34	2.18	1.95	1.35	1.22	0.79	0.69	0.46

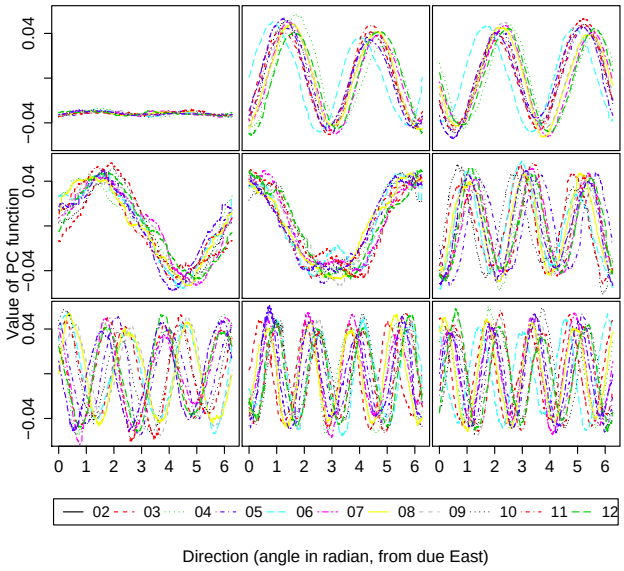

Figure 7: Plot of the first nine principal component functions corresponding to the years 2002-2012 using FPCA based on data of different years. Each panel corresponds to a single principal component function, while each curve in a single plot represents a particular year (2002 to 2012).

4.3 Parametric modeling of contour of region under rainfall

The nature of the PC functions described in the preceding section indicates that they may be approximated by sinusoidal functions with integer frequencies. In other words, a Fourier series approximation of the contours may suffice. To this end, we consider the following functional linear model (Ferraty & Vieu, 2006) for the kth complete contour, $k = 1, \ldots, n_u$,

$$X_k(\theta) = \sum_{i=0}^{d_k} A_{ki} \cos(i\theta) + \sum_{i=1}^{d_k} B_{ki} \sin(i\theta) + e_k(\theta), \tag{6}$$

where d_k is the order of the Fourier series approximation, A_{ki}, B_{ki} are the coefficients of the sinusoids with frequency i and $e_k(\theta)$ is zero mean Gaussian stochastic process for the kth contour. A simpler description of the model is

$$X_k(\theta) = \sum_{i=0}^{d_k} C_{ki} \cos(i\theta - i\phi_{ki}) + e_k(\theta), \tag{7}$$

where $(C_{ki}, i\phi_{ki})$ is the polar representation of the point with Cartesian coordinates (A_{ki}, B_{ki}).

For the kth complete contour, the Fourier coefficients in (6), $\mathbf{A}_k = (A_{k0}, \ldots, A_{kd})$ and $\mathbf{B}_k = (B_{k1}, \ldots, B_{kd})$ were estimated by the least squares method and the value of d_k may be selected by using the risk based data driven technique (see Hart (1997), page 167). Both the mode and the median of $d_1, \ldots, d_{n_u}$ happened to be 6. Therefore, we chose $d_k = 6$ for all k. This selection corresponds to eleven sinusoidal terms in the right hand side of (6), which is comparable to the ten principal component functions that had been deemed adequate in Section 4.2.

Let $(\hat{\mathbf{A}}_k, \hat{\mathbf{B}}_k)$ be the least squares estimators of $(\mathbf{A}_k, \mathbf{B}_k)$, described above, for $k = 1, \ldots, n_u$. For each k, suppose $X_k(\theta; \hat{\mathbf{A}}_k, \hat{\mathbf{B}}_k)$ is the approximation of the k^{th} contour from the parametric model (6), with $d_k = d$, $\mathbf{A}_k = \hat{\mathbf{A}}_k$ and $\mathbf{B}_k = \hat{\mathbf{B}}_k$. Further, suppose $\hat{X}_k(\theta)$ is the equivalent order nonparametric approximation of the k^{th} contour from the

Table 3: Empirical quartiles of $ISE_N^{(k)}$ and $ISE_P^{(k)}$, $k = 1, \ldots, n_u$.

Year	0%	25%	50%	75%	100%
Non-parametric	2.65	31.18	48.48	79.54	640.56
Parametric	2.26	26.32	41.29	69.02	651.74

nonparametric FPCA model (2), i.e.,

$$\hat{X}_k(\theta) = \hat{\mu}(\theta) + \sum_{j=1}^{2d+1} \left(\int_0^{2\pi} (X_k(\zeta) - \hat{\mu}(\zeta)) \varphi_j(\zeta) d\zeta \right) \hat{\varphi}_j(\theta). \tag{8}$$

The integrated squared error for the kth contour using parametric and non-parametric approximation can be computed as follows,

$$\begin{aligned} ISE_P^{(k)} &= \int_0^{2\pi} \left(X_k(\theta) - X_k(\theta; \hat{\mathbf{A}}_k, \hat{\mathbf{B}}_k) \right)^2 d\theta, \\ ISE_N^{(k)} &= \int_0^{2\pi} \left(X_k(\theta) - \hat{X}_k(\theta) \right)^2 d\theta, \end{aligned}$$

for $k = 1, \ldots, n_u$. Table 3 shows that the quartiles of the two quantities (computed from all complete contours over the years 2002-2012) are rather close. In fact, all the three quartiles of the integrated squared error of the parametric approximation are slightly smaller than those of the nonparametric approximation. Figure 8 shows that the histograms of the two sets of integrated squared errors are also comparable. These findings further justify the parametric approximation.

Once the parametric representation of (7) is chosen, the star-hull of every region of rainfall is represented by a set of amplitude and phase parameters. Specifically, the approximation for the kth region is the region bounded by the contour

$$g\left(\sum_{i=0}^{d} \hat{C}_{ki} \cos(i(\theta - \hat{\phi}_{ki})) \right), \quad k = 1, \ldots, n_u, \tag{9}$$

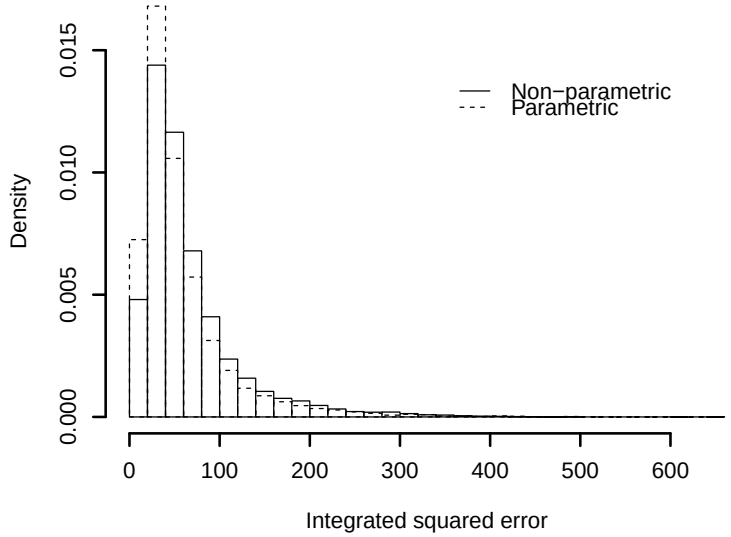

Figure 8: Histograms of integrated squared errors for parametric and non-parametric approximations of contours

where g is the re-transformation function defined in Section 4.2, $(\hat{C}_{ki}, i\hat{\phi}_{ki})$ is the polar representation of the point with cartesian coordinates $(\hat{A}_{ki}, \hat{B}_{ki})$, and $\hat{\phi}_{k0} = 0$.

The distribution of the shape objects is approximately described by the $(2d+1)$-variate distribution of these amplitudes and phases of the sinusoids. In order to appreciate different aspects of this distribution, we consider the marginal distributions of the amplitudes of the different sinusoids, represented in the original (re-transformed) scale. Specifically, for $i = 0, 1, \ldots, d$, we represent the contribution of the i-th summand in (9) by $g\left(\hat{C}_{ki}\right)$. In Figure 9 we have plotted the histograms of the amplitudes (re-transformed as above) computed from all contours of the years 2002-2012. It is not surprising that the distributions of the re-transformed amplitudes are positively skewed. The histogram for $g(\hat{C}_{k0})$ (constant term) has a smaller mode. This is because of the fact that $\hat{C}_{k1}, \ldots, \hat{C}_{k5}$ are constrained to be positive, while $\hat{C}_{k0}$ is not. About half of all the contours have $\hat{C}_{k0} < 0$, and the re-transformed values of these are too small to be attainable by the re-transformed values

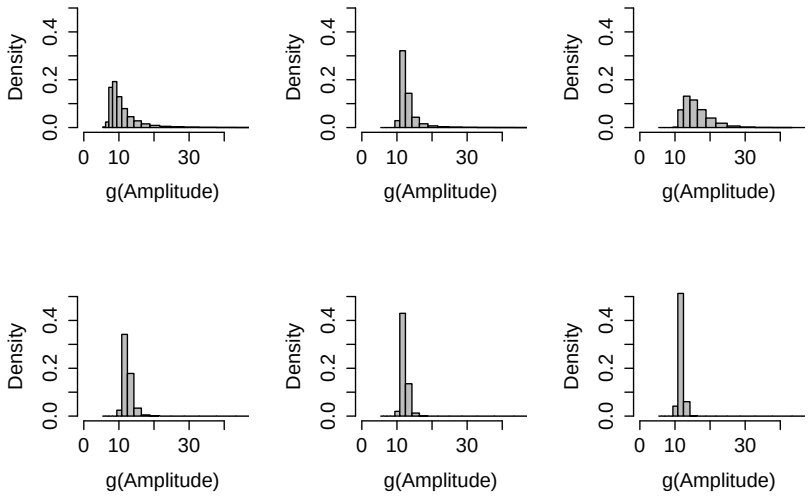

Figure 9: Histograms of re-transformed amplitudes of different sinusoids in (9)

of the other amplitudes. The histograms of $g(\hat{C}_{k1})$ and $g(\hat{C}_{k2})$ (top middle and top right panel of Figure 9) have wider support than the other histograms in this figure, indicating that the corresponding sinusoids in (9) have most diversity in amplitude.

The histogram of $g(\hat{C}_{k2})$ has the widest support. The associated sinusoid approximately corresponds to the second and the third modes of variation identified in Section 4.2. It represents a pattern of elongation/contraction of the mean circular shape that captures the main axiality of the region of rainfall. To identify the modal axis, we consider the bivariate distributions of $(\hat{A}_{k2}, \hat{B}_{k2})$. Each contour contributes a pair $(\hat{A}_{k2}, \hat{B}_{k2})$. We estimate the mode of this distribution by using the kernel density based method of Abraham et al. (2003) with the Gaussian kernel and optimal bandwidth as suggested by Scott (2008). Let us denote $(\hat{A}^*_{k2}, \hat{B}^*_{k2})$ as the estimated bivariate mode. Then the modal version of the second sinusoid in (6) is $\hat{A}^*_{k2}\cos(2\theta)+\hat{B}^*_{k2}\sin(2\theta)$. The corresponding re-transformed contour is the graph of $g(\hat{A}^*_{k2}\cos(2\theta)+\hat{B}^*_{k2}\sin(2\theta))$ against θ in polar coordinates. This contour represents the modal axiality of the region of rainfall, as captured by the second sinusoid in (6). The slope and the length of diameter of this contour describes the direction and extent of the

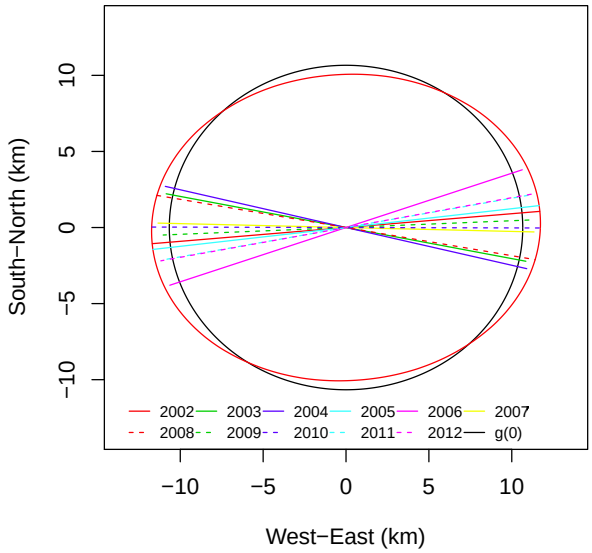

Figure 10: Baseline contour (black solid curve), modal axiality contour for the year 2002 (red solid curve) and diameters of modal axiality contours for the years 2002-2012

modal axiality.

In Figure 10 we have plotted the modal axiality contour $g(\hat{A}^*_{k2}\cos(2\theta) + \hat{B}^*_{k2}\sin(2\theta))$ vs. θ for the year 2002 along with its diameter (red solid curves) and the diameter of the modal axiality contours for each of years 2003-2012 in different colors and different line styles. For reference, we have also included the baseline contour $g(0)$ vs. θ (black solid curve). It may be observed that modal axiality is generally found to be around the East-West line. It is interesting to note that predominance of elongation of convective systems in the East-West axis coincides with that of the Inter Tropical Convergence Zone (ITCZ) of the summer monsoon over India.

We have also analyzed the modal directionality of $g(\hat{A}^*_{k1}\cos(\theta) + \hat{B}^*_{k1}\sin(\theta))$ in a similar manner. However, no strong mode or modal direction has emerged. We omit the details for brevity.

4.4 Reconstruction of contours of regions under rainfall

As an illustration of the nature of approximation of typical contours of regions under rainfall, we show in Figure 11 the approximations through up to ten largest principal components. The top panel shows the contour obtained after re-transforming the right hand side of (8) with $d = 0, 2, 3, 4$. The upper left plot represents only one summand and the other plots of the top row represents successively higher number of terms in the summation. The approximated contour is shown in solid lines, while the actual contour is shown in dashed lines. Inclusion of more and more summands evidently leads to better approximation.

The bottom row of Figure 11 represents the approximation of a region using up to four Fourier basis function in (9). The leftmost plot represents only one summand and the other plots represent successively higher number of terms in the summation. The approximated and actual region are shown in solid and dashed contour, respectively.

Since only ten largest principal component account for 95% of the total variation and we have interpreted nine of them, we examine a few instances of approximation by these principal components in Figure (12). Once again, the star-hull of actual contours of region of rainfall (solid) are compared with the reconstructed contour from the first ten principal components (black dashed contour). The blue dashed contour represents the approximation using six sinusoids. The both approximation is observed to be reasonable.

We now evaluate the performance of the parametric representation in approximating the actual contour (whose star-hull approximation had so far been used for modeling). For each contour we compute the symmetric area difference between the contour of the actual region under rainfall and its parametric approximation using (6) with $d_k = 6$ by aligning the contours at the centroid of their common convex hull. This approximation error is normalized by the area of the corresponding star-hull. Figure 13 shows the histogram of the error distribution. The median approximation error (in terms of area) of the regions of

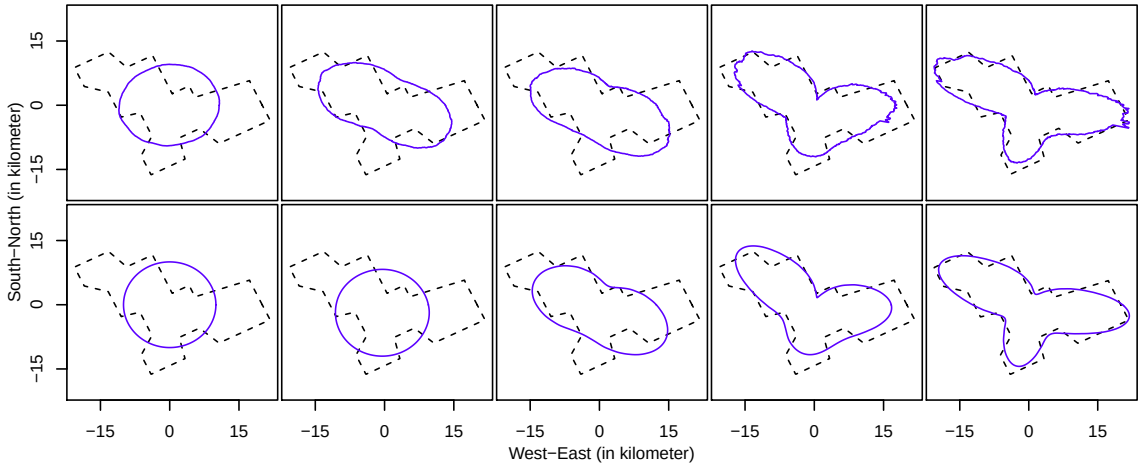

Figure 11: An example of reconstructed contours using cumulative totals of the first few nonparametric (top row) and parametric (bottom row) basis functions.

rainfall is about 9% . This level of overall approximation error appears to be reasonable. As an illustration, we plot in Figure 14 a few instances of actual outlines of region under rainfall (dashed line) and the corresponding parametric approximations (solid line).

5 Concluding remarks

The precipitation areas observed in the TRMM satellite data have diverse and irregular shapes. This fact poses a challenge to modeling, especially for large data sets. In this paper, we have presented an analysis of shapes of contiguous regions of rainfall, through the star-hull envelope of these regions. This approximation, which reduces the shape objects to functional data over $[0, 2\pi]$, makes them amenable to analysis by standard statistical methods.

We carried out functional principal component analysis to analyze these shape objects. The main challenge to this analysis is the presence of a large number (about 25%) of occluded regions. Through a careful analysis of the TRMM monsoon data over Eastern India, we developed a practical yet justifiable method of including the occluded regions in

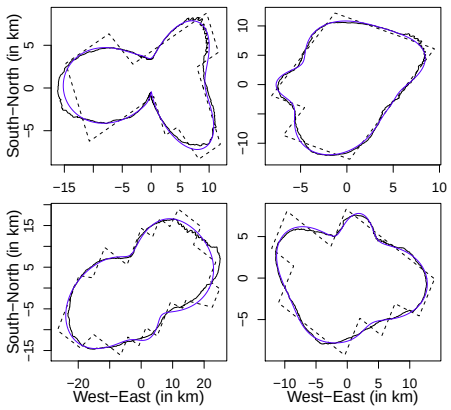

Figure 12: Examples of non-parametrically (black solid line) and parametrically (blue solid line) reconstructed contours along with original star-hulls (dashed line) of some regions under rainfall

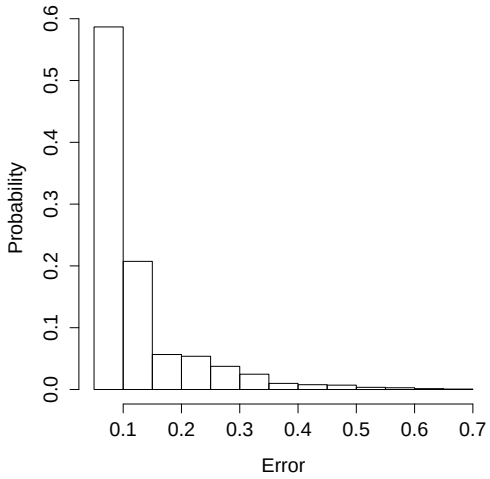

Figure 13: Distribution of normalized error of approximation (through (9)) of regions under rainfall

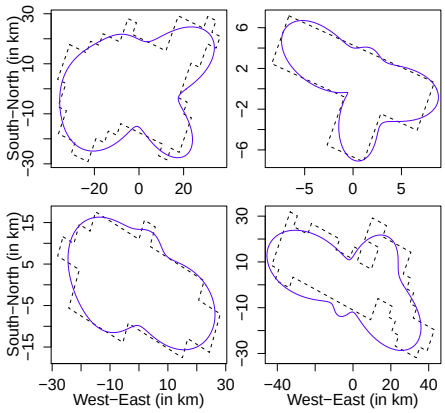

Figure 14: Examples of parametrically reconstructed contours (blue solid line) along with actual contours (black dashed line) of some regions under rainfall

this analysis.

The FPCA (adapted to possibly occluded regions) of the TRMM data set led to the identification of certain dominant modes of variation. These modes of variation are remarkably stable across calendar years. They also happen to be easily interpretable.

Approximation of the FPCA basis functions through sinusoidal functions makes the analysis even more interesting. It turns out that just five sinusoids with eleven associated parameters capture most of the shape information. This finding indicates a good potential for data compression. There are also substantial implications to understanding of the MCS precipitation areas, as each sinusoid makes interpretable contributions to the shape.

With the advent of a method to characterize raining systems based on their shape and size, it might become possible to evaluate the capabilities of Numerical Weather Prediction (NWP) models in simulating such features and to better understand the error characteristics of these models. NWP models are often verified with satellite data such as TRMM data (Chakraborty, 2010), by using different methods (Micheas et al., 2007; Ebert & Gallus Jr, 2009). The phase and amplitudes of the parametric representation of the contours can be used as alternative features of contiguous precipitation areas, which the verification

methods can utilize. Further, the proposed representation might pave the way for studies on how different covariates affect the size and the shape of precipitation areas.

While we used 200 square kilometers as the lower threshold for contiguous areas of positive precipitation, we found the modes of variations to remain stable when the threshold is anywhere between 50 and 1000 square kilometers. Likewise, the pattern did not change much when the threshold for minimum precipitation was raised from 0 to 5 mm per hour. These findings indicate that, though our analysis is limited to a certain range of areas under positive precipitation, our method is possibly scalable. Similar analyses of data sets with other resolution and coverage can expand that range, and possibly lead to a more complete understanding of MCS and other convective systems.

6 Acknowledgement

The first three authors thank Dr. Parthasarathi Ghosh of Geological Studies Unit, Indian Statistical Institute, Kolkata for introducing the TRMM data to them, and for many helpful suggestions. TRMM 2B31 V7 datasets were obtained from the NASA Goddard Space Flight Center Web site.

References

Abraham, C., Biau, G., & Cadre, B. (2003). Simple estimation of the mode of a multivariate density. *The Canadian Journal of Statistics*, *31*(1), 23-34.

Arkin, E. M., Chiang, Y.-J., Held, M., Mitchell, J. S. B., Sacristan, V., Skiena, S. S., & Yang, T.-C. (1998). On minimum-area hulls. *Algorithmica*, *21*(1), 119–136.

Austin, P. M., & Houze, R. A. (1972). Analysis of the structure of precipitation patterns in new england. *Journal of Applied Meteorology*, *11*, 926–935.

Bhattacharya, A., & Bhattacharya, R. (2012). *Nonparametric inference on manifolds* (Vol. 2). Cambridge University Press, Cambridge.

Bosq, D. (2000). *Linear processes in function spaces*. Lecture notes in Statistics.

Chakraborty, A. (2010). The skill of ecmwf medium-range forecasts during the year of tropical convection 2008. *Monthly Weather Review*, *138*(10), 3787-3805.

Delaigle, A., & Hall, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.*, *38*(2), 1171–1193.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. John Wiley & Sons, Ltd., Chichester.

Ebert, E. E., & Gallus Jr, W. A. (2009). Toward better understanding of the contiguous rain area (cra) method for spatial forecast verification. *Weather and Forecasting*, *24*, 1401–1415.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer, New York. (Theory and practice)

Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests*. Springer-Verlag, New York.

Houze, R. A. (1977). Structure and dynamics of a tropical squall–line system. *Mon. Wea. Rev.*, *105*, 1540–1567.

Houze, R. A. (2004). Mesoscale convective systems. *Reviews of Geophysics*, *42*(4). (RG4003)

Jirak, I. L. C., R., W., & L., M. R. (2003). Satellite and radar survey of mesoscale convective system development. *Monthly Weather Review*, *131*(10), 2428–2449.

Johnson, R. H. (1984). Partitioning tropical heat and moisture budgets into cumulus and mesoscale components: Implications for cumulus parameterization. *Mon. Wea. Rev.*, *112*, 1590–1601.

Keith, T. (2016). *An introduction to the theory and practice of plane and spherical trigonometry, and the stereographic projection of the sphere: Including the theory of navigation.* BiblioLife.

Kindratenko, V. V. (2003). On use of functions to describe the shape. *Journal of Mathematical Imaging and Vision*, *18*, 225–245.

Klassen, E., Srivastava, A., Mio, W., & Joshi, S. H. (2004). Analysis of planer shapes of using geodesic paths on shapes spaces. *IEEE transaction on pattern analysis and machine intelligence*, *26*.

Kurtek, S., Srivastava, A., Klassen, E., & Ding, Z. (2012). Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, *107, No. 499*, 1152–1165.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognition*, *31(8)*.

Micheas, A. C., Fox, N. I., Lack, S. A., & Wikle, C. K. (2007). Cell identification and verification of {QPF} ensembles using shape analysis techniques. *Journal of Hydrology*, *343*(34), 105 - 116.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77(2)*.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (Second ed.). Springer, New York.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*(1), 233-243.

Scott, D. W. (2008). *Multivariate density estimation.* John Wiley & Sons, Inc.

Srivastava, A., Klassen, E., Joshi, S. H., & Jermyn, I. H. (2011). Shape analysis of elastic curves in euclidean spaces. *IEEE transaction on pattern analysis and machine intelligence*, *33*.

von Hardenberg, J., Ferraris, L., & Provenzale, A. (2003). The shape of convective rain cells. *Geophysical Research Letters*, *30*(24), n/a–n/a. (2280)

Wang, J. L., Chiou, J., & Muller, H. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*(June), 257-295.

Yao, F., Mller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577-590.

Zipser, E. J. (1982). *Use of a conceptual model of the life cycle of mesoscale convective systems to improve very-short-range forecasts* (K. Browning ed.). Academic Press.