The relationship between Indian monsoon rainfall and low-pressure systems

Kieran M. R. Hunt[1] · Jennifer K. Fletcher[2]

Abstract

Indian summer monsoon precipitation is significantly modulated by synoptic-scale tropical low pressure areas (LPAs), the strongest of which are known as monsoon depressions (MDs). Despite their apparent importance, previous studies attempting to constrain the fraction of monsoon precipitation for which such systems are responsible have yielded an unsatisfyingly wide range of estimates. Here, a variant of the DBSCAN algorithm is implemented to identify nontrivial, coherent rainfall structures in TRMM-3B42 precipitation data. Using theoretical considerations and an idealised model, an effective capture radius is computed to be 200 km, providing upper-bound attribution fractions of 57% (17%) for LPAs (MDs) over the monsoon core zone and 44% (12%) over all India. These results are also placed in the context of simpler attribution techniques. A climatology of these clusters suggests that the central Bay of Bengal (BoB) is the region of strongest synoptic organisation. A *k*-means clustering technique is used to identify four distinct partitions of LPA (and two of MD) track, and their regional contributions to monsoon precipitation are assessed. Most synoptic rainfall over India is attributable to short-lived LPAs originating at the head of the BoB, though longer-lived systems are required to bring rain to west India and east Pakistan. Secondary contributions from systems originating in the Arabian Sea and south BoB are shown to be important for west Pakistan and Sri Lanka respectively. Finally, a database of precipitating-event types is used to show that small-scale deep convection happens independently of MDs, whereas the density of larger-scale convective and stratiform events are sensitive to their presence—justifying the use of a noise-rejecting algorithm.

Keywords India · Monsoon · Depression · Precipitation

1 Introduction

Monsoon depressions (MDs) are synoptic-scale disturbances that typically spin up near the head of the Bay of Bengal, before moving northwestward over peninsular India (Sikka 1977; Hurley and Boos 2015; Hunt et al. 2016a). While it is known that they are associated with both widespread (Godbole 1977; Mooley and Shukla 1989; Stano et al. 2002; Hunt et al. 2016b) and heavy (Ajayamohan et al. 2010; Fletcher et al. 2018; Hunt et al. 2018b) precipitation in central and northwest India, both the fraction of total seasonal rainfall for which they are responsible and its spatial variation (see

✉ Kieran M. R. Hunt
  k.m.r.hunt@reading.ac.uk

[1] Department of Meteorology, University of Reading, Reading, UK

[2] School of Earth and Environment, University of Leeds, Leeds, UK

Jadhav 2002) have garnered divergent estimates: 10% over the Ganges basin area (Dhar and Bhattacharya 1973), 27% for central India (and 14% for all India; Mooley and Shukla 1989), to as much as 50% for all India (Yoon and Chen 2005). MDs are accompanied during the monsoon season by weaker, more numerous disturbances known as monsoon low-pressure areas (LPAs), which are also significant rain-bringers, though less often associated with periods of extreme rainfall. Attribution of monsoonal rainfall to LPAs was considered by Hurley and Boos (2015), who found that the value is sensitive to the choice of radius one uses to ascertain 'association', retrieving numbers in the monsoon core zone from around 40% for a radius of 500 km to over 80% for rainfall within 1000 km of the LPA centre. Praveen et al. (2015) simply attributed all rainfall occurring on days with a LPA present and concluded that 60% of monsoon precipitation was due to such systems.

Recently, Hunt et al. (2016a) developed a method to identify and track Indian monsoon depressions in a way

consistent with definitions used by the India Meteorological Department. Their tracks, along with a 20 years record of high quality satellite observations of tropical rainfall through the tropical rainfall measuring mission (TRMM) multisatellite precipitation analysis (Huffman et al. 2007), make it possible to accurately estimate the contribution of monsoon depressions to both mean and extreme rainfall. However, questions about how to best make this estimate remain. Rainfall around monsoon depressions tends to be highest to the southwest of MD centre (Roy and Roy 1930; Ramanathan and Ramakrishnan 1933; Pisharoty and Asnani 1957; Rajamani and Rao 1981). If a radius of influence is assumed, how large should this radius be?

We approach this question in several ways, from very simple—assuming all rainfall on a day with an MD is caused by the MD—to sophisticated—identifying clusters. We produce a range of estimates with these methods and compare the results to previous results. We also examine the spatial distribution of mesoscale precipitating systems as an indicator of mesoscale organisation of convection in the vicinity of MDs.

Section 2 describes the data we use and the clustering method. In Sect. 3.1 we present upper bounds on mean and extreme rainfall attributable to monsoon depressions using the most naive attribution method. In Sect. 3.2 we show our estimates of MD rainfall attribution using several fixed radii of influence. Section 3.3 shows MD rainfall attribution using the clustering method, and in Sect. 3.4 we explore the types and spatial distribution of precipitating systems associated with monsoon depressions.

2 Data and methodology

2.1 ERA-interim

MD and LPA tracks were obtained using the tracking algorithm described in Sect. 2.3 on the European Centre for Medium-Range Weather Forecasting Interim Analyses (ERA-Interim; Dee et al. 2011).

2.2 TRMM

Our precipitation data is version 7 of the TRMM 3B42 product (Huffman et al. 2007), a multi-satellite product using the TRMM/Global Precipitation Mission (GPM) constellation consisting of the TRMM/GPM core precipitation radars and microwave imagers along with microwave and infrared satellites operated by a range of agencies. TRMM 3B42 is available at three-hourly means on a 0.25° grid for 1998-present, making it ideal for statistical investigation of rainfall associated with synoptic scale systems in the tropics. It can be obtained at https://mirador.gsfc.nasa.gov/.

2.2.1 Database of precipitating systems from TRMM PR

One goal of this study is to characterise how monsoon depressions organise precipitation; we therefore require data that can, to some extent, measure the character and degree in organisation of precipitation. Houze et al. (2015) described a new database of precipitating systems—namely, shallow convective systems, convective cores, and broad-scale stratiform regions—observed by the TRMM precipitation radar (PR) in the 2A25 product. These precipitating systems are classified by the Houze et al. (2015) algorithm according to their reflectivity, vertical and horizontal size, and classification as convective or stratiform by the TRMM 2A23 product. The database can be downloaded from the University of Washington Atmospheric Sciences website (trmm.atmos.washington.edu), and we will refer to it as the UW TRMM database. The database includes the locations of centres of precipitating objects and their time of observation, along with information about their horizontal and vertical sizes and statistics of precipitation within the systems.

The UW TRMM database consists of contiguous areas in the 2A25 product meeting specific thresholds. The types of precipitating systems that we use from the UW TRMM database are as follows: deep convective cores (30 dBZ reflectivity threshold, tops above 8 km); wide convective cores (30 dBZ threshold over a horizontal area at least 800 km^2); deep, wide cores meeting both of the previously stated criteria; and broad stratiform regions (contiguous regions designated stratiform by 2A23 at least 40,000 km^2 in size). These are the moderate thresholds in the UW TRMM database; the database also includes strong thresholds for each category (e.g., 40 dBZ reflectivity, 10 km height for deep convective cores). Houze et al. (2015) argue that moderate thresholds are best applied for systems over the ocean while strong thresholds are best applied for systems over land. However, strong threshold events are relatively uncommon in the Indian monsoon (Houze et al. 2015), making moderate threshold events a better choice for robust statistics.

In this study, we count the total number of each type of UW TRMM precipitating system per MD as a function of distance from MD centre. We estimate the uncertainty on this using the bootstrap method of Efron (1979), subsampling the data to generate 10,000 datasets of the same size as the original, and computing the standard deviation of this.

2.3 Monsoon depression tracks

Several studies have carried out objective, automated tracking on monsoon depressions (Hurley and Boos 2015; Hunt et al. 2016a) and monsoon low-pressure systems (Praveen et al. 2015); in contrast to earlier works where tracking was done manually (e.g. Godbole 1977; Mooley and Shukla 1989; Sikka 2006).

Here, we use the algorithm of Hunt et al. (2016a), which looks for positive anomalies in 850 hPa relative vorticity, as modified by Hunt and Turner (2017) and Hunt et al. (2018a) with one adjustment. It will be necessary to differentiate between depressions and the weaker but more common low-pressure systems (LPAs); to this end, we employ the definitions outlined by the India Meteorological Department (http://imd.gov.in/section/nhac/wxfaq.pdf). A low-pressure area is defined as one closed isobar in surface pressure at 2 hPa intervals, which must be within 3° of the centre when the system is over land, or accompanied by surface winds not exceeding 17 knots (8.7 m s^{-1}) if over sea. A depression (for our purposes, this category also includes 'deep' depressions) must have two to four closed isobars in surface pressure at 2 hPa intervals, which again must be within 3° of the centre when the system is over land, or instead accompanied by surface winds of 17–33 knots (8.7−17 m s^{-1}) if over sea. These categorisation switches are computed directly from the reanalysis data used to perform the tracking.

A point of definition: from hereon, 'monsoon depression' will refer to the whole track of any system that reaches at least depression status (as outlined above); 'low-pressure area' will refer to all parts of all tracks, including those that reach MD strength. Tracking is only carried out on vortices that meet at least the LPA criteria described in the previous paragraph, and only tracks (either part or whole) that occur in the months June to September are considered. Over the entire ERA-Interim archive (1979–2017), that gives us 109 MDs (of which 46 occur during the TRMM period) and 782 LPAs (of which 424 occur during the TRMM period).

2.4 Objective cluster identification

A core objective of this study is to identify coherent areas of precipitation, so that they might be linked (or not) to nearby synoptic activity. This requires the automated partitioning of precipitation footprints into an arbitrary (i.e. not preordained) number of clusters. We should also like our choice of clustering algorithm to work in a non-Euclidean geometry, to allow uneven cluster sizes, and have good scalability over the number of points used. Given these criteria, the most suitable choice of algorithm is the so-called density-based spatial clustering of applications with noise (DBSCAN; Ester et al. 1996). DBSCAN is additionally advantageous in that it permits background noise and highly nonlinear cluster shapes (such as those caused by certain orography) which would be inadequately partitioned by other algorithms.

2.4.1 Glossary

$\forall \chi$	For all members of the set χ.
$\in \chi$	Elements of the set χ.

$\chi_1 := \chi_2$	Replace all elements of set χ_1 with the elements of set χ_2.
$\chi_1 \cup \chi_2$	The union of sets χ_1 and χ_2.
$\mathbb{R}^n$	An n-dimensional coordinate space.
$\lfloor X \rfloor$	The floor of X, that is, its value rounded down to the nearest integer.

2.4.2 General description of the DBSCAN algorithm

All clustering algorithms require some parameter choice by the user, typically the number of clusters. For DBSCAN, two parameters are required: ϵ, the 'neighbourhood radius'; and μ, the minimum number of points to form a cluster core. The prescription is then:

1. Start from some random point P_i that has not already been assigned to some cluster (C_j) or as noise (N).
2. Find all points $P_{j \neq i}$ that are within distance ϵ of P_i. Call this set of points Π_i.
3. Initiate a new cluster, C_i, and assign P_i as a core member of that cluster.
4. If $|\Pi_i| \leq \mu$, assign P_i as noise and return to (i).
5. $\forall P_j \in \Pi_i$, if $P_j \in N$ then assign P_j as an outlier member of C_i; if $P_j \notin N$ then assign it as a core member of C_i.
6. Find all points $P_{k \neq j, k \neq i}$ that are within distance ϵ of P_j. Call this set of points Π_j.
7. If $|\Pi_j| \geq \mu$ then $\Pi_i := \Pi_i \cup \Pi_j$.
8. If unassigned points remain, return to (i).

2.4.3 Application of DBSCAN to precipitation data

Some adaptation is required for this algorithm to be applicable to precipitation data, which are on $\mathbb{R}^3$, rather than $\mathbb{R}^2$, as desired.[1] As clustering ought to depend on only the spatial distribution of the precipitation, and not its magnitude, we can collapse this degree of freedom without loss of generality. The method employed is as follows: consider some gridpoint (i, j), with instantaneous rainfall R_{ij} (in mm hr^{-1}), then distribute $\lfloor R_{ij} \rfloor$ points in the gridbox centred on (i, j). There are two corollaries: firstly, we implicitly reject regions where rainfall is less than 1 mm hr^{-1}, and secondly, we still retain a degree of freedom in how these points are distributed (we choose arbitrarily to place them at random). These choices have a negligible qualitative effect on the outcome.

More important is the selection of the clustering parameters, ϵ and μ. There are objective approaches that can be taken to decide these, most commonly used among clustering applications is the silhouette score, defined thus:

$$S = \frac{1}{N} \sum_i S_i, \tag{1}$$

---

[1] That is to say the points are not only distributed in two dimensions but contain magnitude information.
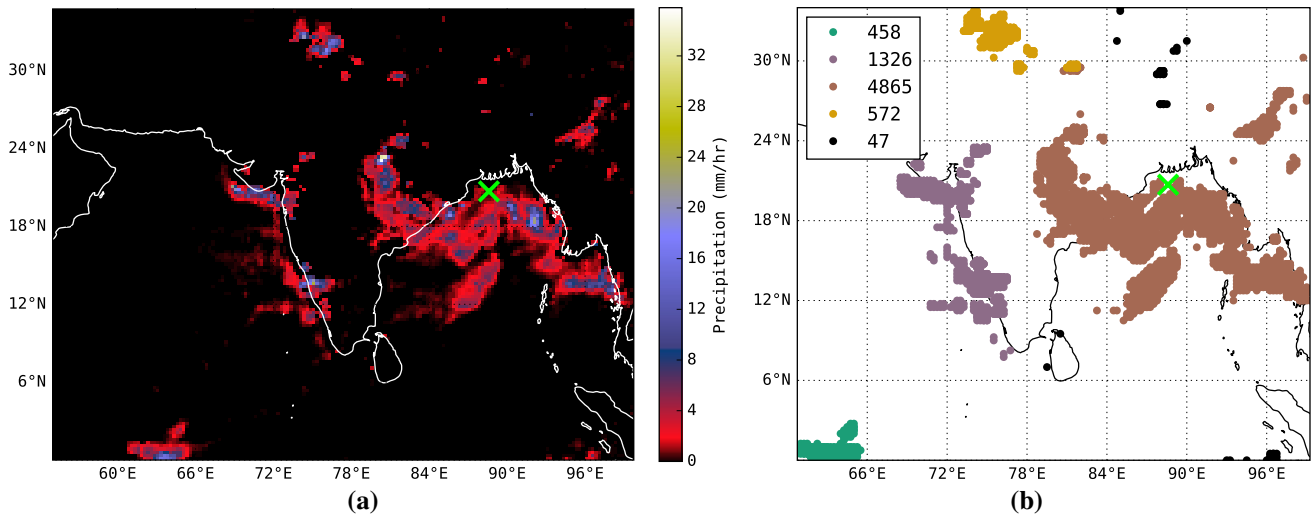
Fig. 1 Case study application for the clustering algorithm, using TRMM 3B42 precipitation data for a nascent depression on 2007-08-06 00UTC, whose centre is marked with a green cross in each figure.

a Instantaneous surface precipitation (mm hr^{-1}); **b** designated clusters demarcated by colour with assigned noise given in black, and the number of pixels in each cluster given in the legend

where N is the total number of points, and

$$S_i = \frac{\Lambda_i - L_i}{\max}\{\Lambda_i, L_i\}, \qquad (2)$$

where L_i is the mean distance between P_i and other members of its cluster, and Λ_i is the mean distance between P_i and the nearest cluster of which it is not a member. This is a simple way to compare cohesion (how close points are to members of their own cluster) and separation (how close they are to members of other clusters), and works particularly well for cases where noise is low.

We computed the mean silhouette score for twenty-four case studies that represented both MD days and non-MD days, representing different points in the monsoon season and diurnal cycle, across a selection of μ and ϵ. The computed optimum values were $\mu = 100$ and $\epsilon = 60$ km.[2] These are the parameters used for the example in Fig. 1b.

3 Results

3.1 Upper bound on rainfall attributed to MDs

We begin with the most naive approach by attributing all rainfall on days with an MD in a given domain to the MD itself—this is also an upper bound on the rainfall attributable to MDs. In order to do this we must identify dates when MDs are near enough to South Asia to influence rainfall

there. Indian MDs usually form in the Bay of Bengal, make landfall in northeast India, and propagate toward northwest India (e.g. Hunt et al. 2016b), i.e., across the monsoon core zone of Rajeevan et al. (2010). We therefore test the effect of MDs on all of South Asia as well as on the monsoon core zone (Fig. 2).
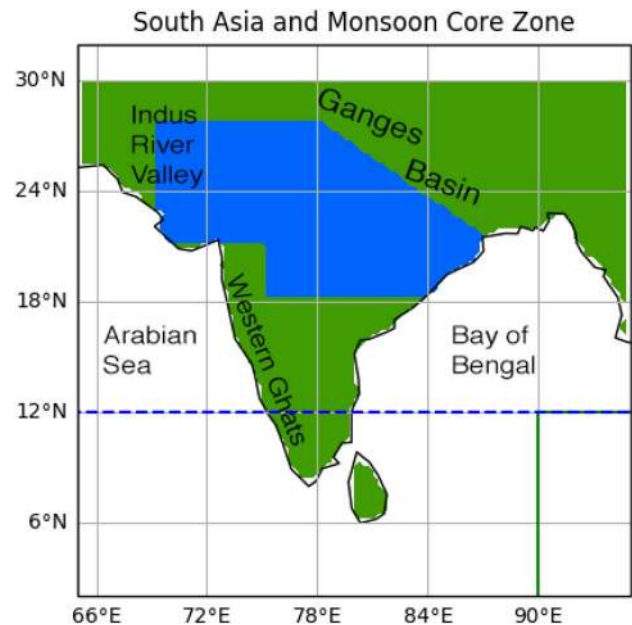
Fig. 2 Green colours indicate regions identified as South Asia, blue colours indicate the Monsoon Core Zone. MDs with centres northwest of the solid green (dashed blue) lines are counted toward MD days for South Asia (MCZ) in the analysis in Sect. 3.1

---

[2] Given our usage of a spherical surface distance metric, this is a great-circle distance.

**Table 1** Mean rainfall (mm hr^{-1}) on MD and non-MD days in South Asia and monsoon core zone (see Fig. 2)

	MD days	non-MD days
South Asia	0.36 ± 1.6	0.29 ± 1.4
MCZ	0.47 ± 1.9	0.28 ± 1.4

Errors indicate one standard deviation

We use the Indian monsoon depression tracks of Hunt et al. (2016b) and classify MD days and non MD days according to the following: if the tracking algorithm identifies an MD north of 12°N or west of 90°E at least once on a given date, the day is classified as an MD day for South Asia. If the algorithm identifies an MD north of 12°N, the day is classified at an MD day for the monsoon core zone. These boundaries are indicated in Fig. 2. This provides an extreme upper bound on the total rainfall that can be attributed to monsoon depressions.

The mean rainfall from TRMM 3B42 on MD days and non MD days is shown in Table 1 for South Asia and the monsoon core zone. In both domains, mean rainfall is higher on MD days than non-MD days—almost twice as high in the MCZ. As expected, MDs have greater impact on the MCZ than on South Asia as a whole, as the MCZ corresponds closely to their typical track (e.g. Hunt et al. 2016b). However, in both regions and for both MD and non-MD days, the standard deviation in rainfall is much greater than the difference between the means. In other words, while on average rainfall is much higher on MD days, the presence or absence of an MD is not a strong predictor of rainfall over even the MCZ, given the large variability in monsoon rainfall. Subsequent results in this section will mostly focus on the MCZ since that is the region where MDs have the greatest effect.

The spatial pattern of mean rainfall on MD days and non MD days is shown in the upper panels of Fig. 3. In the monsoon core zone and the Bay of Bengal, as well as the northern end of the peninsular west coast, MD days have
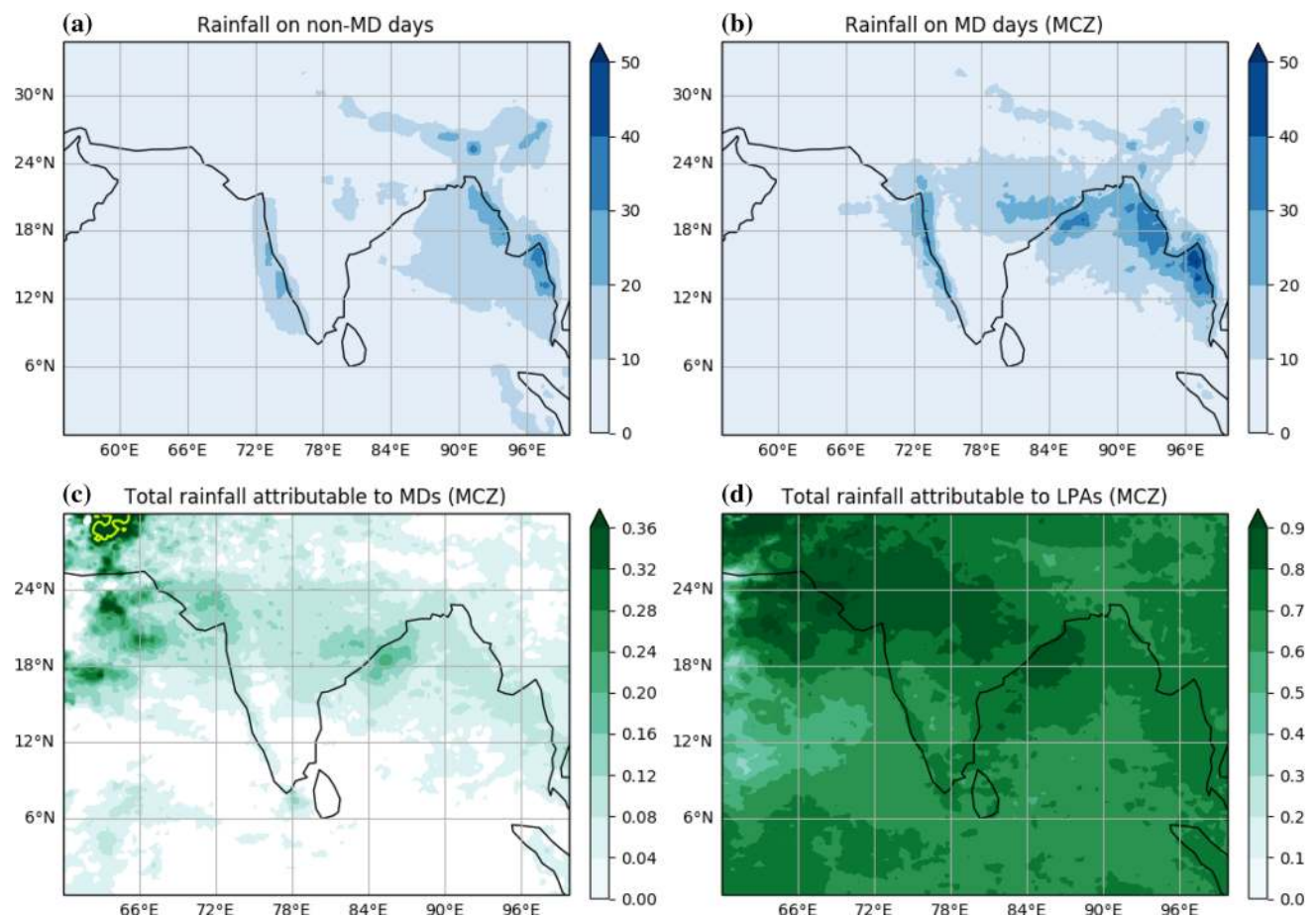
Fig. 3 **a** Mean June–September rainfall on non MD days (mm day^{-1}); **b** mean June–September rainfall on all MD days; **c** fraction of total June–September rainfall that occurred on MD days; **d** fraction of total June–September rainfall that occurred on days with low pressure areas. The yellow contour in **c** indicates 50% of rainfall occurred on MD days

(a) Monsoon depressions

(b) Monsoon low-pressure areas
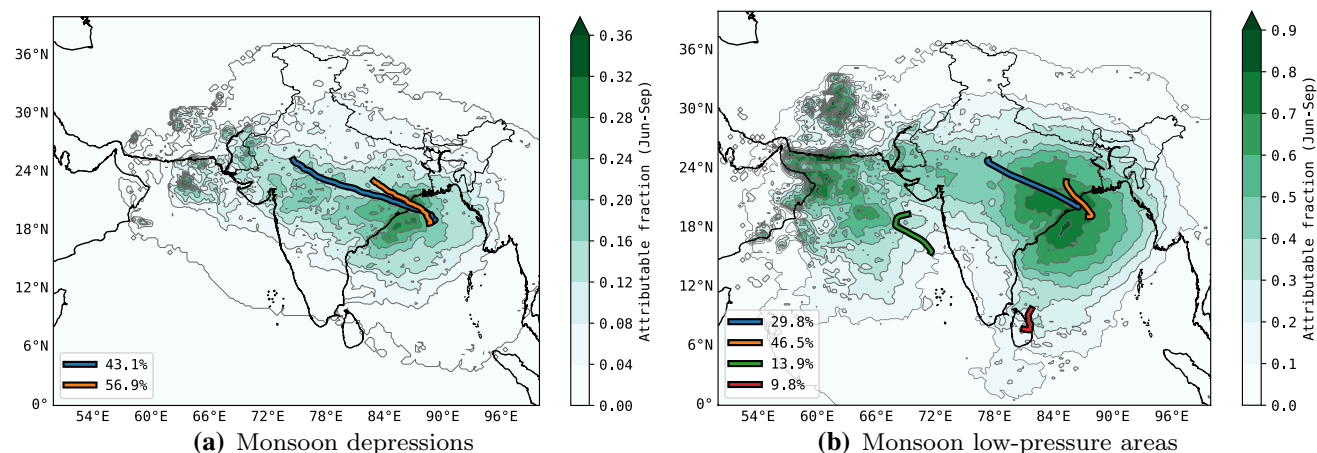
Fig. 4 Fraction of monsoon precipitation attributed to **a** depressions, **b** low-pressure areas using an 800 km radius of influence. For each type of system, *k*-means clustering was used to separate the tracks into statistically significant clusters, whose mean path and population frequency are also shown

considerably higher mean rain rates. However, the total contribution of MDs to rainfall in South Asia is usually less than 20% (Fig. 3c). This is because MDs, as classified by the IMD, are rare, occurring on average three times per monsoon season (e.g., Sikka 1977). The notable exception to this is in the far northwestern region, in the mountain range west of the Indus river in southern Pakistan. In this region MDs contribute around half of the June to September rainfall. Figure 3c suggests that many estimates of the contribution of MD rainfall to total rainfall in India, or various regions within India, are probably too high. Some of these estimates (e.g., Hurley and Boos 2015) included weaker but more frequently occurring monsoon low pressure areas—when all such systems are included (Fig 3d), the maximum rainfall attributable is much higher.

3.2 Assuming a fixed radius of influence

A simple way to compute precipitation attribution is to assign a fixed radius of influence to the systems, and assume that all precipitation occurring within that radius is caused by the system.[3] There are two sources of uncertainty here: firstly, tropical convection is present in the background regardless of cyclone passage, and it is unreasonable to assert that, for example, an isolated convective cell hundreds of kilometres from the centre has been triggered by the system; secondly, determining the correct (or effective) radius of influence is not trivial. That having been said, it provides a useful benchmark given a sensible estimate for the radius

of influence, and has been used in previous studies (Hurley and Boos 2015; Hunt et al. 2018c).

Figure 4 shows the fraction of summer (June–September) precipitation for which (a) monsoon depressions and (b) monsoon LPAs are responsible, assuming a fixed radius of influence of 800 km. In each case, the tracks have been separated into clusters using a *k*-means method, the aim being to produce as many clusters as possible subject to the criterion that they were significantly different from each other. There were 109 tracked depressions in the period 1979–2016, of which 46 existed in the TRMM period (1998–2016). The attribution fraction peaks over the head of the Bay of Bengal, and along coastal areas of Odisha and West Bengal, where it averages about 25% (as much as a third in some places). The fraction diminishes but remains significant over most of the monsoon core zone, and is still over 15% as far west as Gujarat. The depression tracks are separated into two distinct categories by *k*-means, each of which populate about half the catalogue: both originate at the head of the Bay of Bengal; one type then propagates an average of 800 km inland (this type is more prevalent earlier in the season), the other, 1700 km. Both types are significant contributors to the precipitation.

Comparing Fig. 3c with Fig. 4 a reveals an intriguing result: the majority of rainfall over the mountains and plateau of southwestern Pakistan falls on MD days. Though a fraction of MD and LPA tracks do propagate this far, we cannot rule out a more remote influence of monsoon depressions on rainfall in South Asia's arid northwest, which will be examined in future work.

Low pressure areas are considerably more numerous: of the 782 tracked in ERA-Interim (1979–2016), 424 were during the TRMM period (1998–2016). Their 800 km fixed radius contribution to monsoon precipitation is given

---

[3] The 'centre' of a system is here defined as the centroid of its 850 hPa relative vorticity.

(a) Mean cluster frequency

(b) Median cluster size
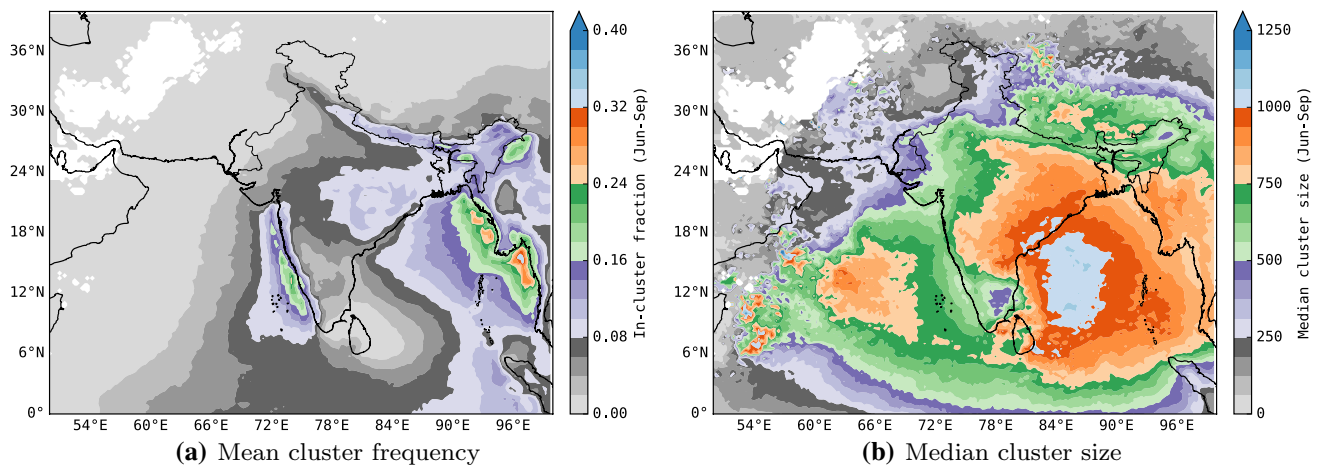
Fig. 5 Climatological precipitation cluster statistics for June–September: **a** fraction of time that a given pixel can be found in an identified precipitation cluster; **b** median cluster size (total number of 0.25° pix-els). Data are 3-hourly from TRMM 3B42, 1998–2016. Please refer to the text for details of the cluster identification algorithm

in Fig. 4b. According to this method, approximately 60% of rainfall in the monsoon core zone can be attributed to these LPAs. This fraction reaches almost 80% for a siz-able area over parts of northwest peninsular India and the Bay of Bengal. The footprint is similar to that of Fig. 4a, although the magnitude is somewhat greater; one key dif-ference, however, is the contribution of precipitation over Pakistan and the Arabian Sea from systems off the west coast of India. Though noisy, some of the values in this area exceed 80%. There are four distinct types of monsoon LPA: two of them, in blue and orange, are analogous to the two types of depression, and comprise more than three quarters of the population. The other two are relatively short-lived and slow-moving systems existing over the east coast of Sri Lanka and in the east Arabian Sea respectively. We have already mentioned the contribution of the latter, but it seems that the former contributes little, if any, precipitation to the monsoon.

3.3 Clustering

The clustering algorithm outlined in Sect. 2 was applied to 19 years of gridded precipitation data (TRMM 3B42, 1998–2016, 0.25° resolution), selected summer climatolo-gies from which are shown in Fig. 5. Figure 5a shows, for June–September, the fraction of time for which an identified cluster is present. We note that the broad structure is quite similar to that of mean summer rainfall (e.g. Sperber et al. 2013): there are maxima upstream of the Western Ghats (India) and the Arakan and Tenasserim ranges (Myanmar), weaker maxima along the Himalayan foothills and in the monsoon core zone, and minima in the south peninsula rain shadow and towards the arid northwest.

Figure 5b shows the climatological median cluster size[4] over the same period. These are most simply interpreted as a metric of the characteristic scale of precipitation organisa-tion at a given point. There are two significant maxima—the larger of which is centered over the Bay of Bengal, spread-ing over much of central India, and the smaller of which is located in the Arabian Sea.

These clusters provide us with coherent mesoscale to syn-optic-scale areas of related precipitation; we assume that it is probable that all precipitation in a given cluster is caused by the same mechanism (e.g. orography, cyclone, or MCS). Now, armed with a database of these precipitation clusters, we can attempt to assign them to depressions and LPAs to determine attribution.

It is not immediately clear how to approach this, so let us examine the relationship between radius-of-influence and attributable fraction for some idealised cases. For each scenario, the mean of 1000 runs is used. Consider a simplified monsoon core zone $(15-25°N, 70-85°E)$, into which a 'depression' centre is randomly placed. In the simplest case, we set rainfall to be homogeneous: 10 units per pixel inside the hypothetical depression, and 1 unit per pixel outside; we also hold the depression radius constant at 300 km. Then, we apply the fixed radius-of-influence technique used in the previous subsection, across a range of radii, computing the attribution fraction that each radius-of-influence gives. This is given by the solid green line in Fig. 6. It grows quadratically with radius, as expected, until a gradient discontinuity at the depres-sion radius (300 km); thereafter it grows as a much slower

---

[4] This is given in units of in-cluster pixels, which have an area of 0.25° × 0.25° here.
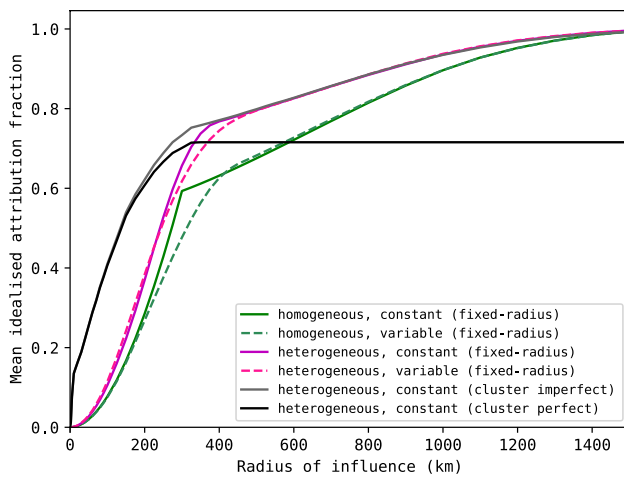
Fig. 6 Attribution fraction as a function of radius of influence for the naïve fixed-radius and cluster-based attribution techniques applied to selected idealised cases. Details of the algorithms, idealisations, and applications are given in Sect. 3.3


Fig. 7 Comparison of mean values of attributable fraction of rainfall in the monsoon core zone (June–September) for MDs (red) and LPAs (blue), and cluster-based (cross) and fixed-radius (circle), for a range of capture radii

quadratic, slowly becoming linear (and eventually asymptotic to 1) as boundary effects become appreciable. The correct attribution fraction, ~ 0.6, which can be read off the right-hand limit of the perfect-cluster line, is at the discontinuity, which here is also the prescribed depression radius. Let us now add an element of complexity, by allowing the depression radius to randomly vary between 200 and 400 km. The resulting change is given by the dashed green line in Fig. 6, and is slight: the only significant change is a smoothing of the discontinuity of the constant-radius case. Again, for reasons that should now be clear, the correct attribution fraction (which is slightly higher than the previous case) is found at the minimum of the second derivative, the knee of the curve.

Of course, precipitation is not homogeneous, if it were we would not be able to cluster it. Therefore, we next replace our flat rainfall with some simple heterogeneous blobs, prescribed as follows: two 'stratiform' blobs of radius 100 km and rain rate 5 units per pixel with centres inside the depression radius; and ten 'convective' blobs of radius 10 km with rain rates of 10 units per pixel, placed randomly in the domain. This was subject to the same computation as for homogeneous rainfall, and is given by the solid (constant depression radius) and dashed (variable radius) magenta lines in Fig. 6. In this case, variance of the depression radius no longer exhibits any significant control on the attribution function because the area-integrated rainfall is fixed in the setup. The correct attribution fraction in this example is about 0.71, but there is no way to extract this value from the naïve attribution function: using the second derivative as we did before now produces an overestimate, even knowing the depression radius is not useful—the two lines intersect the
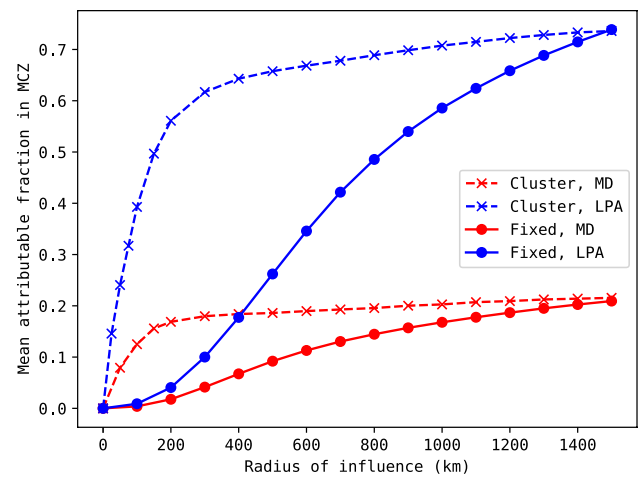
correct fraction at radii of 332 km and 351 km respectively. This naïve method is clearly inadequate for heterogeneous rainfall.

So, how does the more 'intelligent' clustering algorithm fare at attribution? Firstly, each individual cluster is tagged, and if any part of any cluster falls within the radius of influence, then the whole cluster is attributed. This calculation is given by the grey line in Fig. 6. Secondly, a slight change—only those clusters with a large area are attributed, the smaller 'convective' blobs are not, mimicking a perfect noise-removing cluster technique (as discussed in Sect. 2.4), this is given by the black line. The imperfect clustering technique outperforms the naïve method, estimation of the attribution function from the knee of the curve yields a 2% overestimate, an improvement on the earlier 6%. However, linear growth persists for overly large radii of influence. The perfect clustering method, by definition, is asymptotic to the correct attribution fraction; the extremum of its second derivative is therefore a slight underestimate ($\sim 2\%$). In reality, our clustering algorithm falls somewhere in between the two (more likely closer to the imperfect regime than not), but in the case of heterogeneous rainfall, will outperform the fixed radius of influence method.

Now that we can correctly interpret these attribution functions, let us compare them using real data for monsoon LPAs and depressions. Figure 7 shows the mean attributable fraction in the monsoon core zone for both LPAs (blue) and depressions (red) at a range of radii of influence. Monsoon rainfall in general is extremely heterogeneous, and during periods of synoptic activity, also embedded in large clusters. This is clear in the difference between the fixed radius and cluster attribution methods for each type of system, and

(a) Monsoon depressions   **(b)** Monsoon low-pressure areas
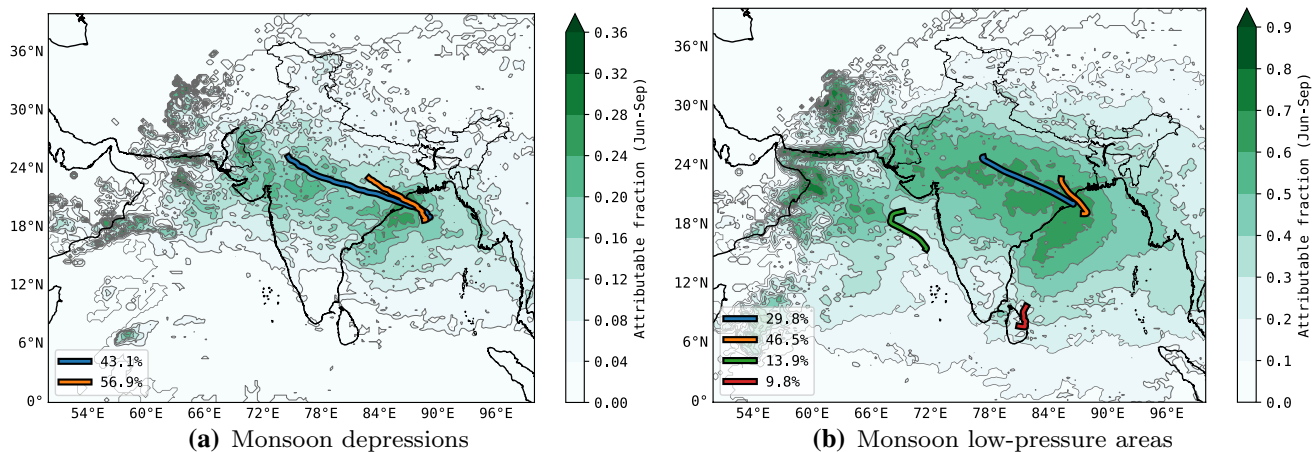
Fig. 8 As Fig. 4, but determining attribution using the clustered precipitation method. All clusters that come within 200 km of the system centre are attributed. Note that the colour scales differ between figures

Table 2 Precipitation-weighted mean attribution fractions for LPAs and MDs in three domains (the monsoon core zone, all India, and the domain) computed using both the fixed-radius and cluster methods

	MCZ	All-India	Full dom.
LPA (fix. 800 km)	0.524 (0.486)	0.371 (0.338)	0.222 (0.158)
LPA (clust. 200 km)	0.572 (0.561)	0.444 (0.425)	0.316 (0.224)
MD (fix. 800 km)	0.154 (0.145)	0.100 (0.093)	0.052 (0.029)
MD (clust. 200 km)	0.172 (0.169)	0.119 (0.115)	0.077 (0.048)

Values in square brackets indicate area-averaged fractions (i.e. no weighting)

indicates that the naïve fixed radius technique could produce quite poor estimates of precipitation attribution fraction.

Using the analysis from Fig. 6, we can read off the radii-of-influence for depressions and low pressure areas in Fig. 7. Doing so, we retrieve a fairly conservative value of 200 km and hence attribution fractions over the monsoon core zone of 17% and 56%; these are necessarily upper bounds. This intuitively seems like a small radius-of-influence, but we recall that once any part of a labelled precipitation cluster falls within this radius, the whole cluster is attributed. Thus, it is best to interpret this figure in the context of Fig. 5b—for example the mean cluster size over the BoB and MCZ is about 800 quarter-degree pixels, which corresponds to a length scale of ∼780 km. Adding this number in quadrature with the stated 200 km returns the 800 km value suggested by the fixed-radius method.

Figure 8 shows how the attribution is distributed spatially for each type of system, along with the main track types for each system (see text associated with Fig. 4 for discussion on these). Spatially, the pattern is similar to that retrieved with the fixed radius-of-influence method, though slightly noisier, slightly lower in magnitude, and with smaller gradient. The last is a result of removing (or at least, mitigating)
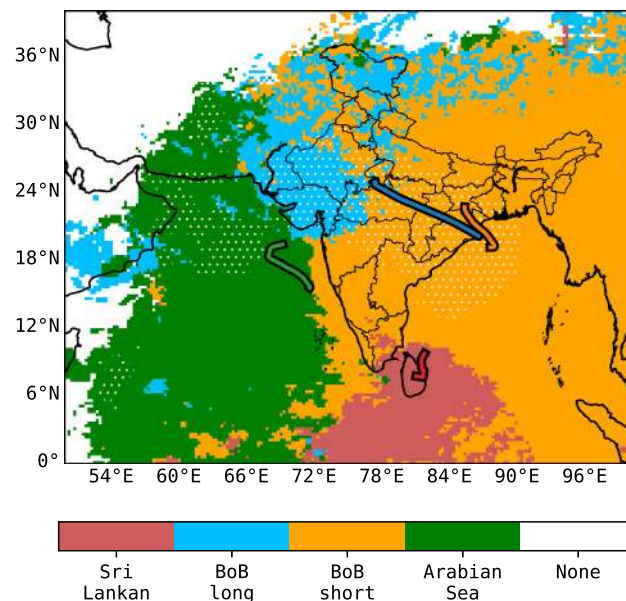
Fig. 9 Regions coloured by the LPA type responsible for the most precipitation. White stippling indicates where more than half the monsoon precipitation is attributed to LPAs; the white area indicates where no monsoon precipitation is attributed to LPAs

the effect of convolution with the track density function. In the case of depressions, the maximum, at the head of the Bay of Bengal, reaches a little over 30%; for LPAs, the maximum also extends reasonably far inland and across the peninsula, where its value is about 70%.

Mean values of the LPA and MD attribution fractions for a selection of domain choices are given in Table 2, for both the fixed-radius and cluster methods. These are given primarily as precipitation-weighted means (indicating the fraction of total precipitation in the domain that is attributable) as well as just an area-average attribution fraction (which

has less physical meaning). The cluster-based attribution method indicates that over 90% of monsoonal precipitation in the MCZ is caused by LPAs (30% of which is caused by MDs); whereas 65% of monsoon rainfall over all India is caused by LPAs (25% of which is caused by MDs). These values are comparable to the bounds computed in earlier work (Hurley and Boos 2015; Praveen et al. 2015). Table 3 provides a state-by-state, cluster-by-cluster breakdown of the attribution values.

We finalise this discussion by looking at the relative importance of the type of LPA in monsoon precipitation. Using the track types outlined in Fig. 4b, which we shall refer to by their genesis basin (i.e. Sri Lankan, BoB long, BoB short, and Arabian Sea), Fig. 9 identifies the category of LPA reponsible for bringing the most precipitation to each region. For reference, these values are tabulated in the "Appendix" section. The resulting partitions are intuitive: the short-lived Sri Lankan LPAs dominate the synoptic rainfall over Sri Lanka and parts of Tamil Nadu; the Arabian Sea systems are the major source of synoptic precipitation over almost the entirety of the Arabian Sea, as well as much of southwest Pakistan and Afghanistan, where they are responsible for over half of all precipitation; the common, short-lived BoB systems are associated with most precipitation across central and north India, as well as over the head of the Bay of Bengal; however, the less common but longer lived Bay of Bengal systems deliver the majority of rainfall to northwest India. While it has been commonplace in previous studies to separate systems by their genesis basin (i.e. Bay of Bengal, Arabian Sea, or land), we have not pursued that here as there is no evidence to suggest that genesis location is a better predictor for rainfall than the whole track. For the curious reader, however, we give the fractions of each cluster whose tracks have geneses over land: Sri Lankan—40.5%; BoB long—67.2%; BoB short—38.6%; Arabian Sea—24.1%.

3.4 Characteristics of precipitating systems around MDs

The heavy rainfall associated with monsoon depressions suggests that the synoptic forcing within depressions organises deep convection. One therefore might expect more organised convection within the radius of influence of MDs. We use the UW TRMM database of precipitating systems as seen by the TRMM PR as a rough objective identification of convective organisation. In particular, we expect broad stratiform regions to occur more frequently where synoptic and mesoscale flows have organised convection. Deep convective cores are expected to occur both as disorganised 'popcorn' convection and embedded within mesoscale convective systems.
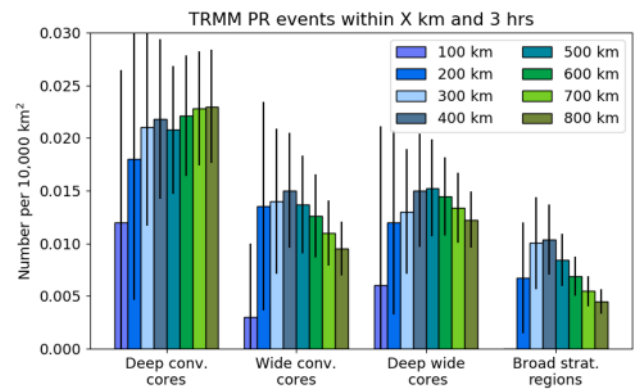
Fig. 10 Number of UW TRMM precipitating system types within a given distance from MD centre, separated by type, per 10,000 km². See Sect. 2.2.1 for descriptions of types. All types identified occurred within three hours of the time at which the MD location was identified. Confidence intervals calculated by bootstrapping

Figure 10 shows the number density of precipitating system types within a range of distances from MD centre. The TRMM PR had a swath width of 247 km, meaning that many events are missed; the reader should therefore focus on the change in the number with radius rather than the values on the y axis. For all radii, deep convective cores are most common and broad stratiform regions are least common. For a radius of 100 km, few precipitating systems of any category are observed, and most categories have the highest density within about 400 km of MD centre, slightly further than the conservative radius of influence determined in Sect. 3.3.

The density of deep convective cores changes little with radius beyond 100 km, suggesting that these types occur under a range of synoptic conditions and contribute to the noise which is filtered out in the clustering algorithm. The density of BSRs—the type most likely to be associated with organised convection—is highest at about 300 or 400 km distance and drops off with increasing radii after that. This is consistent with the expectation that monsoon depressions will organise convection near—but not at—MD centre. The density of wide convective cores also decreases with radius, but with larger uncertainty.

4 Discussion

Monsoon depressions (MDs) and their weaker—but more numerous—counterparts, monsoon low-pressure areas (LPAs) are the canonical rain-bringers of the Indian summer monsoon. Despite several previous efforts to quantify the fraction of monsoonal precipitation for which the former are responsible, there has been a failure to reach consensus, with estimates ranging from 15–50%. The singular

study attempting to quantify the latter found only that the result was dependent on the selected 'radius of influence', an unknown quantity.

In this study, we started with the premise that the radius of influence must also be a sought quantity, before exploring a number of different attribution techniques to determine bounds on the actual fraction of monsoon rainfall that can be attributed to both MDs and LPAs. Using a database of 109 MDs and 782 LPAs, we first approached a solution to the upper bound by making the approximation that all precipitation that falls while a system is active in the domain is caused by that system. The resulting spatial maps of attribution fraction for MDs (LPAs) revealed substantial inhomogeneity: a maximum of almost 20% (80%) in the Bay of Bengal, between 10 and 20% (60–80%) over much of the monsoon core zone (MCZ); and despite being negligible almost everywhere else, was in excess of 50% over parts of Pakistan.

We then refined this estimate by imposing a fixed radius-of-influence, assuming that all precipitation occurring within that distance from a system centre is attributable to that system. Using theoretical considerations, we showed that an appropriate choice of radius is ~ 800 km, which suggests that MDs and LPAs are responsible for 15% and 52% of all monsoonal precipitation in the MCZ respectively, and 10% and 37% over all India, respectively.

The fixed-radius method is subject to substantial noise from unrelated small-scale convective events and orographic precipitation, as well as potential under-counting where large-scale features extend beyond the chosen radius-of-influence. To mitigate against these sources of uncertainty, we introduced a precipitation clustering technique that groups together contiguous and almost-contiguous areas of rainfall, while rejecting smaller scale features that are typically not caused by synoptic-scale circulations. A climatology of these clusters revealed that the region in which they are largest (and thus, presumably, where synoptic organisation has the largest effect) is in the central Bay of Bengal, with a secondary maximum over the Arabian Sea, where the typical radii were ~ 400 km and ~ 350 km respectively.

This technique also required a choice of radius-of-influence—differing from the previous case in that if any part of a precipitation cluster falls within the radius, the whole cluster is attributed. We found that an appropriate radius, which provided an upper bound for attributable precipitation, was 200 km, whose selection has substantially less error than for the fixed-radius technique. This clustering method indicated that MDs and LPAs are responsible for 17% and 57% of all

monsoonal precipitation in the MCZ respectively, and 12% and 44% over all India, respectively.

To more clearly highlight regional contributions, we employed a *k*-means partitioning technique to separate monsoon LPA tracks into four distinct categories. Short-lived but numerous systems originating in the Bay of Bengal dominate the contribution over almost the entire Indian peninsula, distinct longer-lived systems whose genesis is also in the Bay of Bengal are the major precipitation source over northwest India; whereas systems arising in the Arabian Sea are only of particular importance over south Pakistan, and those with genesis near Sri Lanka produce a moderate contribution to rainfall only there and over some parts of Tamil Nadu. For a full inventory, the reader is encouraged to refer to the table in the Appendix.

Finally, we used the University of Washington TRMM database of precipitating events to ascertain how rainfall is organised around MDs; the density of deep convective cores were found to vary little with radial distance from the MD centre, suggesting that they likely exist regardless of the presence of an MD. Conversely, the densities of wide convective cores, deep and wide convective cores, and especially broad stratiform regions were found to vary significantly with radius, suggesting that these are synoptically organised. This lends further observational evidence to support the clustering attribution method.

Future work will look at, in particular, the manner in which low-pressure areas can remotely trigger precipitation over arid areas of Pakistan and Afghanistan; as well as how the results presented here change in the context of extreme precipitation.

Appendix: Table of attribution by system/region

See Table 3.

Table 3 Precipitation-weighted mean attribution fractions by region/country and system type, computed using the clustering method and a 200 km radius-of-influence

Region/country	MD	LPA	L1	L2	L3	L4
Andaman & Nicobar	.079	.332	.009	.060	.230	.013
Andhra Pradesh	.131	.466	.025	.115	.306	.036
Arunachal Pradesh	.040	.256	.003	.047	.195	.004
Assam	.052	.282	.004	.063	.205	.008
Bihar	.081	.407	.006	.110	.281	.007
Chandigarh	.078	.274	.002	.131	.132	.019
Chhattisgarh	.162	.597	.012	.183	.390	.035
Dadra & Nagar Haveli	.211	.469	.016	.207	.189	.079
Daman & Diu	.180	.447	.011	.209	.158	.097
Goa	.085	.352	.017	.083	.199	.062
Gujarat	.193	.549	.009	.298	.171	.111
Haryana	.101	.370	.003	.175	.174	.047
Himachal Pradesh	.061	.222	.001	.106	.109	.016
Jammu & Kashmir	.044	.124	.001	.057	.052	.023
Jharkhand	.126	.549	.011	.164	.367	.017
Karnataka	.064	.310	.041	.067	.164	.042
Kerala	.039	.228	.039	.043	.102	.049
Lakshadweep	.021	.178	.024	.020	.063	.072
Madhya Pradesh	.174	.589	.007	.262	.310	.035
Maharashtra	.157	.484	.018	.170	.255	.064
Manipur	.071	.313	.006	.065	.226	.010
Meghalaya	.057	.348	.004	.069	.266	.006
Mizoram	.096	.369	.007	.084	.254	.022
Nagaland	.054	.287	.006	.065	.204	.009
NCT (Delhi)	.105	.432	.000	.220	.184	.057
Orissa	.190	.617	.009	.191	.409	.027
Pondicherry	.050	.293	.040	.046	.172	.034
Punjab	.067	.268	.001	.112	.135	.040
Rajasthan	.151	.526	.005	.267	.230	.060
Sikkim	.068	.321	.003	.089	.218	.008
Tamil Nadu	.019	.212	.065	.032	.091	.022
Tripura	.097	.365	.007	.081	.258	.022
Uttar Pradesh	.091	.441	.007	.177	.251	.016
Uttarakhand	.071	.314	.004	.141	.164	.015
West Bengal	.101	.436	.006	.113	.299	.011
Bangladesh	.088	.379	.005	.085	.272	.012
Bhutan	.054	.313	.006	.067	.229	.008
India	.120	.448	.013	.156	.258	.035
Maldives	.006	.116	.036	.014	.031	.031
Myanmar	.083	.332	.006	.075	.222	.010
Nepal	.065	.315	.003	.099	.205	.009
Pakistan	.055	.203	.001	.107	.072	.043
Sri Lanka	.011	.223	.122	.025	.049	.031

Here L1–4 refer to the LPA types: Sri Lankan, BoB long, BoB short, and Arabian Sea respectively (as defined in Fig. 9). Red and orange text indicate the most and secondmost attribution fractions by LPA type, grey boxes indicate where the attribution fraction is less than 5%, and green where it is more than 40%

References

Ajayamohan RS, Merryfield WJ, Kharin VV (2010) Increasing trend of synoptic activity and its relationship with extreme rain events over central India. J Climate 23(4):1004–1013

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. https://doi.org/10.1002/qj.828

Dhar ON, Bhattacharya BK (1973) Contribution of tropical disturbances to the water resources of Ganga basin. Vayu Mandal 3:76–79

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26. https://doi.org/10.1214/aos/1176344552

Ester M, Kriegel HP, Sander J, Xu X et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd 96:226–231

Fletcher JK, Parker DJ, Hunt KMR, Vishwanathan G, Govindankutty M (2018) The interaction of Indian monsoon depressions with northwesterly midlevel dry intrusions. Mon Weather Rev 146(3):679–693

Godbole RV (1977) The composite structure of the monsoon depression. Tellus 29:25–40. https://doi.org/10.1111/j.2153-3490.1977.tb00706.x

Houze RA, Rasmussen KL, Zuluaga MD, Brodzik SR (2015) The variable nature of convection in the tropics and subtropics: a legacy of 16 years of the Tropical Rainfall Measuring Mission satellite. Rev Geophys 53(3):994–1021

Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF (2007) The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J Hydrometeorol 8:38–55. https://doi.org/10.1175/JHM560.1

Hunt KMR, Turner AG (2017) The representation of Indian monsoon depressions at different horizontal resolutions in the met office unied model. J Climate 143(705):1756–1771. https://doi.org/10.1002/qj.3030

Hunt KMR, Turner AG, Inness PM, Parker DE, Levine RC (2016a) On the structure and dynamics of Indian monsoon depressions. Mon Weather Rev 144(9):3391–3416. https://doi.org/10.1175/MWR-D-15-0138.1

Hunt KMR, Turner AG, Parker DE (2016b) The spatiotemporal structure of precipitation in Indian monsoon depressions. Q J R Meteorol Soc 142(701):3195–3210. https://doi.org/10.1002/qj.2901

Hunt KMR, Turner AG, Shaffrey LC (2018a) The evolution, seasonality, and impacts of western disturbances. Q J R Meteorol Soc 144(710):278–290. https://doi.org/10.1002/qj.3200

Hunt KMR, Turner AG, Shaffrey LC (2018b) Extreme daily rainfall in Pakistan and north India: scale-interactions, mechanisms, and precursors. Mon Weather Rev 146(4):1005–1022

Hunt KMR, Turner AG, Shaffrey LC (2018c) Representation of western disturbances in CMIP5 models. J Climate. https://doi.org/10.1175/JCLI-D-18-0420.1

Hurley JV, Boos WR (2015) A global climatology of monsoon low pressure systems. Q J R Meteorol Soc 141:1049–1064. https://doi.org/10.1002/qj.2447

Jadhav SK (2002) Summer monsoon low pressure systems over the Indian region and their relationship with the sub-divisional rainfall. Mausam 53(2):177–186

Mooley DA, Shukla J (1989) Main features of the westward-moving low pressure systems which form over the Indian region during the summer monsoon season and their relation to the monsoon rainfall. Mausam 40(2):137–152

Pisharoty PR, Asnani GC (1957) Rainfall around monsoon depressions over India. Indian J Meteorol Geophys 8:15–20

Praveen V, Sandeep S, Ajayamohan RS (2015) On the relationship between mean monsoon precipitation and low pressure systems

in climate model simulations. J Climate 28(13):5305–5324. https://doi.org/10.1175/JCLI-D-14-00415.1

Rajamani S, Rao KV (1981) On the occurrence of rainfall over southwest sector of monsoon depression. Mausam 32:215–220

Rajeevan M, Gadgil S, Bhate J (2010) Active and break spells of the Indian summer monsoon. J Earth Syst Sci 119(3):229–247

Ramanathan KR, Ramakrishnan KP (1933) The Indian southwest monsoon and the structure of depressions associated with it. Mem Ind Meteorol Dept 26:13–36

Roy SC, Roy AK (1930) Structure and movement of cyclones in the Indian seas. Beitr Phys Atm 26:224–234

Sikka DR (1977) Some aspects of the life history, structure and movement of monsoon depressions. Pure Appl Geophys 115:1501–1529. https://doi.org/10.1007/BF00874421

Sikka DR (2006) A study on the monsoon low pressure systems over the Indian region and their relationship with drought and excess monsoon seasonal rainfall. Center for Ocean-Land-Atmosphere Studies, Center for the Application of Research on the Environment

Sperber KR, Annamalai H, Kang IS, Kitoh A, Moise A, Turner AG, Wang B, Zhou T (2013) The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. Climate Dyn 41(9–10):2711–2744. https://doi.org/10.1007/s00382-012-1607-6

Stano G, Krishnamurti TN, Vijaya Kumar TSV, Chakraborty A (2002) Hydrometeor structure of a composite monsoon depression using the TRMM radar. Tellus A 54:370–381. https://doi.org/10.1034/j.1600-0870.2002.01330.x

Yoon JH, Chen TC (2005) Water vapor budget of the Indian monsoon depression. Tellus A 57(5):770–782. https://doi.org/10.1111/j.1600-0870.2005.00145.x