Computers and Geosciences 122 (2019) 45–53

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

Detection of spatiotemporally coherent rainfall anomalies using Markov Random Fields

Adway Mitra[a,*], Ashwin K. Seshadri[b,1]

[a] Indian Institute of Technology, Bhubaneswar, India
[b] Centre for Atmospheric and Oceanic Sciences and Divecha Centre for Climate Change, IISc, Bangalore, India

A B S T R A C T

Precipitation is a large-scale, spatio-temporally heterogeneous phenomenon, with frequent anomalies exhibiting unusually high or low values. We use Markov Random Fields (MRFs) to detect anomalies in gridded annual rainfall data across India from 1901 to 2005, such that these anomalies are spatio-temporally coherent, but permitting flexibility in size and spatial and temporal extent. MRFs are undirected graphical models where each node is associated with a {location, year} pair, with edges connecting nodes representing adjacent locations or years. Some nodes represent observations of precipitation, while the rest represent unobserved (*latent*) states that can take one of three values: high/low/normal. The MRF represents a probability distribution over the variables, using *potential functions* defined on edges of the graph. In our model, these functions enforce spatial and temporal coherence of the latent variables. Optimal values of latent state variables are estimated by maximizing their posterior probability using Gibbs sampling, conditioned on the observations. From these latent states we can identify spatio-temporally extended rainfall anomalies, both positive and negative. We study various properties of rainfall anomalies discovered by this method, such as spatio-temporal size and intensity. Identification of such rainfall anomalies can help in monitoring and studying floods and droughts in India. Properties of anomalies derived from this approach could be used to test climate models and statistical simulators.

1. Introduction

In many parts of the world, such as India, rainfall plays an important role in the economy and the well-being of millions of people (Gadgil and Gadgil, 2006). Consequently, excess or deficient rainfall can have very significant effects, especially if it occurs continuously over a region such as a watershed or river basin. Hence, identification of such spatio-temporally extended events of excess or deficient rainfall is important in both observed historical data and simulations of future scenarios by climate models. In this work, we call such events "anomalies".

In climate science, "anomaly" of a climatic variable (eg. precipitation) at a particular location and time is defined as the amount of deviation from its climatological value, averaged over many years. Generally such deviations are caused by some climatic processes, which may extend over a considerable area and/or persist for a considerable time. Identification of the exact spatial and temporal extent of such processes based on the observations is often not easy, since the manifestation is often uneven over space and time. The important feature here is spatio-temporal coherence. Anomalies can occur at different spatial and temporal scales, and their occurrence is heterogeneous (the statistics are location-dependent) and anisotropic (not uniform in all directions). The more consequential anomalies could be the ones with significant spatio-temporal extent, and therefore it is important to identify them. Identification of such anomalies of rainfall are very useful in monitoring floods (Dhar and Nandargi, 2003; Yan and Hamid, 2015) and droughts (Cook et al., 2010; Kumar et al., 2013; Dracup, 1991; Rouault and Richard, 2005) in India and elsewhere, as it gives us the information about which regions received excess or deficient rainfall in any given year. Any approach to detection of such anomalies should be general enough to work at any time scale of interest, including "extreme rainfall events" ((Goswami et al., 2006; Bernard et al., 2013; Conrad and Sharma, 2017)) which are special anomalies with very large deviation from the mean value, but spatio-temporally localized. With climate change, the frequency of rainfall extremes may increase, along with changes in the spatial pattern of rainfall (Ghosh et al., 2012). To understand these changes, scientists rely on climate models like general circulation models (GCMs) which simulate global climatic variables including rainfall. Such analysis cannot be done manually because of the large and growing volume of data and simulation results, raising the need for automated procedures such as (Narisma et al., 2007; Sandra Maria Araújo and Celso Augusto Guimarães, 2009; Sharma, 2006).

Anomalies are inherently subjective, depending on definition and detection threshold (Chandola et al., 2009). Anomaly detection in general, and spatio-temporal anomaly detection in particular are considered important research areas in Data Science (Shekhar et al., 2015;

* Corresponding author.
E-mail address: adway.cse@gmail.com (A. Mitra).
[1] Role: Problem formulation, mentoring experiments and simulations, writing the paper.

https://doi.org/10.1016/j.cageo.2018.10.004
Received 1 November 2017; Received in revised form 25 September 2018; Accepted 4 October 2018
Available online 10 October 2018
0098-3004/

Kisilevich et al., 2009). Anomalies can be both positive and negative depending on the sign of deviation of rainfall volume from the long-term mean. However, the magnitude of deviation to be considered as "anomaly" is a design choice. The simplest approach to anomaly detection is based on a predefined threshold. With rainfall, one might consider the time-series of annual mean rainfall at each grid location, estimate its mean and variance, and identify years departing significantly from the mean. However, this approach does not account for its spatio-temporal neighbors. Hence, the anomalies found by this approach are generally localized and incoherent.

Here we consider anomalies in rainfall using a *latent state variable* (Bishop, 2006) at each spatio-temporal location. Each latent variable has three possible values: 1 (positive), 2 (negative) or 3 (normal). Such latent (i.e. unobserved) variables are best estimated through probabilistic methods (Bishop, 2006; Neal, 1993). A graph is constructed with all these spatio-temporal variables as nodes, where pairs of nodes corresponding to neighboring locations are connected by edges. An anomaly is a connected component of such a graph, such that at each node in the component the associated latent variables have equal value. The approach of using local wet/dry conditions along with their spatio-temporal extents for monitoring floods and droughts has been attempted earlier also (Du et al., 2013), though using standard precipitation index instead of discrete variables.

We model these latent variables to be spatiotemporally coherent through parameters of a Markov Random Field. We identify coherent regions where anomalies occur, but this does not necessarily mean that these regions must be large. We estimate these latent variables as the maximum posterior (MAP) solution of a Markov Random Field (MRF) - undirected random graphical models generally used to model joint distributions of several variables (Kindermann and Snell, 1980). MRFs are defined using "potential functions" for nodes and edges to encode spatio-temporal coherence of the states. To identify the MAP configuration of the latent states we use a Markov-Chain Monte Carlo approach (Diaconis, 2009; Robert and Casella, 2013) called Gibbs sampling (Stoehr, 2017; Rue, 2001). A discussion on how the proposed model is related to previous models is provided in the Supplementary Material.

2. Methodology

2.1. Definitions and notation

We consider S locations and T years, and spatio-temporal observations Y_{st} of a geophysical variable such as annual-mean rainfall. Then s indexes location and t indexes time, and Y_{st} signifies rainfall received by location s at time t. Unlike time, 2-dimensional spatial locations have no natural ordering. So we order the spatial locations based on their longitude first, latitude next. Each location in the 2-dimensional spatial grid system has 8 neighbors. For each location s, we denote by $NB(s)$ the set of its neighboring locations, according to the grid system. Thus, for a location having coordinates (lat, lon), its neighbors will be $\{(lat + i, lon + j)\}$, where $i, j \in \{-1,0,1\}$, except $(i = 0, j = 0)$ which is the location itself. This particular way of ordering and indexing the spatial locations has no bearing on the analyses undertaken below, and any other indexing scheme is also equally compatible with it. This is because, the indexing does not indicate any sequence of the spatial locations, it just identifies them. The important thing in our analysis is the neighborhood structure, which is based on the spatial locations of the grids and independent of the indexing scheme.

Let us consider a graph G, where each node is associated with a pair (s, t). Further, for each spatio-temporal location (s, t) we have two nodes, one corresponding to Z_{st} and one for Y_{st}. Z_{st} is a discrete variable which indicates the state of rainfall at location s, time t. While Y is known from the dataset, Z is unknown, and must be estimated. We put edges between pairs of nodes corresponding to Z_{st} and $Z_{s't}$ for each year t if s and s' are neighbouring grid-points, i.e. $s' \in NB(s)$. We call such edges as *spatial edges*. Again, we put edges between pairs of nodes corresponding to Z_{st} and $Z_{s,t+1}$ for each location s, and such edges are called *temporal edges*. Finally, for each spatio-temporal pair (s, t) we have an edge between Z_{st} and Y_{st}, and we call such edges as *data edges*. Thus a spatial edge connects a Z-nodes associated with neighboring locations and same time, a temporal edge connects Z-nodes associated with same location but adjacent times, and a data-edge connects Z-node and Y-node at the same location and time. Thus, we have $2ST$ nodes, ST data edges, $S(T-1)$ temporal edges, and $\sum_s |NB(s)|T$ spatial edges.

We consider each location s to be in one of three possible *states* in any year t-high (1), low (2) or normal (3), which is encoded by Z. This follows the conventional classification of rainfall-years as excess rainfall, deficient rainfall, or normal, at each location. The state is represented by a latent discrete variable Z_{st} taking one of 3 values. In such a graph, an *anomaly* is a connected component of the Z-nodes corresponding to spatio-temporal locations, such that all of the nodes in the component have the same value of Z: either 1 (positive anomaly) or 2 (negative anomaly). The goal of anomaly detection in this work is to estimate these latent variables, from which the connected components can be computed and thus spatio-temporally coherent anomalies identified (Chandola et al., 2009).

2.2. Location-wise analysis (LWA)

A naive solution to anomaly detection is to treat the time-series at each location individually. For each time-series we compute mean μ_s and standard deviation σ_s. We then set $Z_{st} = 1$ (high) for those years where $Y_{st} \geq HIGH_s$, $Z_{st} = 2$ (low) for those years where $Y_{st} \leq LOW_s$, and $Z_{st} = 3$ (normal) for all other years, where $HIGH_s$ and LOW_s are thresholds specific to location s. We call this method Location-Wise Analysis (LWA), since it treats each location independently without considering the state of its neighbours. Corresponding assignments to the latent variables by this method are denoted as $Z0$.

This approach suffers from some major limitations. Firstly, it is not clear how to choose the thresholds, and results vary strongly with the choice. The histogram of annual rainfall in most locations resembles the bell-shaped curve of Gaussian distribution. So, it is reasonable to set $HIGH_s = \mu_s + \sigma_s$ and $LOW_s = \mu_s - \sigma_s$. Through the rest of this paper, we will use this choice. However, an approach that circumvents the need to specify such thresholds is a better solution.

The second major limitation of this approach is of course its neglect of spatial coherence in the latent variable. For example an individual location may be in a certain mode, while all its neighbours are in a different mode in the same year. Such isolated anomalies often do not have any particular significance, and may simply be a local artifact of the essentially chaotic process of rainfall. Spatially extended anomalies are usually more significant and consequental. This is the most important limitation of LWA, and we need a fundamentally different approach to circumvent it.

Finally, the LWA approach also neglects temporal coherence in each of the location-specific time-series. An anomaly may extend over a period, but this may be missed by treating each time-point separately. We need state persistence to achieve temporal coherence of the time-series of latent states. This can be achieved by the method proposed below.

2.3. Modelling by Markov Random Fields

Detecting extended anomalies requires a different lens from LWA, one inducing spatial or temporal coherence during assignment of the Z_{st} -variables. To induce such coherence, we take the approach that assigns probabilities to different configurations of latent Z-variables, with higher weights to configurations where Z-assignments are spatially or temporally coherent. This is achieved by modelling the latent variable as an MRF, along the lines of the drought discovery technique in (Fu et al., 2012). We seek to discover spatial and temporal clusters within

which Z-values are the same.

Markov Random Field is an undirected graphical model, where a probability distribution is defined on an undirected graph. Each node in the graph corresponds to a random variable, and each edge has an associated potential function that depends on the random variables corresponding to the two nodes connected by that edge. The full likelihood of the model is defined as the product of all the edge potential functions.

We have 2 nodes for every spatio-temporal pair (s, t) - corresponding to Z_{st} and Y_{st}. Spatial edges, temporal edges and data edges are defined between pairs of variables as mentioned above. In addition to grid-wise latent states, these can also be defined for the all-India mean, relative to its corresponding distribution across years. The Indian Meteorological Department (IMD) currently makes annual forecasts of spatial aggregate rainfall over India during the summer monsoon months of July–September (JJAS), called Indian Summer Monsoon Rainfall (ISMR). We define an analogous quantity for the entire year, All-India Mean Rainfall (AIMR), and denote by Y_t. Its anomalies are relative to its interannual mean μ and standard deviation σ. Once again, we define discrete latent variable Z_t corresponding to AIMR, which can take 3 values.

A Markov Random Field is an undirected graph, with nodes for each (s, t) pair. Corresponding to each (s, t) pair is associated a latent variable Z_{st} and an observation Y_{st}. Each observation node Y_{st} has a single edge, to the corresponding latent variable node Z_{st}. The graph also contains nodes corresponding to each year, associated with latent Z_t and observed Y_t, corresponding to AIMR. For any year t, Z_t is linked by edges to all nodes for that year $\{Z_{st}\}$ for every location s. Large anomalies in ISMR are declared by IMD as excess or deficient rainfall years. However, rainfall is highly heterogeneous spatially. Therefore in order to define anomalies in the aggregate measure of AIMR, we consider not only calculations of Y_t but also the frequencies of local anomalies in the corresponding year. This is achieved by linking the Z_{st} and Z_t nodes. Fig. 1 illustrates the model.

Probabilities are assigned to each configuration of Z using node *potential functions* $\psi^v(Z_{st})$ on each node, edge potentials $\psi^e(Z_{st}, Z_{s't'})$ on each edge occurring between spatio-temporal nodes and $\psi^f(Z_{st}, Z_t)$ on each edge occurring between spatio-temporal nodes and AIMR nodes. Edge potentials influence spatial and temporal coherence and node potentials influence the threshold for anomaly detection. Edge potentials describe prior probabilities that the nodes connected by the edge are in the same state. The node potential functions can be interpreted as describing the prior probability distribution across different states.

The precipitation amount at any location and year, given by Y_{st}, is modelled using a Gaussian distribution with parameters (μ_{sk}, σ_s) specific to the location s and latent state $Z_{st} = k$. These conditional distributions can be interpreted as edge potentials on the $Z_{st} - Y_{st}$ data edges connecting the latent and observed states respectively.

2.4. Spatial and temporal coherence through MRF

The spatio-temporal rainfall volume Y_{st} is modelled as a multi-modal Gaussian distribution, and $Z_{st} = p$ specifies the mode (1:high,2:-low,3:normal). The parameters (μ_{sp}, σ_s) of this distribution depend on the latent state p as well as location s, and are estimated from data. Similarly for spatial mean rainfall Y_t we use a Gaussian distribution with state-specific parameters (μ_p, σ). Initial estimates of these parameters can be made from the dataset using LWA to assign states.

We define **edge potential functions** so that if two vertices connected by an edge have same values of Z then the corresponding edge potential is larger than if the values were different. Since the likelihood function involves products of these edge potentials, this encourages spatial and temporal neighbours to have same state, leading to spatial and temporal coherence. For each edge between location state node Z_{st} and the corresponding AIMR state node Z_t for the same year, the edge potential influences the extent to which the local state is sought to be made coherent with the aggregate state. We define potential functions for different edges as follows:

$$\begin{aligned}
&\psi(Z_{st}, Z_{s't}) = exp(C(s, s')) \quad \textbf{if } Z_{st} = Z_{s't}, = exp(D) \quad \textbf{otherwise};\\
&\textbf{where } s' \in NB(s) \quad \psi(Z_{st}, Z_{s,t+1}) = P \quad \textbf{if } Z_{st} = Z_{s,t+1},\\
&\qquad\qquad = 1 - P \quad \textbf{otherwise};\\
&\qquad\qquad \psi(Z_{st}, Z_t) = exp(1/S) \quad \textbf{if } Z_{st} = Z_t,\\
&\qquad\qquad = 1 \quad \textbf{otherwise};
\end{aligned} \tag{1}$$

To emphasize spatial coherence, D is a small constant compared to $C(s, s')$. The latter describes edge potentials if spatial neighbours are in the same state. Two neighbouring grid-locations need not be highly correlated, for e.g. on either side of a narrow mountain range (such as the Western Ghats). Therefore unlike the MRF estimated by (Fu et al., 2012), where all edges between neighbouring pairs have the same potential function, here the potentials on edges are estimated from data and are location-dependent.

The value of edge potential P, for edges connecting nodes with neighbouring years, lies between zero and one. It induces temporal coherence, and hence is called the temporal coherence parameter. Higher values induce a higher emphasis on temporal coherence.

The third set of edge potentials describes behaviour of edges between the location nodes in any given year and the AIMR-node for that year. It is defined using the exponential, so that the contribution

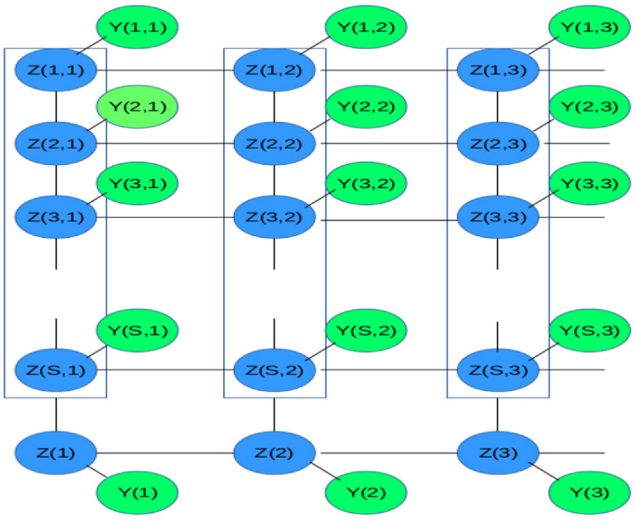

Fig. 1. Proposed Markov Random Field for Anomaly Detection. Each column represents one year and each row represents one location. The horizontal edges are "temporal edges", vertical ones are "spatial edges", angular ones are "data edges". For simplicity, only one or two spatial edges have been shown per location. The latent variables are shown in blue, observed ones in Green. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

depends on the total number of locations whose states coincide with the state assigned to the spatial mean node. S is the total number of locations. The edge potential is higher when the location nodes are in the same state as the spatial mean node.

A discussion of certain aspects of this model such as spatio-temporal separability is provided in the *Supplementary Materials*.

2.5. Anomaly detection by Markov Random Fields

According to the model proposed above, the joint distribution of the random variables Z and Y is:

$$L(Z) \propto \prod_{s,t} \psi^v(Z_{st}) \prod_{e} \psi^e(Z_{st}, Z_{s't'}) \prod_{f} \psi^f(Z_{st}, Z_t) \prod_{s,t} \mathcal{N}(Y_{st}; \mu_{sZ_{st}}, \sigma_s) \prod_{t} \mathcal{N}(Y_t; \mu_{Z_t}, \sigma) \tag{2}$$

Having defined the likelihood function, we carry out inference on the latent variables Z and estimate parameters $(\mu_{sp}, \sigma_s, C(s, s'))$ for locations s, corresponding neighbours s' and conditioned on latent state p. Unlike the maximum likelihood estimation of (Fu et al., 2012) that is based on integer programming, here we carry out inference by Gibbs Sampling, which is computationally simpler (Rue, 2001; Brown and McMahan, 2017; Haran et al., 2003). Each latent variable Z_{st} is initialized based on location-wise analysis described earlier, and corresponding parameters are estimated. Gibbs Sampling is an iterative process, in which we visit each latent variable (say Z_{st}) one by one, compute its conditional distribution (conditioned on the current assignments of the other variables) $p(Z_{st} = k|Z_{-s,-t}, Z_t, Y_{st})$ and sample a new value for Z_{st}. The procedure is repeated for several iterations, after each iteration the parameter estimates are updated according to the contemporary assignments of the latent variables, and samples are collected at regular intervals. The stationary distribution of this Markov chain Monte Carlo procedure is the posterior distribution on the latent variables. The maximum a-posteriori (MAP) estimate of Z-variables can then be made from the samples.

The Gibbs Sampling equation for any latent variable Z_{st} or Z_t is given by:

$$\begin{aligned} p(Z_{st} = k|Z_{-s,-t}, Z_t, Y_{st}) &\propto p(Z_{st} = k, Z_{-s,-t}, Z_t, Y_{st}) \text{ Bayes Theorem} \\ &\propto p(Z_{st} = k, Z_{s't}, Z_{st'}, Z_t, Y_{st}) \text{ Using Markov independence} \\ &= p(Z_{st} = k, Z_{s't}, Z_{st'}, Z_t) p(Y_{st}|Z_{st} = k) \text{ Using Markov independence} \\ &= \psi^v(Z_{st} = p) \psi^f(k, Z_t) \textstyle\prod_{s',t'} \psi^e(k, Z_{s't'}) \mathcal{N}(Y_{st}; \mu_{sk}, \sigma_s) \text{ By definition} \end{aligned} \tag{3}$$

$$\begin{aligned} p(Z_t = q|Z_{-t}, Z_{st}, Y_{st}) &\propto p(Z_t = q, Z_{-t}, Z_{st}, Y_{st}) \text{ Bayes Theorem} \\ &\propto p(Z_t = q, Z_{st}, Z_{t'}, Y_t) \text{ Using Markov independence} \\ &= p(Z_t = q, Z_{st}, Z_{t'}) p(Y_t|Z_t = q) \text{ Using Markov independence} \\ &= \psi^v(Z_t = q) \textstyle\prod_{t'} \psi^f(q, Z_{t'}) \prod_{s,t} \psi^e(q, Z_{st}) \mathcal{N}(Y_t; \mu_q, \sigma) \text{ By definition} \end{aligned} \tag{4}$$

where s' refers to neighbours of s, t' to the previous and next years, i.e. $(t-1)$ and $(t+1)$, the state $k \in \{1,2,3\}$, and $Z_{-s,-t}$ means all the Z-variables except Z_{st}. Markov independence refers to the property of Markov Random Fields that any node is conditionally independent of all other nodes, given its neighbors. While applying this equation, we do not consider variables corresponding to spatio-temporal locations that are not neighbours of Z_{st}, since the Markov property of MRF holds that each node is conditionally independent of all non-neighbouring nodes conditioned on the neighbouring nodes. The Gibbs Sampling proceeds by drawing samples for each Z_{st} and each Z_t from Equation (3), and the optimal value for each latent variable is estimated from the distribution across these samples.

After estimating the latent-variable-set Z, we identify anomalies by discovering spatially and/or temporally coherent sets of spatio-temporal locations. Spatiotemporal anomalies are estimated as connected components of the MRF, such that each node of the connected component has the same value of Z. These values of Z can be either 1 or 2, corresponding to positive and negative anomalies respectively. Due to coherence, the clusters thus identified can be at a single location but extending over several continuous years, or spatially contiguous locations in a single year, or both. Clearly, the spatio-temporal extents of the anomalies discovered this way are not fixed by prior specification.

3. Experimental results

We now come to our experimental study based on the model proposed in the previous section. First of all, we carry out experiments on simulated datasets where we have planted anomalies of different sizes, and we show that our method performs well. This experiment establishes that the results produced by the proposed model are related to the nature of the dataset, and not merely an artifact of the parameters. For space constraints, these results are placed in the Supplementary Materials.

Next, we come to experiments on the real data. We use a dataset of $1^\circ - 1^\circ$ gridded rainfall data measured all over India, for the period 1901–2005. This grid system has 357 locations over India ($S = 357$). The data is available at daily scale, but for the analysis in this paper we compute annual aggregate values. First we do a test of the method based on the all-India state variables. We find that these variables indeed agree with IMD's classification of years as excess/deficient/normal, and also carry additional information about grid-scale conditions. We also study the effects of varying the edge potential function parameters on the latent state assignments. The details of these experiments are given in the *Supplementary Materials*.

Next, we analyze properties of the detected anomalies. Two important questions are how widespread and persistent positive and negative rainfall anomalies are, and how much different the rainfall volumes are from the long-term climatology during each anomaly. We address these questions in this section.

3.1. Anomaly statistics

The *spatio-temporal size* of each anomaly is the size of the corresponding connected component in the graph, i.e. the number of nodes present in it. We measure the STS: mean spatio-temporal size of all anomalies, including all years; and similarly the STSP: mean spatio-temporal size of all positive anomalies; and STSN: mean spatio-temporal size of all negative anomalies. We define the *spatial size* of an anomaly as the number of distinct spatial locations included in the nodes covered by it. The *temporal size* of an anomaly is similarly defined as the number of distinct years included in it. We thereby estimate mean spatial size of all anomalies (SS), only positive (SSP) and only negative (SSN) anomalies. Similarly we measure (TS, TSP, TSN) for corresponding mean temporal sizes.

Each state of Z at each location is associated with a distribution over rainfall values. Fig. 2 shows the mean rainfall values for each of the locations and each state of Z. Mathematically, these are $mean_{t:Z_{st}=1}(Y_{st})$ for positive anomalies, and $mean_{t:Z_{st}=2}(Y_{st})$ for negative anomalies. Two different settings of the MRF are considered: using spatio-temporal coherence with temporal coherence parameters $P = 0.7$ and $P = 0.9$, and the "prop" setting of spatial coherence where edge potentials are proportional to the number of years in which locations have the same state (See Supplementary Material). The plots in Fig. 2 show that these mean rainfall fractions for the different states (shown by green, blue and red plots) are clearly well-separated in most locations.

To quantify the *severity* of each anomaly quantitatively, we first compute the ratio of the rainfall received at each spatio-temporal location covered by the anomaly, and the long-term mean rainfall over each of these locations. We define the *intensity parameter* of the anomaly as the mean of these ratios. Mathematically, let A be the set of spatio-temporal locations affected by a particular positive or negative anomaly a. For each $(s, t) \in A$, we compute $F_a(s, t) = \frac{Y_{st}}{\mu_s}$. Then the intensity of

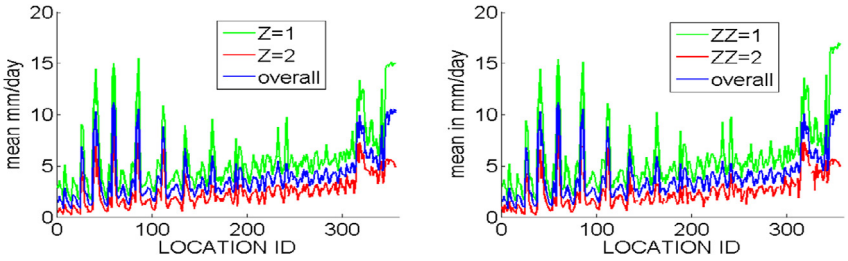

Fig. 2. The mean rainfall at each of the locations in the two anomaly states, and overall, for MRF settings using P = 0:7 (left) and P = 0:9 (right), along with spatial coherence.

Table 1
Mean spatial, temporal, spatio-temporal sizes and mean intensities of positive and negative anomalies in different settings of edge potentials of MRF. A trade-off between the spatial and temporal sizes of anomalies is inherent to anomaly detection; and illustrated here by varying the temporal coherence parameter. Spatial coherence effect in the MRF leads to larger spatial size of detected anomalies, which correspondingly have shorter mean temporal size. Larger temporal coherence parameter leads to longer mean temporal size and correspondingly smaller mean spatial size. Also, the anomalies become more intense (high intensity for positive and low intensity for negative) as the spatio-temporal coherence are increased.

Method	#Anomalies		S-T SIZES		Spatial sizes		Temporal sizes		Intensity	
	NP	NN	STSP	STSN	SSP	SSN	TSP	TSN	IP	IN
LWA	1085	1163	5.3	5.0	5.0	4.4	1.1	1.2	1.34	0.69
MRF-SC	519	472	11.5	12.5	10.8	11.0	1.1	1.2	1.37	0.70
MRF-TC-0.5	1083	1155	6.9	6.6	6.3	5.8	1.2	1.2	1.24	0.80
MRF-TC-0.75	1000	1105	6.7	6.1	5.7	5.1	1.3	1.2	1.26	0.78
MRF-TC-0.90	795	825	6.5	5.9	4.8	5.0	1.6	1.5	1.30	0.76
MRF-TC-0.99	472	365	6.7	6.8	3.4	2.8	2.3	2.6	1.34	0.73
MRF-STC-0.5	550	459	10.8	12.6	10.0	11.1	1.2	1.2	1.32	0.75
MRF-STC-0.75	401	317	11.5	13.5	10.0	11.0	1.3	1.3	1.37	0.71
MRF-STC-0.90	303	137	9.5	15.3	7.5	9.3	1.4	1.8	1.44	0.68
MRF-STC-0.99	208	75	7.0	15.8	3.9	6.4	1.9	2.7	1.47	0.67

anomaly a is given by $I_a = mean_{(s,t)\in A} F_a(s, t)$.

3.2. Effect of MRF settings

We consider location-wise analysis (LWA), and using MRFs under different settings. These settings include only spatial coherence (SC), only temporal coherence with parameter P ($TC - P$) and both spatial and temporal coherence ($STC - P$). Results are shown in Table 1. The different groups of columns show the number of anomalies, spatio-temporal size, spatial size, temporal size and intensity respectively, each one separately for positive and negative anomalies.

The results indicate complex relationships involving spatial and temporal scales of anomalies. As expected, with LWA, the number of anomalies is much larger and their mean sizes much smaller, in comparison to versions of the MRF where various constraints of coherence are present. In the absence of spatial coherence, as the temporal coherence parameter is increased, the spatial size of anomalies becomes smaller. Larger temporal coherence parameter selects for more long-lived anomalies and hence these tend to become smaller in spatial extent. The spatio-temporal size decreases as the temporal coherence parameter is increased. The aforementioned effect is also present when spatial coherence is included in the MRF. The selection for longer but spatially less extended anomalies when the temporal coherence parameter is increased creates a trade-off between spatial and temporal extents. Such a trade-off is intrinsic to spatio-temporal anomaly detection: with a larger emphasis on a certain type of coherence (spatial or temporal) the corresponding size of anomalies increases while the other size decreases.

In Table 1, we also study the mean intensity of the anomalies under different settings of MRF. Clearly, as either type of coherence is increased, the mean intensity parameter of positive anomalies increases, and that of negative anomalies decreases, and given the aforementioned definition of this parameter the selected anomalies are more "intense". This is a welcome result, indicating that use of spatio-temporal coherence helps us to identify severe anomalies, rejecting mild ones.

3.3. Variations among anomalies

The above discussion pertained to parameter-based tradeoffs in mean spatial and temporal sizes of anomalies. However, even for fixed parameter settings of the MRF, there is substantial variation in size and intensity of the detected anomalies. Such variation of spatial and temporal sizes is shown in Fig. 3 for two realizations of the MRF. It is seen that generally larger anomalies tend to be shorter-lived, but there are individual exceptions. There is a large range of temporal sizes for a known spatial size, for both positive and negative anomalies. In Fig. 3 we also plot the variation of intensity with spatio-temporal size of the anomalies in two realizations of MRF. Here the correlation is even weaker.

We compute the correlations between these statistics of individual anomalies. Once again, this is done separately for each setting of the MRF considered in Table 1, and separately for positive and negative anomalies. The results are shown in Table 2. It shows that in almost all the settings the correlation between spatio-temporal size and spatial size is very strong, though it reduces as the temporal coherence parameter P is increased (i.e. the mean temporal size of the anomalies increase). The correlation between spatio-temporal and temporal sizes is less strong, though it increases slightly with P. The spatial and temporal sizes are less well correlated. There is no noticeable correlation between spatio-temporal size and intensity.

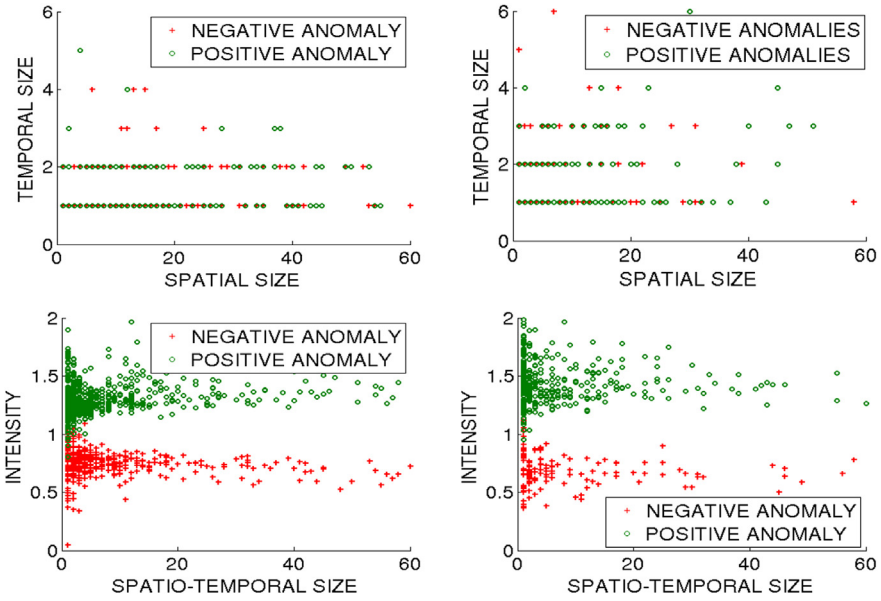

Fig. 3. Above: Temporal versus spatial sizes of individual positive and negative anomalies, in dexed parameter settings. Spatial coherence (prop) is used with two choices of the temporal coherence parameter (left: P = 0:50, right: P = 0:90). Larger anomalies tend to be shorter-lived, but there are individual exceptions and large variability exists in the sizes of individual anomalies. Below: Spatio-temporal size versus intensities of the same set of anomalies.

4. Case studies of some anomalies

In this section we investigate some of the anomalies individually, which were discovered using MRF with spatio-temporal coherence, with temporal coherence parameter $P = 0.9$.

We first consider a positive anomaly that occurred in the states of Odisha and Jharkhand along the eastern coast (Fig. 4A), in the year 1994. This anomaly covered 20 grid-locations, but persisted for only 1 year (spatial size 20, temporal size 1). The long-term mean annual rainfall over the concerned 20 grid-locations is 4.18 mm per day per location, but that year the mean rainfall over these locations was 5.84 mm per day per location (anomaly intensity of 1.4). Overall, the year 1994 was classified as a positive anomaly year in terms of AIMR, with mean rainfall of 4.23 mm per day per location, compared to the long-term mean of 3.94 mm per day per location (intensity of 1.1). The map of locations having local positive and negative anomalies in 1994 are shown in Fig. 4B, which indicates that the Odisha anomaly was quite significant. The LWA-based local anomalies are shown in Fig. 4C. Another major anomaly occurred roughly in the same area (Fig. 6D) in 2001, covering 11 locations. The mean rainfall that year over this anomaly was 5.5 mm per location per day, compared to the long-term mean of 4.1 mm per location per day (intensity of 1.3). The year 2001 was classified as normal at all-India scale, and the map in Fig. 5E shows the locations under positive and negative anomalies according to MRF.

A significant negative anomaly occurred around a stretch of Central India (Fig. 5A) in 2000, which was classified as an all-India negative anomaly year. The anomaly map by MRF of the year is shown in Fig. 5B. The anomaly covered 22 locations which receive 3.88 mm per day per location rainfall on average, but in that year they received only 2.17 mm (anomaly intensity of 0.56). Again, around 10 locations in Odisha near the eastern coast (Fig. 5D) had a negative anomaly in 2002, which was a major drought year in terms of AIMR. These locations, which receive 4.18 mm on average, received only 2.4 mm in 2002 (intensity of 0.57). The MRF-based anomaly map for 2002 is shown in Fig. 5E, while Fig. 6F shows the local anomalies by LWA.

Some anomalies are temporally extended, i.e. they cover several years. A good example is a positive anomaly that covered 5 years from 1987 to 1991, over the Meghalaya and Southern Assam region, covering 24 locations (Fig. 6A). The mean annual rainfall over these locations is 6.35 mm per location per day, but in these 5 years, the mean rainfall volumes were 6.99, 8.53, 6.92, 7.14 and 7.45 mm per location, per day. Among these years, only 1988 and 1990 were classified as positive anomaly at all-India scale, while the other three years were classified as normal. The MRF-based anomaly map of 1987 is shown in Fig. 6B. Again, 11 locations in the south-western state of Kerala (Fig. 6D), one of the wettest parts of India, suffered a negative anomaly stretching over 1985-87, all of which were classified as normal years. The mean rainfall over these locations is 6.15 mm per location per day,

Table 2
Pearson Correlation Coefficients between different pairs of statistics for individual positive and negative anomalies, computed by different methods.

Method	Temp size vs Spat. size		Spat-temp size vs Spat size		Spat-temp size vs Temp size		Intensity vs Spat-temp size	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG
LWA	0.42	0.40	0.99	0.94	0.49	0.61	0.15	−0.1
MRF-SC	0.43	0.33	0.99	0.93	0.51	0.61	0.01	0.1
MRF-TC-0.50	0.45	0.44	0.99	0.96	0.50	0.60	0.23	−0.2
MRF-TC-0.75	0.43	0.44	0.98	0.96	0.52	0.59	0.19	−0.2
MRF-TC-0.90	0.38	0.37	0.95	0.82	0.57	0.69	0.18	−0.1
MRF-TC-0.99	0.33	0.29	0.80	0.70	0.59	0.65	0.06	−0.1
MRF-STC-0.5	0.40	0.26	0.99	0.94	0.48	0.53	0.14	−0.2
MRF-STC-0.75	0.38	0.31	0.97	0.87	0.52	0.68	0.06	−0.1
MRF-STC-0.90	0.34	0.18	0.94	0.79	0.57	0.67	0	0
MRF-STC-0.99	0.32	0.22	0.87	0.65	0.60	0.76	−0.1	0

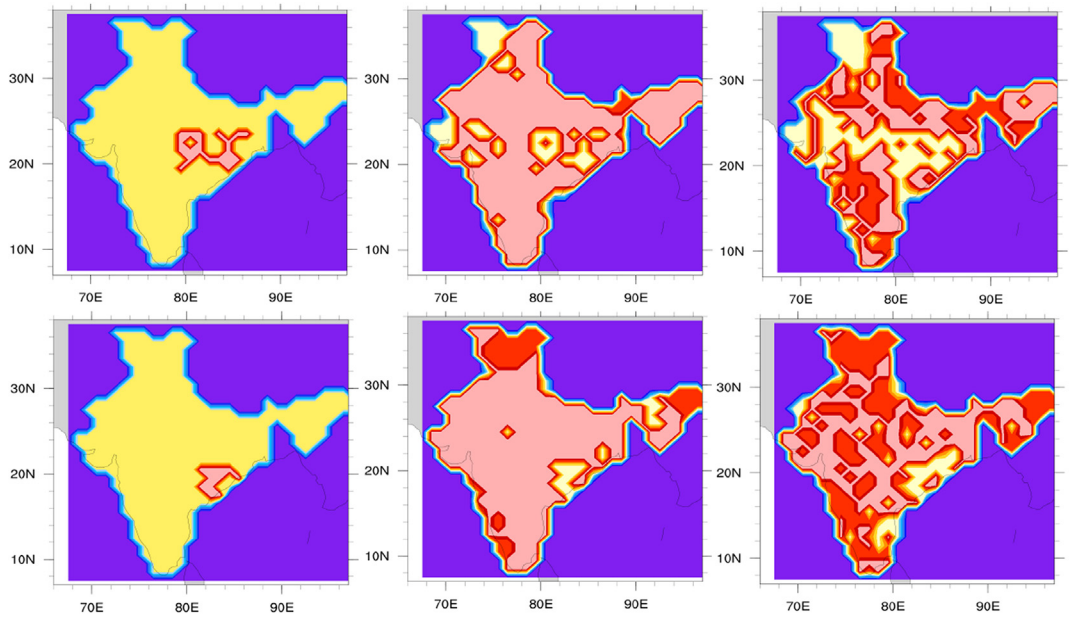

Fig. 4. ABOVE: 4A: a set of locations that formed a positive anomaly in 1994, shown in pink. 4B: MRF-based Anomaly map for 1994 (yellow: positive, red: negative). 4C: LWA-based local rainfall anomaly map for 1994. BELOW: 4D: a set of locations that formed a positive anomaly in 2001, shown in pink. 4E: MRF-based Anomaly map for 2001 (yellow: positive, red: negative). 4F: LWA-based local rainfall anomaly map for 2001. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

but during these three years, this mean was 4.83, 4.59 and 4.9 respectively. The MRF-based anomaly map of 1985 is shown in Fig. 6E.

5. Conclusions

This paper describes a method for coherent anomaly detection using Markov Random Fields (MRFs), where each node is associated with a location and year. Coherence is emphasized because it is an inherent property of rainfall, and we introduce this constraint in order to identify connected components where anomalies are extended over some spatio-temporal size. However, this does not mean that the method can only be used to detect large anomalies, as it is flexible enough to detect anomalies having a wide range of spatial and temporal scales. For applications where a larger scale is of interest, for example involving droughts over a river basin or food-producing region, the method can be adapted to emphasize not only coherence but also spatial extent. While anomalies are subjective by their nature, as they are identified by a procedure that is chosen by the analyst, the general feature of coherence is not unique to anomalies detected through our approach, and is also elicited in anomaly detection methods such as wavelets. The approach presented here has the additional feature of not privileging any particular shape or size. In this sense, MRFs offer an approach to

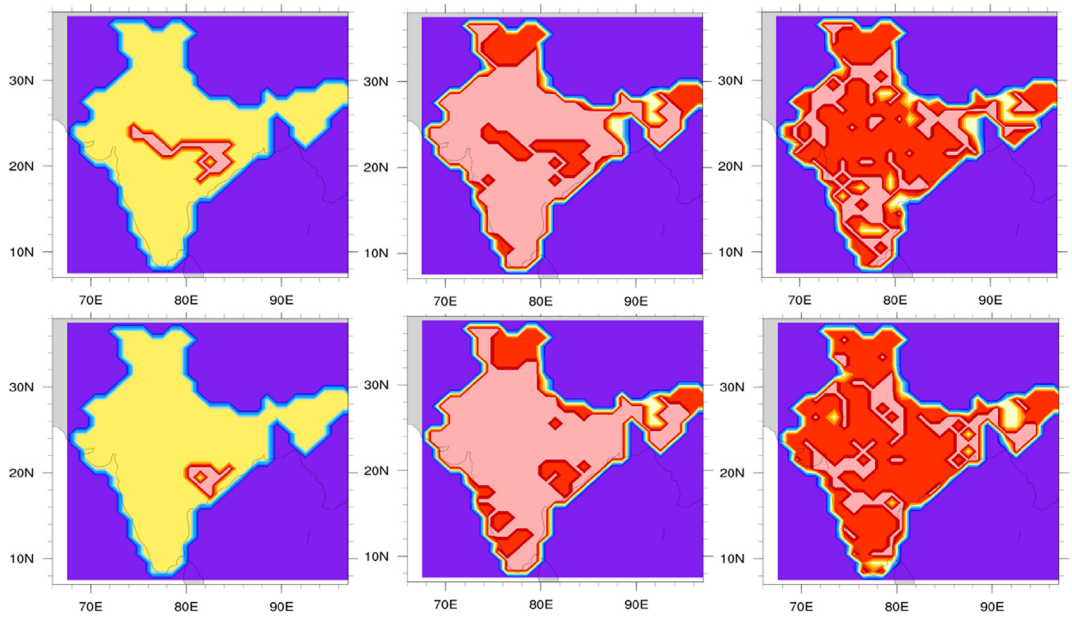

Fig. 5. ABOVE: 5A: a set of locations that formed a negative anomaly in 2000, shown in pink. 5B: MRF-based Anomaly map for 2000 (yellow: positive, red: negative). 5C: LWA-based local rainfall anomaly map for 2000. BELOW: 5D: a set of locations that formed a negative anomaly in 2002, shown in pink. 5E: MRF-based Anomaly map for 2002 (yellow: positive, red: negative). 5F: LWA-based local rainfall anomaly map for 2002. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

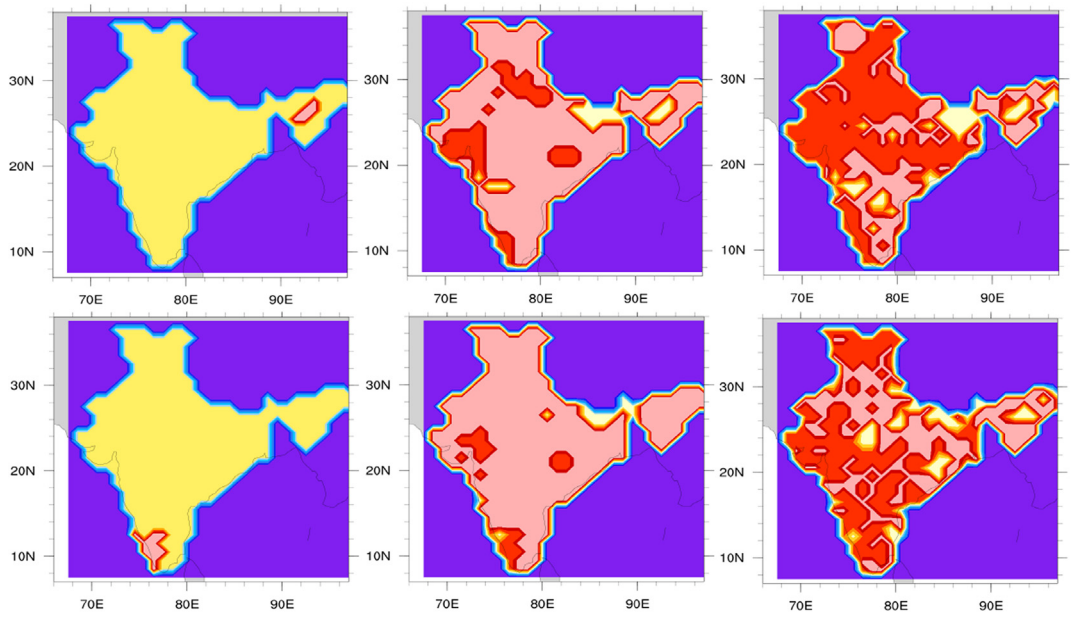

Fig. 6. ABOVE: 6A: a set of locations that formed a positive anomaly in 1987-91, shown in pink. 6B: MRF-based Anomaly map in 1987 (yellow: positive, red: negative). 6C: LWA-based local rainfall anomaly map for 1987. BELOW: 6D: a set of locations that formed a negative anomaly in 1985-87, shown in pink. 6E: MRF-based Anomaly map in 1985 (yellow: positive, red: negative). 6F: MRF-based Anomaly map in 1985. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

handling the heterogeneity and anisotropy in the occurrence of anomalies, where more traditional methods such as wavelets may be limited by particular assumptions about shape or size. We have tested the method on synthetically generated anomalies and shown that the MRF-based approach performs well on a variety of measures as compared to location-wise thresholding based methods, for a variety of different anomaly statistics.

The anomaly states are represented as latent random variables, so probabilistic methods are required for their estimation. The resulting MRF is able to identify more coherent anomalies compared to traditional analysis using location-specific thresholds. The method can discover intense positive and negative anomalies of various sizes, without requiring any thresholds for detecting individual points as belonging to anomalies.

Furthermore, as noted earlier, anomaly detection involves a trade-off between intensity and spatio-temporal extents. The intensities and spatial and temporal extents of the discovered anomalies would depend primarily on the statistics of the data. But the free parameters in the MRF model leaves some flexibility for emphasizing some aspects over others, according to the goals of the application in question. The simulation studies of the Supplementary Material clearly show that sizes of the discovered anomalies in the three datasets depend mainly on the nature of the data, in particular the anomalies themselves, but additionally changing of the temporal coherence parameter (P) also causes subtle differences in results. As pointed out in Section 3.2 of the paper, one can manifest the inherent trade-off between the spatial and temporal extents of the discovered anomalies using this parameter, as demanded by the application. If we are analyzing the rainfall distribution at a single time-point (say, one year) then temporal extent of the anomaly events is not important, and we should focus on the spatial extents. Then we may set the parameter P to low value. On the other hand, if we want to detect regime changes in the time-series, then it makes sense to set a higher value of P. Likewise, the choice of node potentials is also a way by which the users can prioritize one kind of anomalies over another kind, according to their requirements. Formally node potentials describe prior probabilities of the nodes having various states.

Overall, this study provides some understanding of heterogeneities in rainfall over Indian region. The results also raise the question of whether the anomalies discovered by this method are relevant for understanding hydrological floods and droughts, which are based on considering multiple variables, including soil moisture. A natural extension of this work would be to infer anomaly states based on the inclusion of additional climatic and hydrological variables. This method can be used for anomaly detection on other datasets as well. For example, the technique is being applied to a daily maximum temperature dataset to identify heat-waves that cover large areas for several days. In this case, we prioritize positive anomalies over negative ones through node potentials.

Acknowledgments

This research was partially supported by Airbus India Postdoctoral Fellowship for International Center for Theoretical Sciences (ICTS-TIFR), Bangalore, India, and also by Divecha Center for Climate Change, Indian Institute of Science, Bangalore, India. We are thankful to Dr. J. Srinivasan and Dr. V.Venugopal for valuable inputs.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cageo.2018.10.004.

References

Bernard, E., Naveau, P., Vrac, M., Mestre, O., 2013. Clustering of maxima: spatial dependencies among heavy rainfall in France. J. Clim. 26 (20), 7929–7937.
Bishop, Christopher M., 2006. Pattern Recognition for Machine Learning.
Brown, D Andrew, McMahan, Christopher S., 2017. Sampling Strategies for Fast Updating of Gaussian Markov Random Fields. arXiv preprint arXiv:1702.05518.
Chandola, Varun, Banerjee, Arindam, Kumar, Vipin, 2009. Anomaly detection: a survey. ACM Comput. Surv. 41 (3).
Conrad, Wasko, Sharma, Ashish, 2017. Global assessment of flood and storm extremes with increased temperatures. Sci. Rep. 7 (7945).
Cook, E.R., Anchukaitis, K.J., Buckley, B.M., DArrigo, R.D., Jacoby, G.C., Wright, W.E., 2010. Asian monsoon failure and megadrought during the last millennium. Science 328 (5977), 486–489.
Dhar, O.N., Nandargi, Shobha, 2003. *Hydrometeorological Aspects of Floods in India*, Flood Problem and Management in South Asia. pp. 1–33.

Diaconis, Persi, 2009. The Markov chain Monte Carlo revolution. Bull. Am. Math. Soc. 46 (2), 179–205.
Dracup, J.A., 1991. Drought monitoring. Stoch. Environ. Res. Risk Assess. 5 (4), 261–266.
Du, Juan, Fang, Jian, Xu, Wei, Shi, Peijun, 2013. Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province, China. Stoch. Environ. Res. Risk Assess. 27 (2), 377–387.
Fu, Qiang, Banerjee, Arindam, Liess, Stefan, Snyder, Peter K., 2012. Drought detection of the last century: an MRF-based approach. In: SIAM International Conference on Data Mining (SDM).
Gadgil, Sulochana, Gadgil, Siddhartha, 2006. *The Indian Monsoon, GDP and Agriculture*, Economic and Political Weekly. pp. 4887–4895.
Ghosh, Subimal, Das, Debasish, Kao, Shih-Chieh, Ganguly, Auroop R., 2012. Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. Nat. Clim. Change 2 (2), 86–91.
Goswami, Bhupendra Nath, Venugopal, V., Sengupta, D., Madhusoodanan, M.S., Xavier, Prince K., 2006. Increasing trend of extreme rain events over India in a warming environment. Science 314 (5804), 1442–1445.
Haran, Murali, Hodges, James S., Carlin, Bradley P., 2003. Accelerating computation in Markov random field models for spatial data via structured MCMC. J. Comput. Graph Stat. 12 (2), 249–264.
Kindermann, Ross, Snell, Laurie, 1980. Markov Random Fields and Their Applications.
Kisilevich, Slava, Mansmann, Florian, Nanni, Mirco, Rinzivillo, Salvatore, 2009. Spatio-temporal Clustering. Springer.
Kumar, K.N., Rajeevan, M., Pai, D.S., Srivastava, A.K., Preethi, B., 2013. On the observed variability of monsoon droughts over India. Weather and Climate Extremes 1, 42–50.
Narisma, Gemma T., Foley, Jonathan A., Licker, Rachel, Ramankutty, Navin, 2007. Abrupt changes in rainfall during the twentieth century. Geophys. Res. Lett. 34 (6).
Neal, Radford M., 1993. Probabilistic Inference Using Markov Chain Monte Carlo Methods.
Robert, Christian, Casella, George, 2013. Monte Carlo Statistical Methods.
Rouault, Mathieu, Richard, Yves, 2005. Intensity and spatial extent of droughts in southern Africa. Geophys. Res. Lett. 32 (15).
Rue, Håvard, 2001. Fast sampling of Gaussian Markov random Fields. J. Roy. Stat. Soc. B 63 (2), 325–338.
Sandra Maria Araújo, Ideião, Celso Augusto Guimarães, Santos, 2009. Analysis of precipitation time series using the wavelet transform. Rev. Sociedade Natureza 1 (1).
Sharma, Aditi, 2006. Spatial Data Mining for Drought Monitoring: an Approach Using Temporal NDVI and Rainfall Relationship. Master thesis. International Institute for Geoinformation Science and Earth Observation.
Shekhar, Shashi, Jiang, Zhe, Ali, Reem Y., Eftelioglu, Emre, Tang, Xun, Gunturi, Venkata, Zhou, Xun, 2015. Spatiotemporal data mining: a computational perspective. ISPRS Int. J. Geo-Inf. 4 (4), 2306–2338.
Stoehr, Julien, 2017. A Review on Statistical Inference Methods for Discrete Markov Random Fields. arXiv preprint arXiv:1704.03331.
Yan, Hongxiang, Hamid, Moradkhani, 2015. A regional Bayesian hierarchical model for flood frequency analysis. Stoch. Environ. Res. Risk Assess. 29 (3), 1019–1036.