

Precipitation Regionalization Using Self-Organizing Maps for Mumbai City, India

Amit Sharad Parchure, Shirish Kumar Gedam

Centre of Studies in Resources Engineering, Indian Institute of Technology (IIT), Bombay, India

Email: amit231279@gmail.com, Shirish@csre.iitb.ac

How to cite this paper: Parchure, A.S. and Gedam, S.K. (2018) Precipitation Regionalization Using Self-Organizing Maps for Mumbai City, India. *Journal of Water Resource and Protection*, 10, 939-956.
<https://doi.org/10.4236/jwarp.2018.109055>

Received: August 10, 2018

Accepted: September 26, 2018

Published: September 29, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The detailed analysis of individual rain events characteristics is an essential step for improving our understanding of variation in precipitation over different topographies. In this study, the homogeneity among rain gauges was investigated using the concept of “rain event properties,” linking them to the main atmospheric system that affects the rainfall in the region. For this, eight properties of more than 23,000 rain events recorded at 47 meteorological stations in Mumbai, India, were analyzed utilizing seasonal (June-September) rainfall records over 2006-2016. The high similarities among the properties indicated the similarities among the rain gauges. Furthermore, similar rain gauges were distinguished, investigated and characterized by cluster analysis using self-organizing maps (SOM). The cluster analysis results show six clusters of similarly behaving rain gauges, where each cluster addresses one isolated class of variables for the rain gauge. Additionally, the clusters confirm the spatial variation of rainfall caused by the complex topography of Mumbai, comprising the flatland near the Arabian Sea, high-rise buildings (urban area) and mountain and hills areas (Sanjay Gandhi National Park located in the northern part of Mumbai).

Keywords

Minimum Inter-Event Time, Self-Organizing Map, Rain Event, Dendrogram

1. Introduction

Rainfall is an essential boundary condition for the design and operation of urban drainage systems. Compared with natural hydrology, detailed knowledge of the distribution of precipitation characteristics such as duration, depth, and intensity, is more vital for operating such systems due to small spatial-scale involvement and short reaction times from rainfall to runoff. In Mumbai, the occur-

rence of severe floods almost every year has resulted in considerable property, network, and infrastructure damage. Furthermore, the extreme rainfall event that occurred on 26 July 2005 demonstrates the high spatial variability of the rainfall event in different areas over a 24 h period [1]. Therefore, it is vital to represent the time series of rainfall amounts in the form of entities (rain events) that can be employed in specific applications [2]. Such individual rain events could also overcome the problem of a lack of long-term precipitation data at a specific temporal and spatial resolution [3] [4].

Based on a comprehensive review, the time series of rain gauge records is broken down into separate rain events by a defined dry period between rains referred to as the minimum inter-event time (MIT), which is often combined with minimum event depth [2] [5]. Dunkerley [6] noted that MIT values ranged from 3 min to 24 h, with minimum event depths ranging from 0.1 mm to 13 mm. Additionally, statistical techniques, such as autocorrelation analysis [7] [8], and multifractal and self-organized criticality theories [9] have been used. Many researchers [4] [10] [11] have used the assumption that the probability density of inter-event times (IETs) can be adequately represented by an exponential distribution using the coefficient of variation (COV) method.

Therefore, the objective of this research was to study the spatial characterizes of individual events. Hence, to identify individual events at each station using MIT values calculated by the COV method. In addition to MIT values, numerous properties of individual events, such as total depth, total duration, mean rain rate, maximum rain rate, peak timing, previous inter-event time, a fraction of event depth until event maximum and fraction of intra-event rainless periods, can be computed. The interdependence between these properties and IETs can be studied to understand precipitation characteristics in various regions, and this interdependence has been analyzed in various countries by researchers (e.g., Blue Nile River, Ethiopia, [4]; Italy, [12]; Illinois, [13]; the Czech Republic, [10]; Malaysia, [14] [15]; Australian drylands, [16]). In this study, we have used eight properties (used by the many researchers) to characterize rainfall.

The identified properties used for rainfall characterization can further be employed to cluster the rain gauges. Cluster analysis techniques have been used by researchers in various parts of the world including India, e.g., correlation analysis [17] [18], principal component analysis [19], cluster analysis [20] [21], neural networks [22] and shared nearest neighbor (SNN) clustering [23]; however, most of these studies have been conducted on a national and not a regional scale. Apart from the abovementioned techniques, self-organizing maps (SOMs) are a promising technique for cluster analysis [24] [25]. In this study, cluster analysis involved the use of SOMs to cluster rain gauges in groups (regions) such that the comparability of rain gauges within a region is augmented while the similitude of those between regions is limited.

The research presented in this paper is motivated by the fact that, most of the studies on this region used data from two Indian Metrological Department

(IMD) stations (Colaba and Santacruz) to investigate the formation and prediction of extreme events and design rainfall estimation, and that, until now, there have not been any reports related to MIT, rain event property analysis and rain gauge clustering using sub-hourly data. For example, Sherly *et al.* [26] proposed a structure for design precipitation estimation utilizing a multivariate semiparametric approach, while Sen *et al.* [27] used the IMD rainfall dataset of 43 years and performed on-site design rainfall estimation for quantifying spatiotemporal variability used in [27]. Additionally, Nayak and Ghosh [28] used support vector machine and statistical techniques based on machine learning to predict extreme rainfall. Singh *et al.* [29] noted that the unusual pattern of spatiotemporal behavior of extreme rainfall might be related to various physical variables, such as the Indian Ocean Dipole, the El Nino-Southern Oscillation along with the East Atlantic Pattern and the high-speed wind blowing from the Arabian Sea conveying abundance dampness to the surrounding regions. Their results were based on a study of the spatiotemporal characteristics of extreme rainfall using two years of data from 26 rain gauges. Therefore, considering that previous studies have not focused on rain event properties, the present study conducts an effective investigation of rain event properties and their interdependence on IETs, with Mumbai, India, as the study region. This study also seeks to answer the following questions:

- 1) What is the appropriate MIT for the entire Mumbai region using the COV method?
- 2) How can we analyze and understand the spatial characteristics of rainfall properties using SOMs?
- 3) Which rain gauges have similar properties, *i.e.*, lie in the same cluster?

To answer the above questions, the sub-hourly data from a dense gauge network in Mumbai were analyzed. The paper is structured as follows. Section 2 provides a brief description of the study area, data source, and rain gauge characteristics. Section 3 describes the methodology used for the analysis of MIT, rain event properties and clustering using SOM. Section 4 presents the primary findings and their significance. Finally, Section 5 concludes the paper.

2. Study Area and Data Description

Mumbai is situated on the western coast of India and extends between 18.00° - 19.20°N and 72.00° - 73.00°E. It has a total land area of 468 km², with ~68 km² making up the city and 370 km² making up the suburbs [26]. The region has a humid, tropical climate, with monsoons occurring from June to September that is brought in from the southwest Indian Ocean. Initially, Mumbai was composed of a group of islands, which have now been reclaimed to manage the demand for space. This has resulted in the region having a complex topography that comprises flatlands (below or at mean sea level), urban areas, and mountain and hills areas [Sanjay Gandhi National Park located in the northern part with an elevation of up to 450 m above mean sea level (a.m.s.l.)].

Sub-hourly precipitation data (R_p , 15-min. interval) were acquired from the Municipal Corporation of Greater Mumbai (MCGM). These data were collected from 60 rain gauges for the southwest monsoon period (June–September) from 2006 to 2016. However, after accounting for missing data; newer installations; and the reliability, consistency and operation period of each rain gauge, the present study considered data from 47 stations. **Figure 1** shows the unique codes and locations of the gauges with Shuttle Radar Topography Mission (SRTM), and **Table 1** presents the basic characteristics of the rain gauges at the meteorological stations.

3. Methodology

This section describes the methodology used to achieve the objectives of this study.

3.1. Data Preparation and Determination of MIT

In this study, continuous rainfall data recorded by the rain gauges at the meteorological stations were analyzed a sub-hourly basis. Furthermore, to ensure the

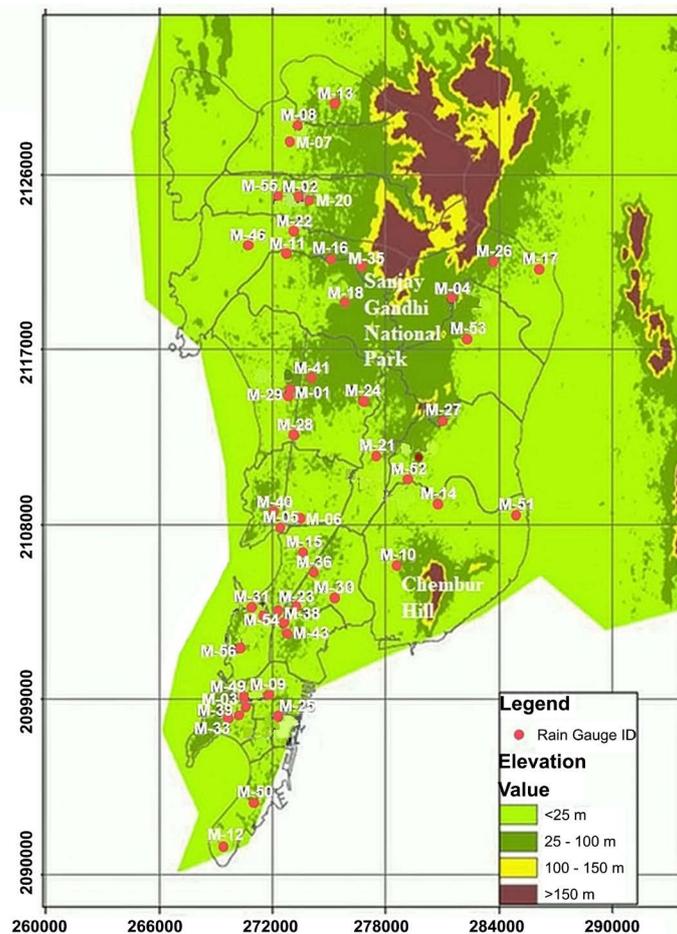


Figure 1. Unique codes and locations of rain gauges at meteorological stations throughout Mumbai with SRTM.

Table 1. Basic characteristics of rain gauges at meteorological stations throughout Mumbai.

Unique ID	Station_Name	Number of years in operation	Height (a.m.s.l.) m	Average annual rainfall (mm)	Final MIT (h)	Average annual number of
M-01	Andheri	10	28.1	2228	5	80
M-02	BMC 1	7	19.7	1793	4	78
M-03	BMC 2	7	27.1	1882	4.25	83
M-04	Bhandup Complex	8	85.7	2137	4	94
M-05	Bandra	10	11.7	2165	4.25	89
M-06	Bandra-Kurla Complex	11	28.2	2185	5.25	72
M-07	Borivali 1	9	15.4	2055	5.25	63
M-08	Borivali 2	5	13.9	2229	4.75	56
M-09	Byculla	11	24.7	2089	4.5	87
M-10	Chembur	11	23.7	2190	7	53
M-11	Chincholi	7	24.8	2123	4.25	101
M-12	Colaba	11	31.3	1990	4.75	76
M-13	Dahisar	8	24.5	2617	4.75	65
M-14	Deonar	10	13.8	2218	4.75	65
M-15	Dharavi	10	12.3	2044	7.25	49
M-16	Dindoshi	6	35	2066	4.5	55
M-17	Gawan Pada	11	10.7	2475	4	84
M-18	Goregaon	9	35.9	2228	5.75	66
M-20	Kandivali	9	25	2202	6.5	25
M-21	Kurla	11	22.5	2598	3.2	70
M-22	Malad	10	19	1940	6.75	57
M-23	Dadar	11	19	2092	7.5	44
M-24	Marol	7	28.3	2139	6.17	81
M-25	Memonwada	11	23.8	2026	4.75	68
M-26	Mulund	10	34.1	2395	7	41
M-27	Vikhroli	8	29.6	2487	2.75	27
M-28	Vile-Parle WL	4	20.3	2393	5.25	77
M-29	Vile-Parle W	10	22.8	2143	5.75	70
M-30	Wadala	11	26.4	2137	3.4	41
M-31	Worli	11	21.6	2034	5	75
M-33	Nana chawk (D Ward)	6	28.8	1892	3.75	90
M-35	C Ward	4	50.1	1905	4.75	41
M-36	F North	4	26.7	2269	2	108
M-38	F South	4	21.5	2177	4.25	76
M-39	Grant Road Eye	4	13.3	2005	4.5	74
M-40	H west Ward	4	31.3	1955	2.75	100
M-41	K East Ward	4	37.2	1981	5.5	54

Continued

M-43	KEM Hospital	4	25.1	1602	4.75	60
M-46	Malvani	4	12.7	2158	3.25	86
M-49	Nair Hospital	4	20.2	1935	4.75	64
M-50	Nariman Fire	4	14.9	1792	4.5	67
M-51	Octroi Mankhurd	4	8.8	2429	7.5	32
M-52	Rawali Camp	4	9.9	2232	3	101
M-53	S Ward	4	32.3	2594	4	73
M-54	SWD Workshop Dadar	4	15.3	2252	5	121
M-55	Workshop Kandivali	4	19.6	2284	7.75	32
M-56	Worli Data Center	4	13.3	1892	4.75	68

reliability of the data and to identify suspect or incorrect values, a validation method was used. In the validation process, the data were not modified but supplemented with an appropriate flag, namely, “Suspect” and “Missing” followed by the application of a range test [30] and double mass curve. Only stations with more than three years of data were considered in the process. The entire procedure resulted in the inclusion of 47 stations, with recorded data time varying from 4 to 11 years. To assure, a near-uniform spatial distribution of regions, marginally deficient records with minor gaps in the sub-hourly precipitation records (splits of up to a few days) was utilized.

Precipitation is recorded using a rain gauge, which measures the amount of water, precipitated by clouds, reaching the ground. Based on homogeneity, the precipitation records can then be clustered into independent events. To Separate and interpret independent events from the time series rain gauge records is subjective. For example, in **Figure 2**, for groups A, B, and C, either each group can be considered as an independent one or all three groups may be part of one event. Another probability is to club A and B together and consider it as an event and Group C to be another event. Hence, to calculate the start and end times of such independent events, an MIT estimate is required. Accordingly, two tempests isolated by a rainless period with values less than a specified MIT value are considered as a single event, as shown in the left-hand corner of **Figure 2** and vice versa on the right side of **Figure 2**. In the present study, optimal MIT was estimated using the COV method proposed by Restrepo-Posada and Eagleson [31] and applied by several researchers. This method proposed a simple check using the COV of IETs. Assuming that the distribution of IETs to be roughly exponential, which implies equal mean and standard deviation, the COV ought to be 1. Therefore, MIT values are systematically modified, and the MIT is prompting $COV = 1$ is identified as ideal.

3.2. Self-Organized Maps (SOM)

The vital information contained in each rain events should be extracted using a

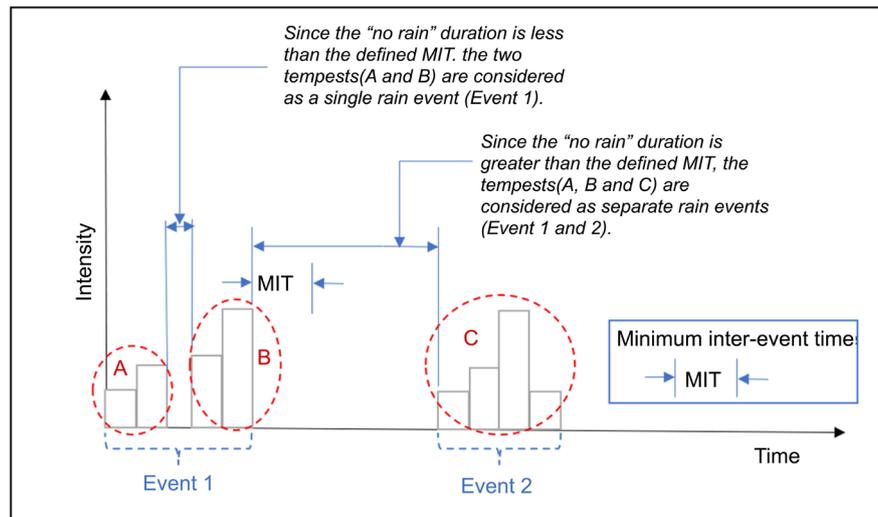


Figure 2. Diagram illustrating the classification of rain events using inter-event times.

limited set of well-chosen properties. However, there is no standard or generally acknowledged list or a particular set of properties that can be utilized to precisely depict and abridge an event. Based on literature reviews, the often-used properties in various studies are duration, depth, mean rain rate, maximum rainfall intensity and intra-event dry periods and hydrology studies have used event peak intensity for an aggregation scale at different time steps [10] [14] [24] [32]. We also include the property describing the fraction of intra-event rainless periods, previous inter-event time and the position of the peak in the event (time of peak and fraction of event depth until event maximum), which are relevant overland flow generation, runoff, and infiltration [33] [34] [35] [36]. In this study, eight properties were obtained for each of the rain events derived from the 47 rain gauges, and these are listed in **Table 2**.

The event characteristics at each station described by the properties (**Table 2**) can be seen as a vector that has a high-dimensional data space. We used a flexible, data-mining, SOM method [37] [38] for exploratory data to analyze such high-dimensional data spaces of the studied rain gauges. A SOM is an unsupervised learning algorithm based on artificial neural networks that produce a low-dimensional representation of a high-dimensional input dataset. SOMs can be used for a variety of operations in exploratory data analysis, such as clustering, data compression, non-linear projection and pattern recognition.

A SOM comprises cells that are sorted out on a regular grid. Each cell is drawn by a d -dimensional weight vector and associated with nearby cells by a relation, which decides the structure, *i.e.*, the topology of the resulting SOM. The SOM is then created through iterative training; input vectors relating to data samples in the given data matrix are randomly picked in each turn, and the distances between them and all weight vectors of the SOM are computed. The cell that has a weight vector nearest to the input vector in question is the input vector's best-matching unit (BMU). After obtaining the BMU, the weight vector is

Table 2. Variables are characterizing rain events in this study.

Sr. No.	Symbol	Variables/Properties	Calculation	Unit
1	D	Event duration	$D = T_{\text{end}} - T_{\text{begin}} + 15[\text{min}]$	min
2	D_i	Intra-event dry periods	$D_i = \sum_{t=T_{\text{begin}}}^{t=T_{\text{end}}} R_d$ with $I_d = \begin{cases} 15, & \text{if } I_d = 0[\text{mm}] \\ 0 & \text{else} \end{cases}$	min
3	D_p	Total Depth	$D_p = \sum_{d=T_{\text{begin}}}^{d=T_{\text{end}}} R_d$	mm
4	R_m	Mean event rain rate	$R_m = \frac{1}{D} \sum_{t=T_{\text{begin}}}^{t=T_{\text{end}}} R_t$	mmh ⁻¹
5	R_{max}	Maximum rain rate	$R_{\text{max}} = \max(R_t)$	mmh ⁻¹
6	IET	Previous inter-event time	$IET = T_{\text{begin}}(\text{currentevent}) - T_{\text{end}}(\text{Previousevent} + 15)$	min
7	T_{mx}	Time to peak	$T_{\text{mx}} = \frac{t^{\text{mx}}}{D}$	min
8	D^{mx}	The fraction of event depth until event maximum	$D^{\text{mx}} = \frac{\sum_{d=T_{\text{begin}}}^{d^{\text{mx}}} D_d}{D_p}$	
9	$D_{d\%}$	The fraction of intra event rainless periods	$D_{d\%} = \frac{D_i}{D}$	%

updated so that the BMU and its neighbors are moved towards the input vector. The SOM is then trained with the net effect of the whole dataset by the batch algorithm, which computes an average of the data samples weighted by the neighborhood function of each data sample at its BMU.

In this paper, we run the SOM tool in MATLAB to create a SOM using the SOM algorithm as described in Vesanto *et al.* [39]. The training dataset considered in this analysis comprises eight properties, along with three geographic coordinates—x, y and z positions—of each rain gauge. The final dataset has an overall dimension of 11 attributes and a size of about 23,000 data samples. Prior to the training, the data were normalized to a [0 1] interval. To avoid the disproportionate influence of the high index values on the training, linear transformation of the data is carried out. On the other hand, geospatial data posed some important features of nature that occurs at all scale, also perceived by its gradual, fuzzy or vague changes raise some issues in the utilization of the SOM algorithm. However, these issues can be resolved using recommended approaches such as data pre-processing (normalization and attribute weighting) and geo-initialization [40] [41]. Therefore, before starting the analysis, the data were normalized, and to preserve the geospatial aspects, the x variable was scaled down using the ratio max(x)/max(y) and was weighted with a value of eight.

There are two phases in which the SOM algorithm training is usually performed, viz. rough and fine-tuning. Relatively large initial learning rate (0.5) and neighborhood radius (3) are used in the rough tuning phase while in the fine-tuning phase, both the above values are taken smaller right from the begin-

ning (0.05 and 1). But, before the algorithm is applied and analyzed, there are few parameters (**Table 3**), that need to define. All these parameters together play an important role in the SOM algorithm as they could influence the result obtained. The parameters used for training the SOM are summarised in **Table 3**. The MATLAB package proposed for the data a map of size 49×15 , based on the hexagonal lattice. However, for an easier presentation of the map, we built a smaller map of size 16×5 in this study. This brought about a decay of the quantization error from 0.420 to 0.559 and enhancement of the topological error from 0.126 to 0.052.

A perfect SOM analysis creates such apparent outcomes that envisioned maps could be dependably deciphered simply by taking a look at them, even though extra apportioning that utilizes SOM as a halfway step is often prescribed to obtain more precise outcomes [39] [42]. In this study, popular visualizations of SOMs, such as U-matrix, component plans, assignment of rain gauges to neurons and the distribution of index properties represented by bar charts, are used. Furthermore, the study uses hierarchical clustering, an unverified method, for clustering the SOM [43]. The approach begins with single data points as individual clusters, and at each progression, each cluster consolidates with the nearest pair of clusters until one cluster remains. Hence, the approach is also known as the agglomerative approach and calls for the definition of cluster proximity. In this investigation, cluster proximity is characterized by the average pairwise proximity among all sets of points in various groups and is represented by the average group distance. The outcome is called a dendrogram which is a tree-like diagram. A dendrogram shows both the cluster and sub-cluster relationships and the order in which the clusters were consolidated. The closeness of the clusters can be depicted by lengths of the limbs, and the data items can be clustered by cutting the dendrogram.

Table 3. Parameters for SOM analysis.

Parameter	Standard SOM
The size of a map	16×5 (49×15)
Grid	Hexagonal
Type	Toroid
H	Gaussian
Training mode	Batch
Rough initial	0.5
Fine initial	0.01
Number of epochs at the rough tuning	1000
Learning rate at the rough tuning	0.5
Initial radius of training	3
Number of epochs at the fine-tuning	2000
Learning rate at the fine-tuning	0.05
Final radius of training	1

4. Results and Discussion

The results obtained in this study are presented and discussed in this section, which is further divided into three subsections.

4.1. Minimum Inter-Event Time (MIT)

The MIT values were estimated using the COV method. In this approach, for all stations, the MIT values varied from 15 min to 24 h. Furthermore, events for each MIT were identified, and the COV of the IETs was obtained. The outcomes are shown in **Figure 3**, where COV is noted to decrease with increasing MIT. The approximate values of MIT for which $COV = 1$ was assessed; **Table 1** shows the appropriate MIT values for all stations. The average MIT for the study area was noted to be 5 h appear reasonable as it was reported that MIT with less than 6 h could be suggested for the urbanest application [4] [14] [16], with the minimum value being 2h at F North station and the maximum values being eight at Workshop Kandivali station. The average annual number of storm events that occurred in Mumbai during the southwest monsoon season in the studied period at each station is given in **Table 1**; the average annual number of storms events varies from 25 to 121.

4.2. Self-Organized Maps (SOM)

The resulting dataset considered in this analysis comprises eight essential properties, along with three geographic coordinates—x, y and z positions—of each

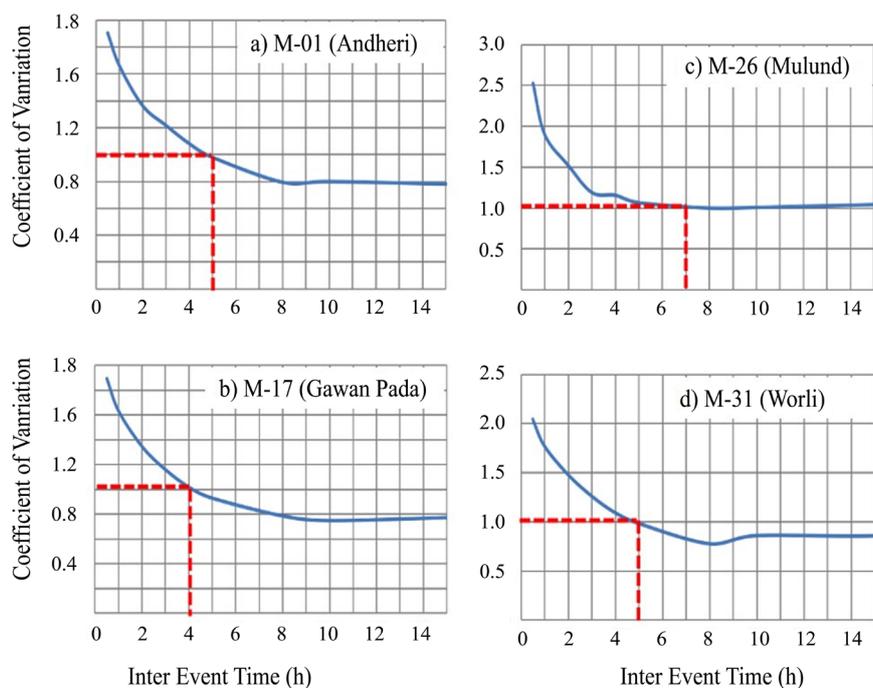


Figure 3. Coefficient of variation against interevent time for station Andheri (a), Gawan Pada (b), Mulund (c) and Worli (d) in Mumbai.

rain gauge. The final dataset has an overall dimension of 11 attributes representing the characteristics of about 23,000 rain events at 47 rain gauges. The rain gauges with similar response are grouped by training a SOM and implementing hierarchical clustering on the SOM.

The U-matrix (**Figure 4(a)**) demonstrates two distinctive parts of the map: blue-colored areas in the southwest part indicate units with a high level of similitude, which can be viewed as discrete clusters. A column of red and orange colors at the focal point of the south side isolates the neurons from the rest of the map and forms a cluster border. Further, we can visualize that there are at least 4 - 5 clusters in the data. The SOM grid with the assignment of 47 available rain gauges to neurons (**Figure 4(c)**) indicates that the gauges marked together on one neuron can be considered as having comparable behavior and form the smallest unit of a cluster. The SOM reduced the variability of the 47 rain gauges to 40 neurons. Additionally, dendrogram cluster analysis was applied to the SOM to reduce the number of clusters further. By assigning the rain gauges to neurons (**Figure 4(c)**), we can group the rain gauges into six sensible clusters—C1 to C6. The spatial distribution of the clustered rain gauges (**Figure 5**) shows that rain gauges with similar response behavior are grouped together.

A marginal rain gauge is identified as a rain gauge that is labeled to neighboring

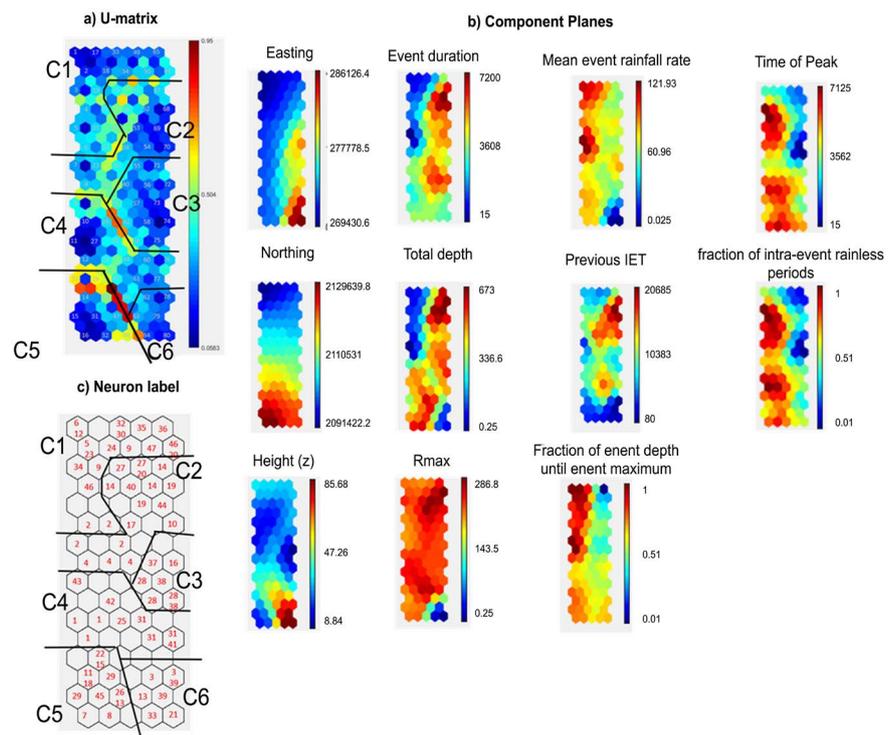


Figure 4. Representations of a SOM: (a) U-matrix: neurons of the SOM are labeled by numbers, indicating structures formed by visualizing distances between neighboring neurons and, on additional hexagons between neurons, medium distances between two neurons; (b) Component planes for each index display mean values of each vital property on the neurons of the SOM; (c) Assignment of rain gauge using neurons labels; gauge IDs correspond to their BMU.

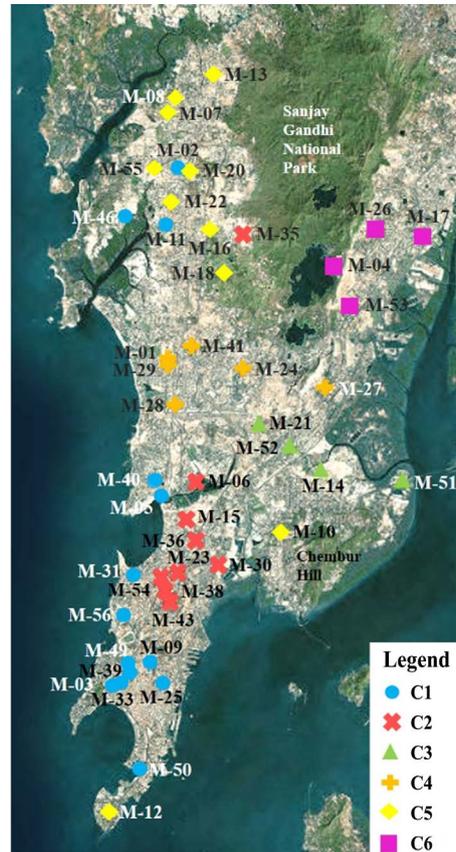


Figure 5. Location and clustering of each rain gauge using SOM.

neurons with a not exceeding medium distance and belonging to different clusters. As a definitive measure of marginal rain gauges, we utilized their quantization errors to the BMU and the neuron of the neighboring cluster. Rain gauges with a similar quantization error to different neurons can be perceived as having a place with two neurons and thus may have a place with two clusters. With this strategy, we enhanced the clustering and areal grouping of rain gauges (Figure 5) and identified three marginal cases (rain gauges M-05, M-12, and M-38) from the clusters through response behavior. These rain gauges belonged to neurons at the edge of a cluster with medium distance to a neighboring neuron of another cluster. In their first 4 BMUs, there was a frequent change between both clusters with slightly increasing quantization errors. In this manner, we assigned the rain gauges to two clusters—M-05 to C1 and C2, M-12 to C5 and C6, and M-38 to C1 and C2.

4.3. Cluster Analysis

In this study, we assumed that two or more rain gauges behave similarly and are grouped by training a SOM and implementing hierarchical clustering. With the information from the component planes (Figure 4(b)) and the distribution of properties for each neuron, we can characterize each cluster by a unique combination of aspects of the rain event characteristics, listed in Table 4.

Table 4. Characteristics of clusters C1 to C6 based on rain event properties.

Cluster	Number of rain gauges	Total depth of rainfall	Mean event rainfall rate	Maximum event rain rate	The fraction of event depth until event maximum	Previous IET	The total duration of the event	the fraction of intra-event rainless periods	Time to peak
C1	14	--	++	+	++	-- to -	--	+	[] to ++
C2	9	++	[]	++	-	++	+	+	[]
C3	4	+	[]	++	-	-	[]	--	-- to -
C4	6	+	[]	++	[]	-	- to []	[] to +	++
C5	10	++	+	++	+	--	[]	+	++
C6	4	--	--	--	+	--	[]	[]	[]

Note: ++, very high, +, high, [], medium, -, low, --, very low.

These clusters are discussed as below:

Cluster 1 (C1) consists of 14 stations located in the area near the Arabian Sea, mainly covering the southwest area of Mumbai. It is characterized by high-intensity, low-depth and short-duration events. A large number of storms were observed in this area, as it has a low previous IET. The component plane (**Figure 4(b)**) of a fraction of event depth until event maximum and time of peak highlighted that the events in this area received the maximum amount of rain before reaching high intensity with peaks in the second or third quartile. This is a result of the southwestern trade winds that carry significant moisture inland from the Arabian Sea in the south. Due to the high moisture, this region experiences heavy monsoonal rainfall and hence has a typical tropical monsoon climate.

Cluster 2 (C2) comprises nine sites, covers the southwest and extends further inland. This cluster represents lowland, urban area with high-rise buildings and is characterized by a large duration, very high intensity and high amount of rainfall. This region experiences peaks of rainfall in the first quartile of the duration (low time of peak value), with most of the rainfall occurring after the peak. This may be due to the high-rise buildings in this region.

Cluster 3 (C3) consists of four sites located between the Sanjay Gandhi National Park and Chembur hills. While this cluster is not similar to cluster 2, it exhibits similar characteristics such as average duration, very high intensity and high amount of rainfall. This cluster does not present a clear picture of the rainfall peaks within an event, which may be due to the funneling action by the hills.

Cluster 4 (C4) is located predominantly in the western part of the city between the Arabian Sea and Sanjay Gandhi National Park. It is characterized by low previous IET, high precipitation intensity, large duration and average depth events, with maximum intensity during the first quartile of the event. Furthermore, the value of the property (fraction of event depth until event maximum) varies in this region between 0.4 - 0.7 indicates that most of the storm events may receive an equal amount of rainfall before and after a high-intensity, which

may be attributed to the high wind gusts caused by the funneling action by the hills.

Cluster 5 (C5) consists of 10 sites located in the flatland that lies between the Arabian Sea and the windward side of the Sanjay Gandhi National Park predominantly in the northern part of the city. The characteristics of this cluster are similar to those of cluster 3, except that this cluster has a very low IET, and have a high time of peak value.

Cluster 6 (C6) comprises four sites, located predominantly in the northeastern part of the Sanjay Gandhi National Park on the leeward side of the hills. Results demonstrate that this region experiences minimal rainfall regarding frequency, duration, and intensity (see **Table 4**). Additionally, the low value of properties (R_{max} , a fraction of event depth until event max and depth) highlighted that the events received by this region are smooth with small numbers of sharp peaks. This may be due to the shadowing effect of the hills, resulting in a decrease in rainfall amount.

The results confirm the effect of complex topography, namely, the flatland near the Arabian Sea, high-rise buildings (urban area), mountain and hills areas (Sanjay Gandhi National Park located in the northern part) on the spatial variation of rainfall. The results highlighted the rain gauges within the cluster 2 (C2) located in the urban pocket received the intense precipitation, which supports the findings of Paul *et al.* [44], shows the urban signature for extreme precipitation will be reflected on rainfall recorded by the stations only when the stations are located within the urban pockets affected by intense precipitation. Considering the following factors, we did not regroup the sites further:

- 1) To evade loss of valuable information on precipitation during the analysis of dissimilar sites (M-10, M-27, and M-35), we did not exclude these sites.
- 2) To evade extraordinary events is unlikely since we consider it to be an intolerable hindrance for the process of evaluating it independently.

5. Conclusions

The primary objective of this research was to study the spatial characteristics of rain events in Mumbai. This was achieved by computing the SOM with 11 properties, represents the characteristics of rain events, analyzed using sub-hourly rain gauge data from 2006-2016. The following results and conclusions can be drawn as:

- 1) This study emphasizes the need for event analysis over the traditional usage of sample analysis and brings out the advantages of event analysis in the case of limited data availability. The sample analysis involving excessively long integration times (hours or days) usually consists of rainy and clear sky periods in the same sample. Furthermore, it consists of distinct physical processes of the rainfall event. Such long integration times can lead to the mixing of rainy and clear sky observations. On the other hand, the data acquired in a short time (minutes) are sensitive to the sensor's characteristics (detection threshold, sensor area, and

noise). Therefore, by analyzing rain events, individual relationships between the properties of rain events can be identified.

2) The various properties derived from rain gauge data can be utilized to portray precipitation in rain time-series studies, which are then used to study various topics including meteorology, hydrology, and climate.

3) We can conclude that the usage of a combination of the selected properties has helped in describing the depth of rain events by providing a relatively accurate summary. These properties are the as total depth, total duration, mean rain rate, maximum rain rate, peak timing, previous inter-event time, a fraction of event depth until event maximum and fraction of intra-event rainless periods.

4) The dependence of the event depth and previous IET, confirms that for events interrupted by no rain periods, an independent rain event cannot be considered as an alternative to rain time series in a forecast,

5) The results show that the gauges can be clustered into six distinct groups using SOM. Furthermore, the clusters confirm the spatial variation of rainfall as a result of the complex topography of Mumbai, namely, the flatland near the Arabian Sea, high-rise buildings (urban area), and mountain and hill areas (Sanjay Gandhi National Park located in the northern part).

The limitations of the study are the followings:

1) In the present analysis, uncertainties associated with the MIT value selection criteria has not been addressed, this is a potential area of future research.

2) The hypothesis of clusters needs to be further tested using the climatological data (temperature, wind speed, wind direction, etc.) measured by MCGM at the 47 stations and radar data obtained by the Indian Metrological Department, Mumbai; this can be a potential area of future research.

3) As this study was based on the analysis of observation made in Mumbai having complex orography and coastal area, the significance of this hypothesis remains to be confirmed through the studies of multiple cases across different cities; this can be a potential area of future research.

Acknowledgements

The authors thank the MCGM for providing rain gauge data and the Indian Metrological Department, Mumbai, for their comments and suggestions, which significantly contributed to improving the clarity of the paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Lokanadham, B., Gupta, K. and Nikam, V. (2012) Characterization of Spatial and Temporal Distribution of Monsoon Rainfall over Mumbai. *ISH Journal of Hydraulic Engineering*, **15**, 69-80. <https://doi.org/10.1080/09715010.2009.10514941>
- [2] Brown, B.G., Katz, R.W. and Murphy, A.H. (1984) Statistical Analysis of Clima-

- logical Data to Characterize Erosion Potential: 4. Freezing Events in Eastern Oregon/Washington. *Oregon Agricultural Experiment Station Spec. Rep. No. 689*, Oregon State University.
- [3] Diodato, N. and Bellocchi, G. (2007) Estimating Monthly (R)USLE Climate Input in a Mediterranean Region Using Limited Data. *Journal of Hydrology*, **345**, 224-236. <https://doi.org/10.1016/j.jhydrol.2007.08.008>
- [4] Haile, A.T., Rientjes, T.H.M., Habib, E., Jetten, V. and Gebremichael, M. (2011) Rain Event Properties at the Source of the Blue Nile River. *Hydrology and Earth System Sciences*, **15**, 1023-1034. <https://doi.org/10.5194/hess-15-1023-2011>
- [5] Larsen, M.L. and Teves, J.B. (2015) Identifying Individual Rain Events with a Dense Disdrometer Network. *Advances in Meteorology*, **2015**, Article ID: 582782. <https://doi.org/10.1155/2015/582782>
- [6] Dunkerley, D. (2008) Rain Event Properties in Nature and Rainfall Simulation Experiments: A Comparative Review with Recommendations for Increasingly Systematic Study and Reporting. *Hydrological Processes*, **22**, 4415-4435. <https://doi.org/10.1002/hyp.7045>
- [7] Adams, B.J. and Papa, F. (2000) *Urban Stormwater Management Planning with Analytical Probabilistic Models*. John Wiley & Sons, New York.
- [8] Nojumuddin, N.S., Yusof, F. and Yusop, Z. (2018) Determination of Minimum Inter-Event Time for Storm Characterisation in Johor, Malaysia. *Journal of Flood Risk Management*, **11**, S687-S699. <https://doi.org/10.1111/jfr3.12242>
- [9] Medina-Cobo, M.T., Garcia-Marin, A.P., Estevez, J. and Ayuso-Munoz, J.L. (2016) The Identification of an Appropriate Minimum Inter-Event Time (MIT) Based on the Multifractal Characterization of Rainfall Data Series. *Hydrological Processes*, **30**, 3507-3517. <https://doi.org/10.1002/hyp.10875>
- [10] Hanel, M. and Máca, P. (2014) Spatial Variability and Interdependence of Rain Event Characteristics in the Czech Republic. *Hydrological Processes*, **28**, 2929-2944.
- [11] Joo, J., Lee, J., Kim, J.H., Jun, H. and Jo, D. (2014) Inter-Event Time Definition Setting Procedure for Urban Drainage Systems. *Water*, **6**, 45-58. <https://doi.org/10.3390/w6010045>
- [12] Balistrocchi, M. and Bacchi, B. (2011) Modelling the Statistical Dependence of Rainfall Event Variables by a Trivariate Copula Function. *Hydrology and Earth System Sciences Discussions*, **8**, 429-481. <https://doi.org/10.5194/hessd-8-429-2011>
- [13] Gyasi-Agyei, Y. and Melching, C.S. (2012) Modelling the Dependence and Internal Structure of Storm Events for Continuous Rainfall Simulation. *Journal of Hydrology*, **464**, 249-261. <https://doi.org/10.1016/j.jhydrol.2012.07.014>
- [14] Shamsudin, S., Dan'azumi, S. and Aris, A. (2010) Effect of Storm Separation Time on a Rainfall Characteristics—A Case Study of Johor, Malaysia. *European Journal of Scientific Research*, **45**, 162-167.
- [15] Chin, R.J., Lai, S.H., Chang, K.B., Othman, F. and Jaafar, W.Z.W. (2016) Analysis of Rainfall Events over Peninsular Malaysia. *Weather*, **71**, 118-123. <https://doi.org/10.1002/wea.2723>
- [16] Dunkerley, D. (2008) Identifying Individual Rain Events from Pluviograph Records: A Review with Analysis of Data from an Australian Dryland Site. *Hydrological Processes*, **22**, 5024-5036. <https://doi.org/10.1002/hyp.7122>
- [17] Saikranthi, K., Rao, T.N., Rajeevan, M. and Bhaskara Rao, S.V. (2013) Identification and Validation of Homogeneous Rainfall Zones in India Using Correlation Analysis. *Journal of Hydrometeorology*, **14**, 304-317.

<https://doi.org/10.1175/JHM-D-12-071.1>

- [18] Sinha, P., Mohanty, U.C., Kar, S.C., Dash, S.K., Robertson, A.W. and Tippett, M.K. (2013) Seasonal Prediction of the Indian Summer Monsoon Rainfall Using Canonical Correlation Analysis of the NCMRWF Global Model Products. *International Journal of Climatology*, **33**, 1601-1614. <https://doi.org/10.1002/joc.3536>
- [19] Nair, A., Mohanty, U.C. and Acharya, N. (2013) Monthly Prediction of Rainfall over India and Its Homogeneous Zones during Monsoon Season: A Supervised Principal Component Regression Approach on General Circulation Model Products. *Theoretical and Applied Climatology*, **111**, 327-339. <https://doi.org/10.1007/s00704-012-0660-8>
- [20] Ahuja, S. and Dhanya, C.T. (2012) Regionalization of Rainfall Using RCDA Cluster Ensemble Algorithm in India. *Journal of Software Engineering and Applications*, **5**, 568. <https://doi.org/10.4236/jsea.2012.58065>
- [21] Bharath, R. and Srinivas, V.V. (2015) Regionalization of Extreme Rainfall in India. *International Journal of Climatology*, **35**, 1142-1156. <https://doi.org/10.1002/joc.4044>
- [22] Saha, M., Mitra, P. and Nanjundiah, R.S. (2017) Deep Learning for Predicting the Monsoon over the Homogeneous Region of India. *Journal of Earth System Science*, **126**, 54. <https://doi.org/10.1007/s12040-017-0838-7>
- [23] Kakade, S.B. and Kulkarni, A. (2017) Seasonal Prediction of Summer Monsoon Rainfall over Cluster Regions of India. *Journal of Earth System Science*, **126**, 34. <https://doi.org/10.1007/s12040-017-0811-5>
- [24] Dilmi, M.D., Mallet, C., Barthès, L. and Chazottes, A. (2017) Data-Driven Clustering of Rain Events: Microphysics Information Derived from Macro-Scale Observations. *Atmospheric Measurement Techniques*, **10**, 1557-1574. <https://doi.org/10.5194/amt-10-1557-2017>
- [25] Manaán, A., Chaudhary, S., Dhanya, C.T. and Swamy, A.K. (2017) Regionalization of Rainfall Characteristics in India Incorporating Climatic Variables and Using Self-Organizing Maps. *ISH Journal of Hydraulic Engineering*, **24**, 147-156.
- [26] Sherly, M.A., Karmakar, S., Chan, T. and Rau, C. (2015) Design Rainfall Framework Using Multivariate Parametric-Nonparametric Approach. *Journal of Hydrologic Engineering*, **21**, Article ID: 04015049. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001256](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001256)
- [27] Sen, S., Vittal, H., Singh, T., Singh, J. and Karmakar, S. (2013) At-Site Design Rainfall Estimation with a Diagnostic Check for Nonstationarity: An Application to Mumbai Rainfall Datasets. *Proceedings of Hydro*, Madras, 4-6 December 2013.
- [28] Nayak, M.A. and Ghosh, S. (2013) Prediction of Extreme Rainfall Event Using Weather Pattern Recognition and Support Vector Machine Classifier. *Theoretical and Applied Climatology*, **114**, 583-603. <https://doi.org/10.1007/s00704-013-0867-3>
- [29] Singh, J., Sekharan, S., Karmakar, S., Ghosh, S., Zope, P.E. and Eldho, T.I. (2017) Spatio-Temporal Analysis of Sub-Hourly Rainfall over Mumbai, India: Is Statistical Forecasting Futile? *Journal of Earth System Science*, **126**, 38. <https://doi.org/10.1007/s12040-017-0817-z>
- [30] Estevez, J., Gavilan, P., García-Marin, A.P. and Zardi, D. (2015) Detection of Spurious Precipitation Signals from Automatic Weather Stations in Irrigated Areas. *International Journal of Climatology*, **35**, 1556-1568. <https://doi.org/10.1002/joc.4076>
- [31] Restrepo-Posada, P.J. and Eagleson, P.S. (1982) Identification of Independent Rainstorms. *Journal of Hydrology*, **55**, 303-319. [https://doi.org/10.1016/0022-1694\(82\)90136-6](https://doi.org/10.1016/0022-1694(82)90136-6)

- [32] Lowe, R., Madsen, H. and McSharry, P. (2016) Objective Classification of Rainfall in Northern Europe for the Online Operation of Urban Water Systems Based on Clustering Techniques. *Water*, **8**, 87. <https://doi.org/10.3390/w8030087>
- [33] He, Z., Zhao, W., Liu, H. and Chang, X. (2012) The Response of Soil Moisture to Rainfall Event Size in Subalpine Grassland and Meadows in a Semi-Arid Mountain Range: A Case Study in Northwestern China's Qilian Mountains. *Journal of Hydrology*, **420**, 183-190. <https://doi.org/10.1016/j.jhydrol.2011.11.056>
- [34] Huang, J., Zhang, J., Zhang, Z. and Xu, C.Y. (2012) Spatial and Temporal Variations in Rainfall Erosivity during 1960-2005 in the Yangtze River Basin. *Stochastic Environmental Research and Risk Assessment*, **27**, 337-351. <https://doi.org/10.1007/s00477-012-0607-8>
- [35] Meusburger, K., Steel, A., Panagos, P., Montanarella, L. and Alewell, C. (2012) Spatial and Temporal Variability of Rainfall Erosivity Factor for Switzerland. *Hydrology and Earth System Sciences*, **16**, 167-177. <https://doi.org/10.5194/hess-16-167-2012>
- [36] Ran, Q., Su, D., Li, P. and He, Z. (2012) Experimental Study of the Impact of Rainfall Characteristics on Runoff Generation and Soil Erosion. *Journal of Hydrology*, **424-425**, 99-111. <https://doi.org/10.1016/j.jhydrol.2011.12.035>
- [37] Kohonen, T. (1995) Self-Organizing Maps. Third Edition, Springer, Berlin. <https://doi.org/10.1007/978-3-642-97610-0>
- [38] Kohonen, T. (2013) Essentials of the Self-Organizing Map. *Neural Networks*, **37**, 52-65. <https://doi.org/10.1016/j.neunet.2012.09.018>
- [39] Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankagas, J. (2000) SOM Toolbox for Matlab 5, Report A57. <http://www.cis.hut.fi/projects/somtoolbox/>
- [40] Fincke, T., Lobo, V. and Bação, F. (2018) Visualizing Self-Organizing Maps with GIS. GI Days. https://www.researchgate.net/profile/Fernando_Bacao/publication/228660313_Visualizing_self-organizing_maps_with_GIS/links/0fcfd50a2730443654000000.pdf
- [41] Lourenço, F.C., Lobo, V.S. and Bação, F.L. (2014) Exploratory Geospatial Data Analysis Using Self-Organizing Maps. Case Study of Portuguese Mainland Regions. <https://www.researchgate.net/publication/237226357>
- [42] Polzlbauer, G., Dittenbach, M. and Rauber, A. (2006) Advanced Visualization of Self-Organizing Maps with Vector Fields. *Neural Networks*, **19**, 911-922. <https://doi.org/10.1016/j.neunet.2006.05.013>
- [43] Tan, P.N., Steinbach, M. and Kumar, V. (2006) Introduction to Data Mining. Pearson Education, Addison Wesley, Boston, 769.
- [44] Paul, S., Ghosh, S., Mathew, M., Devanand, A., Karmakar, S. and Niyogi, D. (2018) Increased Spatial Variability and Intensification of Extreme Monsoon Rainfall Due to Urbanization. *Scientific Reports*, **8**, Article No. 3918. <https://doi.org/10.1038/s41598-018-22322-9>