Detection of Climate Zones using Multifractal Detrended Cross-Correlation Analysis: A Spatio-temporal Data Mining Approach

Monidipa Das
School of Information Technology
Indian Institute of Technology, Kharagpur
West Bengal-721302, India
Email: monidipadas@hotmail.com

Soumya K. Ghosh
School of Information Technology
Indian Institute of Technology, Kharagpur
West Bengal-721302, India
Email: skg@iitkgp.ac.in

***Abstract*—There has been a significant change in climate throughout the last few decades, resulting into the phenomenon of global warming with all its adverse effects on human life and activities. In this context, detection of climate zone is an important issue, since this may help to avert, or to take adequate measures against, any unprecedented natural calamity. Most of the existing works for this purpose are limited only to the independent study of different climate variables featuring a climate zone. In this paper, we have described a novel approach based on *Multifractal Detrended Cross-correlation Analysis (MF-DXA)* between each pairs of such climate variables of interest. In this approach, the spatio-temporal pattern of any location, as determined by the multifractal correlation study, has been exploited by a *K-means* based clustering technique, which can accurately detect various climate zones over a large region. The approach has been evaluated with the daily time series data of the year 2013 for *land surface temperature* and *precipitation rate*, collected from 73 different locations over the entire *Eastern and North-Eastern region* of *India*. The high resemblance of the identified climate zones with the *World Map* of *Köppen-Geiger climate classification* proves the accuracy and efficacy of the proposed approach.**

Keywords*—Climate zone, Multifractal cross-correlation, Spatio-temporal pattern, Data mining, Generic climate classification.

I. Introduction

With the phenomenon of global warming in the background, climate change study has become a major thirst area in recent time. Climate change, which may be due to several natural factors or human activities, can have severe impacts on the ecological, agricultural, industrial, and socio-economic life of any country throughout the world. In the current scenario, a proper identification of climate patterns on a global as well as regional scale may help to initiate appropriate measures for adaptation and alleviation at any point of time.

Till date, several research works [1]–[5], specially involving clustering as data mining technique, have been proposed to discover climate patterns on global and/or regional basis. Irina Mahlstein and Reto Knutti [3] have proposed a clustering algorithm to define regions having similar average climate and change pattern, using various information about regional climatic features. A functional clustering procedure [5] has been presented by E. Di Giuseppe to identify homogeneous climate zones in Italy. The method combines time series interpolation with smoothing penalized B-spline and performs a medoid based partitioning. Vipin Kumar et al. in [1] have used various existing data mining techniques, like, clustering, association analysis etc., to discover interesting spatio-temporal patterns from the time series of earth science data. In [4], different data mining techniques have been proposed to discover the global changes in climate system in terms of carbon cycle variation. The paper also presents a technique to identify regions of uniform spatio-temporal behavior by determining climate indices by cluster analysis. Yurdanur Unal et al. [2] have used the Ward's method of hierarchical cluster analysis to redefine the climate zone in Turkey.

However, in most of the above cases the cluster analyses have been applied on individual climate variables (like, temperature, precipitation, humidity etc.) or on a combined set of them without considering their inter-relationships which play a significant role to determine the climate patterns of any region.

In this paper, we have presented a novel data mining approach to detect the climate zones, i.e. regions with similar spatio-temporal pattern, using *multifractal detrended cross-correlation analysis (MF-DXA)*. The approach is based on a *K-means* based clustering algorithm which is performed on correlation data between each pair of climate variables.

A. Problem Definition and Contributions

Given the spatio-temporal data of any region R in terms of time series of various climate variables featuring the zone: $v_1, v_2, \cdots, v_n$, where n is the total number of climate variables of interest. The problem is to identify different possible climate zones: $c_1, c_2, \cdots, c_k$ in R, where k is the total number of climate zones identified.

Our proposed approach consists of two main steps. In the first step, the approach uses the well-established *MF-DXA* method [6] to estimate the cross-correlation between each pair of climate variable v_i and v_j $(v_i, v_j \in \{v_1, v_2, \cdots, v_n\}, i \neq j)$. This step helps to capture the correlation pattern between each v_i and v_j over different time scales in terms of multifractal scaling exponents. In the second step, we perform a *K-means* based cluster analysis over all the pair-wise correlation data as obtained in the first step. The approach has been evaluated

978-1-4799-7458-0/15/$31.00

with the time series data[1] of the year 2013 for *land surface temperature* and *precipitation rate*, collected from 73 different locations over all the 12 states in the entire *Eastern and North-Eastern region* of *India*, to detect the various climate zones in the region. The high resemblance of the identified zones with the *World Map* of *Köppen-Geiger climate classification* [8] [9] proves the efficacy of the approach. Thus, the main contributions of this work can be summarized as follows:

- Analyzing the multifractal cross-correlation between the $temperature$ and $precipitation$ time series over 73 different locations in East and North-East India by using *MF-DXA*.
- Identifying all the possible climate zones in *Eastern and North-Eastern region* of India.
- Proposing an approach for detection of spatio-temporal patterns using multifractal analysis.
- Demonstrating the effectiveness of the fractal-based proposed data mining approach with respect to the *Köppen-Geiger climate classification*.

The rest of the paper is organized as follows: Section II provides a brief description of the theoretical background behind the work. The proposed data mining approach using multifractal analysis has been illustrated in section III. Section IV reports the experimental results and analysis. Finally, the conclusions have been drawn in section V.

II. Background

A. Different Climate Zones over the World

Climate zones are the areas with distinct climate patterns which generally occur in East-West direction around the world, and can be classified using different climatic variables characterizing the zone. One of the most widely used climate classification systems is the Köppen climate classification which was first proposed by Russian German climatologist Wladimir Köppen [8], and later modified to the Köppen-Geiger climate classification system [9]. The system is based on the *vegetation distribution* over the globe.

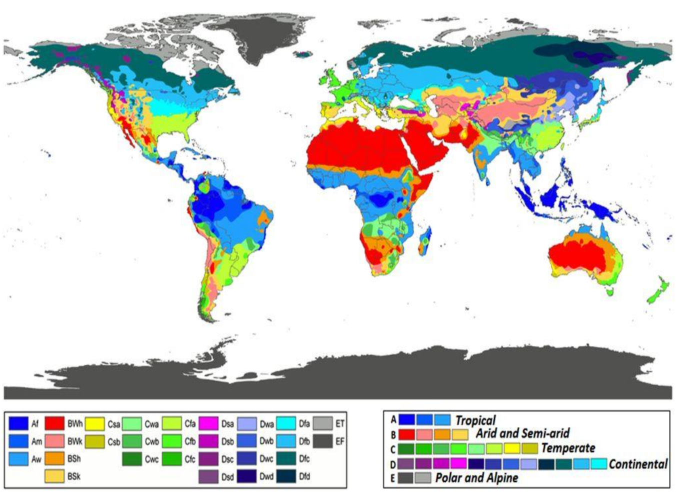

Fig. 1. World Map of Köppen-Geiger climate classification [8]

[1]Data Source: *FetchClimate Explore* of *Microsoft Research* [7]

According to the original Köppen Classification, there are five major climate zones: A (*Tropical*), B (*Arid and Semi-arid*), C (*Temperate*), D (*Continental*), and E (*Polar and Alpine*), which are further subdivided into several sub-zones. Each location is assigned to a specific zone based on *temperature* and *precipitation patterns*. Fig. 1 shows a typical *world map* of the *Köppen-Geiger climate classification*. All 31 climate classes have been illustrated with different colors and corresponding codes. However, two of these classes do not occur anywhere in the global map [9].

B. Multifractal Detrended Cross-Correlation Analysis (MF-DXA)

The *MF-DXA* method was proposed by Zhou et al. [6] to quantify long-range cross-correlations between two non-stationary time series. This method is a multifractal modification of the *detrended cross-correlation analysis (DXA)* [10]. Calculations involved in various stages of the *MF-DXA* method are as follows: Let u_i and v_i be two time series of length L with zero mean.

Stage-1: Cover both the series with $L_l = \lfloor L/l \rfloor$ number of non-overlapping boxes each with side length l.

Stage-2: Now, Estimate the "profile" of both the series within the b-th box as follows:

$$U_b(k) = \sum_{j=1}^{k} u([b-1]l+j) \quad ,and \quad V_b(k) = \sum_{j=1}^{k} [v([b-1]l+j)] \quad (1)$$

where, $k = 1, \cdots, l$.

Stage-3: Detrend the profiles in b-th box for each of the series using the respective local trends $\tilde{U}_b(k)$ and $\tilde{V}_b(k)$ respectively.

Stage-4: Calculate the detrended covariance of each box as follows:

$$F_b(l) = \frac{1}{l} \sum_{k=1}^{l} [U_b(k) - \tilde{U}_b(k)][V_b(k) - \tilde{V}_b(k)] \quad (2)$$

Stage-5: Determine the q-th order detrended covariance:

$$F_{uv}(q,l) = \left[\frac{1}{L_l} \sum_{b=1}^{L_l} F_b(l)^{q/2} \right]^{1/q} \quad (3)$$

where, $q \neq 0$, and

$$F_{uv}(0,l) = \exp \left[\frac{1}{2L_l} \sum_{b=1}^{L_l} \log_e F_b(l) \right] \quad (4)$$

Stage-6: Estimate the scaling-behavior of the detrended covariance by analyzing log-log plots of $F_{uv}(q,l)$ vs. l for each q-value. If the series u_i and v_i are long-range power-law correlated, $F_{uv}(q,l)$ increases with large values of l as a power-law,

$$F_{uv}(q,l) \sim l^{h_{uv}(q)} \quad (5)$$

Where, $h_{uv}(q)$ is the multifractal scaling exponent. Long-range cross-correlations between two series imply that each series has long memory of its own previous values and additionally has a long memory of the previous values of the other series. If the large and small covariance scale differently, then only

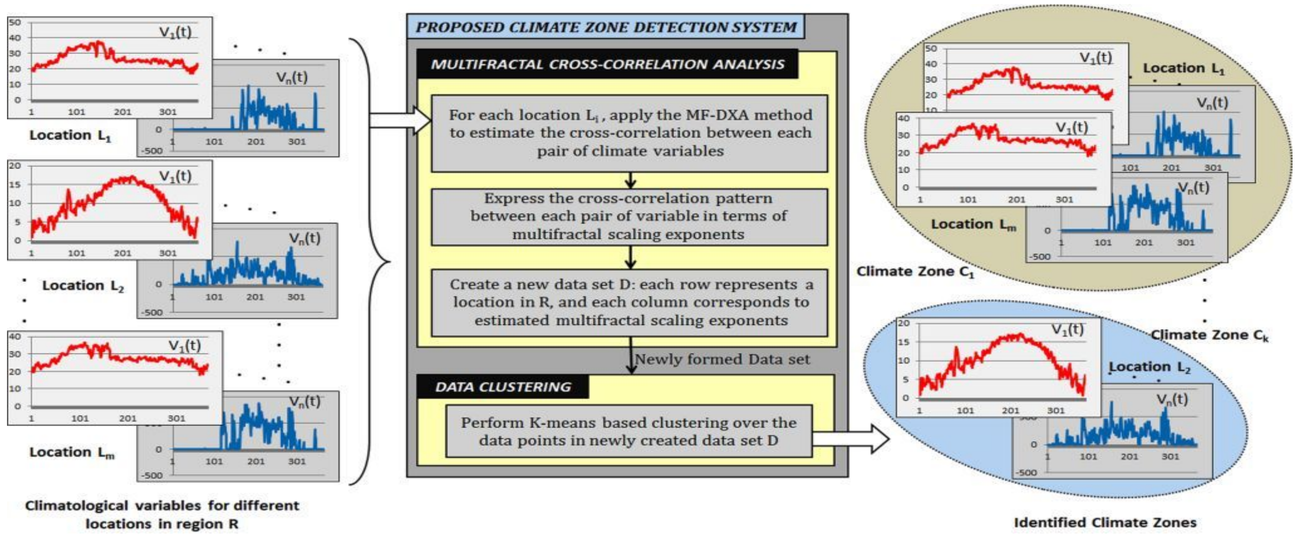

Fig. 2. Block diagram of the proposed approach: Climate zone detection using multifractal analysis

there will be a significant dependence of $h_{uv}(q)$ on q which characterizes a multifractal data series.

This paper presents a new data mining approach based on *MF-DXA* to detect the various climate zones in a large region on the basis of similarity in spatio-temporal patterns among different locations in the region. The approach has been illustrated in the following section.

III. Proposed Data-Mining Approach based on Multifractal Analysis

Fig. 2 shows a basic block diagram of the proposed approach for detection of climate zones of a large area R by mining the associated spatio-temporal data. The system takes as input the daily time series data $v_1(t), v_2(t), \cdots, v_n(t)$ of different climate variables of interest: $v_1, v_2, \cdots, v_n$, and finally generates all the best possible climate zones $c_1, c_2, \cdots, c_k$ in terms of cluster pattern. Each cluster contains a subset of locations $\{l_1, l_2, \cdots, l_m\}$, belonging to the region R. As depicted in the figure, the entire system encompasses two major steps, namely, (a) *Multifractal Cross-correlation Analysis*, and (b) *Data Clustering*.

A. Multifractal Cross-correlation Analysis

This step helps to capture the long-range correlation pattern between each pair of climate variables v_i and v_j over different time scale. In general, correlations indicate a predictive relationship between two or more random variables which may show spatial variation depending on the measurement locations of the variables. Therefore, the cross-correlations can play a vital role to characterize the climate pattern of any particular zone.

Since most of the real world time series data are fractal or multifractal in nature, the correlation pattern between any two climate variables is supposed to have fractal or multifractal properties over different lengths of time scale. Therefore, the main objective here is to analyze the fractal characteristics of the correlation pattern between each v_i and v_j, and find out the multifractal scaling exponents considering different positive and negative order of covariance. For this purpose we exploit the *MF-DXA* method as described in section II. The q-th order detrended covariance between v_i and v_j over a time scale length of l, is calculated using equation 3 and 4 as follows:

$$F_{v_iv_j}(q,l) = \left[\frac{1}{S}\sum_{s=1}^{S} F_s(l)^{q/2}\right]^{1/q} \qquad (6)$$

when $q \neq 0$ and

$$F_{v_iv_j}(0,l) = \exp\left[\frac{1}{2S}\sum_{s=1}^{S} \log_e F_s(l)\right] \qquad (7)$$

where, S is the total number of l-length time segments with respect to the daily time series data.

In case the correlation pattern between v_i and v_j is multifractal, the *log vs. log* plots of q-th order covariance $F_{v_iv_j}(q,l)$ vs. time-scale length l show different slope for different values of q. The set of these slopes is termed as multifractal scaling exponent $h_{v_iv_j}(q)$. The values of $h_{v_iv_j}(q)$ are estimated from the log-log plot of $F_{v_iv_j}(q,l)$ vs. l using equation 5 as follows:

$$h_{v_iv_j}(q) = \frac{\log\left[F_{v_iv_j}(q,l)\right]}{\log l} + C \quad , \qquad (8)$$

where, C is a constant term.

$h_{v_iv_j}(q)$ shows a significant dependence of on q only if the large and small fluctuations scale differently, indicating the multifractal characteristics of the cross-correlation between the considered time series.

In the same fashion we determine the set $h_{v_iv_j}(q)$ for each pair of v_i and v_j in $\{v_1, v_2, \cdots, v_n\}$. Now, on the basis of these estimated multifractal scaling-exponent sets, a new database is created where each tuple/row corresponds to a particular location in the region of interest R, and each of the field/column becomes the scaling exponents, obtained from multifractal cross-correlation analysis between each pair of climate variables considering different positive and negative orders of q. For example, if L_1, L_2, L_3, L_4, and L_5 are five locations in R; X, Y, and Z are three variables of interests;

and q_1, q_2, and q_3 are different orders considered for q, then the new database becomes as shown in TABLE I., where, $h_{XY}^{L_i}(q_1)$

TABLE I. NEWLY FORMED DATA SET WITH MULTIFRACTAL CROSS-CORRELATION INFORMATION

Lonations	$Attr_1$	$Attr_2$	$Attr_3$	$Attr_4$	$\cdots$	$Attr_9$
L_1	$h_{XY}^{L1}(q_1)$	$h_{XY}^{L1}(q_2)$	$h_{XY}^{L1}(q_3)$	$h_{YZ}^{L1}(q_1)$	$\cdots$	$h_{ZX}^{L1}(q_3)$
L_2	$h_{XY}^{L2}(q_1)$	$h_{XY}^{L2}(q_2)$	$h_{XY}^{L2}(q_3)$	$h_{YZ}^{L2}(q_1)$	$\cdots$	$h_{ZX}^{L2}(q_3)$
L_3	$h_{XY}^{L3}(q_1)$	$h_{XY}^{L3}(q_2)$	$h_{XY}^{L3}(q_3)$	$h_{YZ}^{L3}(q_1)$	$\cdots$	$h_{ZX}^{L3}(q_3)$
L_4	$h_{XY}^{L4}(q_1)$	$h_{XY}^{L4}(q_2)$	$h_{XY}^{L4}(q_3)$	$h_{YZ}^{L4}(q_1)$	$\cdots$	$h_{ZX}^{L4}(q_3)$
L_5	$h_{XY}^{L5}(q_1)$	$h_{XY}^{L5}(q_2)$	$h_{XY}^{L5}(q_3)$	$h_{YZ}^{L5}(q_1)$	$\cdots$	$h_{ZX}^{L5}(q_3)$

indicates the q_1-th order scaling exponent in the multifractal correlation analysis between climate variables X and Y at the i-th location L_i. This newly formed data set is now fed to the next step for the required cluster analysis.

B. Data Clustering

As shown in the Fig.2, this step takes as input a new data set, created based on the scaling exponents obtained through the multifractal cross-correlation analysis between each pair of climate variables in the first step. Now we apply the *K-means* clustering algorithm over this new set of data. The clustering analysis is performed based on the fact that any two locations l_i and l_j will be in the same climatic zone if the correlation pattern among the same set of climate variables in both the location shows a high degree of similarity.

While applying the *K-means* algorithm, our approach assumes that the K-value, i.e. the number of climate zones, in known a priori, and the initial data points are selected using random method. Since the data set is already built with the pairwise correlation information for all the variables, we have chosen the Euclidean distance as the similarity measure between the data points.

The output of this data clustering step are $K = k$ number of groups $(c_1, c_2, \cdots, c_k)$ of locations, where all the locations in the same group show the same spatio-temporal pattern in terms of time series data for $v_1, v_2, \cdots, v_n$. The different steps in the entire data mining approach are shown in Algorithm 1.

Algorithm 1 : Detection of various climate zones for a region R

Input: Daily time series data of the considered climate variables $v_1, v_2, \cdots, v_n$, for each of the location $l_1, l_2, \cdots, l_m$ in R; where n is the total number of variables of interest and m is the total number locations considered.
Output: k number of climate zones: $c_1, c_2, \cdots, c_k$ of R, where each zone contains a subset of locations.

V=Set of all considered climate variables=$\{v_1, v_2, \cdots, v_n\}$.
L=Set of all considered locations in R=$\{l_1, l_2, \cdots, l_m\}$.
$h_{v_j v_r}^{l_i}(q)$=q-th order scaling exponent in the multifractal correlation analysis between climate variables v_j and v_r at the i-th location l_i.
p= Total number of q orders considered.
D=An initially empty data set

for each location l_i in L **do**
  for each pair of variables v_j, v_r $(j \neq r)$ in V **do**
    Apply *MF-DXA* method to calculate the q-th ($q \in \{q_1, q_2, \cdots, q_p\}$) order detrended covariance $F_{v_i v_j}(q, l)$ between v_j and v_r over different time scale length l;
    if $q \neq 0$ **then**
      $F_{v_j v_r}(q, l) = \left[\frac{1}{S_l}\sum_{s=1}^{S_l} F_s(l)^{q/2}\right]^{1/q}$
    else
      $F_{v_j v_r}(0, l) = \exp\left[\frac{1}{2S_l}\sum_{s=1}^{S_l} \log_e F_s(l)\right]$
    end if
    /* S_l= Total number of time segment of length l in the annual extent */
    Calculate the multifractal scaling exponent $h_{v_j v_r}^{l_i}(q)$ from the log-log plot of $F_{v_j v_r}(q, l)$ vs. l:
    $h_{v_j v_r}^{l_i}(q) = \frac{\log [F_{v_j v_r}(q,l)]}{\log l} + C$,C is a constant
  end for
  Add the l_i-th tuple $< h_{v_1 v_2}^{l_i}(q_1), h_{v_1 v_2}^{l_i}(q_2), \cdots, h_{v_n v_1}^{l_i}(q_p) >$ in D
end for
/* Apply *K-means* algorithm to cluster the data points in current D*/
while there is no change in cluster centroid (or the change is negligible) **do**
  Select $K = k$ data points from the newly generated data set D as the initial centroids.
  Assign all the other data points to their nearest centroid.
  Re-compute the centroid of each current cluster.
end while
Return the k number of clusters $c_1, c_2, \cdots, c_k$ as the detected climate zones for the region R.

IV. Results

In this section we have evaluated our proposed approach of climate zone detection, in comparison to the generic climate zone classification model of the world. The section starts with a brief description of the working data set and the implementation specifications. Rest of the section discusses about the ultimate results of climate zone detection along with different performance measures.

A. Data

The experiment has been carried out with the daily time series data of the year 2013 for two major climate variables: *land surface temperature* and *precipitation rate*, collected from the *Eastern and North-Eastern* region of *India*. As depicted in Fig. 3(a), this region of India consists of 12 states, namely, *Bihar*, *Jharkhand*, *Orissa*, *West Bengal*, *Sikkim*, *Assam*, *Arunachal Pradesh*, *Meghalaya*, *Tripura*, *Manipur*, *Nagaland*, and *Mizoram*. The spatio-temporal data has been collected from 73 different locations over all these states as pointed in Fig. 3(b). All the data has been obtained from the *FetchClimate Explorer* site [7] of Microsoft Research. TABLE II lists all these twelve states along with the number of sample locations, randomly chosen from each of the states based on the state-area.

TABLE II. NUMBER OF SAMPLE POINTS CHOSEN FROM EACH STATE IN THE STUDY REGION (EASTERN AND NORTH-EASTERN INDIA)

States	Area (sq. km)	Number of sample location chosen
Arunachal Pradesh	83,743	8
Assam	78,438	7
Bihar	94,163	7
Jharkhand	79,714	8
Manipur	22,327	5
Meghalaya	22,429	5
Mizoram	21,081	4
Nagaland	16,579	5
Orissa	155,707	8
Sikkim	7,096	4
Tripura	10, 486	4
West Bengal	88,752	8

B. Experimental Setup

The entire experiment has been done using MATLAB 7.12.0 (R2011a) in Windows 2007 (32-bit Operating System, 2.40 GHz CPU, 2.00 GB RAM), and R-tool version 3.1.1 (32 bit). The multifractal cross-correlation analysis in the first

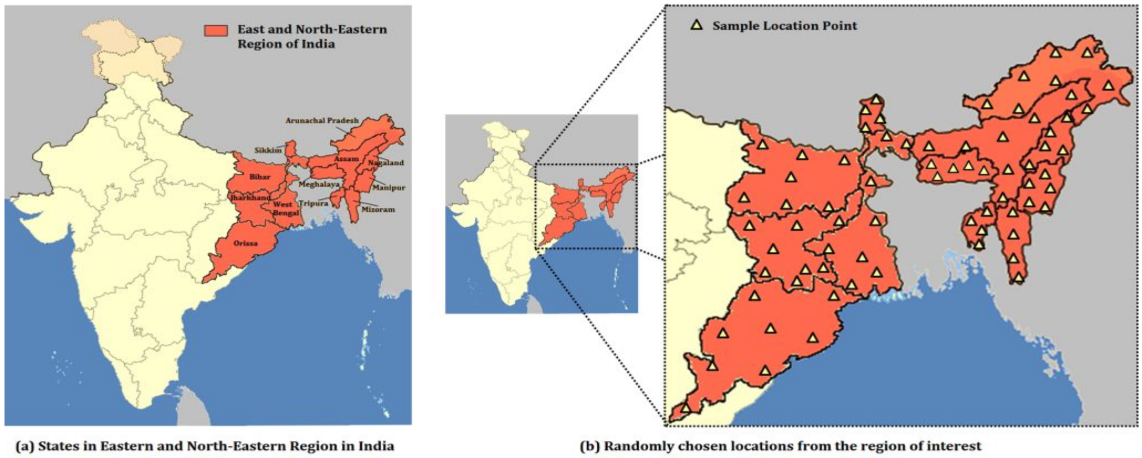

Fig. 3. The region of Case Study: *Eastern and North-Eastern* India

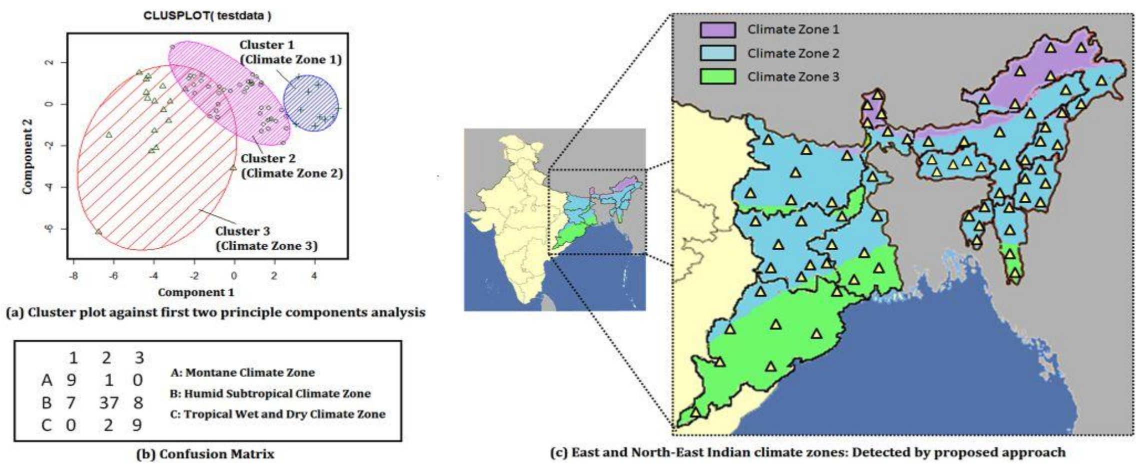

Fig. 4. Results of proposed data-mining technique to detect climate zones (Study region: *Easter and North-Eastern* India)

step of our approach has been performed in the MATLAB environment. Different time scale of size l (in terms of days), and different values of q-orders have been used for the experimental purpose. The R-tool has been used for the cluster analysis stage in the proposed mining approach.

During the correlation analysis step, the log-log plot of q-th order detrended covariance ($F_{uv}(q,l)$) vs. scale size (l) has been studied for identifying the pattern of multifractal correlation between the time series of each pair of climate variable u and v. This multifractal study reveals that the multifractal nature of the correlation pattern changes with the location.

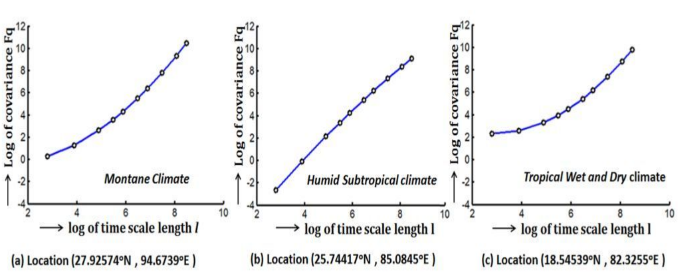

Fig. 5. Spatial variation of the correlation pattern (log-log plot for a single q-value)

It may be observed from Fig. 5 that different climate zones have different nature of plots. To capture the multifractal pattern associated with the location l, we have utilized the set of scaling exponents $h^l_{uv}(q)$ which are estimated as the slope of the log-log plot for each values of q. Each location l is now represented by its corresponding $h^l_{uv}(q)$ values. These new information are then used in the cluster analysis step.

In the clustering step we assume that the number of climate zones is known a priori. According to the generic classification model of *Köppen-Geiger* [8], our study region (Eastern and North-Eastern part of India) composed of three major climatic zones: *Montane*, *Tropical wet and dry*, and *Humid Sub-tropical*. We therefore set $K = 3$, during the *K-means* based cluster analysis. Fig. 4 shows the final output of the proposed approach in terms of a 2D cluster plot in Fig. 4(a), confusion matrix in Fig. 4(b), and in the form of map regions in Fig. 4(c) respectively.

C. Performance Measures

Once the clustering step is over, the best possible set of climate zones, obtained as cluster output, are mapped to the world map of *Köppen-Geiger climate classification* [8] (Fig. 6). This helps in performing a comparative study between the two, with respect to several performance measures.

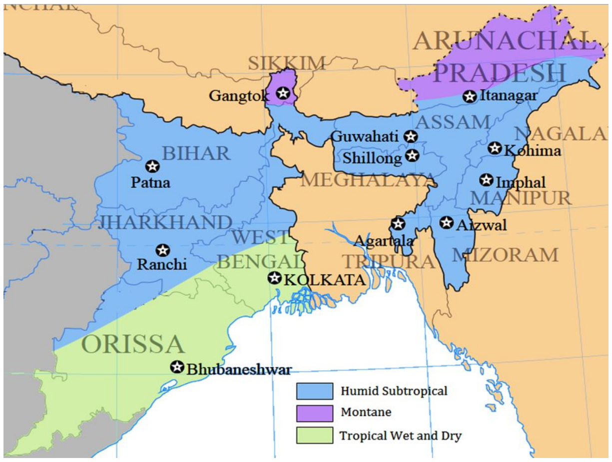

Fig. 6. East and North-East Indian climate zones (based on *Köppen-Geiger climate classification* [8])

Considering the *Köppen-Geiger climate classification* to be the benchmark, three most popular performance measures (*Rand measure*, *F measure*, and *Fowlkes-Mallows index*) have been used in TABLE III to externally evaluate our proposed approach. The values of *true positives (TP)*, *true negatives (TN)*, *false positives (FP)*, and *false negatives (FN)*, have been calculated using the obtained confusion matrix (Fig.4(b)).

TABLE III. EXTERNAL EVALUATION OF THE PROPOSED APPROACH

Class	Rand measure (RI): $\frac{TP+TN}{TP+FP+TN+FN}$	F measure (F_1): $\frac{2.P.R}{P+R}$	Fowlkes-Mallows index (FM): $\sqrt{P.R}$
Montane	0.89	0.70	0.71
Humid-sub-tropical	0.75	0.80	0.81
Tropical-wet and dry	0.86	0.64	0.66

In TABLE III, P is the *precision rate* $=\left(\frac{TP}{TP+FP}\right)$, and R is the *recall rate* $=\left(\frac{TP}{TP+FN}\right)$.

D. Discussions

By analyzing the different outcomes (as depicted in Fig. 4 and in TABLE III) , the following inferences can be drawn:

- From the Fig. 4(c) and Fig. 6, it is evident that the proposed approach has identified all the three different climate zones properly. In Fig.4(c), the indentified climate zone-1 is basically the *Montane* type of climate zone according to the Köppen-Geiger climate classification (Fig. 6). Similarly, the climate zone-2, and climate zone-3 are equivalent to *Humid sub-tropical* and *Tropical wet and dry* climate zone respectively. The accuracy of identification is clearly visible from the TABLE III entries.

- It can also be noted from the Fig. 4(c), Fig. 6, and the confusion matrix in Fig. 4(b) that, some alteration have been introduced in the output of our proposed technique. As expressed through the confusion matrix, some of the humid sub-tropical zones have been identified as the tropical wet and dry zone (e.g. Southern and South-Western part of Bihar, North-Eastern part of Jharkhand, Southern part of Mizoram etc.). These indicate a recent climate pattern change in those regions. The *5-th assessment report* of IPCC [11] supports our outcomes, and demonstrates the efficacy of the proposed approach.

From the above analysis, it is evident that the consideration of multifractal characteristics of cross-correlation between each pair of climate variables, has made the proposed data mining approach able to efficiently capture the spatio-temporal pattern of any location, and finally helps to accurately identify the various climate zones over a large region.

V. CONCLUSIONS

This work presents a *MF-DXA*-based data mining approach, to detect the various climate zones in a large region, on the basis of similarity in spatio-temporal pattern among different locations within the region. It has utilized the multifractal cross-correlations between each pair of climate variables captured for different length of time scales. A case study has been performed with the *temperature* and *precipitation* data of 73 locations in *Eastern and North-Eastern region* of *India* as collected from *FetchClimate Explorer* of *Microsoft Research*. The high resemblance of the identified zones with the *World Map* of *Köppen-Geiger climate classification*, establishes the efficacy of the proposed data mining technique using multifractal analysis. In future, the work can be extended to deal with unknown number of climate zones as well.

REFERENCES

[1] V. Kumar, M. Steinbach, P.-N. Tan, S. Klooster, C. Potter, and A. Torregrosa, "Mining scientific data: Discovery of patterns in the global climate system," in *Joint Statistical Meeting*, 2001.

[2] Y. Unal, T. Kindap, and M. Karaca, "Redefining the climate zones of turkey using cluster analysis," *International Journal of Climatology*, vol. 23, no. 9, pp. 1045–1055, 2003.

[3] I. Mahlstein and R. Knutti, "Regional climate change patterns identified by cluster analysis," *Climate dynamics*, vol. 35, no. 4, pp. 587–600, 2010.

[4] P. Zhang, M. Steinbach, V. Kumar, S. Shekhar, P. Tan, S. Klooster, and C. Potter, "Discovery of patterns of earth science data using data mining," *New generation of data mining applications*, 2005.

[5] E. Di Giuseppe, G. J. Lasinio, S. Esposito, and M. Pasqui, "Functional clustering for italian climate zones identification," *Theoretical and applied climatology*, vol. 114, no. 1-2, pp. 39–54, 2013.

[6] W.-X. Zhou, "Multifractal detrended cross-correlation analysis for two nonstationary signals," *Physical Review E*, vol. 77, no. 6, p. 066211, 2008.

[7] Microsoft-Research, "FetchClimate," http://research.microsoft.com/en-us/um/cambridge/projects/fetchclimate, 2013, [Online; July-2014].

[8] Wikipedia, "Köppen climate classification," http://en.wikipedia.org/wiki/Koppen_climate_classification, 2014, [Online; Accessed 27-July-2014].

[9] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the köppen-geiger climate classification updated," *Meteorologische Zeitschrift*, vol. 15, no. 3, pp. 259–263, 2006.

[10] B. Podobnik and H. E. Stanley, "Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series," *Physical review letters*, vol. 100, no. 8, p. 084102, 2008.

[11] IPCC, "Climate Change 2013: The Physical Science Basis," http://www.ipcc.ch/report/ar5/wg1/docs/, 2013, [Online; Accessed 18-Aug-2014].