

Research Article

Recurrence Based Similarity Identification of Climate Data

Anita Bai,¹ Swati Hira,² and S. Deshpande Parag¹

¹Department of Computer Science & Engineering, VNIT, Nagpur, India

²Department of Computer Science & Engineering, RCOEM, Nagpur, India

Correspondence should be addressed to Swati Hira; hiras@rknec.edu

Received 2 December 2016; Revised 23 April 2017; Accepted 15 May 2017; Published 19 July 2017

Academic Editor: Alicia Cordero

Copyright © 2017 Anita Bai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Climate change has become a challenging and emerging research problem in many research related areas. One of the key parameters in analyzing climate change is to analyze temperature variations in different regions. The temperature variation in a region is periodic within the interval. Temperature variations, though periodic in nature, may vary from one region to another and such variations are mainly dependent on the location and altitude of the region and also on other factors like the nearness of sea and vegetation. In this paper, we analyze such periodic variations using recurrence plot (RP), cross recurrence plot (CRP), recurrence rate (RR), and correlation of probability of recurrence (CPR) methods to find similarities of periodic variations between and within climatic regions and to identify their connectivity trend. First, we test the correctness of our method by applying it on voice and heart rate data and then experimentation is performed on synthetic climate data of nine regions in the United States and eight regions in China. Finally, the accuracy of our approach is validated on both real and synthetic datasets and demonstrated using ANOVA, Kruskal–Wallis, and z -statistics significance tests.

1. Introduction

No location on the earth will have exactly the same climate as another; many do have very similar climatic characteristics, which depend on various factors such as latitude and longitude of the region, temperature, humidity, air pressure, wind, cloudiness, and nearness of sea and vegetation. Temperature is one of the important factors in climate change. The temperature of a region affects humans and biological and physical systems in all continents [1]. Currently, temperature is rising all over the world because greenhouse gases are trapping more heat in the earth's atmosphere. Effective and urgent solutions are needed to identify its impact on agriculture, energy, water supplies, health, plants, animals, ecosystems, forests, recreation, and so forth. Decisions about temperature change are complex and costly and have long-term implications. It is therefore vital that such decisions are based on the best available evidence. We need to understand the quality and provenance of that evidence and to find whether any assumptions have been made in generating it. Understanding temperature change patterns and their periodic variations across time (such as yearly, monthly, and

daily changes) and their changes across environmental space is of great significance.

Climate change adjustment refers to dealing with the present or future expected impacts of climate change. There are various ongoing research efforts on the collection of climate data, the analysis of climate changes, and the modeling of climate processes. Finding correlations or similarities among climate data is one of the central themes of many scientific analyses. A good example is climate data analysis to understand temperature changes over wide ranges of time. It has many applications to agriculture [2, 3], fisheries, ecosystems, water resources, energy infrastructure, business [2, 4], food industry [2, 5, 6], and disaster planning. In addition, climate impacts have also been assessed in potato [7], maize [8], coffee [9], rice [10], sugar [11], wine grape production [12], and so forth. For example, climate change eventually increases the additional price for the agricultural crops such as rice, wheat, maize, and soya beans, which tends to cause a higher substantial fall in cereals consumption.

To identify such climate change pattern, we proposed a recurrence based approach to analyze temperature variations

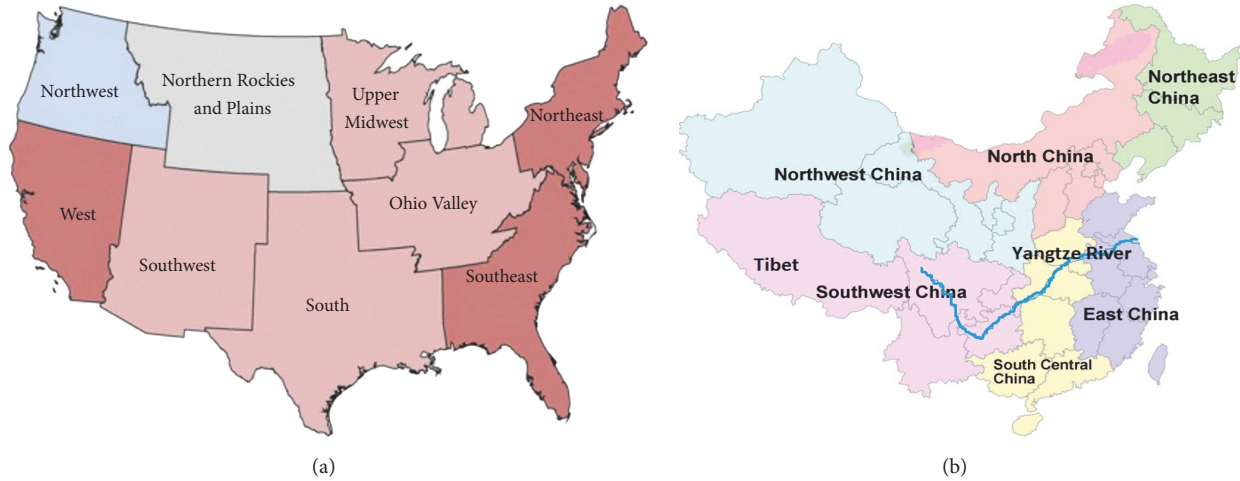


FIGURE 1: (a) Regions in the United States. (b) Regions in China.

in various climatic regions. The major contributions of our approach are summarized as follows:

- (1) Identify the periodic variations of temperature by analyzing trends using recurrence plot (RP) for time series data.
- (2) Discover the climate differences or similarities of the periodic variations existing between two regions using cross recurrence plot (CRP), for example, nine regions in the United States (Ohio Valley, Upper Midwest, Northeast, Northwest, South, Southeast, Southwest, West, and Northern Rockies and Plains) and eight regions in China (South China, the middle and lower reaches of the Yangtze River, North China, Northeast China, the east of Northwest China, the west of Northwest China, Tibet, and Southwest China), shown in Figures 1(a) and 1(b).
- (3) Extract similarities among regions by using techniques such as recurrence rate (RR), RP, CRP, and correlation of probability of recurrence (CPR) on temperature data.
- (4) Calculate the number of connections in each bin over time by binning CPR into three categories of relatedness (weak, moderate, and strong).

We have demonstrated the potentials of this approach for nine regions in the US, eight regions in China, and synthetic, voice, and heart rate data. In this paper, we analyzed nine connected regions in the US using monthly temperature data spread over 120 years. Using RP, we proved the periodic behavior of the variations. Using CRP, we provided the method to test whether two regions have similar variations. RR and CPR are used to show the correlation of probability of occurrence of temperature points between time series of two regions. Substantial experiments indicate that the proposed approach successfully provides useful interpretation of similar or dissimilar patterns between and within regions related to climate change.

2. Literature Review

Climate data (temperature) are usually multidimensional arrays of floating-point numbers. These arrays typically have one temporal dimension and two or three spatial dimensions, which describes the evolution of climate parameters in a time span. The volume of climate data is expanding exponentially day by day and it brings about some challenges for climate data archiving, sharing, and analyzing. A lot of research has been done to analyze financial, stock, economic, and other time series data using recurrence plots, but very few analyses are available for climate data based on recurrence analysis [13–15]. Climate data is analyzed using several other techniques. We are explaining some of them.

Sukharev et al. [16] presented a correlation analysis for time varying multivariate climate datasets. They used k -means clustering method and graph partitioning algorithm to find patterns and connections. The correlation of a single or a couple of variables is also analyzed using pointwise correlation coefficients and canonical correlation analysis. Liu et al. [17] proposed a lossless compression algorithm for the time-spatial climate floating-point arrays. They used adaptive prediction, XOR differencing, and multiway compression to eliminate more data redundancy efficiently and also tried to exploit the correlations among the multidimensions to remove more data redundancy. Sap and Awan [18] used kernel methods for unsupervised partitioning of data to find spatiotemporal patterns in complex and nonlinearly separable climate data. Hendrix et al. [19] described a methodology for capturing and identifying the estimation of a climate network. They performed this by splitting the climate data into a set of overlying decadal time intervals and creating a network for each of these datasets representing the complex interdependencies in the climate system over a particular decade.

RPs and RQA have been successfully used in a large number of scientific disciplines [20] and are particularly used for modeling financial and economic time series. In recent years, several researchers concentrated on RPs and RQA techniques to study deterministic dependencies in financial data. These

techniques are used in various fields such as stock market [21], exchange rates [22], electricity prices [23], and heart beat interval [24]. Furthermore, synchronicity and convergence are also examined among member nations of the Euro region for GDP using cross recurrence analysis [25]. Silva et al. [26] gave an overview of recurrence plots as a representation domain for time series classification, in which Campana-Keogh (CK-1) and Kolmogorov complexity based distances are used to measure the closeness between recurrence plots and to estimate image similarity, respectively.

From the above literature we observed that various analyses are performed using recurrence plots to find hidden data relationships in a sequence of time series datasets, such as stocks, exchange rates, financial data, heart rates, voice, and electricity processes, but very little research has been done to find and visualize the interrelationship in temperature periodic trends for climate data using RP, CRP, and CPR. At the same time, we also observed that no significant research has been done to indicate the difference between and within the temperatures of various climatic regions. So, we decided to use recurrence based methods RR, CRP, and CPR to identify the differences in climate on the basis of temperature for nine US and eight Chinese regions and also the probabilistic correlations between time series. The accuracy of the recurrence based approach is validated on real and synthetic datasets and analyzed using ANOVA, Kruskal–Wallis, and z -statistical significance tests [27, 28].

The rest of the paper is organized as follows. In Section 3, we discuss the brief introduction of terms used in our approach. Section 4 represents the proposed approach. Section 5 shows the experimentation results on various datasets and the validation of the proposed method by applying the significance test. Section 6 concludes the work done in this paper.

3. Materials and Methods

3.1. Recurrence Plots (RPs). Recurrence plots are used to analyze periodic data by visualizing the recurrent behavior of dynamical systems which does not stay constant and changes periodically. It is also applicable to analyze the behavior of nonlinear dynamical and nonstationary systems, for example, temperature. It is used for the study of difference or similarity within a process on time series data. A recurrence plot (RP) is a visual tool that shows the recurrence patterns of a dynamical system [29]. Recurrence is defined as return of the trajectory of a system to a previous state. Recurrence occurs when the system returns to the neighborhood of an earlier point in the phase space. The distributions of recurrence points and diagonal lines along the main diagonal provide an evaluation of the similarity of the phase space trajectories of both dynamical systems.

A cross recurrence plot (CRP) is a tool for nonlinear data analysis, which can be used for the study of differences between dynamical systems. The basic idea of this approach is to compare the phase space trajectories of two tasks in the similar phase space. CRP can be used in order to study the similarity of two different phase space trajectories. On the other hand, the CRP discloses all possible times when the

phase space trajectory of the first system visits approximately a similar area in the phase space trajectory of the second system. The data length of both systems can differ in the CRP matrix which leads to a nonsquare matrix.

We use an extension of the method of recurrence plot to the method of cross recurrence plot, which compares the time dependent behavior of two processes, which are recorded in a time series. Here, we have two time series, each one represented by trajectories X_i and Y_j in the phase space. The test for equality of each point of trajectory X_i with each point of trajectory Y_j by taking the embedding dimension m and delay time τ results in an $N_X * N_Y$ array.

$$CR_{i,j} = \Theta(\varepsilon_i - \|X_i - Y_j\|), \quad X_i, Y_j \in R^m, \quad (1)$$

where $X_i = \{x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\}$ and $Y_j = \{y_j, y_{j+\tau}, \dots, y_{j+(m-1)\tau}\}$. i is $1 \dots N_X$, j is $1 \dots N_Y$, N is the number of points, ε is threshold distance, $\Theta(\cdot)$ is the Heaviside function (i.e., $\Theta(a) = 0$ if $a < 0$ and 1 if $a \geq 0$), and $\|\cdot\|$ is a norm. CR is a matrix of $\{0s, 1s\}$ and an RP is a visual representation of CR obtained by marking a black and white dot for every 1 and 0. Embedding dimension m and delay time τ can be obtained by correlation dimension and autocorrelation function. In this paper, we select the fixed embedding dimension $m = 2$ and delay time $\tau = 4$.

Next, we give some measures to analyze RP and CRP using the sine wave. In order to present the idea of RP and CRP, some figures of the sine wave are presented to guide the description. The sine wave is a geometrical waveform which oscillates (moves up or down) periodically (i.e., the same pattern occurs after a particular time interval). The functionality of the sine is used to build models for processes that repeat in cycles or involve oscillations. Examples that present oscillations include the monthly and seasonal cycles of temperature, heart beats, voice, music, population cycles, and tides.

So, we can say that the sine wave follows the periodic pattern and distinct patterns emerge in its RP. If the data is collected from systems having periodic variations, then a distinct pattern can be seen in its RP; for example, data of climate, voice, and heart rate show a distinct pattern in RP. Also, the distance between diagonals indicates the signal periodicity. Therefore, we can visualize and study the motion of the dynamical system and infer some characteristics that generated the time series. To show the periodic variations clearly, first we explain the RP and CRP on the sine wave, based on which further other datasets are considered. Figure 2 shows the plot of different sine waves and their RPs. We have generated different sine waves in MATLAB using the $\sin()$ function. In Figure 2, we can visualize that the same pattern emerges in RP if series are periodic in nature.

If two time series are periodic and exhibit similarity, then the same pattern emerges in CRP also. CRP of sine waves is shown in Figure 3. CRP discovers all possible points when one sine wave (simple sine wave) visits approximately a similar area of the other sine wave (phase shift). We can see that, in Figure 3(a), clearer patterns emerged because both sine waves are almost the same (i.e., same frequency cycle, but differing by 0.8 phase shift). Even in Figure 3(b),

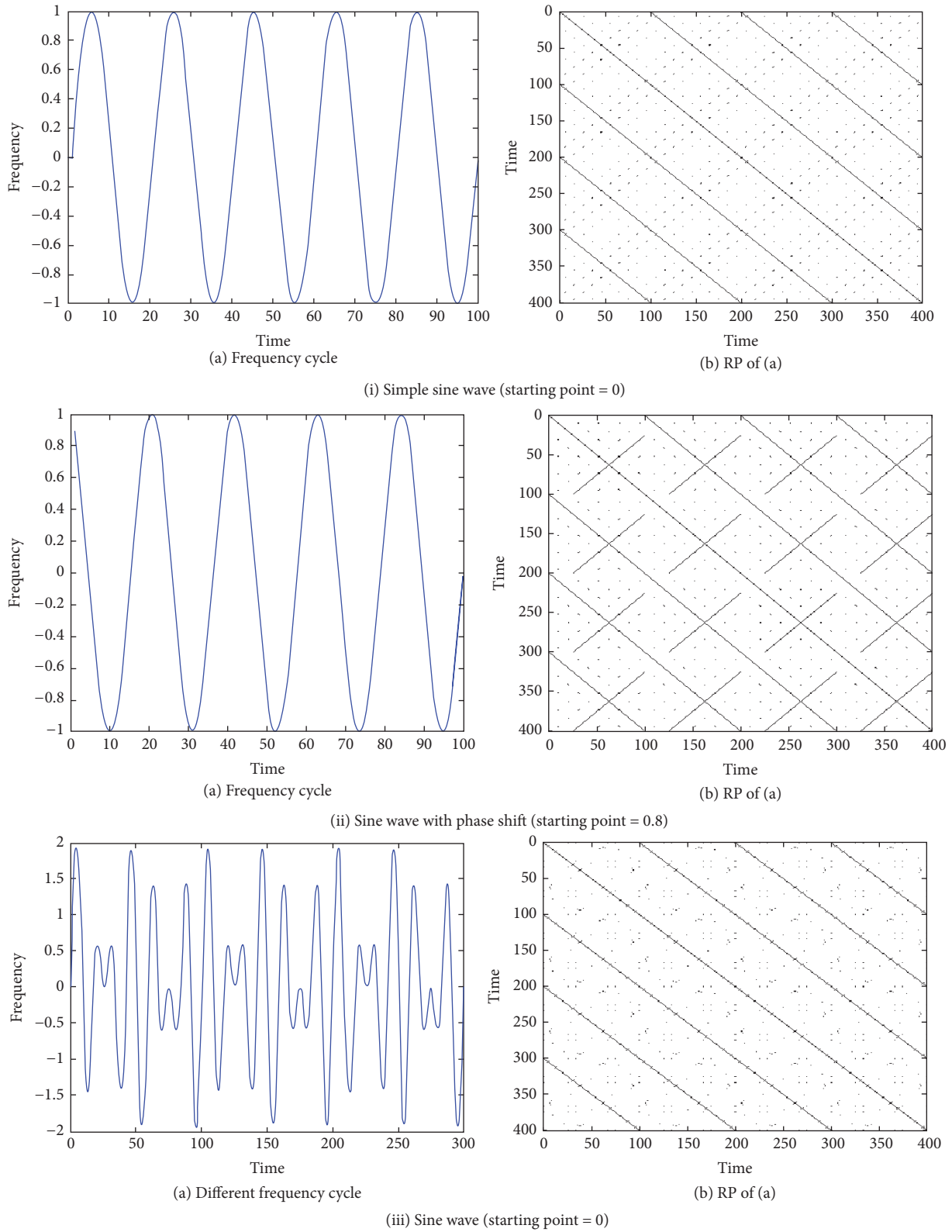


FIGURE 2: Recurrence plots for sine waves.

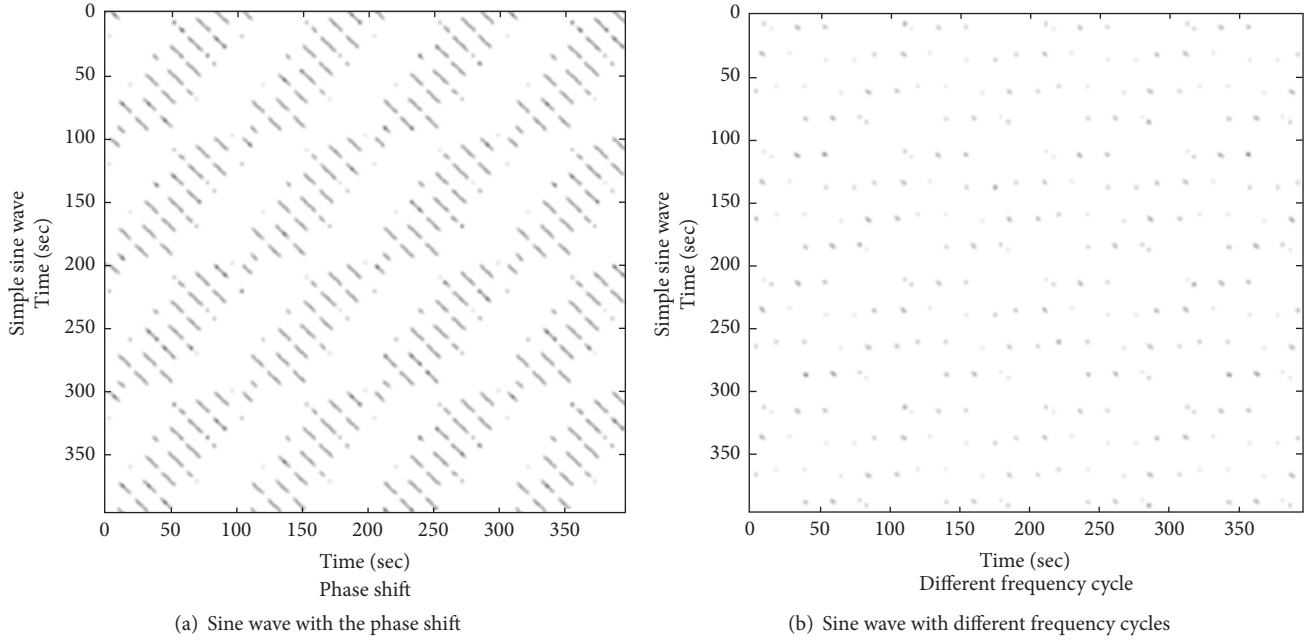


FIGURE 3: CRP for sine wave.

fewer patterns emerged because of their different frequency cycles. So, we can conclude that individual series can follow recurrent behavior, but when they are plotted together, that recurring pattern will not necessarily be followed. Figure 4 shows recurrence rate (RR) graph, described in the next section for all three sine waves shown in Figure 2.

3.2. Probability of Recurrence (τ -Recurrence Rate). The probability of recurrence $p(\tau)$ is the recurrence rate of a diagonal line situated at τ steps from the main diagonal, that is, $CR_{i,j+\tau} \forall i, j = 1, \dots, N - \tau$. This evaluation gives the probability of a $(j + \tau)$ th point falling in the ε -neighborhood of the i th point.

$$RR = p(\tau) = \frac{1}{N - \tau} \sum_{i,j=1}^{N-\tau} CR_{i,j+\tau}. \quad (2)$$

Figure 5(a) represents RP for northeast climate time series and shows a similar probability of occurrence to a prior state for all values of τ in Figure 5(b).

3.3. Correlation of Probability of Recurrence (CPR). CPR is based on RPs and was originally devised to quantify phase synchronization between two systems or stationary time series. It represents the probability of recurrence of the first system and second system. CPR describes the cross-correlation coefficient between the probabilities of recurrence of two trajectories \vec{x} and \vec{y} [20, 30]:

$$CPR = \langle \bar{p}_{\vec{x}}(\tau) \bar{p}_{\vec{y}}(\tau) \rangle, \quad (3)$$

where $\langle \cdot \rangle$ represents the expectation value. All the $p(\tau)$ curves in Figure 5(b) start from $p(0) = 1$, because the calculated recurrence rate always occurs 1 at $\tau = 0$, the main

diagonal. $CPR \approx 1$ implies that two time series variations are periodically synchronized. To predict CPR correctly, we consider an appropriate $p(\tau)$ value where τ is greater than autocorrelation time of the system because mostly a high CPR value is predicted for all trajectories having a similar initial portion of the $p(\tau)$ curve.

$$CPR = \langle \bar{p}_{\vec{x}}(\tau > \tau_s) \bar{p}_{\vec{y}}(\tau > \tau_s) \rangle, \quad (4)$$

where

$$\tau_s = \max(\tau_s(\vec{x}), \tau_s(\vec{y})). \quad (5)$$

Figure 6 illustrates the process involved in evaluating the CPR. The CPR between Ohio Valley and Southeast is calculated using (4) and (5), which is 0.785. In this study, this indicates that two climate time series (Ohio Valley and Southeast) with a high CPR tend to recur at similar times, suggesting some similarity (strong connectivity; refer to Figure 15(a)) in their trends.

3.4. ANOVA. Analysis of variance [28] is used to analyze differences among the group mean values and can identify the significant difference between group means if it exists. ANOVA's F -statistic is calculated as follows.

(i) The variation between the groups is calculated as

Between sum of squares (BS)

$$= m1 (\bar{Z}_1 - \bar{Z})^2 + m2 (\bar{Z}_2 - \bar{Z})^2 + \dots + mi (\bar{Z}_i - \bar{Z})^2, \quad (6)$$

Between mean squares (BM) = $\frac{BS}{\text{degree of freedom}}$.

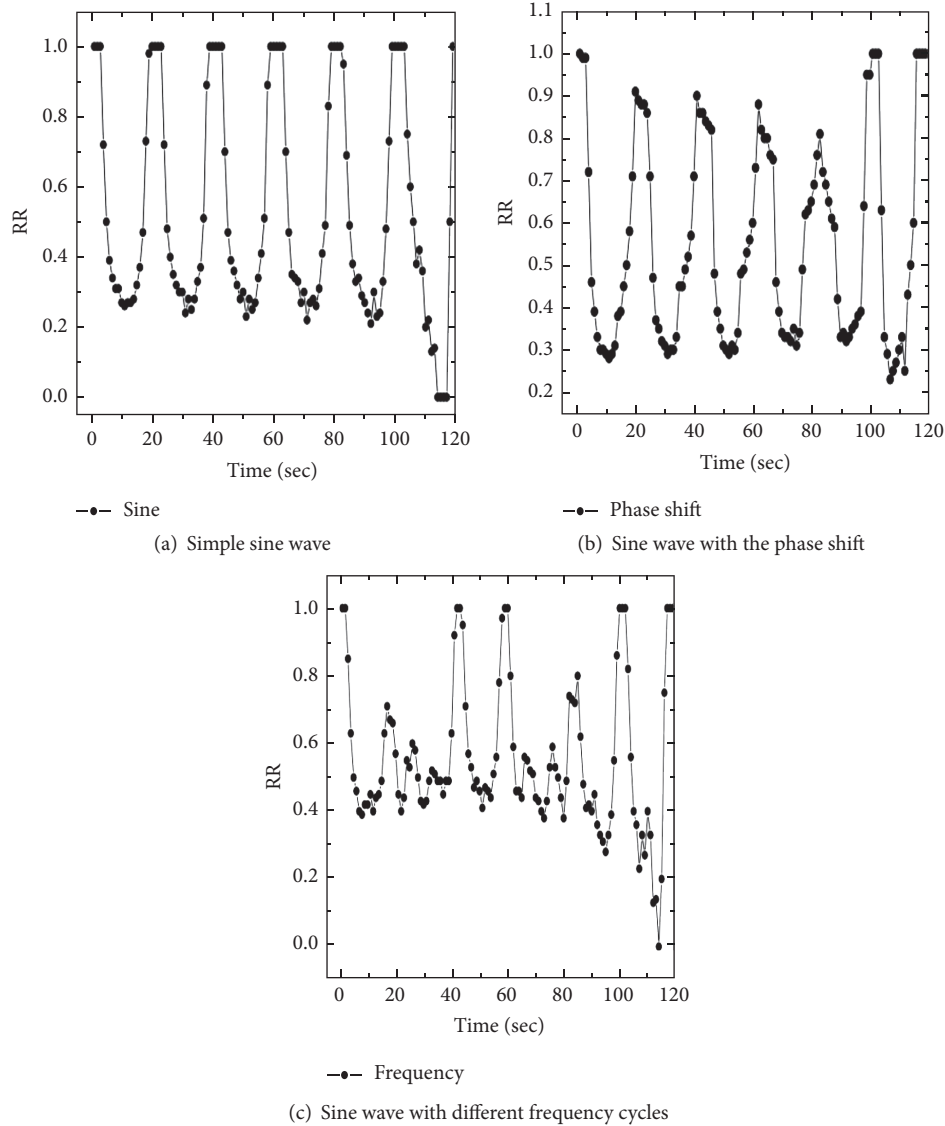


FIGURE 4: Recurrence rate (RR) or $p(\tau)$ curve for sine wave.

(ii) The variation within the groups is calculated as

Within sum of squares (WS)

$$= (m_1 - 1)\sigma_1^2 + (m_2 - 1)\sigma_2^2 + \cdots + (m_i - 1)\sigma_i^2, \quad (7)$$

Within mean squares (WM) = $\frac{WS}{(M - i)}$,

where σ is the standard deviation, M is the number of samples, i is the number of groups, and m_i is the number of samples in group i .

(iii) F -test statistic is calculated as

$$F = \frac{BM}{WM}. \quad (8)$$

3.5. *Kruskal–Wallis Test*. The Kruskal–Wallis test [28] is a rank-based nonparametric test used to determine the

statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. It is used when (1) the data are ordinal and do not meet the precision of interval data, (2) there are serious concerns about extreme deviation from normal distribution, and (3) there is considerable difference in the number of subjects for each comparative group. K -statistics can be calculated as follows:

Computing the K statistic, K

$$= \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1), \quad (9)$$

where n_j is the number of items in sample j , R_j is the sum of the ranks of all items in sample j , and $n = n_1 + n_2 + \cdots + n_k$, the total number of observations in all samples.

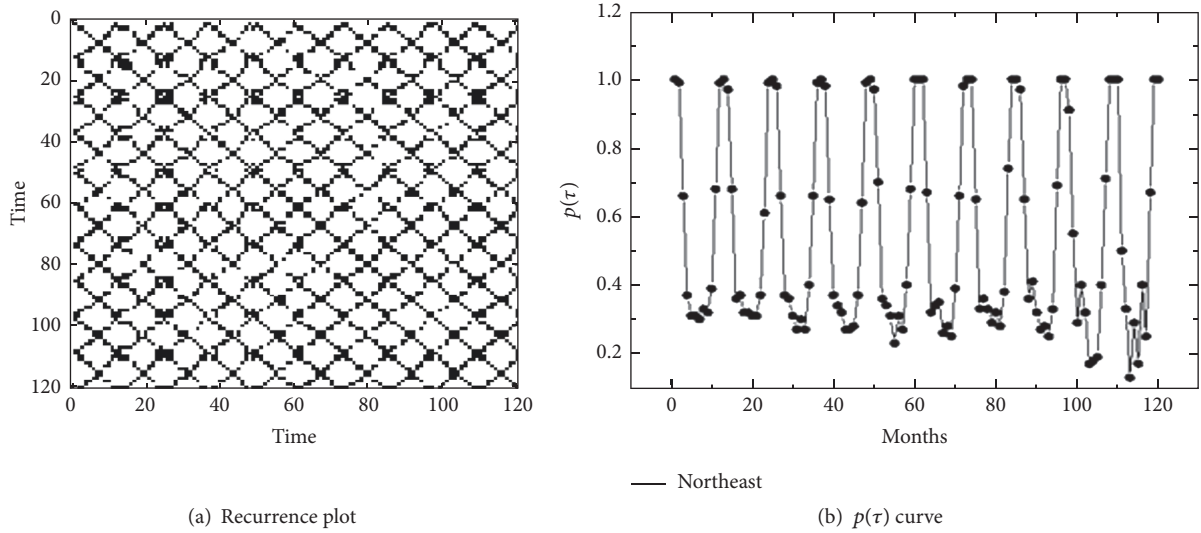


FIGURE 5: Plots for Northeast temperature data.

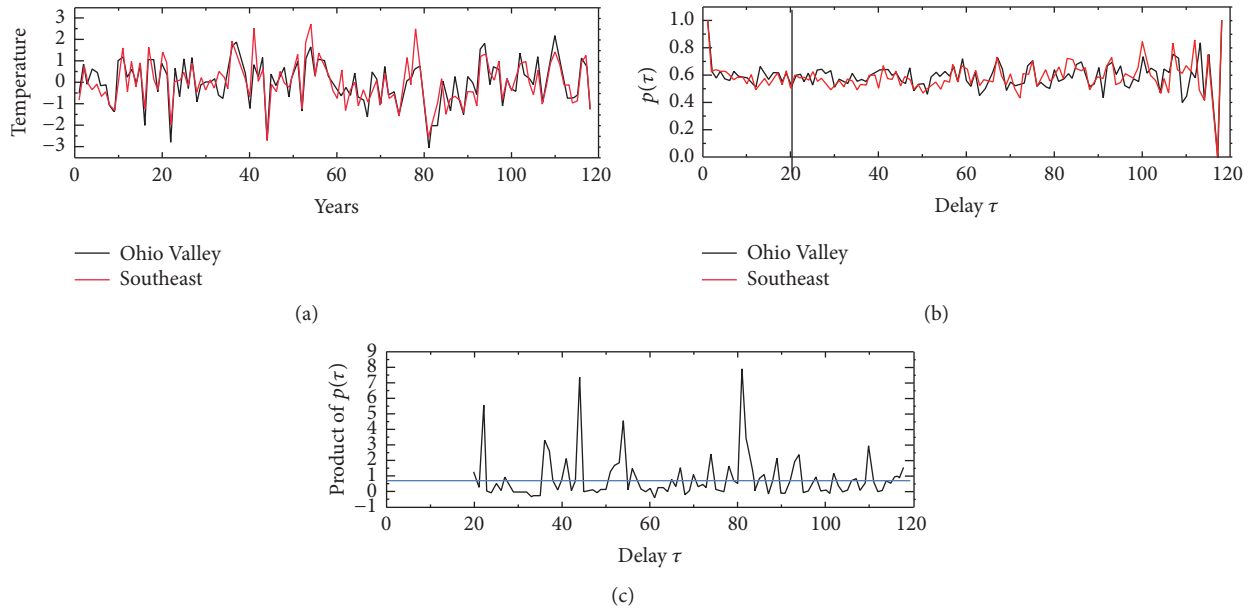


FIGURE 6: Evaluation of CPR from $p(\tau)$. (a) Normalized yearly data for 120 years. (b) $p(\tau)$ curves for these time series with $\tau_s = 20$ are shown with a vertical line. (c) The product of Ohio Valley and Southeast in (b). The horizontal line is the mean of this series, which is the CPR.

4. The Proposed Method

In this section, first we have shown the proposed method in an algorithmic form and then explained each step in detail. The detailed processing of recurrence based similarity identification approach is explained using nine US regions. Experiments are performed on other datasets also.

4.1. Recurrence Based Similarity Identification Algorithm. This section presents a high level summary of the proposed approach, shown in Algorithm 1.

Algorithmic steps involved in the proposed method are explained as follows.

4.2. Data Preparation. In our approach, we used monthly average temperature data for analysis. Since daily data fluctuates more and suffers from estimation error, it is difficult to analyze and compute for 120 years (43800 days). Instead of daily data, monthly data is used because it is approximately normally distributed, fast to compute, and easier to model and it is easier to identify changes in trends and it helps in better strategic decisions.

Sometimes the data collected from available repositories contains missing values, special characters, noise, and outliers. So, first, we clean the data by replacing missing values and special characters, determining presented noise, and removing outliers and then further steps are performed on clean data.

TABLE 1: Monthly average temperature data.

US regions	1895/01	1895/02	1895/03	1895/04	1895/05	1895/06	1895/07	1895/08	...	2014/12
R ₁	18.8	15.6	27.6	43.5	57.1	67.2	65.7	66.7	...	31.6
R ₂	9.5	13.8	28.2	47.3	51.9	58.1	65.6	65.1	...	24.3
R ₃	24.4	29.8	34.3	44.3	49.4	56.1	62.2	62.6	...	33.3
R ₄	24.6	21.9	40.2	55	63	73.5	73.2	74.5	...	37.1
R ₅	39.2	34.8	53.1	63.4	69.2	77	79.3	80	...	46.9
R ₆	43.7	37.3	53.1	61.1	68.3	76.8	77.8	78.5	...	50.5
R ₇	30.3	30.3	40.8	51	58.4	65.1	70.4	70.6	...	36
R ₈	6.9	9.7	26	47.8	56.8	65.8	67.3	67.2	...	24.8
R ₉	35	40.1	44	51.1	58.8	66.2	71.7	71.8	...	31.6

TABLE 2: Normalized temperature values.

US regions	1895/01	1895/02	1895/03	1895/04	1895/05	1895/06	1895/07	1895/08	...	2014/12
R ₁	0.13	0.08	0.27	0.53	0.75	0.91	0.88	0.13	...	0.34
R ₂	0.13	0.19	0.37	0.62	0.68	0.76	0.85	0.13	...	0.32
R ₃	0.22	0.31	0.39	0.56	0.65	0.76	0.87	0.22	...	0.37
R ₄	0.15	0.11	0.38	0.61	0.73	0.89	0.88	0.15	...	0.34
R ₅	0.16	0.08	0.41	0.59	0.7	0.84	0.88	0.16	...	0.3
R ₆	0.2	0.07	0.39	0.56	0.71	0.88	0.9	0.2	...	0.34
R ₇	0.17	0.17	0.35	0.53	0.66	0.78	0.87	0.17	...	0.27
R ₈	0.12	0.16	0.37	0.64	0.76	0.87	0.89	0.12	...	0.35
R ₉	0.2	0.3	0.37	0.5	0.64	0.77	0.87	0.2	...	0.34

R₁–R₉ indicate the notations used for US regions: R₁—Northeast; R₂—Northern Rockies and Plains; R₃—Northwest; R₄—Ohio Valley; R₅—South; R₆—Southeast; R₇—Southwest; R₈—Upper Midwest; R₉—West.

Input ← Multivariate Time series T
Output ← Connectivity trend (Small, Moderate, Weak)
Method:
(1) DP ← Data Preparation (T)
(2) S ← Normalize (DP)
(3) CR ← Cross recurrence (S)
(4) RR ← Recurrence rate (CR)
(5) CPR ← Correlation of probability of recurrence (RR)
(6) CT ← CPR Value

ALGORITHM 1: Recurrence based similarity extraction.

Let T_i represent the daily temperature data for year Y_j .

Step 1. It is difficult to analyze daily temperature data for 120 years (43800 days), so we calculate monthly average temperature to ease the analysis process as follows:

$$\text{Average Temperature, } AT_{\text{month}} = \frac{Y_j \left(\sum_{i=1}^{30} T_i \right)}{30}, \quad (10)$$

where month = Jan, Feb, ..., Dec and $j = 1$ to 120.

Table 1 describes the monthly average temperature data for 120 years.

Data normalization is essential to fit all data in one range for efficient organization of the data. We use $Z_{\text{min-max}}$

normalization method to scale data within the range of (0, 1). Normalized results are shown in Table 2.

$$X_{\text{norm}} = \frac{(X_{\text{mn}} - \min)}{(\max - \min)}, \quad (11)$$

where X_{norm} is the result of the normalized value of temperature, X_{mn} is the temperature value to be normalized, max is the upper bound of the temperature value, and min is the lower bound of the temperature value.

Step 2. In this step, we reduce the time span from 1440 points to 120 points because it is difficult to get a clear visualization of temperature data with large window size (120 years * 12 months = 1440) using RPs. Since seasonal calculation represents the extent of seasonal influence for a particular segment of the year and an average for that particular period trend, therefore, to get an accurate analysis of the temperature data and capture their trends, we compute the average in the period of 4 cycles (spring, summer, autumn, and winter) and each cycle involves 3 months.

Season also indicates regular fluctuations which are repeated from year to year with about the same timing and level of intensity. Seasonal effects are usually associated with climatic changes and their variation is frequently tied to yearly cycles. Therefore, the four seasons with 3-year intervals are considered, which results in a reduced window size of 160 (40 years * 4 months). Figure 7 shows the window size

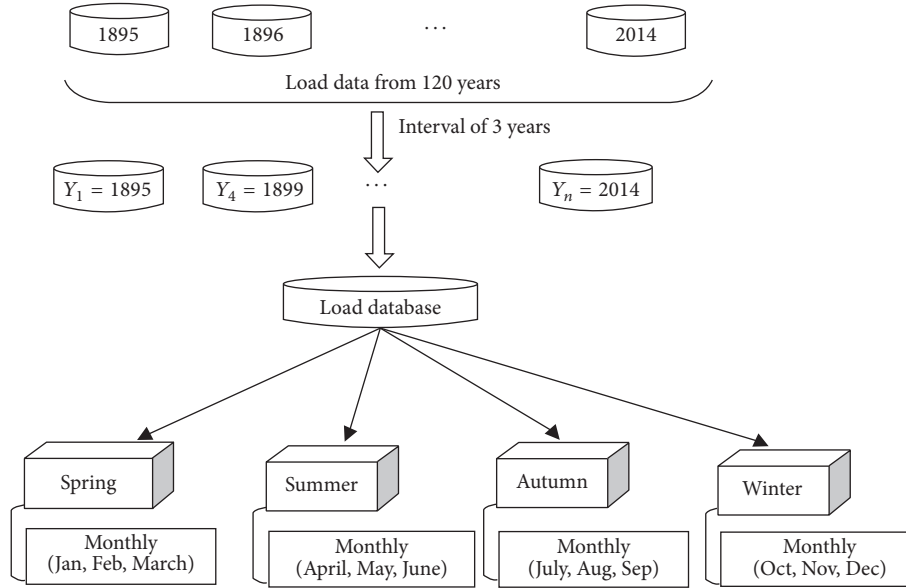


FIGURE 7: Data preprocessing.

reduction process by considering a 3-year time interval and for 4 seasons (spring, summer, autumn, and winter).

4.3. Connectivity Trend Analysis. This connectivity trend analysis enables quantifying a possible similarity and dissimilarity between different regions. It is done by grouping the CPR values into three categories as (a) strongly related ($CT = 1$), (b) moderately related ($CT = 0$), and (c) weakly related ($CT = -1$) using (13). CPR values are calculated using (4) along with the number of pairs for which the CPR value has to be calculated using (12).

The total number of pairs is calculated as follows:

$$\text{number of pairs} = n * \frac{(n-1)}{2}, \quad (12)$$

where n is the number of regions.

Connectivity trend is defined as follows:

$$CT = \begin{cases} 1, & \text{if } |CPR| > \alpha \\ 0, & \text{if } \beta < |CPR| < \alpha \\ -1, & \text{if } |CPR| < \beta, \end{cases} \quad (13)$$

where α and β indicate algorithmic parameters to analyze similarity and dissimilarity. Quantitatively, the similarity between two periodic variations can be computed using CPR.

5. Experimentation

In this section, we perform experiments on both synthetic and real-world datasets to test the performance of the proposed method. All the experiments were conducted on a Windows 7 machine with 2.30 GHz CPU and 4.00 GB RAM. Here, we first describe the dataset, and then the correctness of our approach is shown using heart rate and voice dataset and finally results are calculated on synthetic and climate data.

The algorithmic parameters values are set to $\alpha = 0.8$ and $\beta = 0.5$ for experimentation.

5.1. Dataset. We use our approach on both synthetic and real datasets shown in Table 3. We apply it to data of different kinds but having similar characteristics (i.e., periodic variations) to show the general applicability of the proposed method. To test the applicability of the method, we used datasets such as heart rate and voice where similarity or dissimilarity is known. We compared the outcomes of our method with the known results and found that they are matching with the known results. Although these datasets are unrelated, their characteristics are the same; that is, they have time series data with periodic variations.

Synthetic datasets are generated (named C5Y120, C10Y120, and C15Y120) using *R* fNonlinear package [31]. In this package, to generate different nonlinear time series for 120 points, we used tentSim, logisticSim, and henonSim functions as follows:

- (1) tentSim ($n, n.\text{skip}, \text{parms} = c(a = 2)$): C5Y120
- (2) logisticSim ($n, n.\text{skip}, \text{parms} = c(r = 4)$): C10Y120
- (3) henonSim ($n, n.\text{skip}, \text{parms} = c(a = 1.4, b = 0.3)$): C15Y120

The number of time series points ($n = 120$) is the same for all functions. These functions are used to generate datasets by changing the number of initial values to be skipped from the series ($n.\text{skip}$) and the rest of the parameters are considered by default.

US Climate dataset is obtained from the National Climatic Data Center [32] over the period 1895–2014 for the nine US regions (Ohio Valley, Upper Midwest, Northeast, Northwest, South, Southeast, Southwest, West, and Northern Rockies and Plains).

In China Climate data, the analysis is done on monthly average temperature data to detect abrupt climate changes

TABLE 3: Dataset description.

Dataset	Groups	Number of samples in each group	Remark	Description
<i>Synthetic datasets</i>				
C5Y120	5	120 points	5 (parameters)	Generated using tentSim R fNonlinear package
C10Y120	10	120 points	10 (parameters)	Generated using logisticSim R fNonlinear package
C15Y120	15	120 points	15 (parameters)	Generated using henonSim R fNonlinear package
<i>Real datasets</i>				
US Climate data	9	120 (years)	Samples contain regions' monthly average data	Climate data having monthly average temperature data
China Climate data	8	50 (years)	Samples contain regions' monthly average data	Climate data having monthly average temperature data
Heart rate	4	10 (people)	Samples are grouped based on age	There are a total of forty 120 min ECG recordings for different age groups The duration of these vowel samples was 3 s for healthy voices and 1 s for pathological voices. A 25 or 50 kHz sampling rate was employed to pathological and healthy voices, respectively. All the files were downsampled to 25 kHz and a single 400 ms frame of analysis was considered for each signal
Voice data	6	10 (people)	Samples are grouped based on healthy status and diseases	

in all regions of China for the recent 50 years from 1965 to 2015. China is divided into eight climate regions as follows: (a) South China, (b) the middle and lower reaches of the Yangtze River, (c) North China, (d) Northeast China, (e) the east of Northwest China, (f) the west of Northwest China, (g) Tibet, and (h) Southwest China [32].

In heart rate variability data, the analysis is performed on the PhysioBank [33] dataset for different age groups to analyze the heart rate dynamic properties. There are a total of forty 120 min ECG recordings with 10 people from each group (10–24 years old, 25–40 years old, and 45–67 years old) and 10 elderly people (68–85 years old). The continuous ECG was digitized at 250 Hz.

In voice data, the quantification recurrence measurements are extracted from sustained vowels of speech signals recordings from Disordered Voice Database, Model 4337, developed by Kay Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab [34]. The database includes samples from patients with a wide variety of voice disorders. We present analysis of speech signals to find the difference between healthy voices and voices affected by vocal diseases (normal, Reinke's edema, nodule and vocal cord paralysis, vocal polyps, laryngitis, and contact ulcers). All samples were collected in a controlled environment with the following features: low-noise level, constant microphone distance, direct digital 16-bit sampling, and robust signal conditioning. The selected cases comprise 50 patients with pathological voices (10 with Reinke's edema, 10 with nodule, 10 with laryngitis, 10 with vocal cord paralysis, and 10 with contact ulcers) and 10 patients with healthy voices.

5.2. Correctness of the Proposed Approach. The correctness of our method is demonstrated by applying it to a variety of applications to cope with different situations like heart rate and voice dataset. These datasets are selected because they follow a recurring pattern and have similar characteristics to time series data.

First, we have demonstrated that there occurs a recurring pattern in heart rate which is indicated by the RPs shown in Figure 8(a). If the RP shows a distinct pattern, then there is periodicity in variations. Figures 8(a)(A) and 8(a)(B) indicate RPs of two persons from the same age group (10–24) and their variations are the same so their RPs are similar, while Figures 8(a)(A) and 8(a)(C) indicate RPs of different age groups and their variations are different so both are showing distinct patterns. The similarity and dissimilarity between variations can also be observed using CRP plots. If variations are similar, then CRP plots will show a periodic pattern as indicated by Figure 8(b)(A) and if variations are different then no such pattern is observed as indicated by Figure 8(b)(B).

In other words, we can say that if people belong to the same dataset, their RP and CRP will both follow a recurring pattern but at the same time it is not necessary that if they are from different datasets they follow a recurring pattern in RPs and their CRP will also follow the same, which indicates that there exists a similarity between people belonging to the same age group and difference between people of different age groups (i.e., heart rates of young and old people). Similarly, we can see the similarities and difference between the voice

of healthy people and that of people who have vocal cord paralysis from Figures 9(a) and 9(b).

We also validated our heart and voice data results by the results of previous researchers [35, 36]. They show that heart rate of different age groups differs by plotting RPs. Similarly, the results were observed for voice data.

Table 5 shows the probabilistic correlation of one group of people with another group. It also validates our CRP results by categorizing CPR into three categories. For example, we got a higher number of connections in the strong category and a smaller one in the weak category for a group of people which indicates that the heart rate of one age group or the voice of healthy people does not differ more within a group but it differs with another group of people.

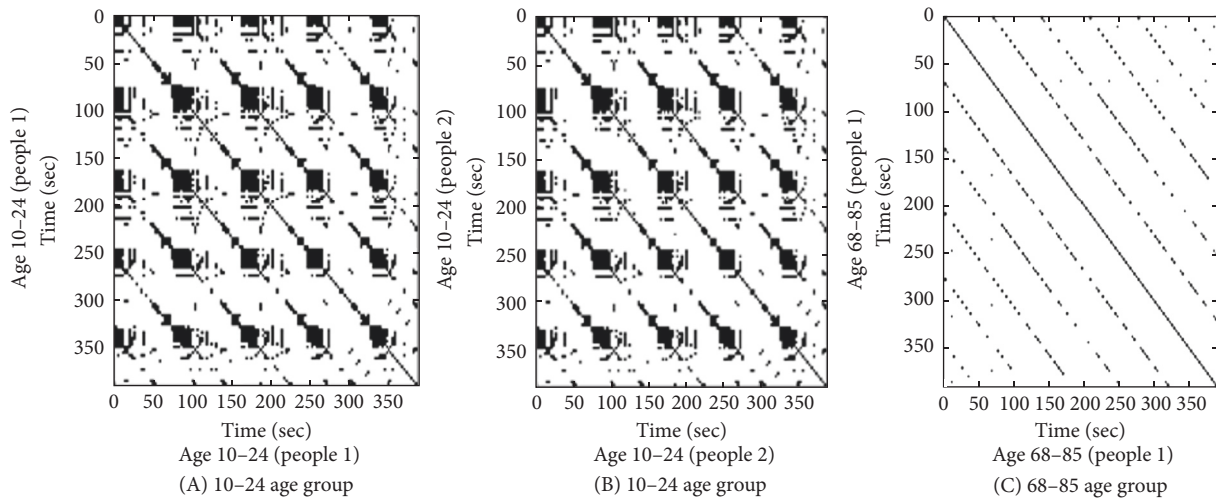
5.3. Climate Data Results. From Figure 10, we can see that RPs of nine US regions have similar periodic variations which indicate that the occurrence of recurrent points along the main diagonal for each region is the same. This means the temperature of one region follows some periodic variation which is different from another region. The recurrence plot for nine regions with 3-year intervals with step size $\tau = 4$ is shown in Figure 10. To show clearer structural changes in the behavior of temperature data and to see the similarities in patterns across the time series, we show the RPs for the first 120 months. RPs for season-wise and 1440 months' data are shown in Appendix 1 (Figures 17 and 18 in the Supplementary Material available online at <https://doi.org/10.1155/2017/7836720>).

Figure 11 shows the CRP between time series of monthly US data for a 10-year time period with lengths of 120 months. We can observe that there is no occurrence of recurrent points along the main diagonal. This means two time series are different and the temperature variation of one region is different from another region.

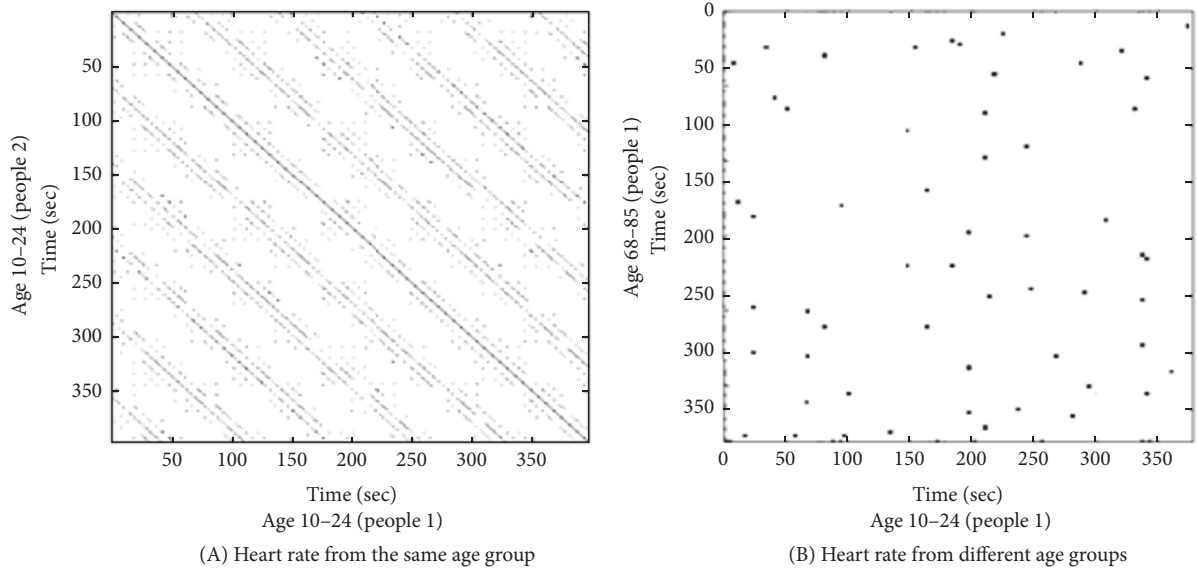
In Figure 11(a), we got more correlated points which show the strong similarity between Northwest and West regions. Similarly, Figures 11(b) and 11(c) show the weak and very weak (less correlated points) connection. Figure 11(d) shows moderate similarity. This relationship can be validated using Figure 1(a). We can also observe that the regions follow similar temperature periodic variations within a region but differ with another region.

These climatic region similarities can also be validated by their longitude and latitude location. For example, in US regions, from Figure 1(a), we can see that Northwest and Southeast regions are situated diagonally, so very low similarity will appear in their temperature. In other words, in the case of Northwest and Southeast regions, sunrays have a direct impact on Northwest and less on Southeast regions, so their temperature differs, which is shown by the smaller number of points in CRP (Figure 11(c)). RPs and CRP for some regions of China are shown in Figures 12 and 13.

Table 4 describes the autocorrelation of US regions for a time period 1895–2014. It is calculated using (2). All region recurrence rate values start from 1 because the probability of recurrence is always 1 at $\tau = 0$. Because of autocorrelation, successive values can be treated as recurrences;



(a) Recurrence plots for heart rate variability data



(b) Cross recurrence plots for heart rate variability data

FIGURE 8

TABLE 4: Probability of recurrence.

US regions	1895/01	1895/03	1895/05	...	1898/01	1898/03	1898/05	...	2014/09	2014/11
R_1	1	0.33	0.67	...	1	0.33	0.66	...	0.57	0.69
R_2	1	0.33	0.68	...	1	0.33	0.73	...	0.71	0.69
R_3	1	0.33	0.68	...	1	0.34	0.69	...	0.71	0.67
R_4	1	0.33	0.69	...	0.99	0.33	0.7	...	0.43	0.66
R_5	1	0.33	0.68	...	1	0.33	0.69	...	0.71	0.7
R_6	1	0.33	0.67	...	1	0.34	0.7	...	0.43	0.7
R_7	1	0.33	0.72	...	0.98	0.33	0.7	...	0.71	0.7
R_8	1	0.33	0.67	...	1	0.33	0.68	...	0.57	0.66
R_9	1	0.34	0.67	...	1	0.35	0.69	...	0.57	0.68

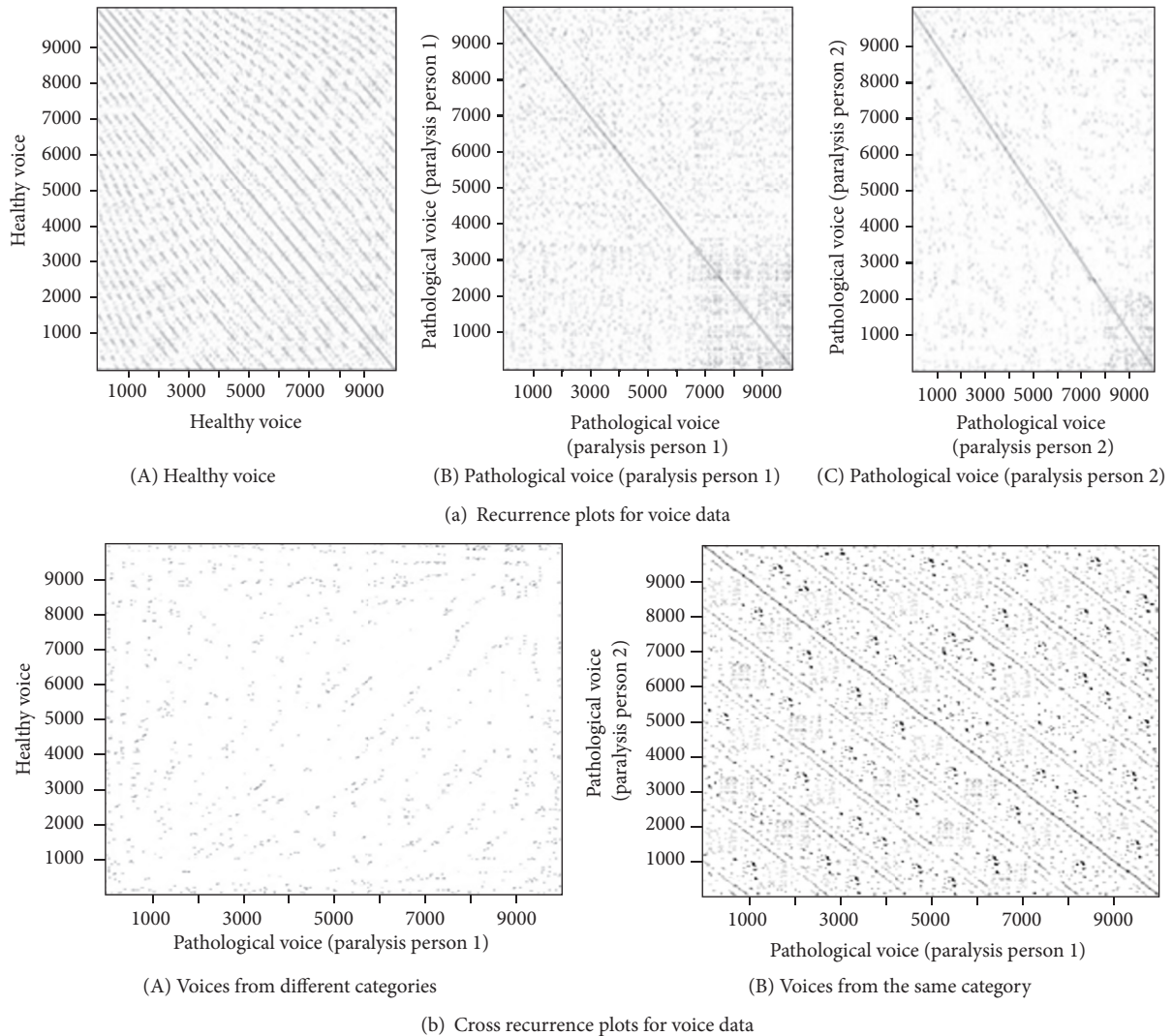


FIGURE 9

thus, a greater recurrence density occurs around the main diagonal.

Figure 14 describes that the spatial distribution of the RR corresponding to the monthly average temperature data of all regions remains unchanged over the 120 months in the US. The figure demonstrates that the monthly average temperature data of each region are similar to the characteristics of the distribution of climate types in the US. This indicates that there exists a slight change in temperature of one region for a particular month (say January) and shows a periodic pattern.

Table 5 describes the number of connections in each bin along time, that is, the results of correlation strength in three categories (i.e., strong, moderate, and weak) calculated using (13). Here, we calculate the CPR (see (4)) for both synthetic and real datasets.

In Table 5, climate region CPR results are shown on monthly data (i.e., correlation between regions). We can interpret that most of the correlation is coming under the weak category for all synthetic, US, and China climate

regions, which indicates that climate change of all regions in the US and China is independent of climate change of other regions. For example, in the US, regions fall under three categories, which are strong (Northwest-West, Ohio Valley-Southeast, and Northeast-Southeast), moderate (Ohio Valley-Upper Midwest, Upper Midwest-South, Northeast-Ohio Valley, Northwest-Southwest, Northwest-South, Northern Rockies and Plains-Southwest, and Northern Rockies and Plains-South), and weak (the rest of the regions) (i.e., the regions having similar temperature variations either latitude- or longitude-wise). Season-wise CPR results for US and China climate regions are shown in Appendix 1 (Table 9).

To show the appropriateness of our climate region results, we represent the correlation connectivity trend enabling quantifying a possible similarity and dissimilarity between US regions for all three categories in Figure 15. From Figure 15(a), we can see that Ohio Valley and Southeast follow an almost similar climate change pattern by moving up and down on the same time periods, which is the indication of

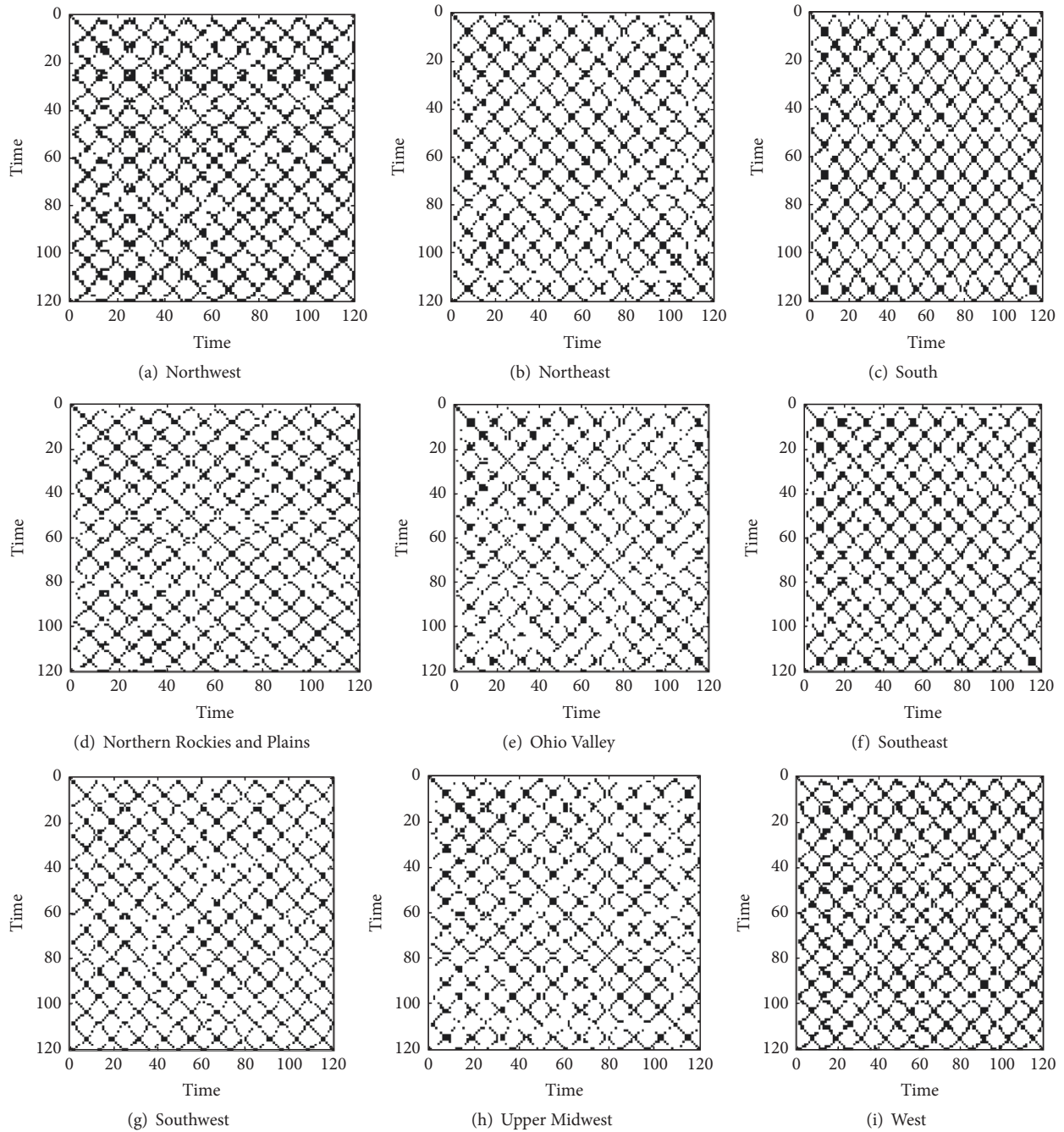


FIGURE 10: Recurrence plots for US regions.

strong connectivity. Similarly, other connectivity trends can be observed.

5.4. Validation Using Significance Testing: ANOVA, Kruskal-Wallis, and z-Test. It is not possible to give an accurate analysis only on the basis of recurrence rate, recurrence plot, and cross recurrence plots on time series. This should be done using a statistical test and an appropriate null hypothesis significance test. So, for the statistical significance of our approach, we are using ANOVA, Kruskal-Wallis, and z-significance test [27, 28] against the analysis obtained

from the recurrence based method on time series data. The statistical methods can be used to analyze similarity between two distinct pieces of data.

Both the Kruskal-Wallis test (often using ordinal data) and one-way ANOVA (typically using interval data) are used to analyze similarity between data series or to determine the statistically significant differences between three or more groups.

If the hypothesis of similarity of means is rejected, this will show that the data does not have a similar pattern and now a query occurs as to which pattern means are distinct.

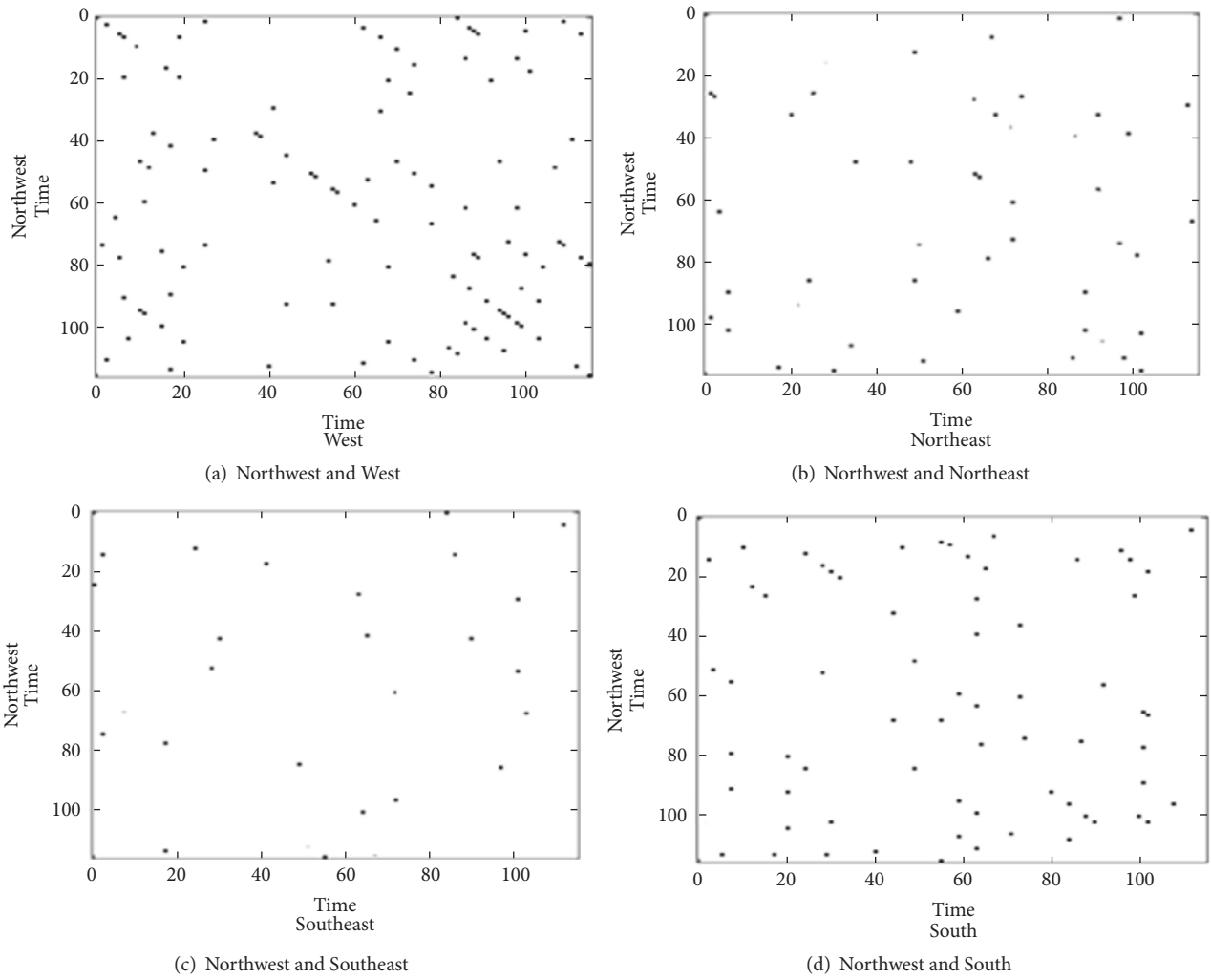


FIGURE 11: CRP for US regions.

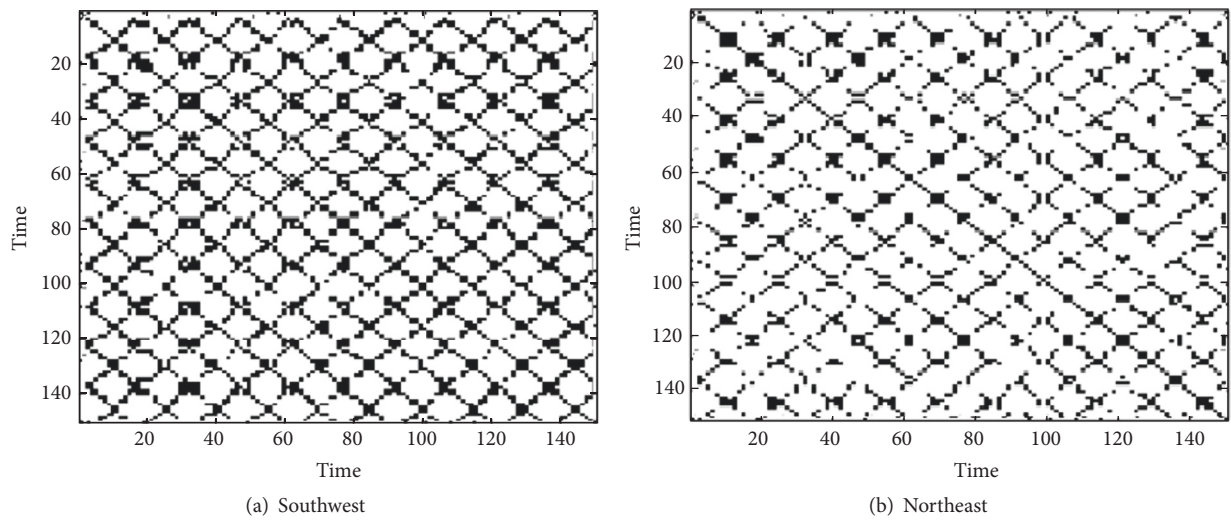


FIGURE 12: Recurrence plots for the regions in China.

TABLE 5: CPR binning table for real and synthetic datasets.

	Strong	Moderate	Weak
<i>Synthetic datasets</i>			
C5Y120	2	1	7
C10Y120	5	4	36
C15Y120	9	11	85
<i>Real datasets</i>			
<i>US Climate data (36 pairs)</i>			
Nine regions	3	7	26
<i>China Climate data (28 pairs)</i>			
Eight regions	4	5	19
<i>Heart rate variability (45 pairs)</i>			
10–24 (age)	36	3	6
25–40	38	5	2
41–67	37	5	3
68–85	32	7	6
<i>Voice data (45 pairs)</i>			
Healthy	38	4	3
Reinke's edema	35	8	2
Nodule	36	5	4
Vocal cord paralysis	40	3	2
Laryngitis	37	5	3
Contact ulcers	33	7	5

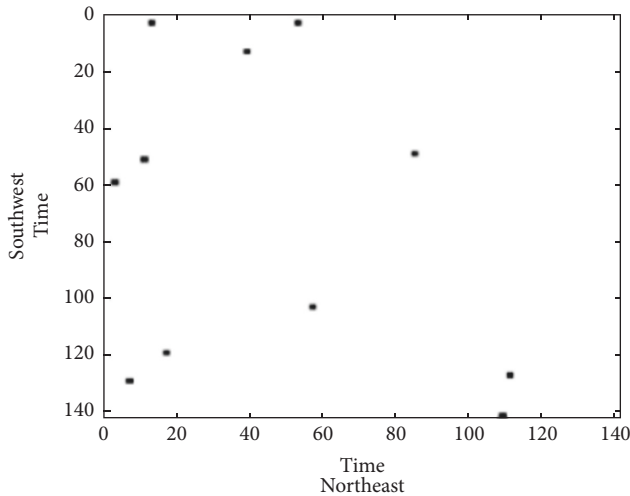


FIGURE 13: CRP for the regions in China: Southwest and Northeast.

The statistical methods used to resolve this query are known as multiple paired comparison procedures. We calculated the hypothesis by using two-tailed significance level estimates (z -test).

The hypothesis is as follows:

$H_0: \mu_1 = \mu_2$, null hypothesis; there are no differences between the groups (9 US regions, 8 climate regions, 6 types of voice, and 4 age groups of heart rate variability).

$H_1: \mu_1 \neq \mu_2$, alternative hypothesis; differences exist.

Validation Using ANOVA Test. Table 6 describes the ANOVA results representing BM (between mean squares of all groups) and WM (within mean square of a group) and their F -statistics value for each dataset.

We observed that the F -test statistic exceeds the significance level of F_{crit} (0.05) in all cases except healthy voice data and heart rate variability data (10–24). For example, for US Climate dataset as seen by the ANOVA test, the calculated F -test statistic (5.10) is greater than F -critical. From this test, it is evident that there are statistically significant differences existing in temperature values between the nine US regions. Similarly, for regions of China, the result indicates that there exists a significant gap between the eight climate regions of China, which demonstrates the effectiveness and reliability of the recurrence based approach in significant climate change recognition in the US and China. In the case of healthy voice data and heart rate variability data (10–24), the calculated F -test statistic is less than F -critical. This F -value indicates similarity within a group (i.e., the same group of people have fewer changes in voice and heart rate).

Validation Using Kruskal–Wallis Test. Table 7 describes the calculated p value of the Kruskal–Wallis test for all datasets, which is less than 0.05 except for healthy voice data and heart rate variability data (10–24). For example, for US Climate dataset, as seen by the Kruskal–Wallis test function, the calculated p value is $2.2e - 16$ which is certainly less than the criterion p value ≤ 0.05 . From this test, it is evident that there are statistically significant differences existing in temperature values between the nine US regions. In the case of healthy voice data and heart rate variability data (10–24), the calculated p value is greater than 0.05, which indicates similarity within a group (i.e., the same group of people have fewer changes in voice and heart rate).

Validation Using z-Test. We calculated the hypothesis at $\alpha = 0.05$ level that there is no difference between all groups versus one group for all datasets. The results are shown in Table 8, which describes the results of testing for similarity of means between all groups for each dataset. Each group is significantly different from the group of all.

From Table 8, we observed that, most of the time, our hypothesis is rejected except for voice data. This indicates that each group has an identity (i.e., there exists less similarity between groups on the basis of temperature). We observed that, for each dataset, there is some exception. For example, in US Climate data, we got only one exception as Southwest, which indicates that there exists less similarity between groups on the basis of temperature. Similarly, we got the exceptions for other datasets. This is a strong indication of each group identity. In case of voice data, most of the time, our hypothesis is accepted which indicates similarity in a group of the same type of people.

5.5. Scalability. To evaluate the scalability of the proposed approach on the CPR value and number of pairs, experiments are done with varying numbers of years. Figure 16(a) represents the correlation of Ohio Valley with Southeast and of

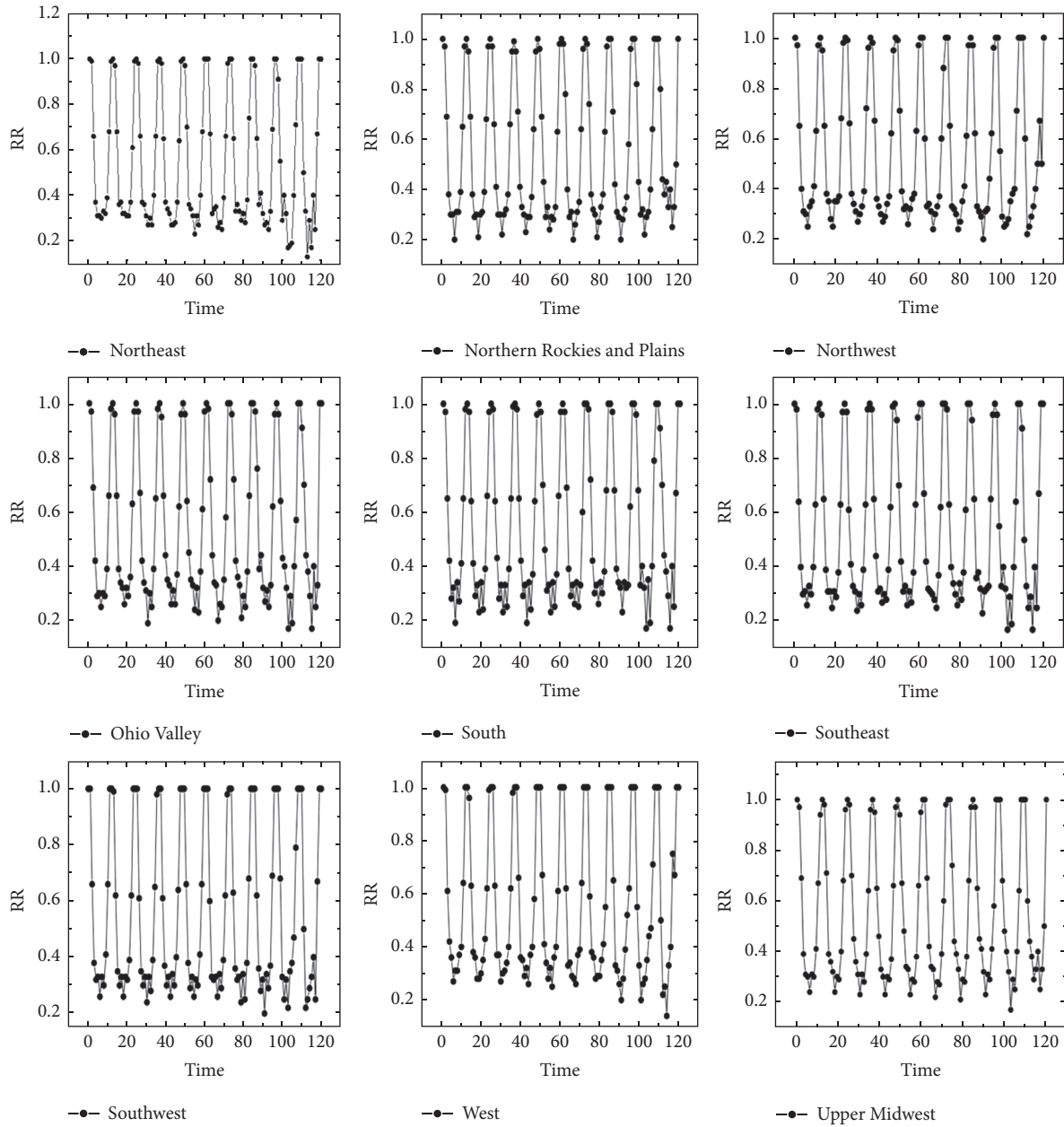


FIGURE 14: Recurrence rate (RR), $p(\tau)$ of each region in the US for 120 months.

TABLE 6: ANOVA test.

Dataset	BM	WM	F -statistics	F -critical
C5Y120	9.45	1.85	5.10	2.13
C10Y120	8.78	3.48	2.52	2.13
C15Y120	6.47	1.18	5.48	2.13
US Climate data	0.32	0.06	5.10	2.13
China Climate data	1.26	0.17	7.41	2.13
Heart rate variability data (age 10–24)	2.45	1.97	1.24	2.13
Heart rate (4 groups)	3.27	1.26	2.59	2.13
Healthy voice data	1.34	0.82	1.63	2.13
Voice data (6 groups)	1.47	0.26	5.65	2.13

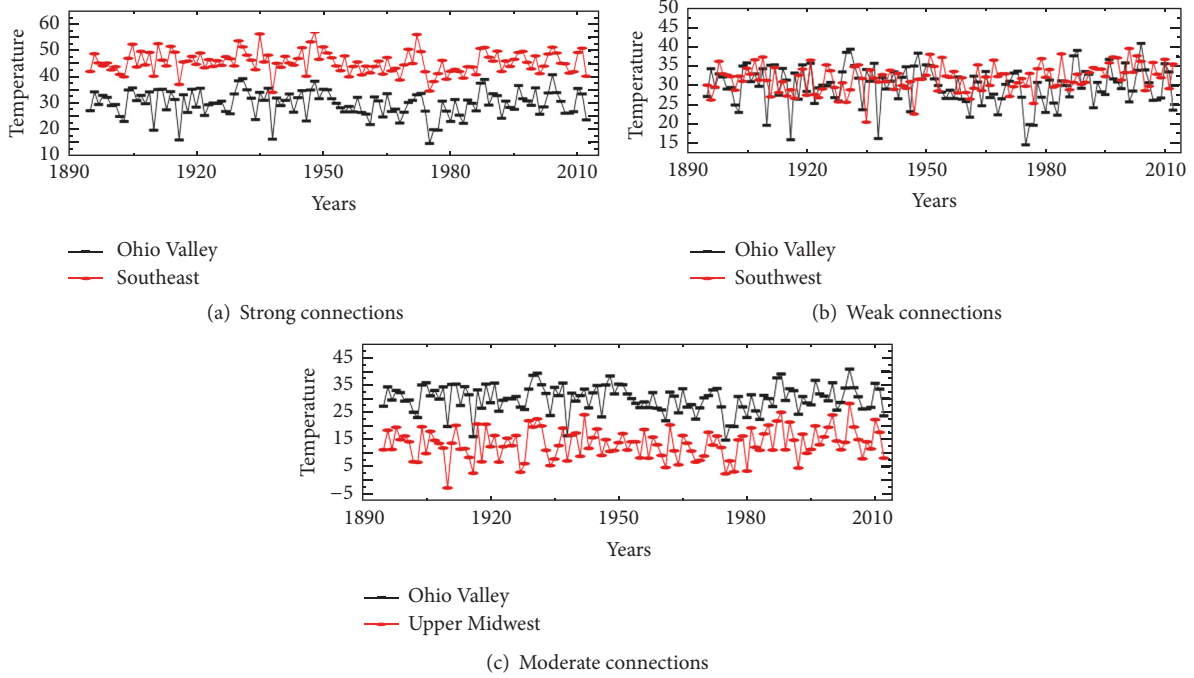


FIGURE 15: Patterns of connectivity.

TABLE 7: Kruskal–Wallis test.

Dataset	p value
C5Y120	$1.4e - 13$
C10Y120	$1.4e - 11$
C15Y120	$1.2e - 11$
US Climate data	$2.2e - 16$
China Climate data	$1.9e - 12$
Heart rate variability data (age 10–24)	1.623
Heart rate (4 groups)	$1.2e - 12$
Healthy voice data	1.412
Voice data (6 groups)	$1.8e - 13$

Southwest and Upper Midwest which indicates that there is a slight change in relationship (i.e., CPR values) throughout the years. In Figure 16(b), it could be seen that the category of relatedness in US regions will be the same with a varying number of years (i.e., extracted pairs belonging to a particular category are not varying much).

6. Conclusion and Discussion

In this paper, we have presented a method to find out similarity or dissimilarity between the two pieces of time series data where each series has periodic variations in the data. Normally, if the variations are periodic, then the two time series may provide the same RP though their nature of variations is different. Analyzing only the RP of the two time series is not sufficient to test whether they have similar periodic variations, so we applied CRP to test the similarity of periodic variations. We have also indicated visual similarity

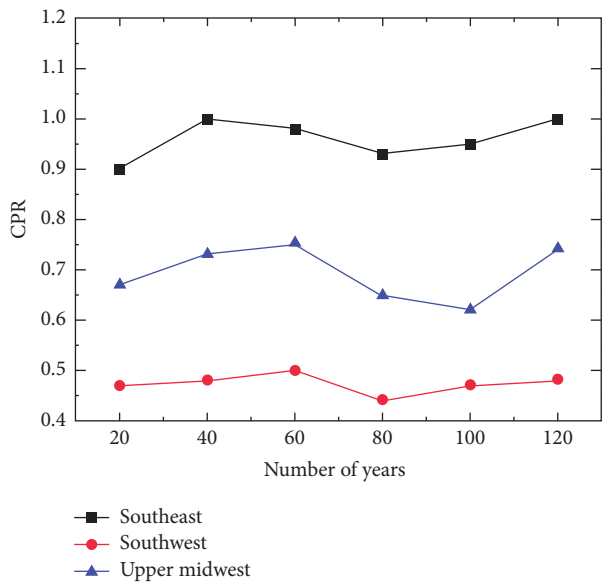
(connectivity) between the two time series using modified CPR.

First, we have tested our approach on the datasets having known similar or dissimilar periodic variations and found that the result of our method matches the already established results and, for that, we have used heart rate dataset, voice dataset, and synthetic datasets generated using the $\sin()$ function. After validating results on the known datasets, we have extended the method for climate data which has similar characteristics (i.e., time series data having periodic variations). We have analyzed data of different climatic regions and inferred that the regions which have similar latitudes and longitudes also have similar variations, while variation patterns change if regions have different latitudes and longitudes which are consistent with the known results. Our outcomes conclude that if the two time series are periodic and have a similar recurrence pattern, this does not mean that the same recurring pattern will emerge in CRP also.

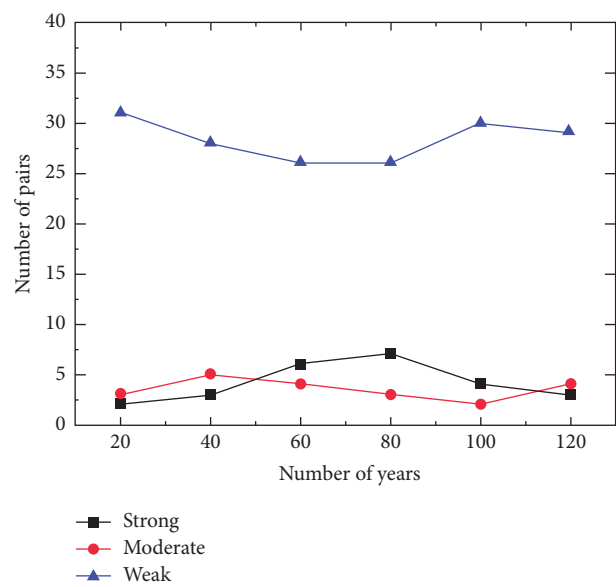
The analysis is done using methods such as quantitative RR, recurrence plots (RPs), cross recurrence plot (CRP), and correlation of probability of recurrence (CPR). First, we have demonstrated the periodicity of temperature variations using the RP plot which is visualized using diagonal lines or square boxes. After that, we have used the modified CPR approach to test similarity (connectivity) of these periodic variations. We have statistically validated our methodology by using ANOVA, Kruskal–Wallis, and z -statistics. This quantitative analysis can effectively recognize the changes of dynamic structure within and between groups. The obtained results suggest that the proposed approach provides a good potential for discrimination between groups of data having periodic variations and can be used to analyze

TABLE 8: Multiple group comparison.

Country 1 versus Country 2	Z-value	Hypothesis results
<i>US, nine regions</i>		
All regions versus Northern Rockies And Plains	-9.13	Hypothesis failed
All regions versus Northwest	-13.47	Hypothesis failed
All regions versus Ohio Valley	-11.33	Hypothesis failed
All regions versus South	2.78	Hypothesis failed
All regions versus Southeast	19.82	Hypothesis failed
All regions versus Southwest	0.85	Hypothesis accepted
All regions versus Upper Midwest	-12.06	Hypothesis failed
All regions versus West	5.79	Hypothesis failed
<i>China, eight regions</i>		
All regions versus South China	-3.009	Hypothesis failed
All regions versus middle and lower reaches of the Yangtze River	-2.263	Hypothesis failed
All regions versus North China	-1.280	Hypothesis accepted
All regions versus Northeast China	2.136	Hypothesis failed
All regions versus east of Northwest China	-1.985	Hypothesis failed
All regions versus west of Northwest China	1.077	Hypothesis accepted
All regions versus Tibet	-2.666	Hypothesis failed
All regions versus Southwest China	1.313	Hypothesis accepted
<i>Voice identification in healthy people</i>		
All persons versus person 1	1.661	Hypothesis accepted
All persons versus person 2	1.302	Hypothesis accepted
All persons versus person 3	1.324	Hypothesis accepted
All persons versus person 4	4.228	Hypothesis failed
All persons versus person 5	0.541	Hypothesis accepted
All persons versus person 6	1.030	Hypothesis accepted
All persons versus person 7	1.44	Hypothesis accepted
All persons versus person 8	1.25	Hypothesis accepted
All persons versus person 9	1.51	Hypothesis accepted
All persons versus person 10	1.293	Hypothesis accepted



(a)



(b)

FIGURE 16: Scalability of the CPR based approach in terms of (a) CPR value. (b) Number of pairs.

the climate data which is provided in the form of time series.

The economic growth of any region or country is different from others on the basis of education, finance, health, environment, business industries, living standard, and agriculture productivity. Somehow, these parameters depend on climate change; the economic time series data shows periodic variations and cycles which may result in differences in economic conditions. Our approach will also help experts to make any climate based policies which affect investment, environment, political stability, technological development, and industrial output such as designing the policies of environmental protection agencies.

In the future, we will try to use recurrence based measures (RQA) with uniform scaling to evaluate recurring patterns with other climate parameters (cloud, weather, pressure, rainfall, etc.) for a better understanding of their interconnection and analyze their effects on economic growth. Furthermore, we can try to parallelize the proposed approach, to reduce the relative time of similarity analysis and memory consumption on extremely large time series datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Anita Bai conceived the idea, designed and supervised the research, and wrote the article. She analyzed and interpreted the data. Swati Hira was responsible for algorithm and manuscript revision for important intellectual content. Swati Hira and Anita Bai were involved in the system design and implementation and drafted the article. S. Deshpande Parag gave valuable pieces of advice on conducting the study and helped in editing the article. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank the Department of Computer Science and Engineering, VNIT, Nagpur, for making the required computing facilities available.

References

- [1] M. L. Parry, O. F. Canziani, P. J. Palutikof, P. J. v. d. Linden, and C. E. Hanson, *Climate Change 2007: Impacts, Adaptation and Vulnerability*, Cambridge University Press, 2007.
- [2] M. K. Linnenluecke, A. Griffiths, and M. I. Winn, "Firm and industry adaptation to climate change: A review of climate adaptation studies in the business and management field," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 4, no. 5, pp. 397–416, 2013.
- [3] S. M. Howden, J.-F. Soussana, F. N. Tubiello, N. Chhetri, M. Dunlop, and H. Meinke, "Adapting agriculture to climate change," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 50, pp. 19691–19696, 2007.
- [4] C. Okereke, B. Wittneben, and F. Bowen, "Climate change: Challenging business, transforming politics," *Business and Society*, vol. 51, no. 1, pp. 7–30, 2012.
- [5] W. Wu, P. H. Verburg, and H. Tang, "Climate change and the food production system: Impacts and adaptation in China," *Regional Environmental Change*, vol. 14, no. 1, pp. 1–5, 2014.
- [6] S. J. Vermeulen, B. M. Campbell, and J. S. I. Ingram, "Climate change and food systems," *Annual Review of Environment and Resources*, vol. 37, pp. 195–222, 2012.
- [7] A. J. Haverkort and A. Verhagen, "Climate change and its repercussions for the potato supply chain," *Potato Research*, vol. 51, no. 3–4, pp. 223–237, 2008.
- [8] D. I. Gustafson, J. W. Jones, C. H. Porter et al., "Climate adaptation imperatives: untapped global maize yield opportunities," *International Journal of Agricultural Sustainability*, 2014.
- [9] O. Ovalle-Rivera, P. Läderach, C. Bunn, M. Obersteiner, and G. Schroth, "Projected shifts in Coffea arabica suitability among major global producing regions due to climate change," *PLoS ONE*, vol. 10, no. 4, Article ID e0124155, 2015.
- [10] Z. Liu, P. Yang, H. Tang et al., "Shifts in the extent and location of rice cropping areas match the climate change pattern in China during 1980–2010," *Regional Environmental Change*, vol. 15, no. 5, pp. 919–929, 2015.
- [11] S. E. Park, "A review of climate change impact and adaptation assessments on the Australian sugarcane industry," in *Proceedings of the 2008 Conference of the Australian Society of Sugar Cane Technologists*, Queensland, Australia, April–May 2008.
- [12] A. Fleming, S. E. Park, and N. A. Marshall, "Enhancing adaptation outcomes for transformation: climate change in the Australian wine industry," *Journal of Wine Research*, vol. 26, no. 2, pp. 99–114, 2015.
- [13] C. L. Webber and J. P. Zbilut, "Recurrence quantification analysis of nonlinear dynamical systems," *Tutorials in Contemporary Nonlinear Methods for The Behavioral Sciences*, pp. 26–94, 2005.
- [14] N. Marwan and J. Kurths, "Cross recurrence plots and their applications," *Mathematical Physics Research at The Cutting Edge*, pp. 101–139, 2004.
- [15] M. I. Coco and R. Dale, "Cross-recurrence quantification analysis of categorical and continuous time series: An R package," *Frontiers in Psychology*, vol. 5, article 510, 2014.
- [16] J. Sukharev, C. Wang, K.-L. Ma, and A. T. Wittenberg, "Correlation study of time-varying multivariate climate data sets," in *Proceedings of the IEEE Pacific Visualization Symposium, PacificVis 2009*, pp. 161–168, Beijing, China, April 2009.
- [17] S. Liu, X. Huang, Y. Ni, H. Fu, and G. Yang, "A high performance compression method for climate data," in *Proceedings of the 12th IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA 2014*, pp. 68–77, Milan, Italy, August 2014.
- [18] M. N. Sap and A. M. Awan, "Finding spatio-temporal patterns in climate data using clustering," in *Proceedings of the International Conference on Cyberworlds (CW'05)*, pp. 155–162, Singapore, Singapore, November 2005.
- [19] W. Hendrix, I. K. Tetteh, A. Agrawal, F. Semazzi, W.-K. Liao, and A. Choudhary, "Community dynamics and analysis of decadal trends in climate data," in *Proceedings of the 11th IEEE International Conference on Data Mining Workshops, (ICDMW '11)*, pp. 9–14, BC, Canada, December 2011.
- [20] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5–6, pp. 237–329, 2007.

- [21] C. L. Webber and J. P. Zbilut, "Recurrent structuring of dynamical and spatial systems," in *Proceedings of the International Meeting*, pp. 101–133, 1997.
- [22] J. Belaire-Franch, D. Contreras, and L. Tordera-Lledó, "Assessing nonlinear structures in real exchange rates using recurrence plot strategies," *Physica D: Nonlinear Phenomena*, vol. 171, no. 4, pp. 249–264, 2002.
- [23] F. Strozzi, E. Gutiérrez, C. Noé, T. Rossi, M. Serati, and J. M. Zaldívar, "Measuring volatility in the Nordic spot electricity market using recurrence quantification analysis," *European Physical Journal: Special Topics*, vol. 164, no. 1, pp. 105–115, 2008.
- [24] J. Zbilut, M. Koebbe, H. Loeb, and G. Mayer, "Use of recurrence plots in the analysis of heart beat intervals," *Computers in Cardiology*, pp. 263–266, 1990.
- [25] P. M. Crowley, "Analyzing convergence and synchronicity of business and growth cycles in the euro area using cross recurrence plots," *The European Physical Journal Special Topics*, vol. 164, no. 1, pp. 67–84, 2008.
- [26] D. F. Silva, V. M. A. De Souza, and G. E. A. P. A. Batista, "Time series classification using compression distance of recurrence plots," in *Proceedings of the 13th IEEE International Conference on Data Mining (ICDM '13)*, pp. 687–696, IEEE, Dallas, TX, USA, December 2013.
- [27] S. Hira, A. Bai, and P. S. Deshpande, "Estimating the similarities of G7 countries using economic parameters," *Advances in Intelligent Systems and Computing*, vol. 435, pp. 59–67, 2016.
- [28] R. I. Levin and D. S. Rubin, *Statistics for Management*, Pearson Prentice Hall, University of North Carolina, 1997.
- [29] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL*, vol. 4, no. 9, pp. 973–977, 1987.
- [30] M. C. Romano, M. Thiel, J. Kurths, I. Z. Kiss, and J. L. Hudson, "Detection of synchronization for non-phase-coherent and non-stationary data," *Europhysics Letters*, vol. 71, no. 3, pp. 466–472, 2005.
- [31] "fNonlinear package," 2015, <https://cran.r-project.org/web/packages/fNonlinear/fNonlinear.pdf>.
- [32] "Climate center data link," 1970, <http://www.ncdc.noaa.gov>.
- [33] D. Hoyer, B. Pompe, H. Herzel, and U. Zwiener, "Nonlinear coordination of cardiovascular autonomic control: The fundamentals of nonlinear dynamics," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 6, pp. 17–21, 1998.
- [34] K. Elemetrics, *Elemetrics Corp. Disordered Voice Database*, Model 4337, 3 edition, 1994.
- [35] H. Ding, S. Crozier, and S. Wilson, "A new heart rate variability analysis method by means of quantifying the variation of nonlinear dynamic patterns," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1590–1597, 2007.
- [36] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering Online*, vol. 6, article 23, 2007.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

