SPATIO-TEMPORAL CHARACTERIZATION WITH WAVELET COHERENCE : A NEXUS BETWEEN ENVIRONMENT AND PANDEMIC

Iftikhar U. Sikder[1] and James J. Ribero[2]

[1]Department of Information Systems, Cleveland State University, USA
[2]IBA, University of Dhaka, Bangladesh

ABSTRACT

Identifying spatio-temporal synchrony in a complex, interacting and oscillatory coupled-system is a challenge. In particular, the characterization of statistical relationships between environmental or bio-physical variables with the multivariate data of pandemic is a difficult process because of the intrinsic variability and non-stationary nature of the time-series in space and time. This paper presents a methodology to address these issues by examining the bivariate relationship between Covid-19 and temperature time-series in the time-localized frequency domain by using Singular Value Decomposition (SVD) and continuous cross-wavelet analysis. First, the dominant spatio-temporal trends are derived by using the eigen decomposition of SVD. The Covid-19 incidence data and the temperature data of the corresponding period are transformed into significant eigen-state vectors for each spatial unit. The Morlet Wavelet transformation is performed to analyse and compare the frequency structure of the dominant trends derived by the SVD. The result provides cross-wavelet transform and wavelet coherence measures in the ranges of time period for the corresponding spatial units. Additionally, wavelet power spectrum and paired wavelet coherence statistics and phase difference are estimated. The result suggests statistically significant coherency at various frequencies providing insight into spatio-temporal dynamics. Moreover, it provides information about the complex conjugate dynamic relationships in terms phases and phase differences.

KEYWORDS

Wavelet analysis, Cross-wavelet power, Wavelet coherence, Covid-19, Singular Value Decomposition

1. INTRODUCTION

Transformation by localized wavelike function called ‘wavelet’ addresses the inefficiencies of Fourier transformation by using waveforms of shorter duration at higher frequencies and waveforms of longer duration at lower frequencies [1]. Fundamentally, wavelets analyze a signal or time series according to scale where high frequency is represented by low scale and low frequency by high scale resulting into better frequency resolution for low frequency events and better time resolution for high-frequency events. Additionally, wavelets capture features across a wide range of frequencies and enables one to analyze time series that contain non stationary dynamics at many different frequencies[2]. Wavelet transformation can be done in a smooth continuous way (continuous wavelet transform - CWT) or in discrete steps (discrete wavelet transform - DWT).

Due to their versatility in handling very irregular complex data series in absence of knowing the underlying functional structure, wavelet transform analysis can be applied to analyze diverse physical phenomena e.g. climate change, financial analysis, cardiac arrhythmias, seismic signal de-noising, video image compression and so forth [3] - [4].

In this paper, we have applied the wavelet transformation to elucidate the interconnection between the environment and the Covid-19 pandemic. The dynamics between the bio-physical or climatic variables specifically the temperature and the diffusion of Covid-19 cases is reported to be ambiguous [5], [6], [7]. There are various claims with regards to the dependency between the incidence or prevalence and environmental variables. It has often been argued that lower (cold) temperature act as a catalyst in significantly increasing the spread of Covid-19 [8], [9]. There also exist alternative claims that warm temperatures slow down the spread of Covid-19 [10]. In contrast to these claims, some scholars assert that temperature does not play any role in the spread of Covid-19 [11]. In this paper, we have examined some specific empirical relationships of such dependencies, namely wavelet coherence and its statistical significance, phases and phase differences using the dataset of the USA.

2. OBJECTIVES AND SCOPE

The primary objective of the paper is to characterize the dynamic relationship of Covid-19 and a time-variant bio-physical parameter namely the temperature. This paper aims to provide an empirical investigation that captures and analyzes the characteristic relationship of these variables. The study area was limited within the United States. The data for Covid-19 cases was collected from the fifty (50) states, and the corresponding data on temperature of the same period was collected from these states. The period covered was between Jan. 21, 2020, till date. Around 40,000 records (20000 Covid-19 data records and 20000 temporal temperature data records) have been collected and used for the research [12], [13].

3. METHODOLOGY

The variables used in the model are featured as time series data, and thus expected to fluctuate with an associated noise. Employing conventional smoothing technique involving amplitude-based statistical analysis would not be appropriate to achieve the research objective. Therefore, we adopt a Wavelet Transform algorithm not only to capture the periodicities of the variables over the time, but also to establish coherence among the variables in the frequency domain.

3.1. Modeling Framework

The Figure 1 below details the process flow of the research:

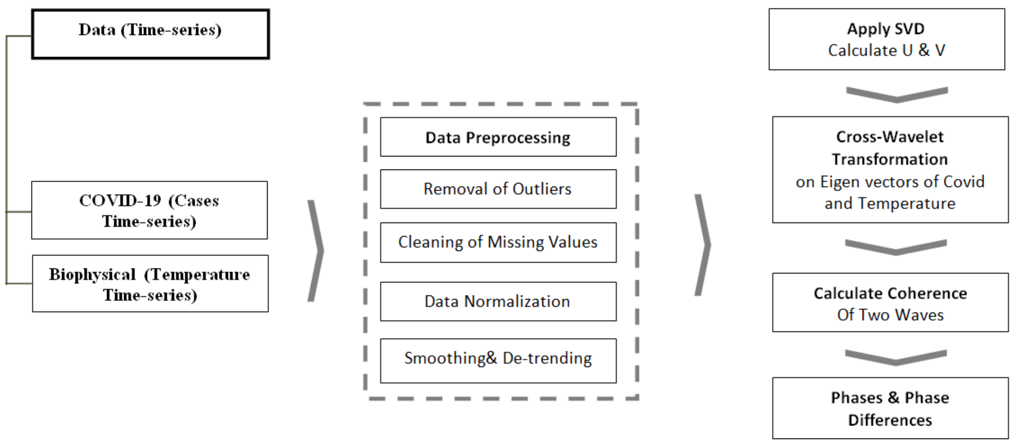

Figure 1. Methodology flow chart

3.1.1. Singular Value Decomposition and Wavelet Transformation

A widely adopted matrix factorization or dimension reduction technique namely Singular Value Decomposition (SVD) was used. SVD is one of the methods for matrix factorizations that generates eigen decomposition of high dimensional data. It enables low-rank approximation of a matrix. Given a data set $\mathbf{X} \in \mathbb{C}^{m \times n}$ columns $X_k \in \mathbb{C}^m$, the SVD generates a unique decomposition:

$$\mathbf{X} = U\Sigma V^* \quad (1)$$

where U is a m × m and V is n × n unitary matrices with orthogonal columns. Σ is real, nonnegative matrix with unique diagonal entries which is known as singular values of **X**. The * denotes complex conjugate transpose. It is possible to represent the decomposition in terms of compact or economy SVD representation where Σ is square diagonal of size r × r , where rank r <= min (m, n). Thus SVD provides low-rank approximations. The approximated or truncated representation is given by the sum of rank-1 matrices:

$$X' = \sum_{k=1}^{r} \sigma_k u_k v_k^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \sigma_3 u_3 v_3^* \quad (2)$$

In this study we selected $r = 3$ which accounts for significant variance of the original data matrix. The selected vectors (v_1, v_2, v_3) represent dominant time series signals which could be subject to further spectral analysis using wavelet analysis. Unlike Fourier analysis which permits to study the cyclical nature of time series in frequency domain without any time localization, wavelet analysis allows to characterize the periodicity of time series over time in terms of orthogonal basis by providing multi-resolution decompositions. The Wavelet function $\psi(t)$ is given by :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a \neq 0 \quad (3)$$

Where a is a scaling or dilation parameter and b is a translation parameter. The scaling involves stretching (when $\lceil a \rceil > 1$) or compressing ($\lceil a \rceil < 1$), while translating means shifting position in time.

3.1.2. Data Pre-Processing

The data sets on Covid-19 cases and temperature across the 50 states accessed in a format that was not readily available for analysis. The pre-processing steps involved the basic data cleaning functions such as removal of irrelevant attributes and missing values, removal of outliers, de-meaning, linear detrending, and data normalization. The data processing was done in R environment using *WaveletComp* package [14]. The outputs of the pre-processing are data frames that were transformed into rectangular matrices, where the rows represent either Covid-19 cases by states or temperature, and the columns represent the date. The continuous wavelet transformation of a time series $x(t)$ with respect to wavelet function ψ is :

$$W_{x,\psi}(a,b) = \int_{-\infty}^{\infty} \boldsymbol{x}(\boldsymbol{t}) \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) dt \quad (4)$$

Thus, continuous wavelet transformation $W_{x,\psi}(a,b)$ is a convolution function that provides time localized frequency information. The wavelet power spectrum (scalogram) is simply given by :$\left|W_{x,\psi}(a,b)\right|^2$. For example, given a synthetic time series $y(t)$,

$$y = \sin\left(\frac{2\pi}{p_1}\right) + \sin\left(\frac{2\pi}{p_2}\right) + \ \epsilon \ (5)$$

which consists of composite of two sinusoids with period p_1 and p_2 and additive Gaussian noise ϵ; for specified periods of $p_1 = 50$, $p_2 = 100$, the time series exhibit characteristic periodicities as shown in Figure 3a. A wavelet transformation of $y(t)$ and subsequent power spectrum clearly identifies two distinct regions of high wavelet power in red color in the scalogram (see figure 3c). The transformation result depends on the choice of wavelet function $\psi(t)$. In this case, we have selected *Morlet Wavelets* function, which is family of functions given by the equation:

$$\psi_{\omega_0}(t) = Ke^{-i\omega_0 t}e^{\frac{-t^2}{2}} \qquad (6)$$

Where, K is the normalizing constant and with the value $K = \pi^{-1/4}$ ensures unit energy of $\psi_{\omega_0}(t)$. Essentially, Morlet Wavelet is the product of a complex time series with a Gaussian function normalized by a constant. Figure 3b shows the Morlet function in real axis as well as in imaginary axis viewed from different orientation.

3.1.3. Calculating Singular Value Decomposition (SVD)

The SVD technique was applied on both the Covid-19 and temperature data sets to compress the data into ortho normal eigen basis to rectangular matrices.Based on the top singular values, top three eigen states for both Covid-19 and temperature were selected, which in combination accounts for significant total variance of the original data. Plots of the transformed data set are displayed in Figure 2 below showing negative correlation (with *correlation coefficient*of-0.34) of cases and temperature.

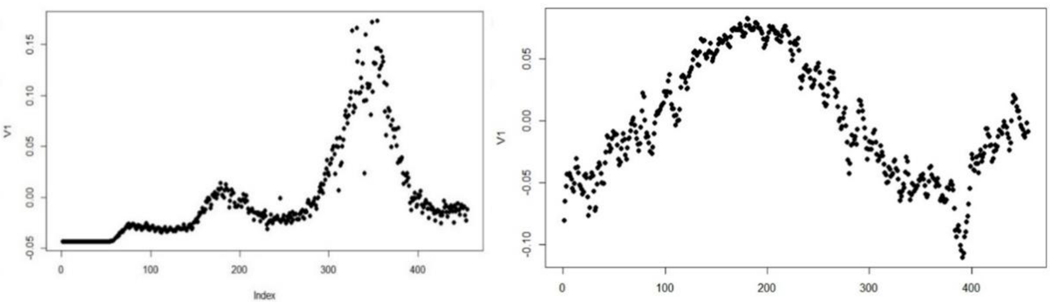

Figure 2. Significant eigen vectors of temperature and Covid-19 cases

3.1.4. Calculating Wavelet Transformation

A Wavelet transform decomposes a time series into a set of wavelets localized in time. The Wavelet transformation was performed on both Covid-19 and temperature time – series of *eigen* vectors. We applied *Morlet* continuous Wavelet transforms to the transformed data using the *WaveletComp* R package (Figure 3).

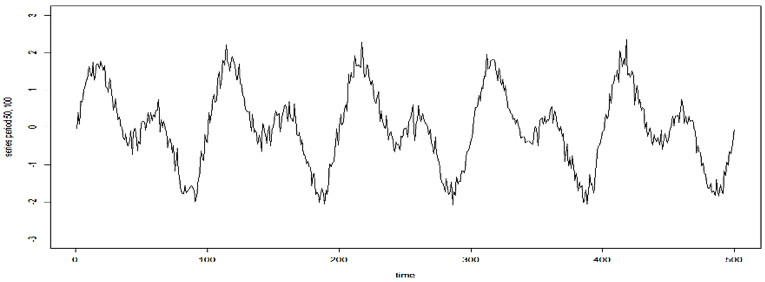

Figure 3a. Synthetic time series with period p_1= 50 and p_2 =100 with additive Gaussian noise

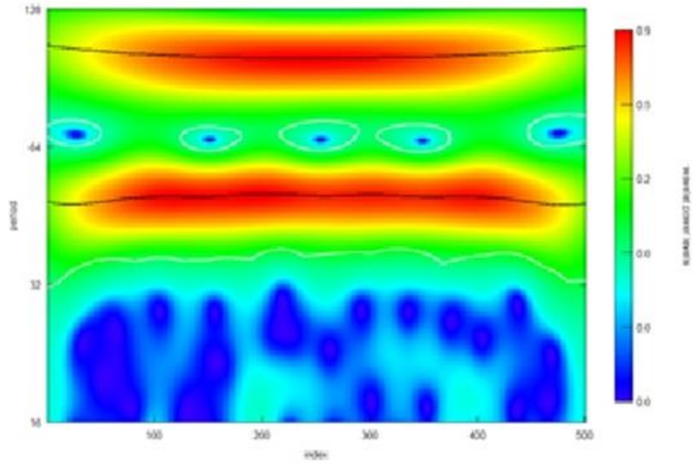

Figure 3b. Wavelet power spectrum (Scalogram) of Synthetic time series with period p_1= 50 and p_2 =100 with additive Gaussian noise

Wavelet transformation leads to a continuous, complex-valued output of the time series that preserves both time and frequency resolution parameters. The transform is separable into its real part and imaginary part providing information on both local amplitude and instantaneous phase. The separation allows for the investigation of coherency between the two time series. Given two time series $X(t)$ and $Y(t)$, and corresponding *wavelet spectrums* $W_x(a, b)$ and $W_y(a, b)$ which could be considered as localized energy spectrum varying with scale a, and translation b, and associated frequency ω and time t. The cross-wavelet transformation $W_{xy}(a, b)$ is associated with complex-valued wavelet coherency [15], [16]:

$$\Upsilon(a, b) = \frac{\langle W_{xy}(a,b)\rangle}{\sqrt{\langle W_x(a,b)\rangle * \langle W_y(a,b)\rangle}} \qquad (7)$$

and the normalizing wavelet power spectra coherence is:

$$\Upsilon^2(a, b) = \frac{\left[\langle Re(W_{xy}(a,b))\rangle\right]^2 + \left[\langle Im(W_{xy}(a,b))\rangle\right]^2}{\langle W_x(a,b)\rangle * \langle W_x(a,b)\rangle} \qquad (8)$$

The angle brackets $\langle \ \ \rangle$ denotes the smoothing operation over time. The squaring of the amplitude component gives us the wavelet power spectrum $0 \leq \Upsilon^2(a, b) \leq 1,$ which is somewhat analogous to conventional correlation coefficient.

The corresponding wavelet phase difference is given by:

$$\varphi_{xy} = \tan^{-1}\frac{Im\langle W_{xy}\rangle}{Re\langle W_{xy}\rangle} \qquad (9)$$

After computing the wavelet power spectrum for each of *eigenvector*, we analyze the coherence of the paired waves of Covid-19 and temperature using the coherence function. The phase lags between the variables were also computed. The cross-wavelet transformation provided cross-magnitude, phase differences, non-stationarity, and coherency between signals. Using these results of the cross-wavelet transformation, a series 'synchronicity at certain periods and across certain ranges of time was analyzed.

4. RESULTS AND MODEL INTERPRETATION

The cross-wavelet analysis generated coherence plot that shows that that there is a coherence (correlation) between Covid-19 and temperature, and these relationships are statistically significant (the region enveloped or bounded by the white line). The phases and phase differences show varied results. Figures 4a, 4b, 4c shows Covid-19 and temperature are out of phase with varying phase lags while Figure 4d and 4e shows that are in phase. Comparing the result of plots 4a and 4e, temperature were both out of phase in 4a, with temperature leading and Covid-19 lagging by 96 days, while from 4e both time series were in phase. Though temperature was leading, the lag period was much narrow (around 5 days) compared to 4a.

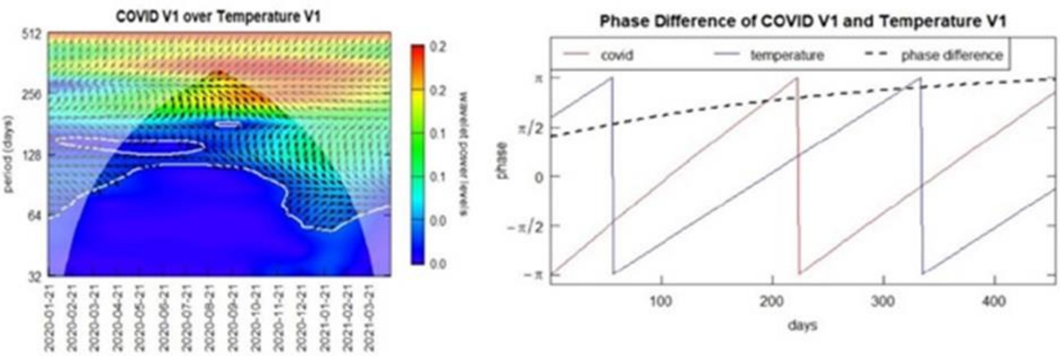

Figure 4a. Cross-wavelet analysis generated coherence plots (4a. Covid-19 V1 & Temperature V1)

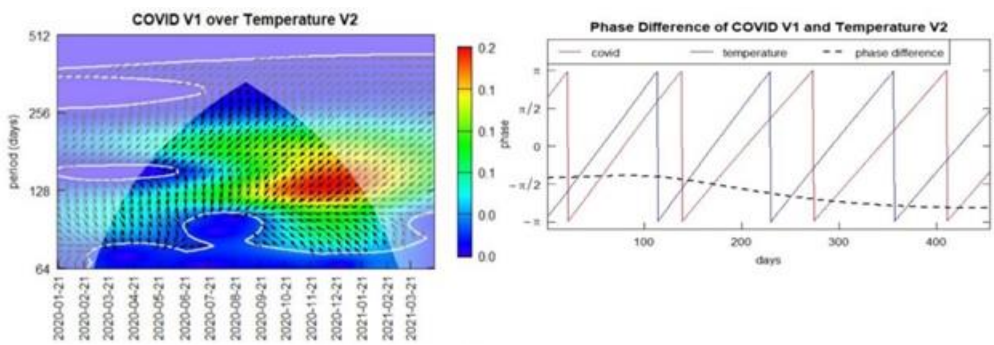

Figure 4b. Cross-wavelet analysis generated coherence plots (Covid-19 V1 & Temperature V2)

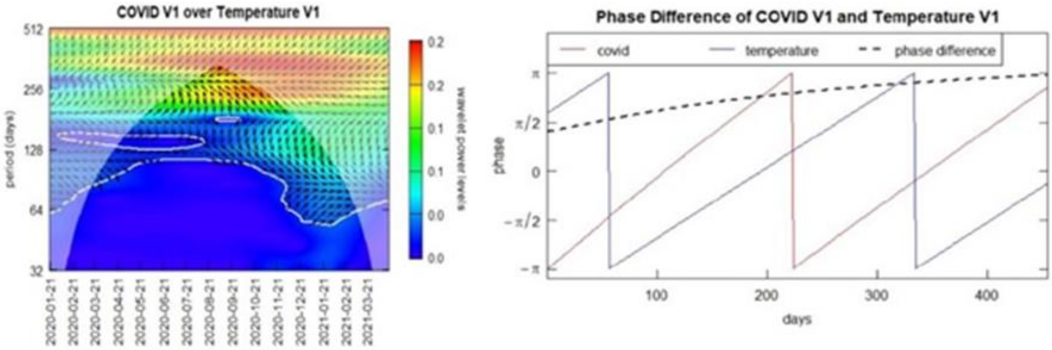

Figure 4c. Cross-wavelet analysis generated coherence plots (Covid-19 V1 & Temperature V3)

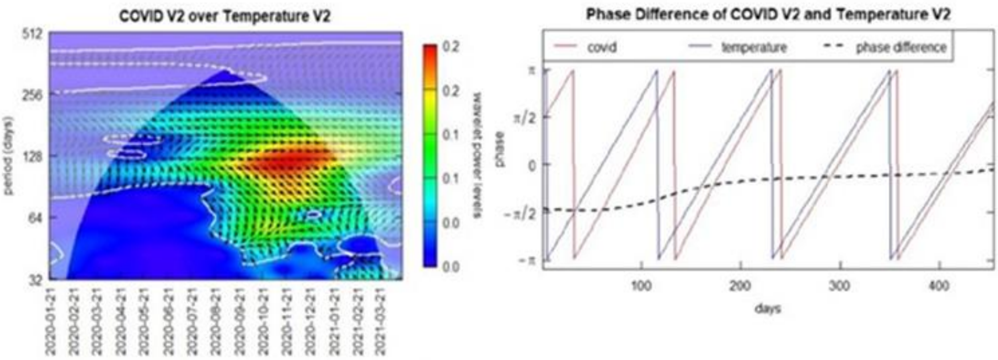

Figure 4d. Cross-wavelet analysis generated coherence plots (Covid-19 V2 & Temperature V2)

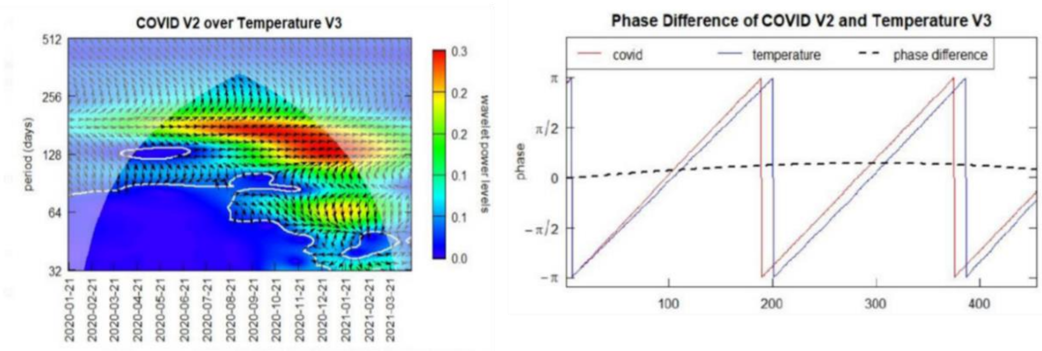

Figure 4e. Cross-wavelet analysis generated coherence plots (Covid-19 V2 & Temperature V3)

5. CONCLUSION

The complex oscillatory interactions of the environmental variables and the incidence of pandemic make it difficult to characterize the subtle synchronization of the coupled system. In contrast with standard method, we used wavelet coherence because it is appropriate for analyzing non stationary signals. The advantage of the proposed methodology is that it does not require the stationarity assumption of time-series which is often very difficult to fulfil. In this research, the space-time dynamics between the two time-series, namely the Covid-19 and the corresponding temperature is investigated in the time-localized frequency domain using SVD and analytic Morlet wavelet transforms. The results from the continuous cross-wavelet transform shows power spectrum strengths and coherence corresponding at various frequencies (periods). The coherence statistics suggest statistically significant relationship. The results also show varying phases and phase lags with leading and lagging behavior showing complex conjugate dynamics. Future studies focusing on spatially explicit mapping of coherence and other signal processing techniques e.g. S*ingular Spectrum Analysis* (SSA), *Empirical Mode Decomposition* (EMD) etc. could provide additional explanatory schemes and better understanding of the spatio-temporal dynamics of the disease.

REFERENCES

[1] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," SIAM journal on mathematical analysis, vol. 15, pp. 723-736, 1984.

[2] I. Daubechies, "Ten lectures on wavelets (SIAM, Philadelphia, 1992)," ed: Crossref, 1992.

[3] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," biometrika, vol. 81, pp. 425-455, 1994.

[4] J. B. Ramsey, "The contribution of wavelets to the analysis of economic and financial data," Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 357, pp. 2593-2606, 1999.

[5] P. K. Sahoo, M. A. Powell, S. Mittal, and V. Garg, "Is the transmission of novel coronavirus disease (COVID-19) weather dependent?," Journal of the Air & Waste Management Association, vol. 70, pp. 1061-1064, 2020.

[6] G. Kroumpouzos, M. Gupta, M. Jafferany, T. Lotti, R. Sadoughifar, Z. Sitkowska, et al., "COVID-19: A relationship to climate and environmental conditions?," Dermatologic therapy, vol. 33, pp. e13399-e13399, 2020.

[7] D. K. Rosario, Y. S. Mutz, P. C. Bernardes, and C. A. Conte-Junior, "Relationship between COVID-19 and weather: Case study in a tropical country," International journal of hygiene and environmental health, vol. 229, p. 113587, 2020.

[8] M. Wang, A. Jiang, L. Gong, L. Luo, W. Guo, C. Li, et al., "Temperature significant change COVID-19 transmission in 429 cities. medRxiv 2020.02. 22.20025791," ed, 2020.

[9] D. N. Prata, W. Rodrigues, and P. H. Bermejo, "Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil," Science of the Total Environment, vol. 729, p. 138862, 2020.

[10] B. Oliveiros, L. Caramelo, N. C. Ferreira, and F. Caramelo, "Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases," MedRxiv, 2020.

[11] T. Jamil, I. Alam, T. Gojobori, and C. M. Duarte, "No evidence for temperature-dependence of the COVID-19 epidemic," Frontiers in public health, p. 436, 2020.

[12] "COVID-19 Dataset," ed.

[13] "Temporal Temperature Dataset," ed.

[14] A. Roesch, H. Schmidbauer, and M. A. Roesch, "Package 'WaveletComp'," The Comprehensive R Archive Network2014, 2014.

[15] D. Maraun, J. Kurths, and M. Holschneider, "Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing," Physical Review E, vol. 75, p. 016707, 2007.

[16] M. Chavez and B. Cazelles, "Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data," Scientific reports, vol. 9, pp. 1-9, 2019.

AUTHORS

Iftikhar U. Sikder is an Associate Professor jointly appointed in the department of Information Systems and the Electrical Engineering and Computer Science at Cleveland State University, USA. He holds a PhD in Computer Information Systems from the University of Maryland, Baltimore. His research interests include soft computing, granular computing, spatial databases, spatio-temporal data mining and adaptive complex systems. His papers appeared in the journal of Knowledge-Based Systems, Risk Analysis, Expert Systems with Applications, Intl. journal of Digital Earth, International Journal of Mobile Communications, and Information Resources Management Journal, Intl. Journal of Management & Decision Making, and Intl. Journal of Aerospace Survey and Earth Sciences and many others journals. He has authored many book chapters and presented papers in many national and international conferences. Dr. Sikder is also currently serving in the editorial board of International Journal of Computational Models and Algorithms in Medicine and Intl. Journal of Computers in Clinical Practice.

James J. Ribero is an Adjunct Faculty at IBA, Dhaka University, Bangladesh. He holds MBBS, MS(Microbiology) and MBA from Dhaka University, Bangladesh. His research interests include applications of Machine Learning and Big Data technologies into medical and life sciences. He has authored book chapters, monographs, and presented papers in many national and international conferences. He was the former executive editor of The Orion (ISSN 1606-9722) medical journal.