چهارمین سمینار آمار فضایی و کاربردهای آن
۵–۶ آبان ماه ۱۴۰۰
27-28 Oct. 2021
4th Seminar on Spatial Statistics and Its Applications

The Functional Clustering of Daily Air Quality Indices in the First Month of Previous and Current Year of COVID-19 Prevalence in Tehran

Mohammad Fayaz[1*], Alireza Abadi[2,3]

[1]Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

[2]Department of Community Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

[3]Social Determinants of Health Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Abstract:

The COVID-19 impacts air quality indices during the lockdown by restricted policies for transformations, working hours and etc. Meanwhile, the different campaigns such as broadcasting advertising, social distancing and etc. help to inform people to control their behaviors. In this study, we focused on the daily air-quality indices before and in the early stage of national lockdowns that means in the first month of COVID-19 in Tehran. We compare the time series of this period with the exact previous year period with two functional clustering methods. The first one is based on the functional principal scores obtained by EM clustering, called EM Cluster and the second one is k-Centers Functional Clustering (kCFC). We study the clustering results with two indices: 1) the correct percentage of clustering and 2) adjusted Rand Index. The correct percentage of clustering for NO_2 is the highest among others, between 80 and 90% and it has decreasing pattern after 20 days. Finally, we conclude that the air pollution of Tehran like other important cities in the world reduced due to the social campaign and national lock-down.

Keywords: COVID-19, Functional Clustering, Air Quality, Tehran.

Mathematics Subject Classification (2010): 60Gxx, 60Hxx, 60Bxx.

---

*Speaker: Mohammad.Fayaz.89@gmail.com

1 Introduction

The impact of COVID-19 on climatology, air quality, air pollution or vice versa was studied with different study design such as questionnaires, sampling and etc. and various statistical methods such as Panel vector auto-regressive models (Panel VAR), Spatio-temporal models, classification methods, Generalized Additive Model (GAM) in Iran and its provinces (Ahmadi , 2020; Asna-ashary , 2020; Barbieri , 2020; Broomandi , 2020; Sahafizadeh , 2020; Srivastava , 2020; Tadbiri , 2020; Zareie , 2020). In this research, we study this effect by clustering the air-quality indices time series to discover the patterns. In this regard, there are various type of statistical methods and we choose the functional data analysis methods because they cluster the whole time series rather than independent multivariate points. (Chiou , 2007; Jacques , 2013).

2 Method

The hourly and daily air quality indices are observed in two periods of time: 1) From 2019-02-19 to 2019-11-04 and 2) From 2020-02-19 to 2020-11-04. There are 22 air stations in Tehran that capture eight following indices: Ozone (O_3), Carbon Monoxide (CO), Nitrogen Dioxide related indices (NO, NO_2 and NO_x), Sulfur Dioxide (SO_2) and suspended particulates smaller than 2.5 and 10 μm in aerodynamic diameter ($PM_{2.5}$ and PM_{10}). In this study, we select the first 30 days of the spreading COVID-19 in Iran and the exact previous year period for comparison purposes. The daily pattern of each index was calculated based on the hourly values and the missing values are omitted. The outliers are detected for each index as the 5^{th} and 95^{th} percentile of the distribution, and these two values are replaced to the lower and upper their values, respectively. The new variable, COVID status, was added that indicates the records from the first period or the COVID-19 period. The aim of this study is: firstly, it clusters daily curves of each index with considering the underlying functional structure of the data and secondly it compares the cluster results with the COVID-19 status and calculate the correct percentage of the clustering result and adjusted Rand Index. We use two algorithms for the first purposes: 1) Functional principal component analysis (FPCA) scores were obtained and the EM-Cluster method is used to cluster them and 2) k-Centers Functional Clustering (kCFC) method that directly clustering the functional data is used. These methods are extended the multivariate clustering to cover the functional properties of the observed time series. The second purpose was obtained by comparing the clustering membership result and the status of the COVID-19 in the contingency table and the correct rate of the table was computed as the number of correctly specified divided on the total observation and the adjusted RAND index is calculated. These indices show the ability of the clustering

method to identify the differences between time-series before and during COVID-19 and also, they can show that how air quality indices are changes due to the early stage of lock-down and social campaigns in the first 30 days of COVID-19. All calculations are done with R 4.1.0 and R studio with the following packages fdapace, EMCluster and pdfCluster. (Chiou , 2007; Jacques , 2013; Carroll , 2020; Chen , 2015; Adelchi, 2018)

3 Result

According to table 1, the correct rate and aRand of the two methods are the highest for NO_2 and they are between 80 and 90% and between 50 and 60% , respectively. The NO, NO_x and SO_2 have correct rate greater than 50% in two methods. These indices have the highest correct rate and it indicates that they are affected by the restriction of the COVID-19 more than others. The EM-Cluster was obtained with the Fraction of Explained Variation (FEV) of at least 90% and kCFC has FEV threshold equals 90%. There is one eigenfunction for some air indices. In the EM cluster method, the highest correct rate and aRand index are for SO_2 that are equal to 96.77% and 87.09% respectively. And the lowest correct rate and aRand are for O_3 that are near chance level and equal to 52.78% and -1.02%, respectively. In the KcFC, the highest correct rate and aRand are for PM_{10} that are equal to 97.06% and 88.23%, respectively. And the lowest correct rate and aRand are for CO and O_2 that are equal to 15.15% and -2.48% , respectively. The functional mean, first and second eigenfunctions of NO, NO_2, NO_x, $PM_{2.5}$ and PM_{10} are plotted in figure 1. The decreasing pattern of all indices after f0 days are observed.

4 Discussion

Previous studies showed that the COVID-19 lock-down has a significant impact on the air quality indices with spatial variations in Tehran (Broomandi , 2020) that is aligned with the most important cities in the world (Venter , 2020). This study shows that in the first 30 days of COVID-19 registries that the national lock-down was in the early stage, the reductions of air pollution indices are observed. Many studies investigate the relationship between COVID-19 transition rate, death and etc. with the weather parameters (Contini , 2020; Fattorini , 2020; Stanam , 2020; Travaglio , 2021; Wu , 2020; Zhang , 2020; Lau , 2020).

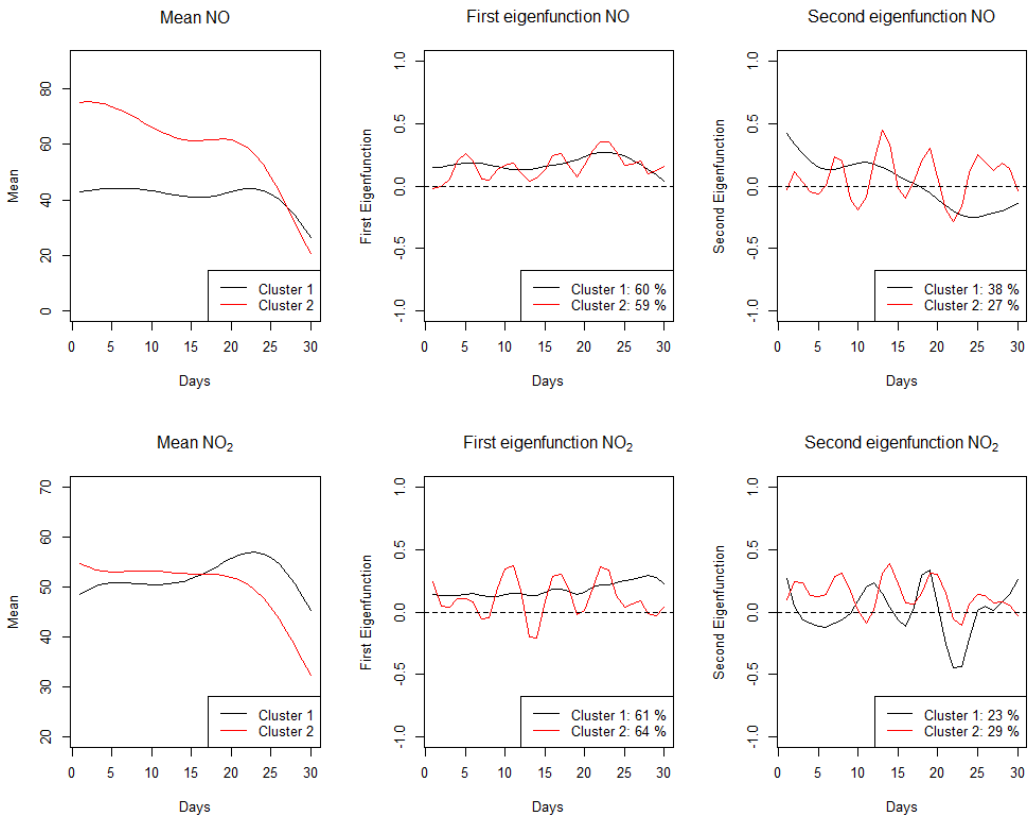

Figure 1: The Result of KcFC clustering.

Table 1: The results of clustering methods group by air quality indices and clustering method.

	EM Algorithm				KcFC					
					Cluster 1		Cluster 2		Correct	aRand
	FEV1	FEV2	Correct	aRand	FEV1	FEV2	FEV1	FEV2		
CO	0.82	0.12	69.70%	12.86%	76%	-	97%		15.15%	46.99%
NO	0.77	0.19	66.67%	8.71%	60%	38%	59%	27%	78.79%	31.05%
NO2	0.61	0.32	87.88%	56.04%	61%	23%	64%	29%	90.91%	65.90%
Nox	0.71	0.26	63.64%	4.82%	90%	-	56%	29%	81.82%	38.66%
O3	0.77	0.16	52.78%	-1.02%	65%	30%	68%	28%	47.22%	-2.48%
PM10	0.92	-	58.82%	1.62%	92%	-	58%	40%	97.06%	88.23%
PM25	0.87	0.07	42.42%	0.73%	83%	-	90%	-	80.65%	35.52%
SO2	0.90	0.09	96.77%	87.09%	70%	18%	72%	-	66.67%	8.51%

References

Ahmadi M, Sharifi A, Dorosti S, Ghoushchi SJ, Ghanbari N. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Science of the Total Environment. 2020;729:138705.

Asna-ashary M, Farzanegan MR, Feizi M, Sadati SM. COVID-19 outbreak and air pollution in Iran: A panel VAR analysis. Joint discussion paper series in economics; 2020.

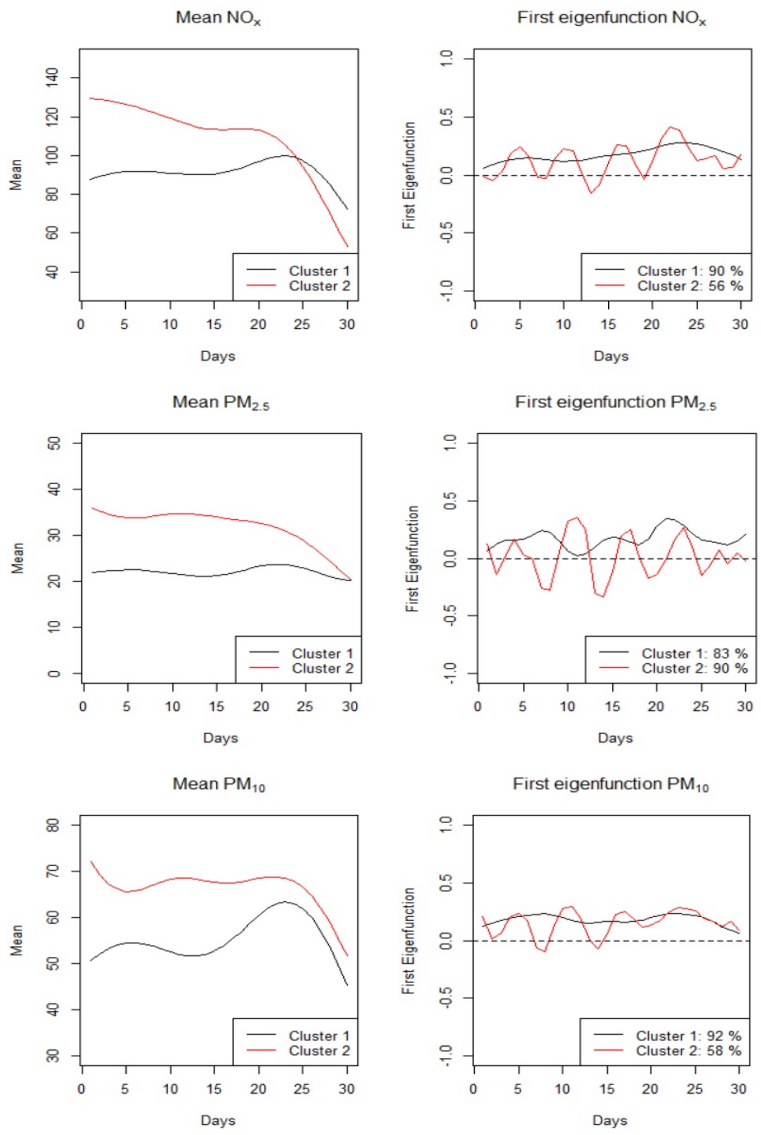

Figure 2: The Result of KcFC clustering.

Barbieri DM, Lou B, Passavanti M, Hui C, Lessa DA, Maharaj B, et al. Survey data regarding perceived air quality in Australia, Brazil, China, Ghana, India, Iran, Italy, Norway, South Africa, United States before and during Covid-19 restrictions. Data in brief. 2020;32:106169.

Broomandi P, Karaca F, Nikfal A, Jahanbakhshi A, Tamjidi M, Kim JR. Impact of COVID-19 event on the air quality in Iran. Aerosol and Air Quality Research. 2020;20(8):1793-804.

Sahafizadeh E, Sartoli S. High temperature has no impact on the reproduction number and new cases of COVID-19 in Bushehr, Iran. medRxiv. 2020.

Srivastava A. COVID-19 and air pollution and meteorology-an intricate relationship: A review. Chemosphere. 2020:128297.

Tadbiri H, Moradi-Lakeh M, Naghavi M. All-cause excess mortality and COVID-19-related deaths in Iran. Medical journal of the Islamic Republic of Iran.2020;34:80.

Zareie B, Roshani A, Mansournia MA, Rasouli MA, Moradi G. A model for COVID-19 prediction in Iran based on China parameters. MedRxiv. 2020.

Chiou JM, Li PL. Functional clustering and identifying substructures of longitudinal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007;69(4):679-99.

Jacques J, Preda C. Funclust: A curves clustering method using functional random variables density approximation. Neurocomputing. 2013;112:164-71.

Carroll C, Gajardo A, Chen Y, Dai X, Fan J, Hadjipantelis P, et al. fdapace: functional data analysis and empirical dynamics. R package version 05. 2020;5.

Chen W, Maitra R. EMCluster: EM algorithm for model-based clustering of finite mixture Gaussian distribution. R Package. 2015.

Azzalini, Adelchi, Giovanna Menardi, and Maintainer Giovanna Menardi. "Package pdfCluster." (2018).

Venter ZS, Aunan K, Chowdhury S, Lelieveld J. COVID-19 lockdowns cause global air pollution declines. Proceedings of the National Academy of Sciences. 2020;117(32):18984-90.

Contini D, Costabile F. Does air pollution influence COVID-19 outbreaks? : Multidisciplinary Digital Publishing Institute; 2020.

Fattorini D, Regoli F. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. Environmental Pollution. 2020;264:114732.

Stanam A, Chaudhari M, Rayudu D. Effects of temperature on COVID-19 transmission. Medrxiv. 2020.

Travaglio M, Yu Y, Popovic R, Selley L, Leal NS, Martins LM. Links between air pollution and COVID-19 in England. Environmental Pollution. 2021;268:115859.

Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States. MedRxiv. 2020.

Zhang Z, Xue T, Jin X. Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities. Science of the Total Environment. 2020;741:140244.

Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Ichii H, Zacharski M, et al. The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. Journal of Microbiology, Immunology and Infection. 2020;53(3):467-72.