A new multivariate functional ANOVA approach for assessing air quality data amid COVID-19 pandemic

Un nuovo approccio funzionale dell'ANOVA multivariata per la valutazione dei dati sulla qualità dell'aria durante la pandemia COVID-19

Adelia Evangelista, Christian Acal, Ana M. Aguilera, Annalina Sarra, Tonio Di Battista and Sergio Palermi

Abstract To reduce the SARS-CoV-2 virus spreading, worldwide governments implemented a series of restriction measures that led to a downturn in several economic sectors. Recent studies have instead documented a respite to the environment. In this work, we evaluate the impact of lockdown on air quality of the urban area of Chieti-Pescara (Abruzzo region, Italy). To this end, we adopt a functional data analysis approach. Specifically, to check the differences between the temporal evolution of different pollutants (PM_{10}, $PM_{2.5}$, NO_2 and benzene) in terms of the location of measuring stations, a novel approach for multivariate FANOVA for independent measures, is proposed. The results obtained reveal changes in pollutants behaviour during the lockdown period.

Abstract *Per ridurre la diffusione del virus SARS-CoV-2, i governi di tutto il mondo hanno implementato una serie di misure restrittive che hanno portato a una recessione in diversi settori. Recenti studi hanno, invece, documentato una tregua per l'ambiente. In questo lavoro, valutiamo l'impatto del lockdown sulla qualità dell'aria dell'area urbana di Chieti-Pescara (regione Abruzzo, Italia). A tal fine, adottiamo un'analisi funzionale dei dati. Nello specifico, per verificare le differenze tra l'evoluzione temporale di diversi inquinanti (PM_{10}, $PM_{2.5}$, NO_2 e benzene) a seconda della posizione delle stazioni di monitoraggio, si propone un nuovo approccio basato sulla FANOVA multivariata per misure indipendenti. I risultati ottenuti rivelano cambiamenti nel comportamento di tutti gli inquinanti durante il periodo di lockdown.*

A. Evangelista, A. Sarra, T. Di Battista
Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University G. d'Annunzio, V.le Pindaro, 42; 65127 Pescara (Italy), e-mail: asarra@unich.it, e-mail: adelia.evangelista@unich.it, e-mail: tonio.dibattista@unich.it
C. Acal, A.M. Aguilera
Department of Statistics and O.R. and IEMath-GR, University of Granada (Spain), e-mail: chacal@ugr.es, e-mail: aaguilera@ugr.es
S. Palermi
Agency of Environmental Protection of Abruzzo (ARTA), V.le G. Marconi, 51; 65127 Pescara (Italy), e-mail: s.palermi@artaabruzzo.it

Key words: Air pollution, COVID-19/Lockdown, FDA/Functional Data Anlysis, Public health, Multivariate ANOVA

1 Introduction

Starting from the first outbreak, first identified in Wuhan (China), in late December 2019, the Coronavirus disease affected the entire world, posing major public health and governance concerns. Due to the wide spread of this pandemic disaster, authorities enforced different measures, resulting in prohibitions of various aspects of human activities. Global and local economy had intense damaging, especially in sectors such as tourism, commodity markets and transportation. On the other hand, several works have reported that highly industrialized zones of the world observed a remarkable reduction in air pollution, essentially due to restrictions placed upon industrial activities and the dropping in road transport. To this regard, [1] detected a significant reduction of air pollution in various areas of China and India; [2] studied the behaviour of the levels of pollutants across USA while [3] reported a sharp reduction in air pollution in large parts of Europe. Following these lines of research, in this paper, we investigate the effects of the quarantine policies adopted by the Italian Government on air quality in the urban area of Chieti-Pescara (Abruzzo region, Italy). By carrying out a functional data analysis (FDA), we compare the behaviour of different air pollutants in two different periods of time: before lockdown and during lockdown days. Starting from the foundations given by [4, 5], the FDA has been extensively used in the last decades, also in environmental studies, because FDA paradigm makes it possible to work with the entire time spectrum of pollutants time series, bringing additional information to be recovered from the data than in the vectorial approach (see, among others, [6], [7]). Within the methodological FDA framework, we proposed here a novel approach based on the Functional Analysis of Variance (FANOVA) for independent measures. The rest of the paper is organised as follows: Section 2 gives the description of the area under study and the air pollutants. Section 3 briefly defines the new methodology proposed while in Section 4 the main results obtained are illustrated. Finally, in Section 6 there are some concluding remarks.

2 Area of study

In this work, we take into account the metropolitan area of Chieti-Pescara (along the Adriatic coast of central Italy), defined of critical importance in terms of environmental pollution. Chieti-Pescara conurbation represents one of the most important industrial pole of the Abruzzo. In the last years, this area has registered an increase of industrial activity, as well as of the urban development. All these aspects make the metropolitan area the locus of growing environmental concerns for the high level of resource consumption, greenhouse gas emissions and air quality pollution.

2.1 Air pollution data

Air pollution data for this analysis consists in hourly measurements of PM_{10}, $PM_{2.5}$, NO_2 and benzene, detected by the Regional Agency for the Environmental Protection (ARTA) of Abruzzo by means of five monitoring stations of the regional air quality network. The air quality monitoring sites of Pescara (Teatro d'Annunzio), Chieti and Francavilla are defined as *Urban Background type* (UB) due to their spatial location not affected by important pollution sources. On the other hand, the monitoring stations of Pescara (Via Firenze) and Montesilvano are named as *Urban Traffic type* (UB) because are nearby roadside, so they are influenced by traffic emissions. The pollutants data collected have been divided in two time intervals: the first is defined *pre-lockdown period* starting from the 1st February 2020 to the 10th of March 2020; the second is named *during-lockdown period* from the 11st of March to the 18th of April 2020.

Table 1 Net and % variation of pollutants concentration levels in the urban area of Chieti-Pescara

	UT			**UB**	
Net variation	**fi**	**mo**	**th**	**ch**	**fr**
NO_2	-13.9	-14.7	-21.2	-10.3	-7.6
PM_{10}	5.1	3.7	5.7	4.3	7.3
$PM_{2.5}$	2.9	2.2	3.1	4.4	4.1
Benzene	-0.31	-0.15	0.22	0.18	0.04
% variation					
NO_2	-57.9	-58.7	-65.2	-54.8	-49.1
PM_{10}	20.5	16.8	22.0	19.4	40.8
$PM_{2.5}$	19.0	15.6	19.8	26.7	34.4
Benzene	-32.57	-27.56	40.06	19.63	4.27

Acronyms of monitoring stations:

fi=Via Firenze; mo=Montesilvano; th=Teatro d'Annunzio; ch=Chieti; fr=Francavilla al Mare

An initial investigation of the net and percentage variations of the four pollutants in the different monitoring sites, before and during lockdown, is given in Table 1. The analysis shows that NO_2 recorded a significant and marked reduction both in traffic and in background monitoring stations. This is in line with our expectations, due to the collapse of vehicular traffic after the measures imposed by the government. The levels of particulate matter (PM_{10} and $PM_{2.5}$) registered an increment in all measuring sites. This could be reasonable knowing the peculiarity of this pollutant. Finally, for benzene, we observed an opposite behaviour: for the UT stations the pollutant decreases, while it increases in the BT stations.

3 Methodological framework

In FDA approaches, the data analyzed are curves, or more typically functions, that varying over time, space, or other continuous support. The first approach proposed in this work has the aim of testing the equality, across two different conditions or periods, of mean functions related to a unique functional variable. Let $X_{jr}(t)$ be the sample functions, where $t \in T = [a,b]$ is the temporal time interval, $j = 1,...,n$ is the sample unit and $r = 1,...,R$ represents the number of the different periods or time (or conditions) to compare. In the current work, we compare only two different periods ($R = 2$). We define with $\mu_r(t) = E[X_{jr}(t)]$ the mean function associated to each functional variable in each condition or time period. The aim is to test $H_0 : \mu_1(t) = \mu_2(t)\ \forall t \in [a,b]$, against the alternative that its negation holds. The two statistics proposed to carry out the hypothesis testing are those introduced by [8] which take into account simultaneously the between and within variabilities. Smaga's statistics are defined as:

$$\mathscr{D}_n = n \int_T \frac{\left(\overline{X}_1(t) - \overline{X}_2(t)\right)^2}{\hat{K}(t,t)}\, dt,$$

$$\mathscr{E}_n = sup_{t \in [a,b]} \left\{ \frac{n\left(\overline{X}_1(t) - \overline{X}_2(t)\right)^2}{\hat{K}(t,t)} \right\},$$

where $\hat{K}(t,t) = \frac{\sum_{j=1}^{n}\left[(X_{j1}(t) - \overline{X}_1(t)) - (X_{j2}(t) - \overline{X}_2(t))\right]^2}{n-1}$.

$\mathscr{D}_n$ and $\mathscr{E}_n$ can be computed by considering the basis expansion. A further stage in our analysis was to test the equality of the multivariate dimensional mean functions for independent groups (g). To comply with this aim, we consider the multivariate FANOVA for independent measures, introduced and described in detail in [9]. An important result about FANOVA has been obtained by [10] which demonstrate that FPCA of $X(t)$ is the same as to apply a Multivariate PCA on the matrix $A\Psi^{1/2}$. The hypothesis to test is:

$$\left\{ H_0 : \mu_1(t) = ... = \mu_g(t)\ \forall t \in [a,b]. \right.$$

against the alternative that its negation holds. In this setting, two problems are encountered. First, the multivariate homogeneity tests do not perform well with high dimensional vectors and the number of basis functions needed for an accurate approximation of sample is usually high. As a solution, we propose to test the multivariate homogeneity on the vectors of the most explicative principal components scores. This new methodology proposes an extension of the novel parametric and nonparametric approaches introduced using Functional Principal Component Analysis for univariate functional data ([11]) to the multivariate case.

4 Results

The impact of lockdown measures on air quality has been investigated with the functional testing procedures described in Sect. 3. To carry out the analysis, we convert the discrete values into curves, by means of cubic B-spline smoothing with 20 basis functions. Firstly, we adopted a FANOVA for repeated measures to statistically prove the results obtained in Table 1. The statistics $\mathscr{D}_n$ and $\mathscr{E}_n$ are used to test the within and between group variability and the p-values obtained by means of permutation tests. The results of the tests are significant for NO_2, PM_{10} and $PM_{2.5}$, suggesting that there are differences for these pollutants in the means curves before and during the lockdown period. Not statistical significance was found, instead, for benzene. However, considering that the results of hypotheses testing for benzene are very close to the limit region and taking into account the small sample size, we can also conclude that there are also differences in the means curves of benzene for the two time periods. A second step of our study consists in the multivariate analysis of variance for independent measures. We tested if there are differences between the temporal evolution of all pollutants in terms of the location of measuring station. According to the results displayed in in Table 2, significant differences were found in terms of the location of the monitoring stations in relation to PM_{10} (before lockdown) and benzene (during lockdown).

Table 2 Multivariate FANOVA for independent measures

p-value	**BL**	**DL**
All pollutants	0.000	0.302
NO_2	0.562	0.272
PM_{10}	0.000	0.306
$PM_{2.5}$	0.889	0.685
Benzene	0.186	0.000

Acronyms:
BL=Before Lockdown; DL=During Lockdown

5 Conclusion

In this work, changes in air pollution during the COVID-19 pandemic have been evaluated by means of a functional analysis of variance with univariate repeated measures and multivariate independent measures. The results has proven a possible misclassification of air monitoring stations in the urban area of Chieti-Pescara, probably due to the NO_2. since the proposed technique failed to discriminate between UB and UT measuring sites, despite the fact that NO2, in urban areas, is a pollutant mostly produced by traffic emissions. This result is of great importance for environmental protection agencies which should identify the presence of redundant or misclassified monitoring sites, in order to reduce the cost of pollution monitoring and ensure the integrity and accuracy of air pollution information.

Acknowledgements The authors thank ARTA for providing the data. This research was funded by project PID2020-113961GB-I00 of the Spanish Ministry of Science and Innovation (also supported by the FEDER program), project FQM-307 of the Government of Andalusia (Spain) and the PhD grant (FPU18/01779) awarded to Christian Acal. The authors also thank the support of the University of Granada, Spain, under project for young researchers PPJIB2020-01.

References

1. Wang, P., Chen, K., Zhu, S., Wang, P., Zhang, H.: Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resour. Conserv. Recy. **158**, 104814 (2020) doi: 10.1016/j.resconrec.2020.104814
2. Berman, J. D., Ebisu, K.: Changes in U.S. air pollution during the COVID-19 pandemic. Sci. Total. Environ. **739**, 139864 (2020) doi:https://doi.org/10.1016/j.scitotenv.2020.139864
3. Sicard, P., De Marco, A., Agathokleous, E., Feng, Z., Xu, X., Paoletti, E., Rodriguez, J. J. D., Calatayud, V.: Amplified ozone pollution in cities during the COVID-19 lockdown. Sci. Total. Environ. **735**, 139542 (2020) doi: https://doi.org/10.1016/j.scitotenv.2020.139542
4. Ramsay, J.O., Silverman, B.W.: Applied Functional Data Analysis: Methods and Case Studies. Springer-Verlag, New York (2002)
5. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, second ed. Springer-Verlag, New York (2005)
6. Escabias, M., Aguilera, A. M., Valderrama, M. J.: Modeling environmental data by functional principal component logistic regression. Environmetrics **16**, 95-107 (2005) doi: https://doi.org/10.1002/env.696
7. Martínez Torres, J., Pastor Peréz, J., Sancho Val, J., McNabola, A., Martínez Comesaña, M., Gallagher, J.: A Functional Data Analysis Approach for the Detection of Air Pollution Episodes and Outliers: A Case Study in Dublin, Ireland. Mathematics **8**, (2020) doi:/doi.org/10.3390/math8020225
8. Smaga, L.: A note on repeated measures analysis for functional data. AStA Adv. Stat. Anal. **104**, 117-139 (2020) doi: https://doi.org/10.1007/s10182-018-00348-8
9. Acal,C., Aguilera, A. M., Sarra, A., Evangelista, A., Di Battista, T., Palermi, S.: Functional ANOVA approaches for detecting changes in air pollution during the COVID-19 pandemic. Stoch. Environ. Res. Risk. Assess. 1-19 (2021) doi: https://doi.org/10.1007/s00477-021-02071-4
10. Ocaña, F. A., Aguilera, A. M., Escabias, M.: Computational considerations in functional principal component analysis. Comput. Stat. **22**, 449-465 (2007) doi:10.1007/s00180-007-0051-2
11. Aguilera, A. M., Acal, C., Aguilera-Morillo, M. C., Jimènez-Molinos, F., Roldàn, J. B.: Homogeneity problem for basis expansion of functional data with applications to resistive memories. Math. Comput. Simul. **186**, 41-54 (2021) doi: https://doi.org/10.1016/j.matcom.2020.05.018