Lietuvos matematikos rinkinys
Proc. of the Lithuanian Mathematical Society, Ser. A
Vol. 64, 2023, pages 36–53
https://doi.org/10.15388/LMR.2023.33658

Vilnius University Press

Air quality investigation using functional data analysis methods*

Akvilė Vitkauskaitė, Milda Salytė

Faculty of Mathematics and Informatics, Vilnius University
Naugarduko 24, LT-03225 Vilnius, Lithuania
E-mail: akvile.vitkauskaite@mif.stud.vu.lt; milda.salyte@mif.stud.vu.lt

Received July 10, 2023; published online November 20, 2023

Abstract. In this research paper, a comprehensive analysis of particulate matter (PM10) and nitrogen dioxide (NO2) pollution concentrations in six different Lithuanian regions is presented. The analysis employs data smoothing, principal component analysis (PCA), exploratory data analysis, hypothesis testing, and time series analysis to provide a thorough examination. Functional data analysis approaches were used to find the origins and effects of these air pollutants by revealing their data patterns. The functional data analysis techniques demonstrate their effectiveness in revealing deep links within large datasets, assisting in the control of air quality problems. This research provides valuable insights into air quality challenges in Lithuanian regions. The study, aimed at comparing air quality across different regions, indicates that there are no significant differences in PM10 and NO2 between the two groups. Notably, reliable forecasts for 2023 data are attainable for PM10 in regions such as Vilnius Old Town, Vilnius Lazdynai, Šiauliai, and Klaipėda. For NO2, successful forecasting can be applied to Vilnius Old Town, Vilnius Lazdynai, and Šiauliai.

Keywords: air pollutants; particulate matter; nitrogen dioxide; functional data analysis; exploratory data analysis, hypothesis testing

AMS Subject Classification: 62P12

Introduction

Air pollution is a global concern with significant implications for human health and ecosystems. Each year, over 4 million premature deaths are attributed to outdoor

* The report was presented at 64th Conference of the Lithuanian Society of Mathematicians.

air pollution, primarily caused by fine particles known as PM2.5. However, the composition and toxicity of PM2.5 vary across different locations and over time. To effectively address this issue, researchers and policymakers need to determine the most hazardous constituents of air pollution in specific regions and prioritize their mitigation [1]. Apart from PM2.5, PM10 is another significant air pollutant. The research by Ottaviano *et al.* [4], investigated the impact of PM10 on upper airway acute (UA) illnesses. This study showed a correlation between levels of PM10 and the frequency of referrals for specific UA disorders, suggesting that PM10 can help forecast and manage health-related effects. The study conducted by Vaičiulis *et al.* [7] found that higher PM10 levels in the air, occurring 5–11 days before acute myocardial infarction, were associated with a significant increase in the risk of fatal AMI (Acute Myocardial Infarction), particularly during the winter-spring period. Furthermore, nitrogen dioxide (NO2) is related to various health problems. In 2020, Ogen [3] proposed a potential connection between NO2 exposure and deaths related to the coronavirus.

In this paper two of the most often detected air pollutants, particulate matter (PM10) and nitrogen dioxide (NO2) are investigated in different regions of Lithuania. These two pollutants have been linked to heart diseases and other health issues. We can identify locations with high pollution levels and adopt policies to enhance the quality of the air, for example reducing emissions from transportation and industry and boosting the use of renewable energy sources. By monitoring these pollutants, we can improve the environment for both people and ecosystems.

Literature review

The 2012 study by Shaadan *et al.* [6] highlighted the advantages of using a functional data approach to assess and compare PM10 pollutant behaviour during extreme haze years in Selangor, Malaysia. The study revealed implicit information, previously unseen with conventional methods, about the different behaviours of PM10 within and between the study years. The analysis provided evidence of the impact of climate change, emission sources, and meteorological conditions on the severity of the PM10 problem and proposed the functional depth method for detecting critical exceedance days. Wang *et al.* [8] used functional data analysis to examine PM2.5 concentrations in China. Within the context of FDA, the methodology comprised the application of roughness penalty to smooth PM2.5 pollution functions, employing adaptive weighting clustering analysis to categorize fluctuations, and conducting functional ANOVA to assess the significance of differences among various regions. The findings revealed significant changes in concentration patterns among locations, emphasizing the significance of improved methods for evaluating PM2.5 data. The importance of multiple modeling approaches was emphasized in practical recommendations for government policy on air quality regulation. Additionaly, Torres *et al.* [2] evaluated several analytical strategies for detecting air pollution events and outliers. The study demonstrated the effectiveness of functional data analysis in detecting patterns and outliers in nitrogen dioxide concentrations, highlighting the limits of traditional approaches as well as the advantages of functional data analysis for comprehensive air pollution control. Rigueira *et al.* [5] studied the application of functional data analysis for detecting outliers and evaluating the impact of the COVID-19 outbreak on air quality in Gijón, Spain in 2022. The study compared the methodology of classical analysis, statistical process control, and functional data analysis, revealing that functional data analysis outperformed the other two methods in spotting outliers in high-variability data. The suggested outlier identification method, which is based on functional directional outlyingness, was effective in detecting abnormalities in air quality data. The results were validated using real-world data from Gijón, Spain, and further research areas were recommended to improve the model's performance and widen its application to other datasets.

Like in Wang *et al.* [8] study, we also used data smoothing technique (and used Fourier basis as it was done by Torres *et al.* [2]) and tested the significance of difference among different regions using ANOVA. We also performed outliers detection as similar outliers' analysis were performed by Torres *et al.* [2] and by Rigueira *et al.* [5].

The cited literature provides a comprehensive basis for understanding the significance of air quality, provides valuable insights into various methods, goals and motives for its analysis. This work is the catalyst for further detailed motivation that explains the main research objectives in the next section.

Motivation and main objectives

Understanding and analysing air quality data is important for human health and overall well-being. The significance of air quality data is exemplified by its inclusion in apartment ads on the popular Lithuanian real estate website aruodas.lt, emphasizing a growing awareness of its importance among the public. With these motivations, our research is driven by two key goals: comparing air quality across different regions to identify potential differences in pollution levels and forecasting average levels of particulate matter (PM10) and nitrogen dioxide (NO2) for each month of 2023. These objectives not only contribute valuable insights to the current understanding of air quality but also address practical concerns related to environmental health and urban planning.

Data and dataset preparation

Our chosen objective was to examine air quality measurements across diferent regions in Lithuania. Thus, we strategically selected six distinct regions:

- Vilnius, Old Town
- Vilnius, Lazdynai
- Kaunas, Noreikiškės
- Klaipėda, City Center
- Šiauliai
- Mažeikiai

Our focus was on two key pollutants, particulate matter (PM10) and nitrogen dioxide (NO2), chosen based on the publicly available data from the Environment Protection Agency Lithuania.

The selection of these regions and pollutants was based on the need for data completeness and comparability. The Environment Protection Agency Lithuania provides daily data from 17 air quality stations, measuring concentrations of various pollutants, including PM10, carbon monoxide, sulfur dioxide, NO2, and ozone. Initially, our priority was to ensure high-quality final dataset. Therefore, we prioritized regions and pollutants where complete data for both PM10 and NO2 were available in order to make possible meaningful comparisons of air quality indicators across selected regions.

The process of collecting this data presented its own set of challenges. The daily data files, stored in PDF format, posed a difficulty in terms of extraction and data consolidation. To overcome this, we developed a data scraping solution using the node.js framework. This solution allowed us to efficiently download PDF files for multiple years from the Environment Protection Agency Lithuania's website, automating a task that would have been time-consuming if done manually.

Subsequently, we implemented a Python script to handle the extraction and joining of data from the downloaded PDF files. This script systematically read and explored the content of each PDF file, specifically searching for identifiable tables. Once identified, the script proceeded to extract and interpret the tabular data, creating a distinct dataframe for each table and then consolidating these dataframes into one.

Additionally, despite our efforts to achieve data completeness, we encountered some missing values (NA) in the obtained data. To address this issue, we conducted data interpolation and selected a timeframe from October 1, 2019, to April 30, 2023. This timeframe was chosen strategically, considering the data quality and completeness, and served as the basis for our analytical methods.

The final dataset is a table. It consists of daily air quality measurements for PM10 and NO2 in the selected regions, ordered chronologically from oldest to newest (data is showed in Fig. 1). On randomly selected dates, we cross-checked our scraped data with information published on the Environment Protection Agency Lithuania's website.

Fig. 1. Interpolated data of PM10 and NO2.

Our dataset, which includes data from six sites on two pollutants, stands out for its comprehensive content and the sophisticated automation used in its gathering. The complex procedure of extracting and consolidating daily data from PDF files from October 2019 to April 2023 highlights the exceptional nature of our dataset. Manual extraction would have been an enormous task that we successfully avoided.

In summary, our dataset creation approach involved strategic region and pollutant selection, automated data scraping, thorough data consolidation, and data validation processes to generate a comprehensive dataset. The dataset provides the groundwork for meaningful analyses focused on understanding air quality issues in Lithuania.

Case study

Smoothing

We used the Fourier basis and harmonic acceleration methods to refine the patterns of air quality data. The Fourier basis is a mathematical method which simplifies complex data by identifying recurring patterns or frequencies. The harmonic acceleration operator helps to smooth out these patterns and reduce unnecessary oscillations or irregularities. This combination was chosen because air quality data often contains repeating seasonal elements that can be effectively captured by these methods. For instance, the Fourier basis allows us to extract seasonal variations in pollutant levels, while harmonic acceleration helps smooth these variations, reducing minor spikes or dips, thus highlighting larger trends more accurately.

The choice of the lambda value, set at 1e6, significantly influences the degree of smoothing. The choice of lambda value at 1e6 is relatively high. It corresponds to strong smoothing, significantly reducing the noise and emphasizing overarching trends within the PM10 and NO2 measurements. Thus, lambda serves as an adjustable factor to capture significant changes without disturbing the overall pattern of the data set.

This data smoothing method aimed to improve the data quality by reducing noise and emphasizing underlying trends. This preparation step allowed for a more comprehensive understanding of the differences in PM10 and NO2 measurements in different regions. The PM10 data from six different regions was plotted on a single graph, offering a clearer view without overlap of NO2 data, while NO2 from these regions was similarly showed in the graph next to it in Fig. 2. This allowed for a better comparative analysis by visualizing each variable's trends distinctly.

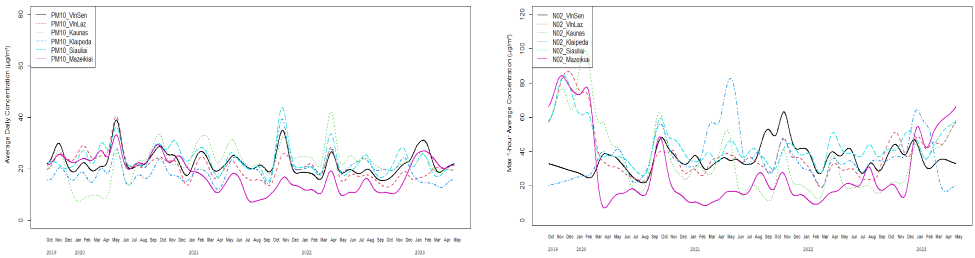

Fig. 2. Smoothed data of PM10 and NO2.

Outliers detection

Outliers detection was conducted using the Modified Band Depth (MBD) method to identify the tendencies and similarities among different regions for PM10 and NO2 data. The MBD assesses how curves in a dataset relate to one another, pinpointing similarities or differences. This step seeks to highlight a region with akin trends, contributing to the primary objective of comparing air quality across distinct regions.After this process, only three regions in each pollutant group showed analogous data tendencies: Vilnius Senamiestis, Vilnius Lazdynai, and Šiauliai for PM10, while Vilnius Senamiestis, Vilnius Lazdynai, and Klaipėda demonstrated similar patterns for NO2. This information was visually depicted in the graph (Fig. 3), facilitating the comparison of these trends.

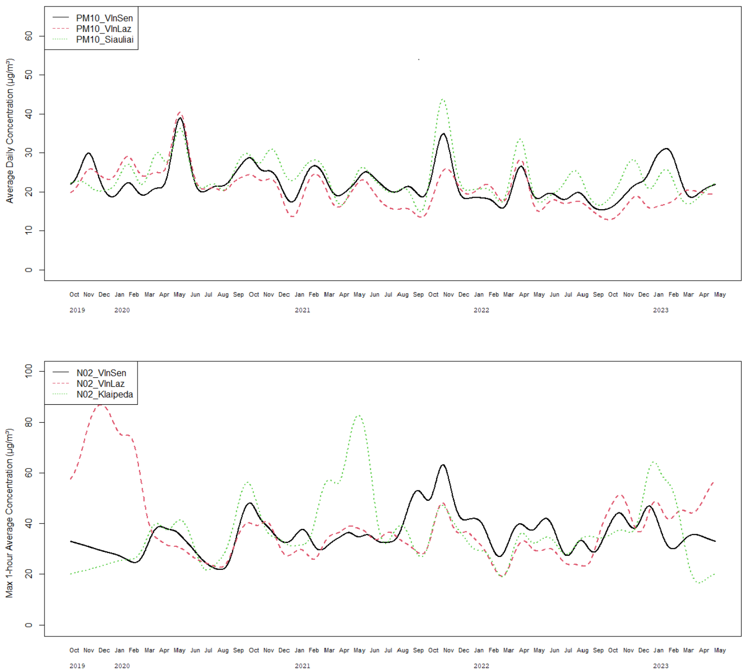

Fig. 3. Smoothed data without outliers.

Principal component analysis

Principal component analysis (PCA) is a statistical method for reducing data dimensionality by identifying underlying patterns, and decreasing the volume-based complexity of data. For PM10 and NO2 across six regions, PCA can uncover connections between the regions and different time points, providing insights into patterns or trends.

Varimax rotation is an approach that simplifies PCA results by increasing the variance of the squared loadings. It makes results more interpretable. It reorients the principal components to ensure each variable aligns more strongly with one principal

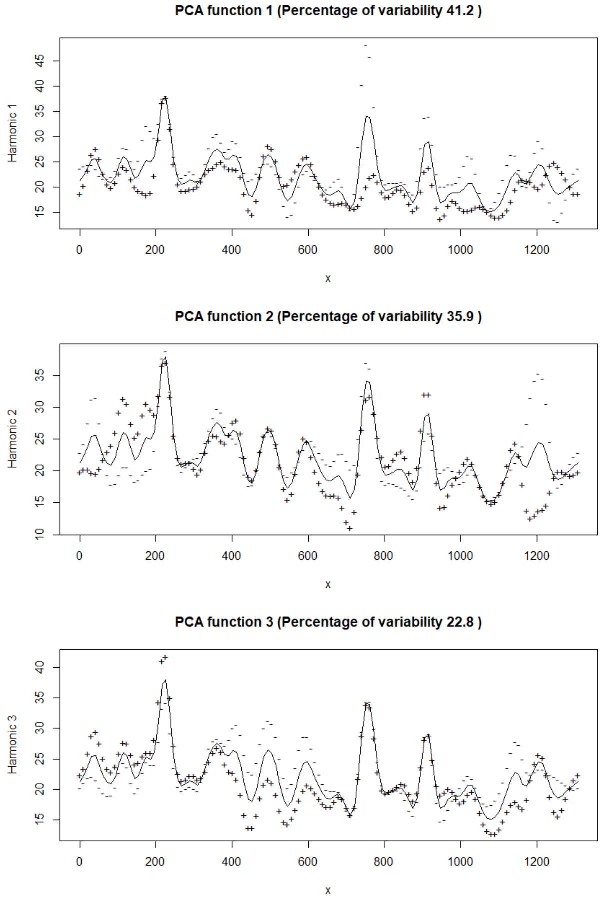

Fig. 4. Principal component analysis on smoothed particular matter data.

component and weakly with others, aiding in clearer interpretations of the relationships between variables. PCA was applied to the smoothed data using the first three harmonics. The graphs illustrate these harmonics and the PCA results in Fig. 4 and Fig. 5. The displayed percentage of variability represents the amount of variance explained by each principal component. Varimax rotation to the PCA object was applied using the varmx.pca.fd function. The resulting plots show the rotated PCA objects. The solid line represents the pollutant concentration, while the dotted and dashed lines illustrate how the addition or subtraction of a multiple of each principal component curve impacts the data.

Statistical comparisons in air quality analysis

Hypothesis testing – ANOVA

Data was splitted into two groups. The first group consisted of the big cities – Vilnius (Old Town), Vilnius (Lazdynai) and Kaunas (Noreikiškės), and the second –

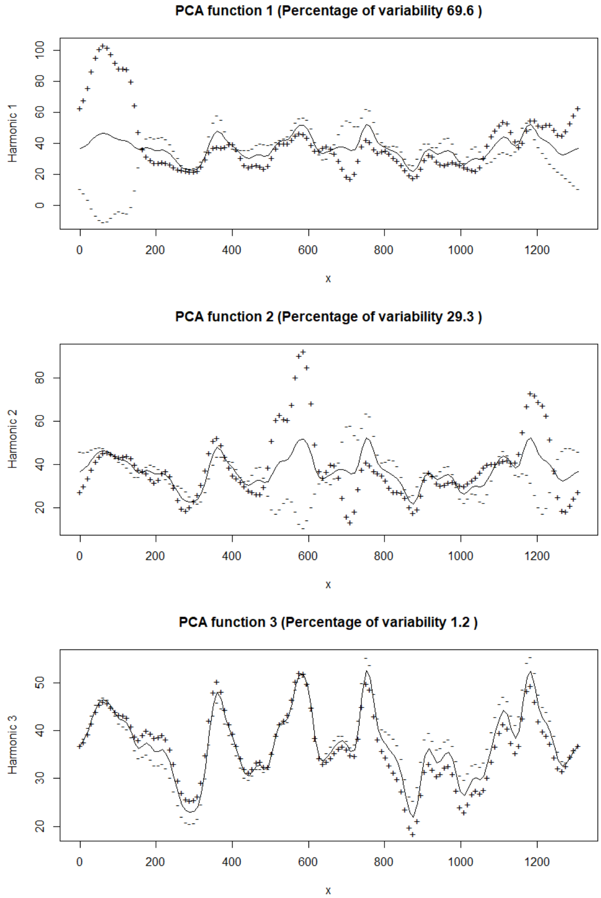

Fig. 5. Principal component analysis on smoothed nitrogen dioxide data.

Klaipėda (City Center), Šiauliai and Mažeikiai. This clustering strategy was aimed at simplifying the analysis by dividing the largest cities for a collective study and grouping Vilnius and Kaunas due to their proximity.

ANOVA (Analysis of Variance) hypothesis testing is a statistical method used to assess whether there are any statistically significant differences between the means of two or more groups. While analyzing air quality data, ANOVA can be applied to examine differences in pollution levels between regions. After formulating the null and alternative hypotheses, ANOVA testing provides a structured method for assessing the significance of differences in air quality measurements and provides a comprehensive understanding of pollution patterns across locations.

Therefore, we formulated hypotheses to examine the potential differences in particulate matter (PM10) and nitrogen dioxide (NO2) levels between two groups. The null hypothesis (H_0) propose that there are no significant differences in the air quality measurements among the groups, while the alternative hypothesis (H_1) suggests the presence of a statistically significant distinction.

$$\begin{cases} H_0: \text{There are no significant differences in } PM_{10} \text{ and } NO_2 \text{ levels among the groups.} \\ H_1: \text{There is a statistically significant distinction in } PM_{10} \text{ and } NO_2 \text{ levels among} \\ \quad \text{the groups.} \end{cases}$$

In both cases we found out that we cannot reject the null hypothesis at the 5% level of significance. This analysis suggests that there are no significant difference in nitrogen dioxide and particular matter levels between the groups (Vln_Kns and Klp_Sl_Maz). Multiple tests, including FP, CH, CS, L2N, L2B, L2b, FN, FB, Fb, and GPF, consistently fail to demonstrate any significant distinctions (Table 1 and Table 2).

Table 1. Analysis of variance for functional data (NO2).

Test	Test statistic	p-value
FP	0.7250366	0.818
CH	743465.6	0.6473
CS	743465.6	0.5358
L2N	371732.8	0.5333609
L2B	371732.8	0.6930538
L2b	371732.8	0.4368
FN	0.7251539	0.5534256
FB	0.7251539	0.7165311
Fb	0.7251539	1
GPF	1.376437	0.7655389
Fmaxb	89.04348	NA
TRP ANOVA	–	0.5173081
TRP ATS	–	0.3580221
TRP WTPS	–	0.76125

Table 2. Analysis of variance for functional data (PM10).

Test	Test statistic	p-value
FP	1.480846	0.218
CH	173283.5	0.3502
CS	173283.5	0.201
L2N	86641.76	0.2166639
L2B	86641.76	0.0882517
L2b	86641.76	0.2933
FN	1.480681	0.2683494
FB	1.480681	0.1461974
Fb	1.480681	1
GPF	2.290124	0.3178891
Fmaxb	98	NA
TRP ANOVA	–	0.811374
TRP ATS	–	0.6274703
TRP WTPS	–	0.6141

Pointwise ANOVA and Group means

Pointwise ANOVA involves assessing the statistical significance of differences at individual points in time within a dataset. Group means, on the other hand, represent the average values over time for distinct groups.

In the context of our air quality analysis, we used Pointwise ANOVA to evaluate smoothed data at various time points, calculating p-values. Simultaneously, the graph (Fig. 6) illustrating group means displayed the average values over time for each group. Notably, all lines remained above the significance level, indicating an absence of significant differences between group means.

Two samples pointwise-test

A two-sample pointwise test is a statistical analysis conducted on two distinct groups of data to assess the feasibility that their means originate from the same population. This test specifically compares the means of two samples, providing insights into potential differences between the groups (Fig. 7).

In this analysis, our results show that we cannot reject the null hypothesis, which indicates that the population means are equal, with a significance level of $\alpha = 0.05$.

All things considered, our statistical comparisons of the air quality analysis included three main components: hypothesis testing using ANOVA, pointwise ANOVA

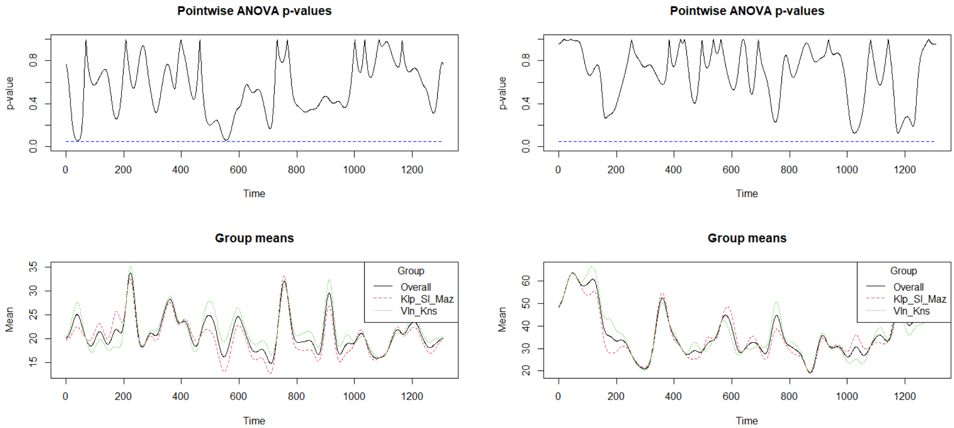

Fig. 6. Comparison of pointwise ANOVA and group means on PM10 and NO2 data.

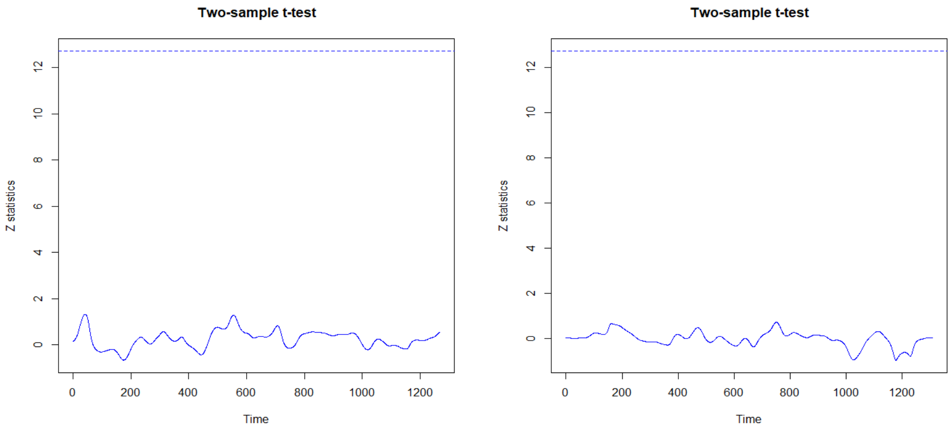

Fig. 7. Comparison of pointwise-tests on PM10 and NO2 data.

and group means, and two-sample pointwise comparisons. An ANOVA test examining differences in air quality between major cities and non-urban areas revealed no statistically significant differences in pollutant levels between these groups. Pointwise ANOVA and analysis of group means further supported this finding, highlighting that no significant differences were observed at individual time points or mean values within individual regions. In addition, a two-sample t test evaluating the mean differences between the two groups confirmed that there were no significant differences. These results, in response to the aim of the study to compare air quality in different regions, state that there are no significant differences between the two groups in PM10 and NO2.

Functional time series

The data transformation and aggregation process involved in this analysis converts daily data into yearly data visualized by month. Only complete years within the period from January 1, 2020, to December 31, 2022, were considered.

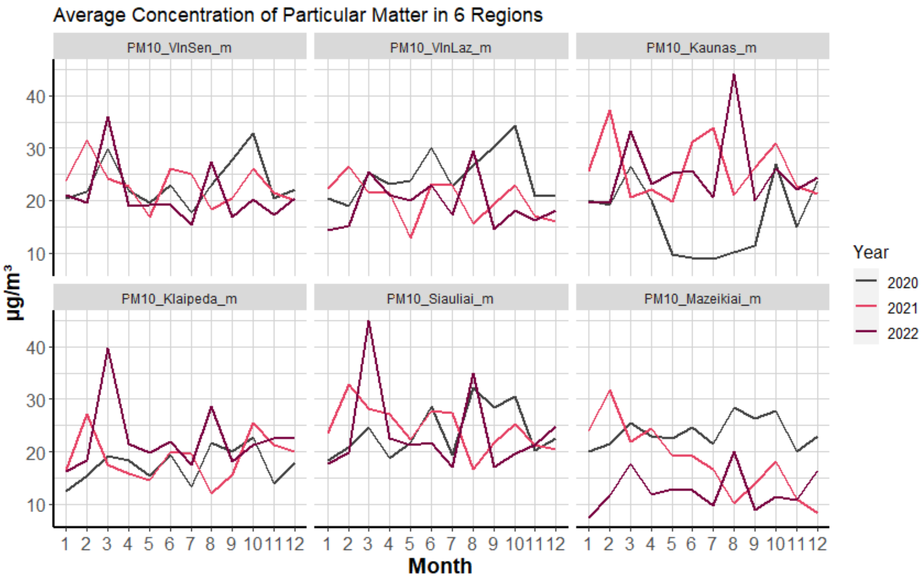

Fig. 8. Average concentration of particular matter in 6 regions.

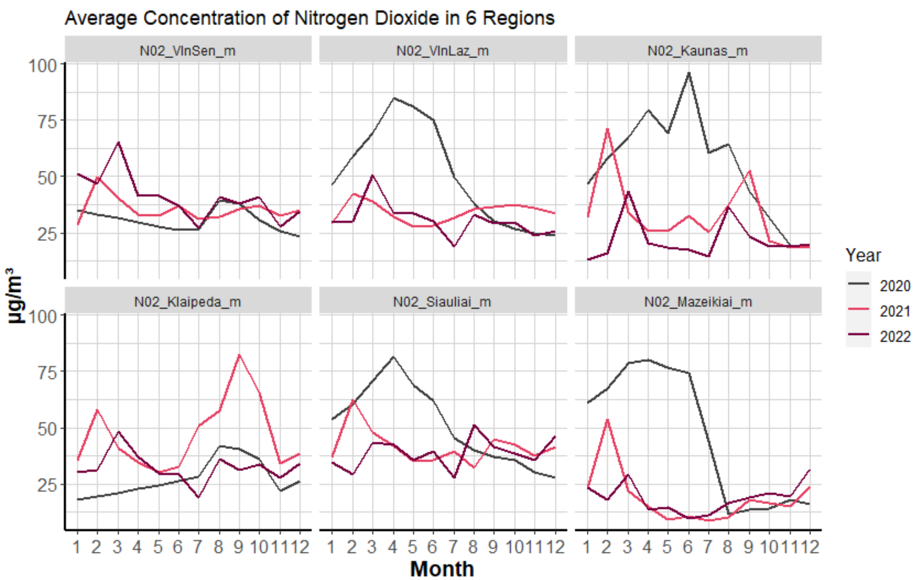

Fig. 9. Average concentration of nitrogen dioxide in 6 regions.

Then we plotted the average concentration of particulate matter (Figs. 8 and 9) in six different regions.

We performed calculations and generated plots of Functional AutoRegressive models for each variable in the dataset. Initially, we transformed the concentration variables by applying a logarithmic function. Subsequently, we iterated through each vari-

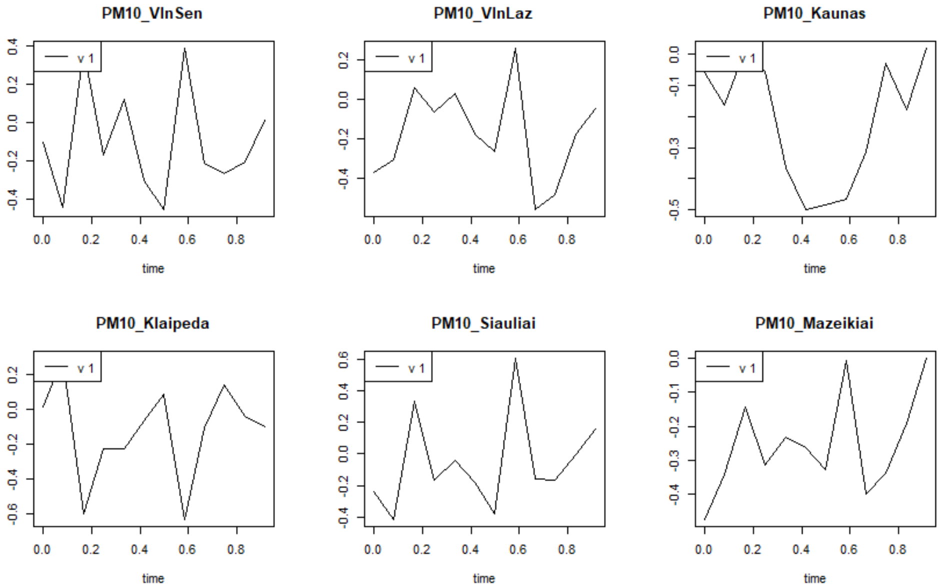

Fig. 10. Time series models for PM10 variables.

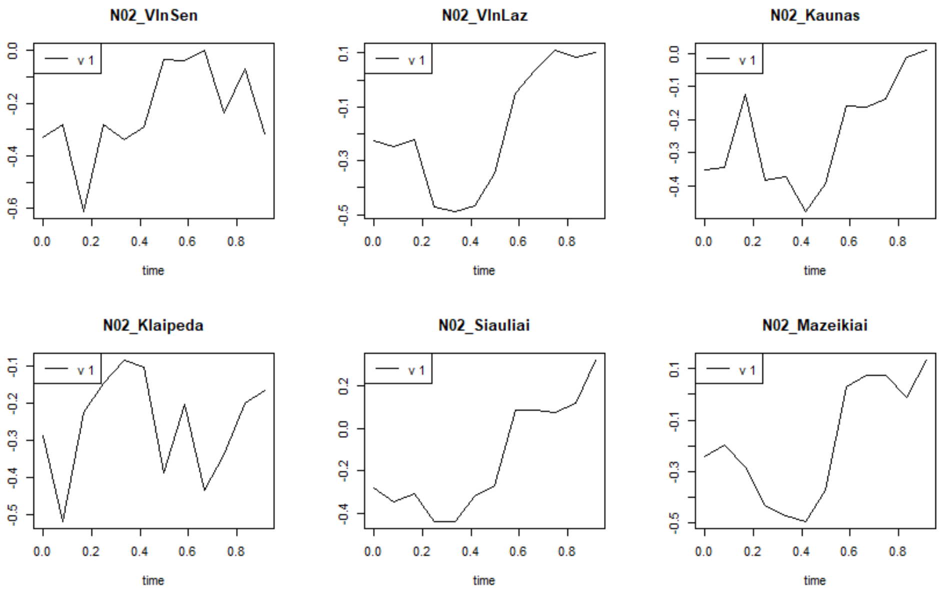

Fig. 11. Time series models for NO2 variables.

able, calculated the corresponding Functional AutoRegressive model, and stored the results. Finally, we visualized the fitted Functional AutoRegressive models (Figs. 10 and 11) for each variable in a grid layout. By examining these models, we gained insights into the temporal patterns of the average concentration of particulate matter across various regions.

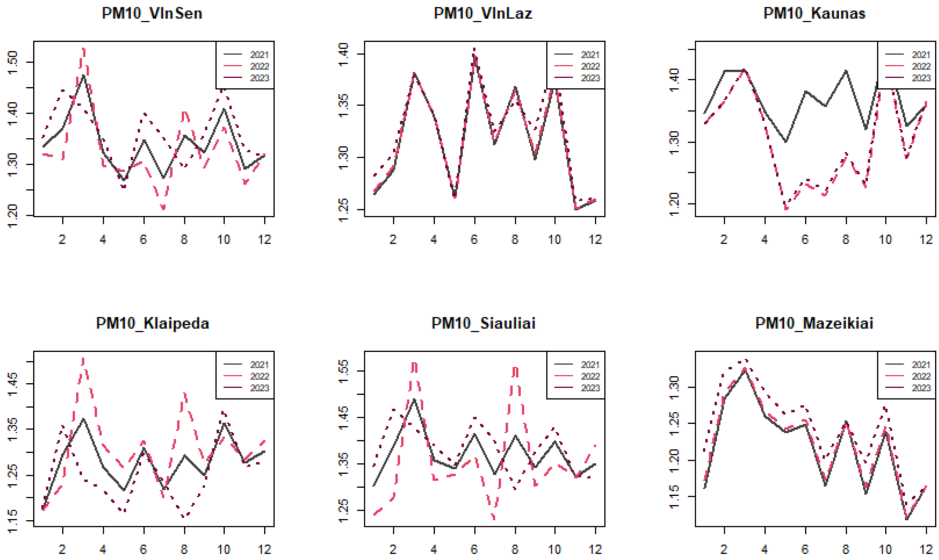

Fig. 12. PM10 predictions.

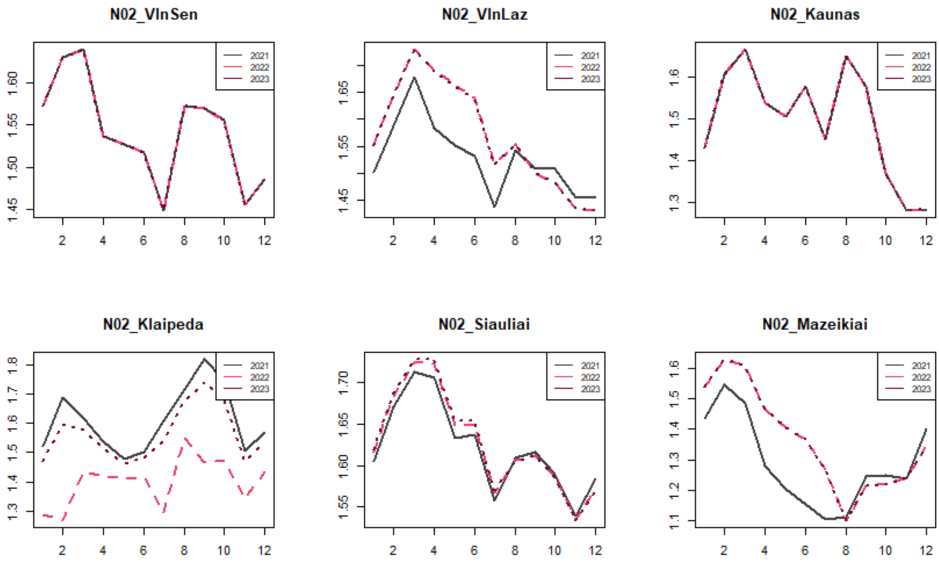

Fig. 13. NO2 predictions.

We generated predictions using Functional AutoRegressive models for each variable. The main aim was to predict data for 2023. The plots (Figs. 12 and 13) displayed the predicted values over time, with different colors representing the variables. This allowed us to visualize the predicted trends for each variable based on the Functional AutoRegressive models.

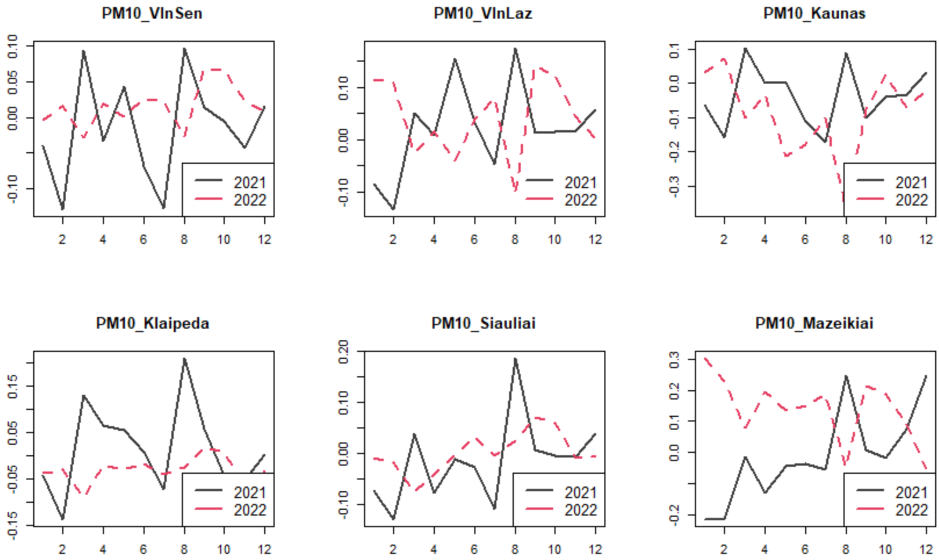

Fig. 14. PM10 errors.

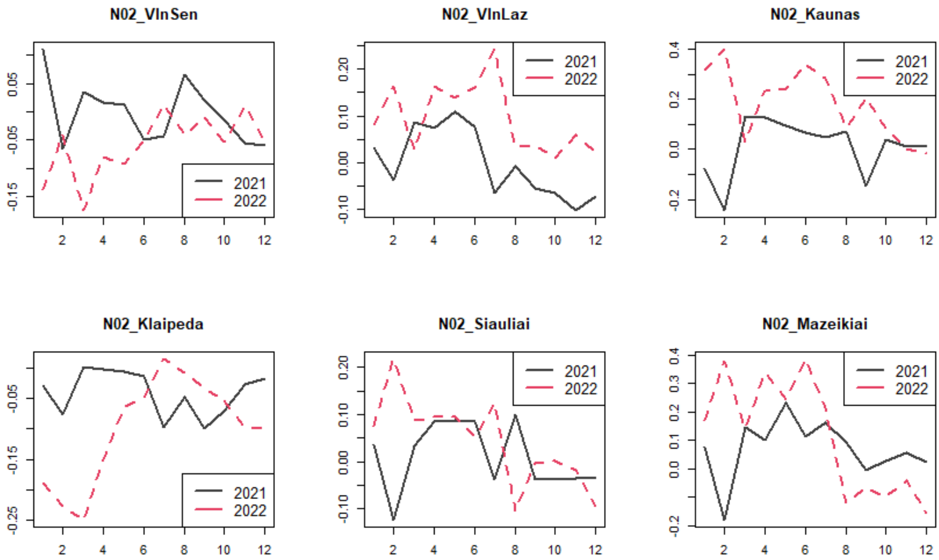

Fig. 15. NO2 errors.

We evaluated the errors of the predictions for each variable (Figs. 14 and 15). For each variable, we generated predictions using the Functional AutoRegressive models and selected the corresponding real values. The errors were then calculated by subtracting the predicted values from the real values.

The observed errors were found to be significant in magnitude. Consequently, we made the decision to enhance our model. This adjustment was deemed necessary to improve the accuracy and performance of the model in predicting the particular matter and NO_2 concentration for 2023.

The model enhancement strategy involved implementing a second model. This allowed to find the best parameter values for the optimization of the prediction model. Our process involved testing different values of parameter k2 (k2 represents a parameter in the model that determines the number of lags considered in the modeling process, refining the accuracy of the forecast model by minimizing the total error sum) to refine the Functional AutoRegressive (FAR) models. For each variable, we systematically evaluated the errors using the predicted and actual values for the years 2021 and 2022. By setting a value of k2 that minimized the total sum of errors, we improved the model's predictive accuracy. After that, we attempted to increase the accuracy of forecasts for the variables PM10 and NO2 using the improved models created by this approach. Incorporating the optimized k2 values specific to each variable, we created updated models to enhance our forecasting accuracy.

The improved models (Figs. 16 and 17) were used to make more accurate forecasts for 2021–2023. period forecasts. These updated forecasts provided better forecasting results for each variable, meaning that the accuracy of our forecasting models improved significantly.

Based on the analysis of model 2, specifically for particulate matter concentration, the forecasted errors (Figs. 18 and 19) for the year 2023 indicate that regions such as Vilnius (Senamiestis), Vilnius (Lazdynai), Šiauliai, and Klaipėda demonstrate an acceptable level of accuracy. This suggests that the model's predictions for particulate matter concentration in these specific areas are reliable for the upcoming year.

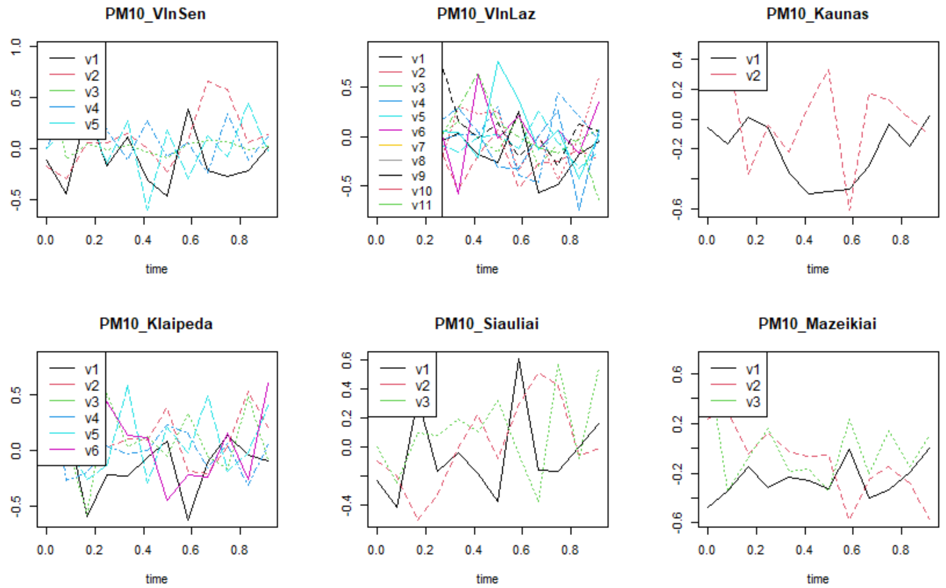

Fig. 16. Time series second models for PM10 variables.

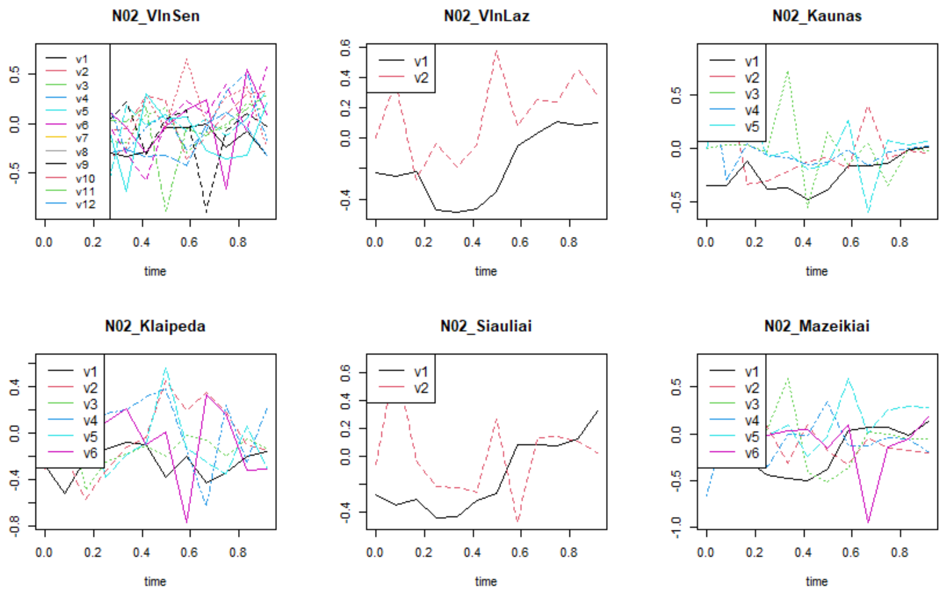

Fig. 17. Time series second models for NO2 variables.

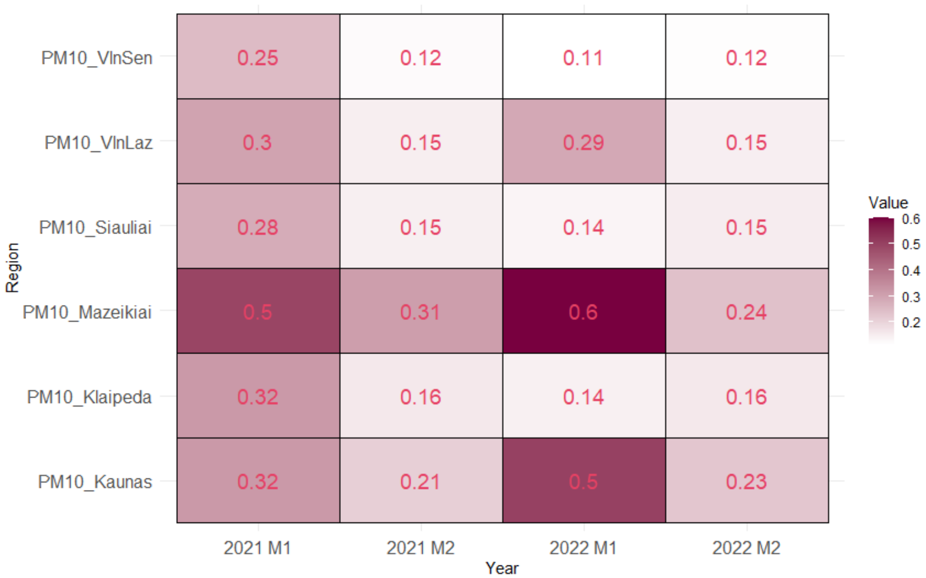

Fig. 18. PM10 errors heatmap.

Considering the evaluation of model 2 for nitrogen dioxide concentration, the projected errors for the regions of Vilnius (Senamiestis), Vilnius (Lazdynai), and Šiauliai in 2023 demonstrate satisfactory precision. In the case of other regions, it is assumed that additional historical data would be necessary to facilitate accurate predictions.

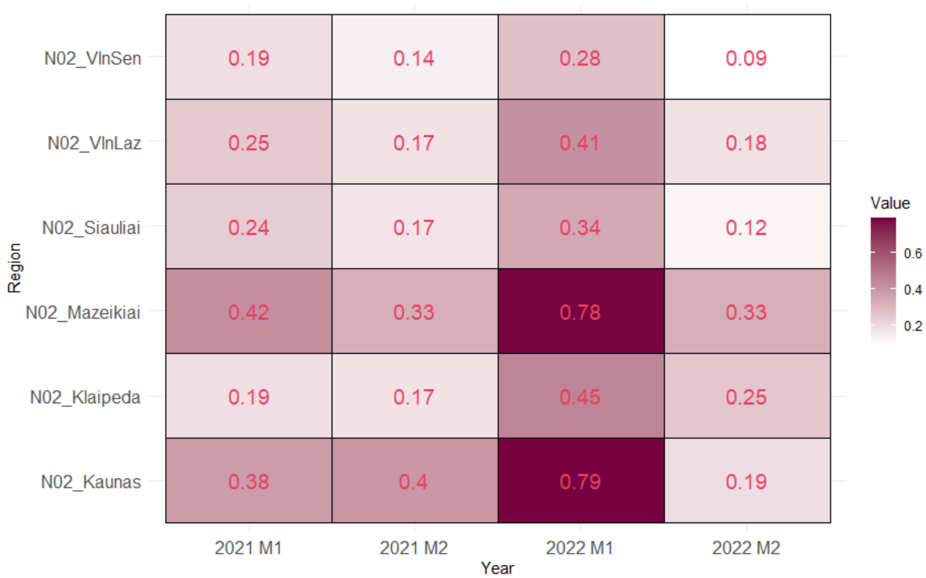

Fig. 19. NO2 errors heatmap.

Limitations

This study provides valuable insights, but is not without its limitations. To improve the analysis, incorporating additional regions and extending the duration of historical data could yield more refined and accurate outcomes. The study encountered a hurdle in extracting data from distinct PDF files on the Environment Protection Agency Lithuania website, underscoring the necessity for a database that is both user-friendly and conducive to analytical processes. A potential resolution involves establishing a new accessible database that consolidates daily data in a centralized location. Additionally, the study acknowledges limitations, including the exclusion of external factors such as industrial activities or traffic, which could offer a more holistic perspective on the factors influencing air quality in the studied regions. Future research endeavors could benefit from exploring the integration of these external factors, unraveling their relationships with pollution sources and industrial or traffic patterns. Moreover, the study recognizes that not all air quality stations have complete data for all pollutants, emphasizing the importance of addressing data completeness gaps for a more comprehensive understanding of air quality patterns.

Conclusions

We have performed data smoothing, exploratory data analysis, hypothesis testing, and principal component analysis (PCA) to explore the link between nitrogen dioxide (NO2) and particulate matter (PM10) concentrations in six different regions of Lithuania. Functional data analysis methods were successful in examining data patterns and possible outcomes. The results did not show considerable differences in

NO2 and PM10 concentrations between locations, underscoring the need to reduce air pollution.

Overall, this work offers insightful information into the dynamics of NO2 and PM10 concentrations, creating well-informed choices for reducing air pollution. For further analysis, more regions and longer duration historical data could be taken to reach even better results. The relationships between pollution sources and industrial or traffic patterns also could be further examined.

References

[1] X. Li, L. Jin, H. Kan. Air pollution: a global problem needs local fixes. *Nature*, **570**(7762):437–439, 2019.

[2] J. Martínez Torres, J. Pastor Pérez, J. Sancho Val, A. McNabola, M. Martínez Comesaña, J. Gallagher. A functional data analysis approach for the detection of air pollution episodes and outliers: a case study in Dublin, Ireland. *Mathematics*, **8**(2):225, 2020.

[3] Y. Ogen. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. *Sci. Total Environ.*, **726**:138605, 2020.

[4] G. Ottaviano, A.L. Pendolino, G. Marioni, M.A. Crivellaro, B. Scarpa, E. Nardello, C. Pavone, M.V. Trimarchi, E. Alexandre, C. Genovois, A. Moretto, M. Marani, P.J. Andrews, R. Marchese-Ragona. The impact of air pollution and aeroallergens levels on upper airway acute diseases at urban scale. *Int. J. Environ. Res.*, **16**(4):42, 2022.

[5] X. Rigueira, M. Araújo, J. Martínez, P. J. García-Nieto, I. Ocarranza. Functional data analysis for the detection of outliers and study of the effects of the COVID-19 pandemic on air quality: a case study in Gijón, Spain. *Mathematics*, **10**(14):2374, 2022.

[6] N. Shaadan, S.M. Deni, A.A. JemaIn. Assessing and comparing PM10 pollutant behaviour using functional data approach. *Sains Malaysiana*, **41**(11):1335–1344, 2012.

[7] V. Vaičiulis, J. Vencloviené, A. Miškinytė, R. Ustinavičienė, A. Dėdelė, G. Kalinienė, D. Lukšienė, A. Tamošiūnas, L. Seiduanova, R. Radišauskas. Association between outdoor air pollution and fatal acute myocardial infarction in Lithuania between 2006 and 2015: A time series design. *Int. J. Environ. Res. Public Health*, **20**(5):4549, 2023.

[8] D. Wang, Z. Zhong, K. Bai, L. He. Spatial and temporal variabilities of PM2. 5 concentrations in China using functional data analysis. *Sustainability*, **11**(6):1620, 2019.

REZIUMĖ

Oro kokybės tyrimas naudojant funkcinės duomenų analizės metodus

A. Vitkauskaitė, M. Salytė

Šiame moksliniame darbe pateikiama išsami kietųjų dalelių (KD10) ir azoto dioksido (NO2) taršos koncentracijų šešiuose skirtinguose Lietuvos regionuose analizė. Šis tyrimas apima duomenų suglodinimą, pagrindinių komponentų analizę (PCA), tiriamąją duomenų analizę, hipotezių tikrinimą ir laiko eilučių analizę. Funkcinių duomenų analizės metodai buvo naudojami siekiant nustatyti šių oro teršalų kilmę ir poveikį, atskleidžiant jų duomenų modelius. Funkcinių duomenų analizės metodai rodo jų efektyvumą atskleidžiant giliuosius ryšius dideliuose duomenų rinkiniuose, padedančius kontroliuoti oro kokybės problemas.Tyrimas, kurio tikslas buvo palyginti oro kokybę skirtinguose regionuose, rodo, kad tarp dviejų grupių nėra reikšmingų KD10 ir NO2 skirtumų. Be to, patikimos 2023 m. KD10 prognozės yra pasiekiamos tokiuose regionuose kaip Vilniaus senamiestis, Vilniaus Lazdynai, Šiauliai, Klaipėda. NO2 sėkmingą prognozavimą galima pritaikyti Vilniaus senamiesčiui, Vilniaus Lazdynams ir Šiauliams.

Raktiniai žodžiai: oro teršalai; kietosios dalelės; azoto dioksidas; funkcinių duomenų analizė; tiriamųjų duomenų analizė; hipotezių tikrinimas