applied sciences

MDPI

Article

Impact of Meteorological Factors on Air Pollution Prediction Based on Multiple Linear Regression with R Program: A Case Study with Comparable-Pollutant Parameters and Fragmented Seasonal Datasets from a Heavy-Urban-Traffic Monitoring Station

Zoltan Kazi, Ljubica Kazi * and Snezana Filip

Technical Faculty "Mihajlo Pupin", University of Novi Sad, 23000 Zrenjanin, Serbia; zoltan.kazi@tfzr.rs (Z.K.); filipsnezana@gmail.com (S.F.)

* Correspondence: ljubica.kazi@tfzr.rs

Abstract

Air pollution data processing is a relevant aspect of urban-life-quality monitoring. This study relates the values of air pollutants to meteorological factors with prediction models created with multiple linear regression (MLR) in the R program. Fragmented data were obtained in the years 2023/24 from the heaviest-urban-traffic location in Belgrade, Serbia. A Serbian-comparable list of air pollutants (PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, and O_3) was created according to an analysis of sensor availability at monitoring stations. With the prediction models and seasonally clustered data (spring and summer data clusters were incomplete), each of these pollutants (except O_3) was related to all measurable meteorological factors (atmospheric pressure, wind speed, relative humidity, and temperature). All independent (meteorological) factors showed a moderate level (Multiple $R^2 < 0.7$) of impact on the prediction of air pollutant values (dependent factors). The second group of prediction models related pollutant triplets, and fourteen of the created models reached $R^2 > 0.7$. The prediction model with the highest relevant R^2 value ($CO \sim PM_{10} + NO_2$) was used for CO pollutant prediction and in-sample fit analysis. The obtained mean absolute error is ~0.13 mg/m^3. This case study contributes to the examination of the role of meteorological factors in urban-traffic air-quality monitoring within a seasonal context.

Keywords: air pollution; multiple linear regression; R programming; data prediction; meteorological factors; season; data clustering; comparable parameters

Check for updates

Academic Editors: Silvia Puiu, Mihaela Tinca Udriștioiu and Liliana Velea

Received: 25 June 2026
Revised: 19 August 2026
Accepted: 20 August 2026
Published: 31 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Air pollution has become one of the major problems in urban areas, especially in regions with intensive traffic or industrial activities. High values of pollutants are seriously affecting human health [1]. Air pollution in urban areas can have harmful effects, such as cancer [2], asthma [3] and low birth weight [4]. It is the fourth leading cause of death, according to Parra et al. [5]. Therefore, significant efforts are being made to detect, evaluate and reduce air pollution [6].

According to Lu et al. [7], methods for predicting air quality can be classified into three categories: statistical, potential and numerical methods. Statistic-based prediction based on linear and non-linear regression methods have attracted special attention because of their

prediction accuracy [8]. The R language has been utilized [9] to develop linear regression and multiple linear regression models for air pollution prediction [10].

Air-quality monitoring stations are usually placed at locations of different types, such as near roads in urban and industrial areas, and their intended location determines their categorization, along with their specific characteristics and air pollutant detection sensitivity [11]. Various technologies are used to integrate traffic air pollution monitoring stations with the internet, with the aim of automating the collection of data and air-quality monitoring processes [12]. It is essential to carefully plan and implement the adequate positioning of air pollution measurement stations, particularly in urban environments [13], with special consideration for near-road air-quality monitoring and data acquisition [14].

Recent studies on the impact of meteorological factors on air pollution prediction suggest that the seasonal clustering of data enables predictions with higher accuracy [15]. Meteorological factors have been combined with spatial variations and ambient influences to demonstrate their associations with the measured values of traffic air pollution [16], but multiple sources of pollution (industrial and traffic) have been considered [17]. In studies on the influence of meteorological factors on air pollution prediction, only several air pollutant types have been addressed [16,18]. In [17], each air pollutant was related to each of the meteorological factors considered within seasonally clustered data, but the joint effect of all the meteorological factors together was not presented among the results. It can be concluded that previous research has not provided results relating air pollutant values to multiple meteorological factors with multiple linear regression methods; further, the lists of air pollutants considered in previous analyses were incomplete, and the measured values were obtained from air-quality monitoring stations located where mixed sources contributed to pollution. The results of the previous research gap analysis guided our research aims and design.

Our research study focuses on air pollution prediction based on seasonally clustered data, and we created MLR-based prediction models that examine the influence of meteorological factors on air pollution value prediction. Compared with other research results, our contribution aims are as follows:

- Comprehensiveness—We address all measurable meteorological factors (wind speed, atmospheric pressure, temperature, and relative humidity) in all seasons and all comparable-air-pollutant types (PM_{10}, $PM_{2.5}$, SO_2, NO_2, and CO, but not O_3) whose sensors are available at Serbian air-quality monitoring stations.
- Preciseness of results—We selected an air pollution monitoring station, "Mostar", located in the area with the most intensive traffic flow in Belgrade, Serbia, and a single source of pollution—urban traffic.
- Alignment with weather as a natural phenomenon—Since all meteorological factors are always measured simultaneously, we created MLR-based prediction models for all measurable meteorological factors and investigated the latter's associations with air pollutant values.

  This study addresses the following research questions:

  - RQ1. What are the comparable parameters, i.e., the types of air pollutants, that can be detected at different air-quality monitoring stations in Serbia?
  - RQ2. How are meteorological parameters related to values of air pollutants in a heavy-traffic area, considering specific seasons?
  - RQ3. How are air pollutants mutually related in a heavy-traffic area?

  The contributions of this manuscript are aligned with the research questions:

1. Selection of comparable parameters for air pollution prediction, based on pollutant sensor availability at monitoring stations in Serbia (RQ1).

2. Utilization of the R program with the multiple linear regression method (MLR) with the purpose of creating air prediction models and supporting air pollution prediction and seasonal analysis of meteorological factors' impact on the prediction process (RQ2 and RQ3).
3. Air pollution prediction models based on MLR (RQ2 and RQ3).
4. Seasonal analysis of meteorological factors' impact on air pollutant values, where each comparable air pollutant (dependent value) is related to all four of the independent values—meteorological parameters (atmospheric pressure, wind speed, relative humidity, and temperature) (RQ2).
5. Analysis of the relation among air pollutant values, where one air pollutant (dependent value) is related to two other air pollutants (independent values) (RQ3).
6. Prediction of air pollutant values based on MLR-based prediction models that include other air pollutant values, not meteorological factors (RQ3).

The rest of the paper is organized as follows: Section 2 presents a literature review of multiple linear regression and hybrid methods, software tools and R programming utilization in air pollution prediction and consideration of meteorological factors in the context of seasonal variations. Section 3 presents materials and methods, with the research methodology, the multiple linear regression method, the R program (created with RGUI, R version 4.6.1) and the research sample. Section 4 presents the results of applying the R program to the data sample obtained from air pollution monitoring station "Mostar", located in the area with the heaviest traffic in Belgrade, Serbia. The results are related to the selection of comparable-pollutant types that are mostly available in Serbia, and these pollutant types are considered in the rest of the research study. An analysis of the impact of meteorological factors on the values of the comparable air pollutants is performed with seasonally clustered data obtained from the air-quality monitoring station, while cross-seasonal prediction is computed with the R program in relation to air pollutant triplets. Section 5 provides a discussion of the results in the context of the research questions, a comparison with previous research and this study's limitations. Section 6 provides conclusions as a summary of the results and future research directions.

2. Related Work

2.1. Multiple Linear Regression and Hybrid Methods in Air Pollution Prediction

Statistical regression methods have been used in many different areas, including in prediction algorithms [19]. The linear regression method has been utilized in air pollution prediction regarding some types of air pollutants and a variety of meteorological, seasonal and geographical factors. Air pollution predictions are based on big data, such as in the research study by Zheng et al. [20], for forecasting air quality at fine-grained level. Sharma et al. [21] presented results of forecasting air pollution in the region of New Delhi, using descriptive and predictive analyses that implemented the linear regression method. Prediction of air pollution was also performed with component analysis and the linear regression model [22].

Hussain et al. [23] analyzed and compared multiple statistical methods and models in the context of the modeling and monitoring of air pollutant values, such as CO_2. Multiple linear regression was utilized in air pollution-related research, such as in the work by Park et al. [24], where a methodology to predict indoor $PM_{2.5}$ concentrations, using residential-environment monitoring data and a time-divided multiple linear regression model, was proposed. Vlachogianni et al. [25] proposed a multiple regression model for forecasting concentrations of NO_x and PM_{10} in Athens, Greece, and Helsinki, Finland.

Hybrid utilization of multiple linear regression and other methods and approaches has been proposed in research on air pollution prediction. Dun et al. [26] proposed a hybrid

model comprising gray multivariable regression and a fractional-order accumulation model to be used with different meteorological factors to predict air pollutants (PM_{10}, $PM_{2.5}$, and NO_2) in Shijiazhuang and Chongqing, China. Roy et al. [27] proposed a hybrid model comprising genetic algorithm and multivariate polynomial linear regression. Linear regression was used for the prediction of gas concentrations, while a genetic algorithm approach was used to minimize errors in linear regression prediction. Predictive equations were formed for O_3, CO and NO_2 based on data values for temperature, relative humidity, benzene, and NO_x.

2.2. Software Tools and R Programming in Air Pollution Prediction

Diverse software solutions that include application of linear regression methods are used to support air pollution research and prediction. Ilijazi et al. [28] explored software-supported visualization and mathematical models for air pollution emissions models. In the study by Özdemir et al. [29], meteorological factors were related to PM_{10} concentrations with utilization of multiple linear regression and artificial neural networks and specifically applied to data from urban and industrial regions in Kocaeli, Turkey. Basurko et al. [30] presented a Multilayer Perceptron that utilizes regression models to forecast hourly O_3 and NO_2 levels in the Bilbao area. Eren et al. [17] applied a machine learning approach to prediction of air pollutants in relation to seasonal analysis, with a case study in Sakarya, Turkey.

The R language has become frequently utilized in computational statistics, visualization, and data science [31,32]. R is a free and open-source programming language [33], utilized with open-source software RGUI for statistical computing and graphics, available from the R foundation (R-project) website [34]. R is primarily intended for statistical computing, but it is a powerful and increasingly popular tool for graphical presentation features [35]. One of the main R programming language features is extendibility and scalability, with a large community of scientists and practitioners being engaged in adding new functions. According to Morandat et al. [36], the basic R language is constantly enhanced by hundreds of packages and libraries with functions that cover a large set of statistical methods and applications, so scientists have an opportunity to use a huge combination of program features for various analyses and research tasks. In particular, R could be used with various multiple regression methods, where the ridge regression method could be applied for estimating the coefficients of multiple regression models when variables are highly correlated [37].

In recent years, the R programming language has been frequently used in ecological scientific research. Guenzi et al. [38] utilized the R programming language in weather monitoring, which allowed them to relate merged data from radar and rain gauges observations. Frischot and Francois [39] applied R packages and functions in the ecology and landscape area. Fisher et al. [10] applied a multiple regression method with R to construct, fit, and compare a complete prediction model related to ecological or environmental predictors while defining full subsets of input information.

Seo and Lee [9] presented the results of using the R programming language as statistical software to analyze a large number of data associated with air pollution in Korea. Results from Setiawan's study [40] are based on utilizing the R language and R Studio for prediction of nitrogen dioxide pollution.

2.3. Meteorological Factors and the Impact of Seasonal Variations on Air Pollution Prediction

Natural factors, such as geographical position, meteorological conditions, seasons and ground quality, can be related to measured and predicted values of air pollution. Syafei et al. [22] applied a linear regression model and considered air pollutants like NO_2, PM and

ozone (O_3), as well as data related to meteorological and temporal factors from monitoring stations in Indonesia. Choi et al. [18] presented the results of creating regression models for prediction of particulate matter (PM_{10}, $PM_{2.5}$, and PM_1) in Gangneung, Korea, based on meteorological parameters (relative humidity, air temperature, and wind speed), PM_{10}, and $PM_{2.5}$ related to non-dust and dust periods in Beijing, China. Tularam et al. [41] created a hybrid regression model of air pollution and land use with incorporated meteorological information from atmospheric dispersion models for predicting SO_2, NO, and PM in the region of Durban, South Africa.

Cao et al. [16] investigated spatial variation in traffic pollution in urban non-motorized lanes, with special consideration of the impact of meteorological and ambient factors on local particle variations. It has been shown that these factors had significant but varied effects, such as riverside winds and roadside greenery. Eren et al. [17] selected Sakarya, Turkey, as a region near the sea with both industrial and urban-traffic areas to analyze the impact of seasonal meteorological factors on air pollution.

Zhao et al. [15] created a multivariate linear regression model to enable short period prediction of $PM_{2.5}$, based on data obtained through remote sensing, meteorological factors from ground monitoring (temperature, wind velocity, and relative humidity) and other gaseous pollutants—SO_2, NO_2, CO, and O_3. The validity of the derived regression models was assessed, and it was shown that annual data prediction had lower fit compared with the seasonal prediction models (the spring and winter prediction models were more accurate).

Data clustering has been proposed to be related to seasons [22], dust/non-dust periods [18] or other grouping according to days with similar meteorological conditions [42]. Hodgsonand Phillips [42] examined seasonal (winter versus summer) air-mass types related to pollutant concentrations of O_3, PM_{10}, NO_x, NO_2 and CO in Birmingham, UK. Cluster analysis was used to make groups of days with similar meteorological conditions, which led to six types of air mass. It was shown that high pollutant extremes were more common in winter, which was related to more favorable meteorological conditions and increased human activities that led to higher emissions during the cold season.

Zeng et al. [43] proposed a method that includes static and dynamic approaches (dual branch) to mapping relationships among variables in air pollution prediction (pollutant concentration, meteorological factors and temporal features), with particular attention paid to seasonal heterogeneity, formulated as patterns. They proposed seasonal-model training based on distinct physical and chemical patterns of pollution across different seasons. For example, O_3 concentrations were higher in summer because of high temperature and photochemical reactions, while the $PM_{2.5}/PM_{10}$ ratio is crucial in winter, when pollutant accumulation is higher.

3. Materials and Methods

3.1. Research Methodology

The research questions of this study are addressed through various activities, their methods, inputs and results:

1. RQ1. What are the comparable parameters, i.e., the types of air pollutants, that can be detected at different air-quality monitoring stations in Serbia? Activities and methods: Analysis of air pollution types that can be detected at all air-quality monitoring stations in Serbia. Selection criteria: Air pollutant sensor is available at more than 50% of all monitoring stations. Inputs: Data from Environmental Protection Agency. Output: List of comparable-air-pollutant types.
2. RQ2. How are meteorological parameters related to values of air pollutants in a heavy-traffic area, considering specific seasons? Activities and methods: The data obtained from the heavy-urban-traffic location are seasonally clustered, and each season's

cluster is used for creating MLR-based prediction models where each pollutant from the comparable list is related to all meteorological factors' values. Input: Data from a monitoring station obtained during one year. Output: MLR-based prediction models.

3. RQ3. How are air pollutants mutually related in a heavy-traffic area? Activities and methods: The annual dataset from a monitoring station in a heavy-traffic area is segmented into three parts (datasets) to conduct in-sample prediction. The first part enables MLR-based prediction models, the second one is an input to prediction, while the third one contains measured values to be compared with the predicted interval values. After this data segmentation, MLR-based prediction models where each pollutant from the comparable list is related to two other pollutants are created. Finally, the MLR-based prediction models are used for prediction of data, and the prediction results are compared with the measured data from the sample. Input: Data obtained from the monitoring station. Output: MLR-based prediction models, the prediction data, the comparison between the predicted data and the measured data, and in-sample fit analysis of MLR-based prediction models.

Figure 1 presents a simplified workflow diagram that connects activity phases related to all research questions and the main inputs and outputs. For better understanding, inputs and outputs are presented with visualizations.

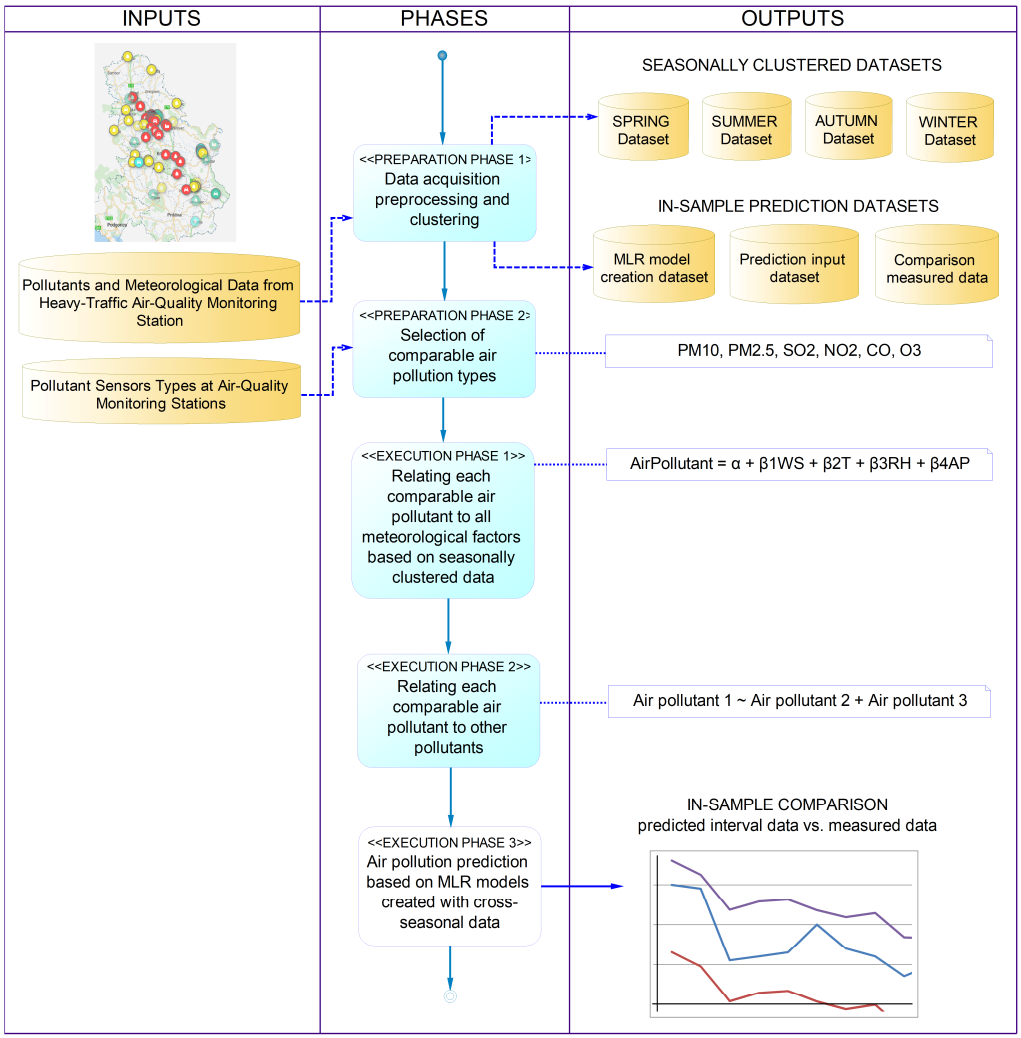


Figure 1. The research method activity diagram with main phases.

Figure 1 presents the research workflow, which consists of two preparation phases and three execution phases. Visualized key input sources are represented with the map

of a specific monitoring station and the country map of all monitoring stations. The right side presents key outputs: datasets created in the data acquisition preprocessing and clustering phase, the extracted list of comparable air pollutants, two types of MLR-based prediction models and a diagram for visual comparison between predicted air pollution interval values and measured values. Figures 2 and 3 present details of specific phases, their activities, inputs and results.

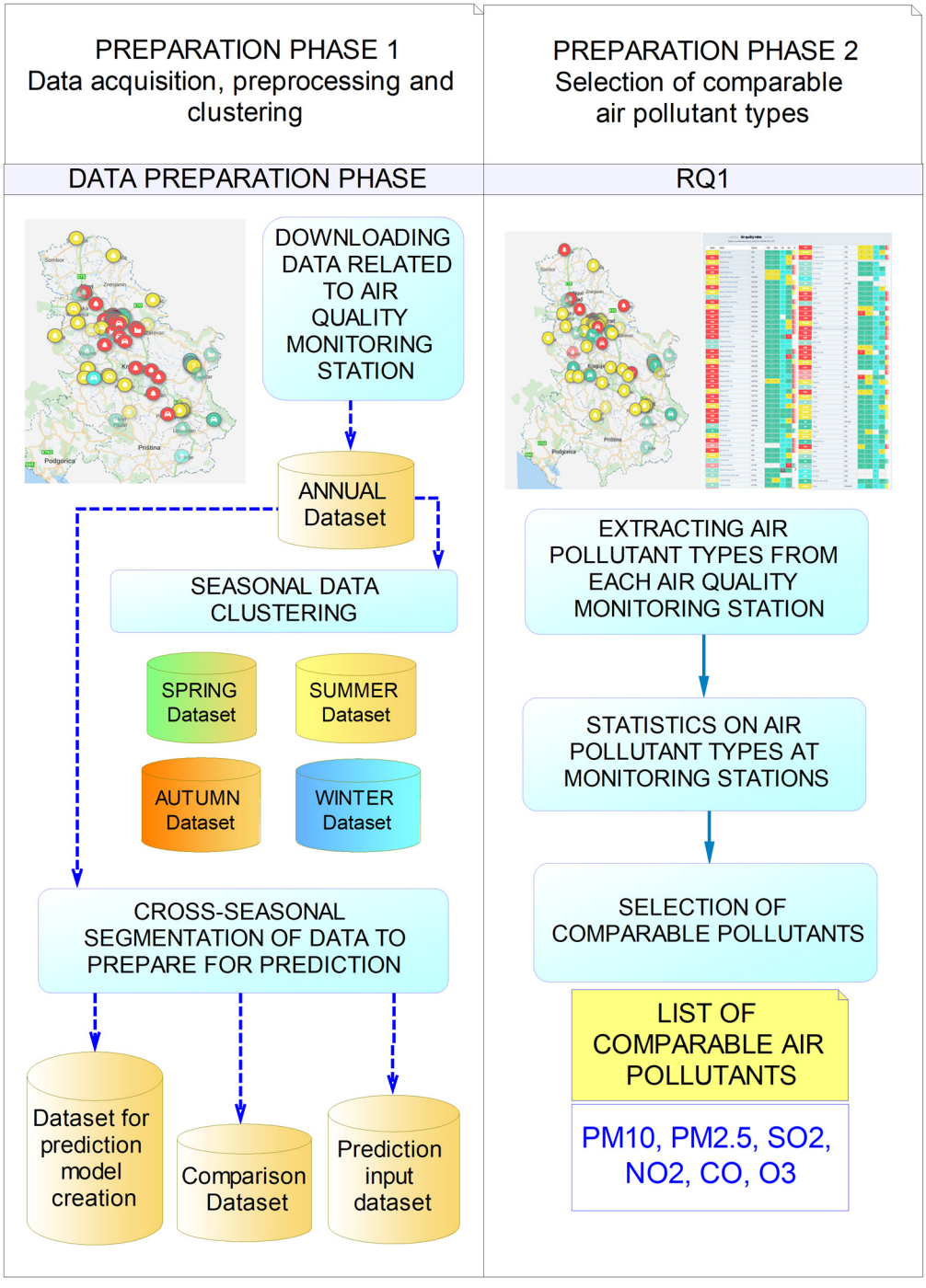


Figure 2. The research method activity diagram with phase 1 and phase 2 details: activities (blue ovals), inputs (presented with images) and outputs (presented as datastores and a list).

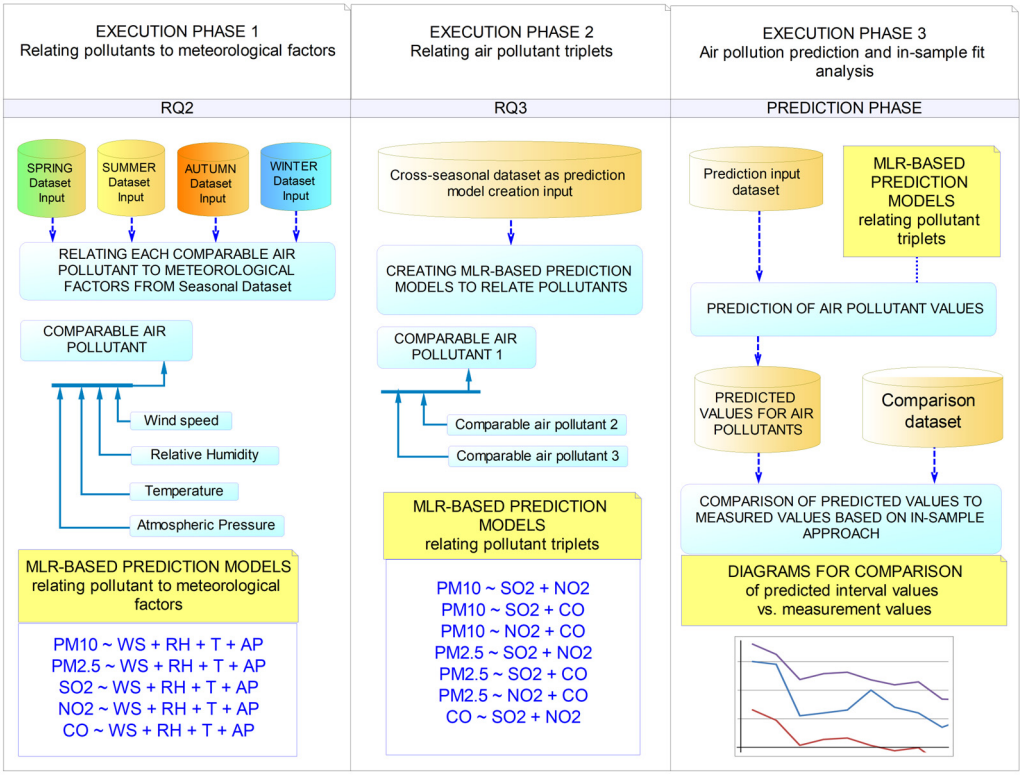


Figure 3. The research method activity diagram with details of phases 3–5: activities (blue ovals), inputs (data stores) and outputs (lists and a diagram).

Figure 2 presents phase 1 and phase 2. Phase 1 is related to data acquisition and preprocessing. Data from a specific air-quality monitoring station are downloaded and segmented in two ways: (1) seasonally, by making separate dataset for each season, and (2) cross-seasonally, for in-sample prediction and in-sample fit analysis of the MLR-based prediction models. Phase 2 visualizes the country map with all air-quality monitoring stations and the data from these stations, as a source of data for extracting air pollutant types and statistics of the most commonly present air pollutant sensors. Figure 3 presents phases 3, 4 and 5. Phase 3 deals with creating MLR-based prediction models for each comparable air pollutant in relation to all four meteorological factors, based on separate seasonally clustered datasets. Phase 4 creates MLR-based prediction models where each pollutant is related to two other pollutants. In phase 5, MLR-based prediction models and cross-seasonal segmented datasets are used to derive prediction data and compare these results with the measured in-sample data.

3.2. Multiple Linear Regression Method

Multiple linear regression (MLR) is a statistical method that models the linear relationship between a dependent variable and multiple independent variables [17]. In this research study, two types of MLR-based prediction models are created:

1. For establishing relations between each pollutant and four meteorological factors, the MLR function is presented as:

$$PO = \alpha + \beta_1 AP + \beta_2 T + \beta_3 W + \beta_4 RH \quad (1)$$

where PO is the dependent variable (air pollutant); the independent variables are AP (atmospheric pressure), T (temperature), W (wind speed), and RH (relative humidity); α is the intercept; and β_i denotes the regression coefficients.

2. For establishing relations between one pollutant and two other pollutants, the MLR function is presented as:

$$PO_1 = \alpha + \beta_1 PO_2 + \beta_2 PO_3 \quad (2)$$

where PO_1 represents the dependent variable (air pollutant concentration value), while the independent variables are air pollutants PO_2 and PO_3.

3.3. R Program

This case study utilizes the MLR method with the R program, which is executed within the software tool RGUI, version 4.6.1., 64 bit, and it is available at [44]. The RGUI software tool enables execution of the R program integrated with datasets loaded from textual files, and outputs can be textual files or graphics. The additional library ggplot2 is available from [45].

The R program for creating an MLR model that relates air pollutant values (in this example, it is SO_2) to meteorological parameters is presented in Listing 1. The dataset for MLR model creation is loaded with the *read.table* command from a separate file for each season (in this example: *"MostarSpring.txt"* placed in the local folder *"C:/Data/AirData/ResultsWeather"*) annotated with the *setwd* command. The command *dataframe < -datatable* enables structural reading from the txt file, assigning columns from the textual file to dimensions' names (used later for vectors). The *LM* command creates a prediction model based on data from a *dataframe*, relating air pollutant value (left parameter before ~) to all meteorological parameters (right parameters). Finally, the computation for obtaining the MLR function coefficients (α and β_i) and statistical accuracy of the created model is performed with the *summary* command.

Listing 1. The R program for creating an MLR prediction model that relates pollutant values to all meteorological factors values (example with the spring dataset and the SO_2 pollutant).

```
>setwd("C:/AirData/ResultsWeather")
>datatable<- read.table("MostarSpring.txt", header = TRUE, fill = TRUE)
>dataframe<- datatable[,c("PM10","PM25","SO2", "NO2","CO", "W", "AP",
"T", "RH")]
>dim(dataframe)
> model1 <- lm(SO2~AP + T+W + RH, data=dataframe)
>summary(model1)
```

The R program (presented in Listing 1) creates an MLR prediction model that establishes relations of a multiple linear function $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$, where Y is the pollutant value and X_i denotes meteorological factors values. With the previously presented R program, in this case study, each comparable pollutant is related to all meteorological parameters. Note that in the R program, "PM25" is used instead of the $PM_{2.5}$ annotation.

Graphical commands are used for visual presentation of the relationship of three dimensions (based on MLR models). Scatterplot diagrams are created with ggplot2 library commands: the *library* command enables using ggplot2 library functions, while the *ggplot* command creates the diagram. In this case study, the ggplot2 commands for creating scatterplots were used to visualize relationships (obtained with the *method = lm* parameter) between pollutant triplets. Finally, the *ggsave* command saves the diagram in a file in the previously selected graphic format. Listing 2 presents the R program that uses ggplot2 commands.

Listing 2. The R program for creating png files with scatterplots created with ggplot 2 commands.

```
>setwd("D:/Ljubica/Rpodaci")
>datatableAnual<- read.table("MostarAnual.txt", header = TRUE, fill =
TRUE)
>dataframeAnual<- datatableAnual[,c("SO2", "PM10", "NO2", "NOX",
"CO", "NO", "PM25","W", "AP", "T", "RH")]
>dim(dataframeSpring)
>library(ggplot2)
>createdDiagram <- ggplot(dataframeAnual, aes(x = NO2, y = PM10, color
= SO2)) + geom_point() + geom_smooth(method = lm, fill = NA)
>ggsave(file.path("D:/Ljubica/Rpodaci", "AnnualSO2_NO2_PM10.png"),
plot = createdDiagram, width = 10, height = 10, units = "in", dpi = 300)
```

The R program presented in Listing 2 establishes an implicit (with method = lm) MLR relation between SO_2 (the dependent value) and the independent values of NO_2 (x-axis) and PM_{10} (y-axis), obtained as vectors from the *dataframe* represented with *dataframeAnual*.

The complete R program for explicitly creating MLR models and prediction of values (based on that model and prediction input data) is presented in Listing 3.

Listing 3. The R program for creating MLR prediction models and prediction processing.

```
#setting the working directory in R and export file
setwd("D:/Ljubica/Rpodaci")
# PREDICTIONMODEL CREATION
# Loading MLR model creating data
datatableMLRcreation<-
read.table("MostarInSample_MLRCreation_CrossSeasonalData.txt", header = TRUE, fill
=TRUE)
dataframeMLRcreation<- datatableMLRcreation[,c("SO2", "PM10", "NO2", "NOX",
"CO", "NO", "PM25","W", "AP", "T", "RH")]dim(dataframeMLRcreation)
#creating MLR-based prediction model
PredModel <- lm(CO~PM10+NO2, data=dataframeMLRcreation)
summary(PredModel)
# PREDICTIONDATA PREPARATION
# Loading prediction input data
datatablePredictionInput<- read.table("MostarInSample_PredictionInput.txt", header =
TRUE, fill = TRUE)dataframePredictionInput<- datatablePredictionInput[,c("SO2",
"PM10", "NO2", "NOX", "CO", "NO", "PM25","W", "AP", "T",
"RH")]dim(dataframePredictionInput)
# Creating input vectors
PM10 <- as.vector(dataframePredictionInput$PM10)
NO2 <- as.vector(dataframePredictionInput$NO2)
# Creating dataframe from two input vectors PM10 and NO2
dataframeNewInput <- data.frame(PM10,NO2)
# PREDICTION COMPUTING-RESULTS
predict(PredModel, newdata=dataframeNewInput, interval = 'prediction')
```

The R program presented in Listing 3 consists of three parts. The first part is related to creating an MLR model that will be used for prediction. The second part relates to

preparation of input data for the prediction, and the third part is related to computation of the prediction, with the *predict* function that includes three parameters (prediction MLR models, new prediction input data in the form of dataframe and the interval type). The *interval* parameter of the *predict* function can take '*confidence*' and '*prediction*' values, and both modalities are related to computing values for the predicted intervals (lower (*lwr*) and upper (*upr*) values). The interval = '*prediction*' results in a wider interval.

3.4. Research Sample

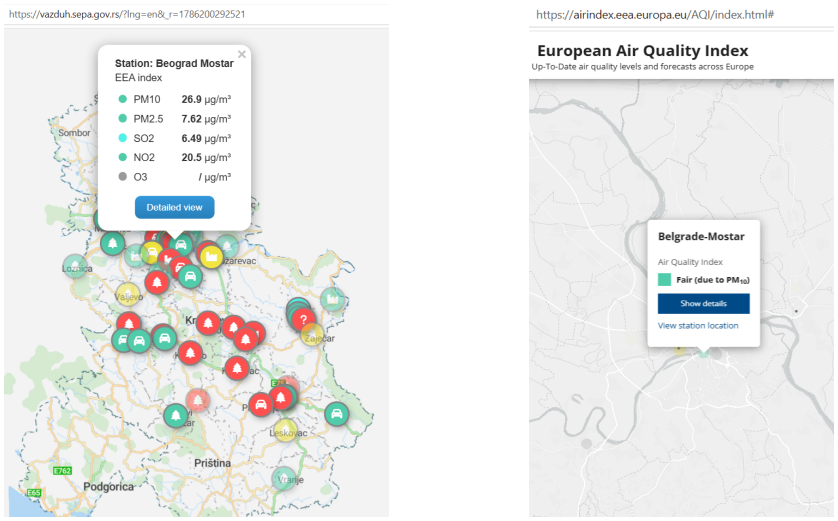
3.4.1. Sample Data Source and Characteristics

Air-quality control in Serbia is conducted according to Serbian and international legal regulations. The Serbian Environmental Protection Agency (SEPA) is affiliated to the Ministry of Environmental Protection, Republic of Serbia, and all relevant legal documents are published on the official website [46]. Relevant publications include Serbian legislation: Environmental protection law ("Official gazette RS", May 2009), Regulation on the content and manner of managing the environmental protection information system ("Official gazette RS", December 2009), Regulation on the National List of Environmental Protection Indicators ("Official gazette RS", May 2011), Air Protection Law ("Official gazette RS", June 2025), Regulation on the establishment of an air-quality control program in the state network ("Official gazette RS", August 2011) and Regulation on monitoring conditions and air-quality requirements ("Official gazette RS", July 2013).

International standards and legal documents are related to Serbian air-quality legal regulations. In the Regulation on monitoring conditions and air-quality requirements, it is stated that quality assessment of air pollution monitoring focuses on measurement procedures and instrument verification, data processing and data presentation according to Serbian standard SRPS ISO/IEC 17025 [47], which is aligned with international ISO/IEC standards. Methods for the measurement of specific pollutant concentrations are defined in separate international standards adopted in Serbia: SRPS EN 14212 (SO_2, NO_2, and NO_x) [48], SRPS EN 12341 (PM_{10}) [49], SRPS EN 14907 ($PM_{2.5}$) [50], SRPS EN 14626 (CO) [51], and SRPS EN 14625 (O_3) [52].

Reports from automated air-quality monitoring stations are aligned with Serbian legislation and international standards. Regular daily reports from all air-quality monitoring stations are published on the official website of SEPA [53], and reports from some of the air pollution stations are also published on the European Air Quality Index (EAQI) official website [54]. For this research study, the data sample is chosen to be related to the Serbian capital city, Belgrade. A specific urban-traffic monitoring station, "Mostar", placed next to the crossroad "Mostar Loop" in the zone of the heaviest traffic, is selected. Figure 4 presents the location of "Mostar" station through graphics on both web portals—SEPA [53] and European Air Quality Index [54].

The European Air Quality Index [54] was developed jointly by the European Commission's Directorate General for Environment and the European Environment Agency and harmonized with the Directive (EU) 2024/2881 on ambient air quality and cleaner air for Europe [55]. According to [11], "Mostar" air-quality monitoring station is categorized as a traffic and urban monitoring station, and it is selected among all Serbian air-quality monitoring stations, since it is aligned with the requirements to be registered and to report data to the European Air Quality Index web portal. It enables continuous data collection and distribution ("more than 75% of valid data out of all the possible data in a year" [11]), and these data are related to all required pollutant types [54]: particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), ozone (O_3) and sulphur dioxide (SO_2). According to [54] (status as of 8 August 2026), "Mostar" station reports on PM_{10}, $PM_{2.5}$, NO_2 and SO_2.



(**a**) (**b**)

Figure 4. The location of "Mostar" monitoring station: (**a**) On the website of the Serbian national network of air-quality monitoring stations [53], where each colored circle presents one air-quality monitoring station; (**b**) on the European Air Quality Index official web portal [54], with a map of Serbia and a pop-up with an air-quality status summary (data obtained on 8 August 2026).

The coordinates (44°47′55″ N 20°27′00″ E) of "Mostar" monitoring station are obtained from the European Air Quality Index website [54]. "Mostar" station is presented in Figure 5 in Google maps satellite view [56] and in Figure 6 in Google Earth view [57].

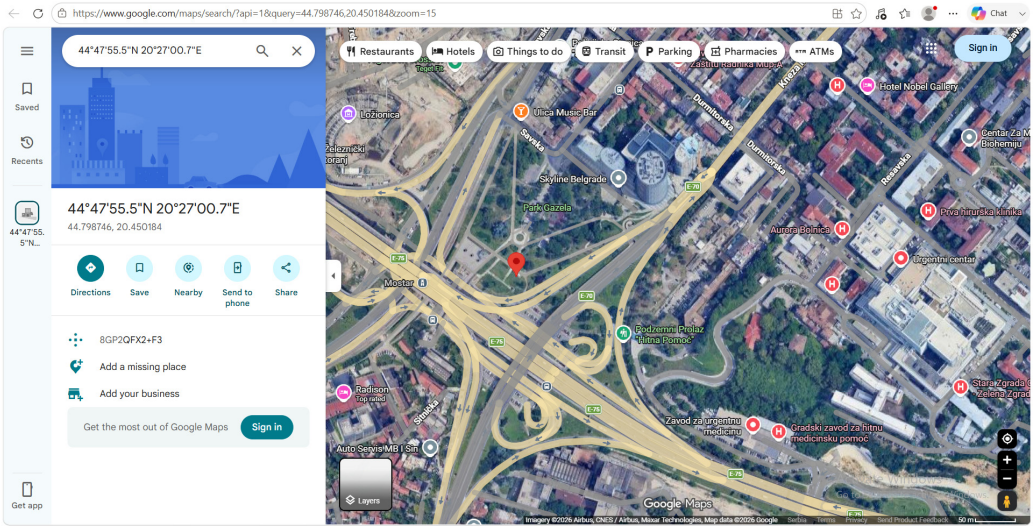


Figure 5. Air pollution monitoring station "Mostar" presented on the Google Maps website [56]; the exact location is marked with the red pin symbol.

Figures 5 and 6 present the exact location of "Mostar" station and the ambient (i.e., surrounding) area. Both figures show the closest surrounding of the "Mostar loop" crossroad and "Mostar" station. The surrounding area is not industrial and not a "very urban" space: three segments have parks, one is empty, and two segments are dominantly residential. Therefore, we can conclude that in this case, the ambient conditions do not have significant influence on the results; air pollution measurements' results from "Mostar" station are closely related to the traffic source, not the industrial or residential source.

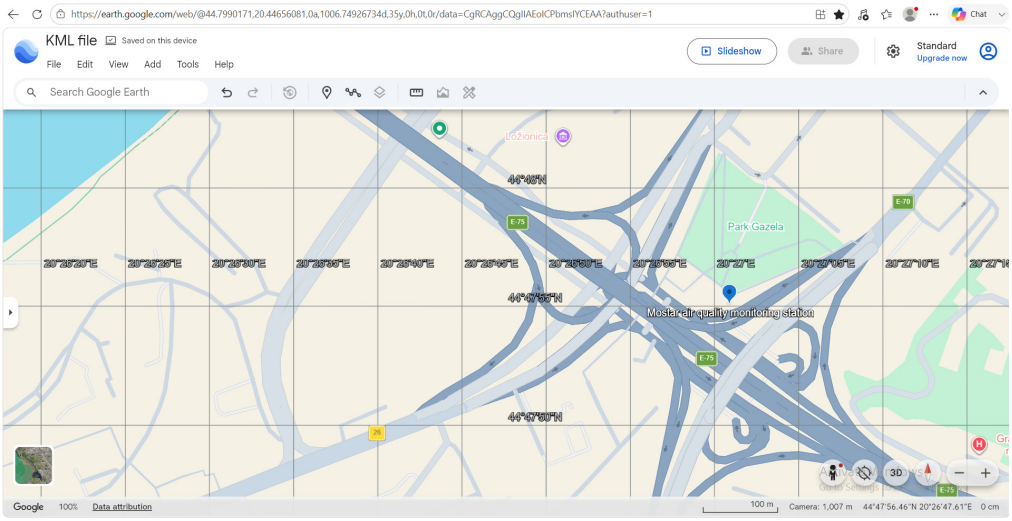


Figure 6. Air pollution monitoring station "Mostar" presented on the Google Earth website [57]; the exact location is marked with the blue pin symbol.

3.4.2. Dataset Acquisition, Preprocessing and Clustering

The case study presents results based on the dataset that was obtained from the SEPA website [53], for the period of 22 May 2023 to 21 May 2024. There are several steps performed from accessing the SEPA website to result processing:

1. *Data acquisition*. The version of the SEPA website from the years 2023/24 had data available at hourly resolution and average daily values, presented as data tables directly on the website pages, after selecting the monitoring station. There was no option for XLS or CSV download (which is currently available on the latest version of this website). Therefore, once a day, the daily average values were copied directly from the web page data table and pasted into a simple MS Excel file. Daily averages for all measured data types (air pollutants and meteorological factors) were included for each day in the MS Excel file as a separate row, while the various types of measured data were separated in columns.
2. *Data preprocessing*. To create a file in a format suitable for processing by the R program, data from the MS Excel file were exported into a single plain textual file (with .txt extension) having tab-delimited structure. The first row of this file contains the names of the measured data types.
3. *Data quality control*. The annually collected data are cross-seasonal, and simple data quality control was conducted by checking: (a) redundancy, (b) data formats and (c) data domains (intervals of valid data values). It was determined that no errors were found regarding these aspects.
4. *Data clustering.* The dataset with annually collected data was split into other datasets and saved to separate text files. Each separate text file (representing a specific type of dataset) had the same data structure as the source file (with cross-seasonal data, i.e., data collected annually). Data were horizontally partitioned, according to the dates of measurements, i.e., the date value from the first column. The segmentation of data was performed for the following:
    (a) *Seasonal clustering.* Based on the analysis of the dates for each row, the annual dataset was partitioned into four files for each season.
    (b) *Cross-seasonal in-sample segmentation for prediction.* The annual dataset was split into three separate files to enable: (a) cross-seasonal creation of MLR-based prediction models that relate pollutant values to other pollutants; (b) input to prediction of air pollutants, based on previously created MLR-based prediction

models; and (c) measured data to be compared with the predicted fit and interval values.

5. Missing data determination and processing:
   (a) The dates of data collection were between 22 May 2023 and 21 May 2024. In the annually collected dataset, missing days are:
   21 June 2023–7 August 2023 (48 days = 53% of summer);
   25 April 2024–28 April 2024 (4 days = 4.44% of spring).
   Summary: A total of 52/366 days (14%) are missing, while 86% were properly collected. Note that 29.2.2024 was a leap day, so the year 2024 had 366 days.
   (b) Reasons for omitted data are related to the organization of data collection, but these data were available on the SEPA website and could have been collected.
   (c) *Spring data partly completed*. Since spring 2023 was incomplete (data collection started on 22 May 2023), the first part of the missing spring season data were compensated with spring data from the year 2024 (21 March–21 May 2024). Still, 25 April 2024–28 April 2024 (4 days) are missing, so the spring dataset is regarded as fragmentary.
   (d) *Summer data incomplete*. Only data from 8 August 2023–20 September 2023 were included, but 21 June 2023–8 August 2023 data are missing. No summer data were collected in the year 2024. In further analysis, the summer dataset is regarded as fragmentary.

3.4.3. Dataset Structure and Characteristics of Clustered Datasets

The first part of the annual dataset textual file, which was created during data preprocessing, is presented as the file screenshot in Figure 7.

DAY	SO2	PM10	NO2	NOX	CO	NO	PM25	W	AP	T	RH
5/22/2023	7.91	8.78	32.49	51.42	0.45	12.35	6.34	1.23	1003.37	21.01	55.88
5/23/2023	5.45	11.18	39.7	80.54	0.49	26.63	6.94	0.94	1002.35	21.78	56.96
5/24/2023	7.81	14.76	42.33	84.57	0.46	27.56	10.31	0.95	1003.49	21.4	65.98
5/25/2023	4.57	12.32	33.36	64.88	0.36	20.62	8.79	0.98	1006.54	20.38	79.66
5/26/2023	4.46	11.6	29.02	61.8	0.36	21.39	9.25	0.95	1006.36	19.71	85.23
5/27/2023	7.48	9.65	30.2	53.5	0.34	15.14	8.14	1.03	1006.59	18.13	85.23
5/28/2023	5.32	9.66	21.1	37.3	0.31	10.58	7.73	0.88	1004.73	19.03	81.16
5/29/2023	5.8	8.93	30.54	57	0.35	17.25	6.85	0.62	1004.38	19.8	75.36
5/30/2023	8.07	10.23	28.12	66.35	0.36	24.91	7.69	0.91	1004.97	20.08	77.42
5/31/2023	4.64	7.79	29.98	54.33	0.32	15.94	6.01	1.06	1007.01	19.53	81.26

Figure 7. Part of annual textual file created during data preprocessing.

Figure 7 presents the structure of each textual data file, i.e., a specific dataset. All seasonal datasets and datasets created for cross-seasonal prediction have the same structure, as presented in Figure 7. There are twelve columns in each dataset, including the following:

- Day (date of data acquisition).
- Air pollutant values (7 columns): sulphur dioxide (SO_2), particulate matter (PM_{10} and $PM_{2.5}$), nitrite oxides (NO_2, NO, and NO_X), and carbon monoxide (CO).
- Meteorological values (4 columns): W (wind speed), AP (atmospheric pressure), T (temperature), and RH (relative humidity).

The measurement units for the data obtained from the monitoring station are as follows:

- Air pollutants: air pollutant values are presented with the $\mu g/m^3$ measurement unit, except for CO, which is presented in mg/m^3.
- Meteorological parameters: T (air temperature, Celsius scale), RH (relative humidity, percentage), AP (atmospheric pressure, mbar), and W (wind speed, m/s)

Data collection was performed during one year from 22 May 2023 to 21 May 2024. The collected and preprocessed data were organized in data clusters according to the date value. Table 1 presents basic data about seasonally clustered data.

Table 1. Time periods and seasonally clustered datasets.

Dates in the Dataset	Season	Seasonally Clustered Dataset	Number of Days (Data Items)
22 May 2023–20 June 2023	Spring 2023	Spring	30 (330)
8 August 2023–20 September 2023	Summer 2023	Summer	44 (484)
21 September 2023–20 December 2023	Autumn 2023	Autumn	91 (1001)
21 December 2023–20 March 2024	Winter 2023/24	Winter	91 (1001)
21 March 2024–21 May 2024	Spring 2024	Spring	58 (638)
		Total	314 (3454)

Table 1 presents seasonally clustered data characteristics in four columns: date intervals, seasons, seasonally clustered dataset and number of days (data rows) and data items. Data items refer to the total number of measured values to be processed, and the number is derived from the number of days (data rows) multiplied by 11 measurement data types/columns (7 pollutants + 4 meteorological factors). In the TOTAL below the table, there is a sum of all days included in this case study, as well as the sum of all data items that were processed.

Cross-seasonal data from the annual dataset were horizontally partitioned into three datasets for the in-sample-based initial prediction of air pollutant values in relation to the values of other air pollutants. Two datasets were chosen to have the last two months' data separated for the purpose of prediction input and comparison of predicted values with measured values (with an equal number of 29 days/data rows). Table 2 presents these three datasets' basic characteristics.

Table 2. Datasets for the initial cross-seasonal pollutant-to-pollutant prediction.

Dates in the Dataset	Dataset Purpose	Number of Days (% of Days) (Data Items)
22 May 2023–20 March 2024	MLR-based prediction model creation	256 (82%) (2816)
21 March 2024–18 April 2024	Prediction input	29 (9%) (319)
19 April 2024–21 May 2024	Comparison measured data–predicted data	29 (9%) (319)

Table 2 presents datasets for the initial prediction of pollutant values based on MLR prediction models (created from the first dataset) and data from the prediction input dataset (the second dataset). The predicted values are compared with the data from the third dataset.

Additional in-sample fit analysis of one of the created MLR-based prediction models (which had the highest R^2 value for relating pollutant triplets) was conducted on the segmented cross-seasonal data with two datasets, as presented in Table 3.

Table 3 presents the horizontal partitioning of the obtained annual dataset into two datasets. The first dataset is used for MLR-based prediction model creation, while the second dataset is used for prediction input (values from the independent variables in the MLR equation—two pollutants) and for the comparison of the predicted values (dependent variable—one pollutant) with the measured values (the independent variable). Relating the input pollutant values of two pollutants to the computed prediction value of a third pollutant is based on the "equality of dates": the same-day measured data of the two independent variables are used for the computation of the predicted value of the third

variable. From the same data row (used to extract values of independent pollutants), the measured value of the dependent variable (the third pollutant) is used to be compared with the predicted value of that third pollutant.

Table 3. Datasets for in-sample fit analysis of one MLR-based prediction model for cross-seasonally relating a pollutant's value to two other pollutants' values.

Dates in the Dataset	Dataset Purpose	Number of Days (% of Days) (Data Items)
22 May 2023–20 March 2024	MLR-based prediction model creation	256 (82%) (2816)
19 April 2024–21 May 2024	Prediction input	29 (9%) (319)
19 April 2024–21 May 2024	Comparison measured data–predicted data	29 (9%) (319)

4. Results

4.1. List of Comparable-Pollutant Types

To establish the comparability and comprehensiveness of results in this research study, the data structure of the obtained dataset was compared with the data structures of other air-quality monitoring stations. All air pollution monitoring stations in Serbia (89) were examined from the monitoring station map [53]. Each monitoring station was analyzed regarding the availability of sensors for specific types of air pollutants. Results are derived from all Serbian monitoring stations with the status of their sensors' availability on April 23, 2026, and the statistics are presented in Figure 8.

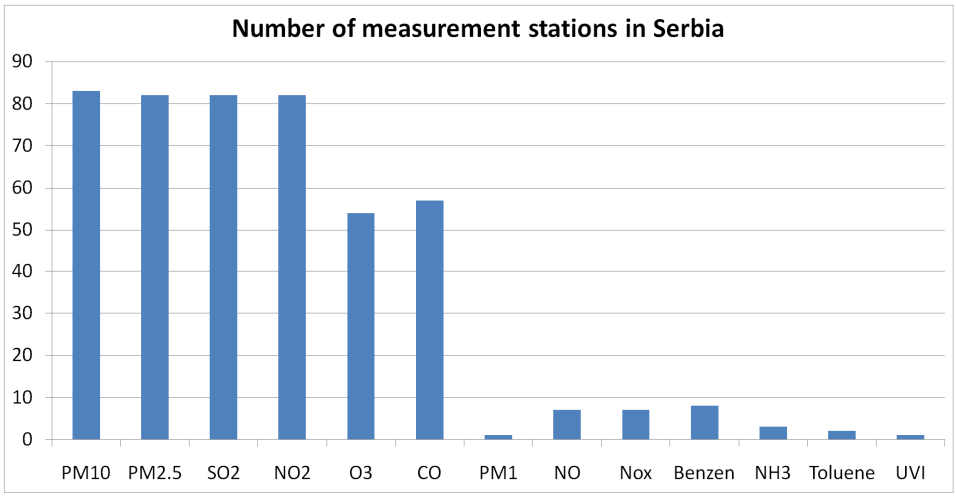


Figure 8. Statistics on air-quality pollutant measurement availability across all 89 air-quality monitoring stations in Serbia (status as of 23 April 2026).

The diagram in Figure 8 presents the numbers of air-quality monitoring stations in Serbia that are able to detect and measure values of specific air pollutants. These numbers of Serbian air-quality monitoring stations are as follows: PM_{10} (83), $PM_{2.5}$ (82), SO_2 (82), NO_2 (82), O_3 (54), CO (57), PM_1 (1), NO (7), NO_x (7), benzene (8), NH_3 (3), toluene (2), and UVI (1).

According to the previously presented situation with air-quality monitoring stations in Serbia, sensors for PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3 and CO air pollutants have the highest availability (at more than 50% of all air-quality monitoring stations). Therefore, these air

pollutants could be categorized as "Serbian-comparable" (in further text also "comparable") parameters/components. Other pollutants (PM_1, NO, NO_x, benzene, NH_3, toluene, and UVI) do not have a significant distributions of available sensors at air pollution monitoring stations in Serbia.

Relating the Obtained Comparable-Pollutant Types to the Case Study Dataset Structure

The selected air-quality monitoring station, "Mostar" (status as of 23 April 2026), was able to detect and measure values of the comparable pollutants of particulate matter ($PM_{2.5}$ and PM_{10}), sulphur dioxide (SO_2), and nitrogen dioxide (NO_2), while detection of carbon monoxide (CO) and ozone (O_3) was not possible at this monitoring station. However, in the period of 22 May 2023–21 May 2024, "Mostar" air pollution monitoring station was able to detect and measure carbon monoxide (CO), nitrogen monoxide (NO) and nitrogen oxides (NO_X). Since CO data are included in previously obtained dataset (from the years 2023/24), they were included in further analysis, together with the meteorological parameters of air temperature, relative air humidity, atmospheric pressure and wind speed. NO and NO_X pollutants were not included in the comparable list, so they were eliminated from the further analysis. Finally, the dataset that was used in further analysis had the following characteristics:

- Obtained in the period of 22 May 2023–21 May 2024 from "Mostar" air-quality monitoring station, located at the heaviest-traffic location in the Serbian capital city, Belgrade, with a total number of data rows (days) of 314.
- Pollutant types that were included in the analysis were the Serbian comparable parameters (pollutants) of PM_{10}, $PM_{2.5}$, SO_2, NO_2 and CO (O_3 data are not measured at "Mostar" station).
- Meteorological factors measurable at "Mostar" station were included in the analysis: atmospheric pressure, wind speed, relative humidity, and temperature.

4.2. Seasonal Analysis of Impact of Meteorological Factors on Air Pollutant Values Derived from Prediction Models

To determine the relation between pollutants values and meteorological factors' values, seasonally clustered datasets were used for creating MLR-based prediction models for each comparable pollutant (except O_3) in relation to all meteorological factors. The results are presented in Table 4 (spring), Table 5 (summer), Table 6 (autumn) and Table 7 (winter). All tables present statistical results for each MLR-based prediction model with *p*-value, F-statistic, residual standard error (RSE), Multiple R-squared (R^2) and Adjusted R-squared, and each model is presented with computed values for α (intercept) and β coefficients in the MLR function $\beta_1\ AP + \beta_2\ T + \beta_3\ W + \beta_4\ RH$.

Table 4. Spring season: MLR relation between pollutant and meteorological factor values.

MLR Model/Coefficients $\alpha + \beta_1 AP + \beta_2 T + \beta_3 W + \beta_4 RH$	***p*-Value**	**F-Statistic**	**RSE**	**Multiple R^2**	**Adjusted R^2**
SO_2~AP + T+W + RH	2.473×10^{-7}	11.24	1.885	0.3513	0.32
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	168.21685	−0.14953	−0.23395	−0.53550	−0.07025
PM_{10}~AP + T+W+RH	0.01617	3.234	9.654	0.1348	0.09315
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	316.24755	−0.27582	−0.64741	−0.93846	−0.22127
NO_2~AP + T + W + RH	$<2.2 \times 10^{-16}$	32.79	9.118	0.6124	0.5938
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	6.51125	0.11125	−1.12055	−9.44614	−0.71546

Table 4. *Cont.*

MLR Model/Coefficients $\alpha + \beta_1 AP + \beta_2 T + \beta_3 W + \beta_4 RH$	*p*-Value	F-Statistic	RSE	Multiple R^2	Adjusted R^2
CO~AP + T + W + RH	2.778×10^{-9}	15.03	0.0951	0.4201	0.3922
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−1.921905	0.002887	−0.011673	−0.080445	−0.004007
$PM_{2.5}$~AP + T + W + RH	0.0784	2.179	5.348	0.09503	0.05142
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	136.46198	−0.11509	−0.31714	−1.57624	−0.08404

Table 5. Summer season: MLR relation between pollutant and meteorological factor values.

MLR Model/Coefficients $\alpha + \beta_1 AP + \beta_2 T + \beta_3 W + \beta_4 RH$	*p*-Value	F-Statistic	RSE	Multiple R^2	Adjusted R^2
SO_2~AP + T + W + RH	0.2564	1.387	4.008	0.1245	0.03471
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	182.74511	−0.17401	−0.17273	−0.62666	0.11563
PM_{10}~AP + T + W + RH	1.598×10^{-7}	15.05	3.543	0.6068	0.5665
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−281.78514	0.25165	1.65010	−1.81415	0.13504
NO_2~AP + T + W + RH	0.4928	0.8662	16.18	0.08159	−0.01261
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	230.5623	−0.1727	−0.6840	2.5568	−0.4758
CO~AP + T + W + RH	4.106×10^{-6}	11.11	0.0628	0.5327	0.4848
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−10.980193	0.011145	0.013316	−0.080172	−0.002552
$PM_{2.5}$~AP + T + W + RH	1.682×10^{-9}	21.76	2.185	0.6906	0.6589
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−288.01333	21.76	2.185	0.6906	0.6589

Table 6. Autumn season: MLR relation between pollutant and meteorological factor values.

MLR Model/Coefficients $\alpha + \beta_1 AP + \beta_2 T + \beta_3 W + \beta_4 RH$	*p*-Value	F-Statistic	RSE	Multiple R^2	Adjusted R^2
SO_2~AP + T + W + RH	0.0003951	5.719	3.605	0.2101	0.1734
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	30.728843	−0.006087	3.605	0.2101	0.1734
PM_{10}~AP + T + W + RH	2.285×10^{-6}	9.429	16.94	0.3048	0.2725
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−461.7564	0.5314	−1.5595	−4.8972	−0.2570
NO_2~AP + T + W + RH	2.046×10^{-12}	21.73	9.808	0.5026	0.4795
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−208.9461	0.2867	0.4132	−7.4467	−0.4847
CO~AP + T + W + RH	6.351×10^{-7}	10.41	0.2694	0.3262	0.2949
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−3.652143	0.005050	−0.024814	−0.197527	−0.004182
$PM_{2.5}$~AP + T + W + RH	3.041×10^{-9}	14.81	13.44	0.4078	0.3803
Coefficients	α:	β_1:	β_2:	β_3:	β_4:
	−336.07725	0.38765	−1.46729	−5.89504	−0.07372

Table 7. Winter season: MLR relation between pollutant and meteorological factor values.

MLR Model/Coefficients α + β1AP + β2T + β3W + β4RH	*p*-Value	F-Statistic	RSE	Multiple R^2	Adjusted R^2
SO_2~AP + T + W + RH Coefficients	0.0009742 α: 4.17437	5.103 β1: 0.02283	5.131 β2: 0.11088	0.1918 β3: −2.31928	0.1542 β4: −0.17192
PM_{10}~AP + T + W + RH Coefficients	2.467×10^{-5} α: −506.6289	7.671 β1: 0.5841	20.94 β2: −0.6138	0.263 β3: −12.1886	0.2287 β4: −0.1941
NO_2~AP + T + W + RH Coefficients	2.304×10^{-10} α: −842.5447	17.12 β1: 0.9139	13.66 β2: 1.5065	0.4434 β3: −6.5192	0.4175 β4: −0.5539
CO~AP + T + W + RH Coefficients	0.0006247 α: -1.259×10^{1}	5.406 β1: 1.337×10^{-2}	0.3327 β2: 7.407×10^{-3}	0.2009 β3: -1.476×10^{-1}	0.1637 β4: 7.082×10^{-5}
$PM_{2.5}$~AP + T + W + RH Coefficients	8.202×10^{-9} α: 8.202×10^{-9}	13.95 β1: 13.95	15.14 β2: 15.14	0.3935 β3: 0.3935	0.3653 β4: 0.3653

To compare the results from the multiple regression models and linear regression (LR) models (relating comparable-pollutant values to meteorological factors), for the purpose of an ablation study, linear regression models were created for pairs of each pollutant and each meteorological factor. Table 8 presents only values for Multiple R^2 for the comparison between LR and MLR results.

Table 8. Seasonal LR prediction models compared with MLR prediction models in Multiple R^2 values.

Seasonal LR Model/ Multiple R^2 PO~[AP, T, WR, H]	Spring	Summer	Autumn	Winter
SO_2~AP	0.09062	0.0007135	0.03159	0.0004422
SO_2~T	0.1132	0.03532	0.009036	0.04983
SO_2~W	0.002011	0.01878	0.1607	0.03351
SO_2~RH	0.06476	0.09826	0.02254	0.1197
SO_2~AP + T + W + RH	0.3513	0.1245	0.2101	0.1918
PM_{10}~AP	0.01537	0.028	0.1382	0.1466
PM_{10}~T	0.03402	0.4791	0.1953	0.09445
PM_{10}~W	0.003351	0.01841	0.01735	0.1848
PM_{10}~RH	0.04777	0.01784	0.05981	0.01524
PM_{10}~AP + T + W + RH	0.1348	0.6068	0.3048	0.263
NO_2~AP	0.04345	0.006039	0.0588	0.05172
NO_2~T	0.0005154	0.002582	0.2691	0.09377
NO_2~W	0.09235	0.02646	0.1748	0.03128
NO_2~RH	0.344	0.06564	0.2727	0.2156
NO_2~AP + T + W + RH	0.6124	0.08159	0.5026	0.4434
CO~AP	0.04925	0.1203	0.1296	0.1359
CO~T	0.02497	0.07146	0.1407	0.03315
CO~W	0.1337	0.1404	0.1001	0.1156
CO~RH	0.09778	0.05111	0.05791	0.005547
CO~AP + T + W + RH	0.4201	0.5327	0.3262	0.2009

Table 8. *Cont.*

Seasonal LR Model/ Multiple R^2 PO~[AP, T, WR, H]	Spring	Summer	Autumn	Winter
$PM_{2.5}$~AP	0.004883	0.001827	0.1415	0.2339
$PM_{2.5}$~T	0.02844	0.2304	0.2956	0.2484
$PM_{2.5}$~W	0.01242	0.1547	0.02968	0.2184
$PM_{2.5}$~RH	0.0109	0.04716	0.1433	0.09404
$PM_{2.5}$~AP + T + W + RH	0.09503	0.6906	0.4078	0.3935

Diagrams representing Multiple R^2 values obtained with seasonal datasets in relating comparable air pollutants to meteorological factors are presented in Figures 9–13.

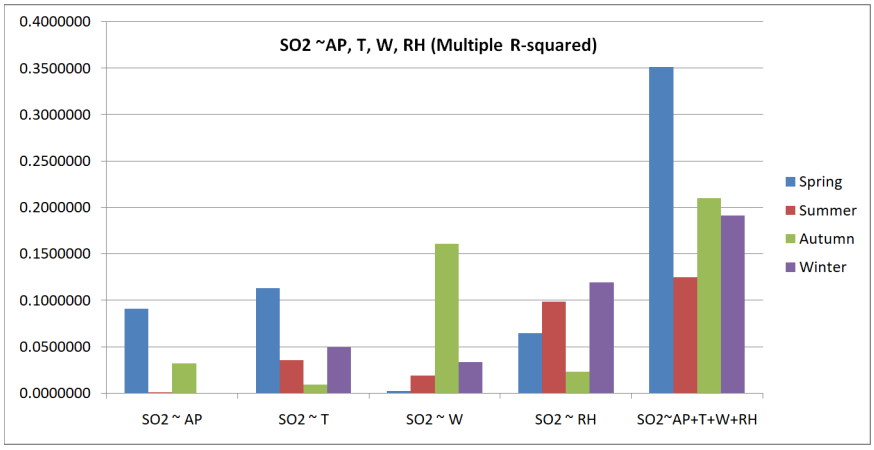


Figure 9. Seasonal impact R^2 values for SO_2 in relation to meteorological factors AP, T, W, and RH.

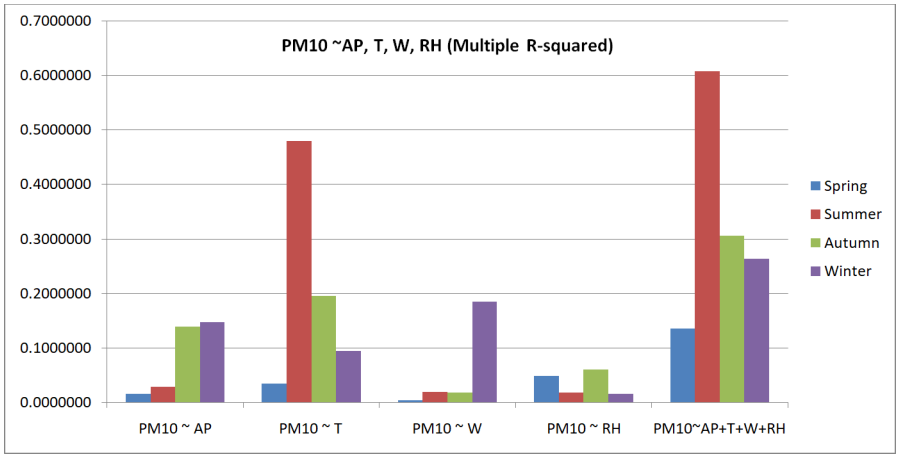


Figure 10. Seasonal impact R^2 values for PM_{10} in relation to meteorological factors AP, T, W, and RH.

Multiple R^2 values for data from Tables 4–8 are visualized in Figures 9–13. They present comparisons of results of linear regression models (relating each pollutant to each meteorological factor) and multiple linear regression models (relating each air pollutant to all four meteorological parameters). By analyzing these results, the following could be concluded:

- Figure 9 shows that the highest Multiple R^2 are reached for the relations of SO_2~AP in spring, SO_2~T in spring, SO_2~W in autumn, and SO_2~RH in winter, while overall, the highest value is that for SO_2~AP + T + W + RH in spring (0.3513).

- Figure 10 shows that the highest Multiple R^2 are reached for the relations of PM_{10}~AP in winter, PM_{10}~T in summer, PM_{10}~W in winter, and PM_{10}~RH in autumn, while overall, the highest value is that for PM_{10}~AP + T + W + RH in summer (0.6068).
- Figure 11 shows that the highest Multiple R^2 are reached for the relations of NO_2~AP in autumn, PM_{10}~T in autumn, NO_2~W in autumn, and NO_2~RH in spring, while overall, the highest value is that for NO_2~AP + T + W + RH in spring (0.6124).
- Figure 12 shows that the highest Multiple R^2 are reached for the relations of CO~AP in winter, CO~T in autumn, CO~W in summer, and CO~RH in spring, while overall, the highest value is that for CO~AP + T + W + RH in summer (0.5327).
- Figure 13 shows that the highest Multiple R^2 are reached for the relations of $PM_{2.5}$~APin winter, $PM_{2.5}$~T in autumn, $PM_{2.5}$~W in winter, and $PM_{2.5}$~RH in autumn, while overall, the highest value is that for $PM_{2.5}$~AP + T + W + RH in summer (0.6906).

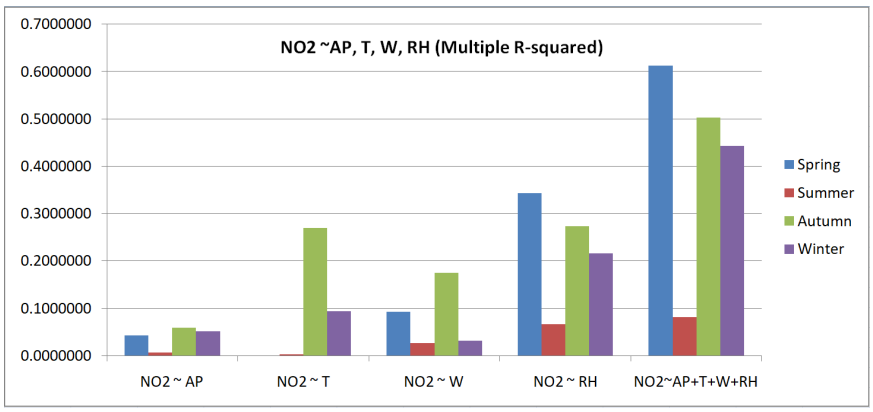


Figure 11. Seasonal impact R^2 values for NO_2 in relation to meteorological factors AP, T, W, and RH.

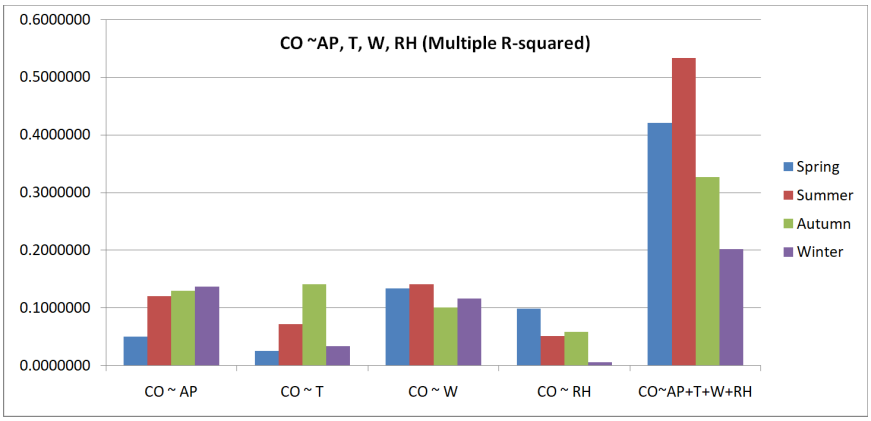


Figure 12. Seasonal impact R^2 values for CO in relation to meteorological factors AP, T, W, and RH.

Figure 14 presents a diagram with R^2 values obtained for prediction functions that relate all comparable-air-pollutant values to seasonally clustered data: each air pollutant is related to each meteorological factor (with the linear regression method) and to all four meteorological factors (with the multiple linear regression method).

By comparing the values in Figure 14, considering different air pollutants, the following could be concluded:

- PM_{10}, $PM_{2.5}$ and CO reach the highest relation values with all four meteorological factors in summer, while SO_2 and NO_2 do so in spring.
- $PM_{2.5}$ has the highest value of Multiple R^2 (0.6906) in relation to all four meteorological factors.

- All comparable parameters in relation to all four meteorological factors (in both linear and multiple linear regression) have values of Multiple $R^2 < 0.7$.

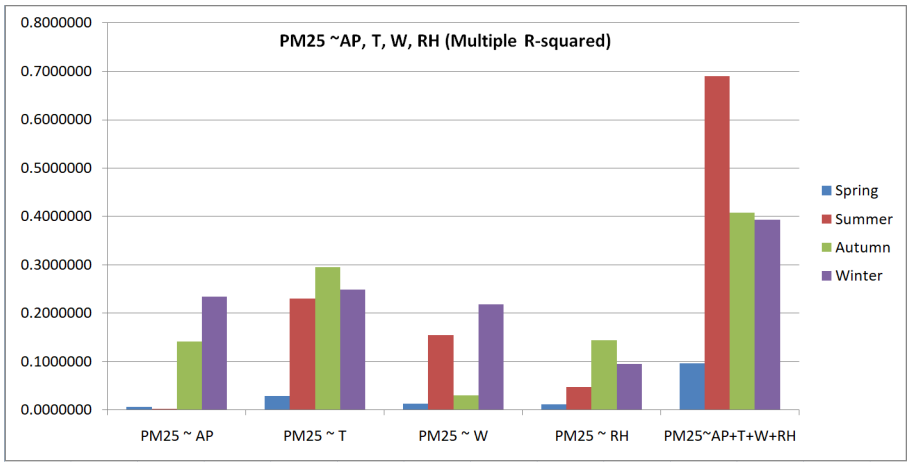


Figure 13. Seasonal impact to R^2 values for $PM_{2.5}$ in relation to meteorological factors AP, T, W, RH.

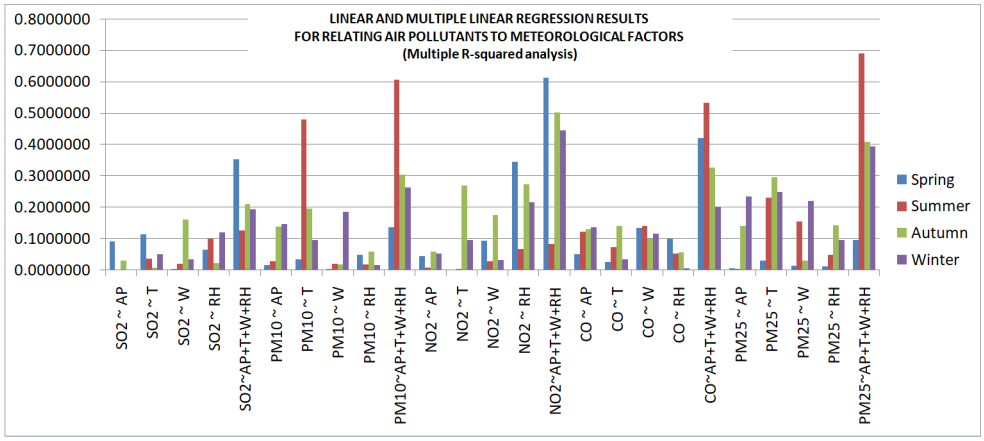


Figure 14. Comparison of linear and multiple linear regression values for R^2 relating air pollutants to meteorological factors (data from "Mostar" station, years 2023/24).

4.3. Results of Cross-Seasonal Analysis of Relations Between One Pollutant's Value and Two Other Pollutants' Values

The results presented in this section are related to the MLR prediction models for establishing relationships for triplets of air pollutants and the analysis of the MLR prediction models' R^2 values. The MLR prediction models were created with the R program and cross-seasonal datasets (subsets of items of the annual dataset).

The relation for triplets of pollutants (PO_1~PO_2 + PO_3) is visualized with ggplot2 commands in R (Listing 2) and presented in Figure 15 as a visual illustration of the relation of an air pollutant triplet: SO_2 values (presented as colored dots) depend on two other pollutants—NO_2 (presented on the *X*-axis) and PM_{10} (presented on the *Y*-axis).

MLR prediction models were created to relate pollutants in triplets of Serbian-comparable-pollutant values. Table 9 presents the whole list of all created correlation models and the highest obtained values (>0.7) for Multiple R^2. In Table 9, MLR prediction models with the following are specially emphasized:

(a) The highest Multiple R^2 values (>0.7);
(b) Both PM_{10} and $PM_{2.5}$ being included in the MLR model triplet.

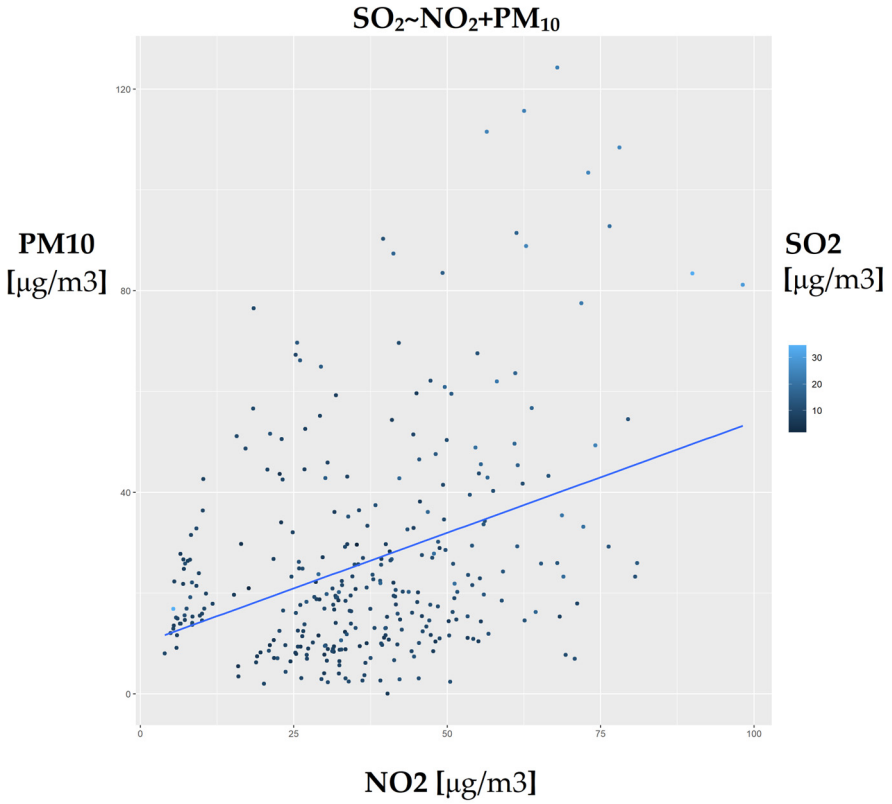


Figure 15. Scatterplot diagram of SO_2 µg/m^3~NO_2 µg/m^3 + PM_{10} µg/m^3 with the annual data (22 May 2023–21 May 2024) obtained from "Mostar" station in the years 2023/24; the line presents a prediction linear function derived from the prediction model relating NO_2 and PM_{10} values to SO_2 values (presented as dots around the line).

Table 9. Cross-seasonally created MLR prediction models for relating pollutant triplets.

No	Correlation	Multiple $R^2 > 0.7$	PM_{10}~$PM_{2.5}$
1	SO_2~PM_{10} + NO_2		
2	SO_2~PM_{10} + CO		
3	SO_2~PM_{10} + $PM_{2.5}$		
4	SO_2~NO_2 + CO		
5	SO_2~NO_2 + $PM_{2.5}$		
6	SO_2~CO + $PM_{2.5}$		
7	PM_{10}~SO_2 + NO_2		
8	PM_{10}~SO_2 + CO	0.8427	
9	PM_{10}~SO_2 + $PM_{2.5}$	0.9621	√
10	PM_{10}~NO_2 + CO	0.8532	
11	PM_{10}~NO_2 + $PM_{2.5}$	0.9603	√
12	NO_2~SO_2 + PM_{10}		
13	NO_2~SO_2 + CO		
14	NO_2~SO_2 + $PM_{2.5}$		
15	NO_2~PM_{10} + CO		
16	NO_2~PM_{10} + $PM_{2.5}$		
17	NO_2~CO + $PM_{2.5}$		
18	CO~SO_2 + PM_{10}	0.853	
19	CO~SO_2 + NO_2		
20	CO~SO_2 + $PM_{2.5}$	0.8416	
21	CO~PM_{10} + NO_2	0.8776	
22	CO~PM_{10} + $PM_{2.5}$	0.8424	
23	CO~NO_2 + $PM_{2.5}$	0.8755	
24	$PM_{2.5}$~SO_2 + PM_{10}	0.9574	√

Table 9. *Cont.*

No	Correlation	Multiple $R^2 > 0.7$	PM_{10}~$PM_{2.5}$
25	$PM_{2.5}$~SO_2 + NO_2		
26	$PM_{2.5}$~SO_2 + CO	0.8096	
27	$PM_{2.5}$~PM_{10} + NO_2	0.9574	√
28	$PM_{2.5}$~PM_{10} + CO	0.9523	√
29	$PM_{2.5}$~NO_2 + CO	0.8396	
√	Prediction model that includes both $PM_{2.5}$ and PM_{10}		

Table 9 shows the 29 created MLR prediction models that relate pollutant triplets. Among them, fourteen MLR prediction models with values of Multiple $R^2 > 0.7$ are emphasized. Among these fourteen MLR models, there are five MLR models that relate values of PM_{10} to $PM_{2.5}$ in the pollutant triplets combined with other pollutants.

According to Abdulah et al. [58], the determination coefficient R^2 is used to measure the extent to which a prediction model fits the data. According to Xu et al. [59], R^2 and R^2 (adjusted) can be used as the criterion numbers to evaluate the regression effect, where R^2 is considered appropriate enough, and even R^2 (adjusted) is used in some human-environment studies instead of R^2, because it does not depend on the number of independent variables. In [59], it is concluded that in the case of small coefficients of determination, the most important aspect is to check if that regression model passes the significance test; then it could be considered appropriate to quantitatively explain the extent of the independent variables' impact on the dependent variables. However, regression models based on low values of R^2 cannot be used to predict the specific value of the dependent variables based on the existing independent variables [59].

Considering the computed MLR prediction models with the highest R^2 to be relevant, we determined that among the created models, there are 14 models with R^2 above 0.7 and 5 models with R^2 above 0.95. These five models relate other air pollutant values to both PM_{10} and $PM_{2.5}$. Xue et al. [60] analyzed chemical characteristics, sources and the relation between concentrations of the air pollutants of $PM_{2.5}$ (particle size smaller than 2.5 μm) and PM_{10} (particle size smaller than 10μm). In [60], the $PM_{2.5}/PM_{10}$ ratios were approximately 0.7, and paired sample *t*-tests confirmed similar sources of $PM_{2.5}$ and PM_{10}, which are, in the case study (Hefei City in China), related to coal combustion and high-tech manufacture industry (heavy-metal pollution). According to findings from [60], regarding PM2.5 and PM_{10} characteristics, in our research, the MLR models that relate them were excluded from further analysis. Therefore, other generated MLR models (9 models with $R^2 > 0.7$) were considered.

4.4. Prediction of Air Pollutant Values Based on MLR Model with the Highest Multiple R^2

To illustrate prediction of values for Serbian-comparable air pollutants, we selected one of the previously nine computed and extracted MLR models, the one with the highest R^2 value, to be used in prediction and the comparison with measured values. The selected MLR model is CO~PM_{10} + NO_2, with R^2 = 0.8776. The MLR model coefficients and statistical evaluation parameters (needed for the significance test) are presented in Table 10 as a result of R program computation.

The values presented in Table 10 are related to the coefficients of the MLR-based prediction model function and the results of prediction model significance testing (performed with R program—*summary* function).

Table 10. Statistics of MLR prediction model with the highest relevant R^2 value.

MLR Model: CO~PM10 + NO2			
Residuals:	Residual standard error: 0.1163 on 253 degrees of freedom		
Min: −0.31442	Multiple R-squared: 0.8776		
1Q: −0.05907	Adjusted R-squared: 0.8766		
Median: 0.00147	F-statistic: 907 on 2 and 253 DF		
3Q: 0.06626	*p*-Value: $<2.2 \times 10^{-16}$		
Max: 0.63975			
Coefficients:			
Estimate	Std. Error	t value	Pr(>|t|)
(Intercept) 0.0353465	0.0162960	2.169	0.031 *
PM10 0.0122536	0.0003631	33.743	$<2 \times 10^{-16}$ ***
NO2 0.0037720	0.0004412	8.550	1.19×10^{-15} ***

Signif. codes: 0, '***'; 0.001, '**'; 0.01, '*'; 0.05, '.'; 0.1, ' '; 1.

MLR-based prediction model function: $CO = \alpha + \beta_1\ PM10 + \beta_2\ NO_2$, where $\alpha = 0.0353465$, $\beta_1 = 0.0122536$, and $\beta_2 = 0.0037720$. Thus,

$$CO = 0.0353465 + 0.0122536\ PM_{10} + 0.0037720\ NO_2$$

A significance test was performed with the obtained statistical values for the regression model (LR and MLR) to quantitatively determine the extent of the impact of the independent variables' values on the dependent variable and the overall model fit. For this MLR prediction function, all relevant statistical indicators (except Min and Max Residuals) had adequate values, so it could be concluded that this multiple linear regression function represents a good fit to the input data that were used for creating this function:

- Residuals: Min, Median, and Max: The median value is close to zero, and small values are expected for good functions. However, Min and Max values do not have similar absolute values (since residuals are the differences between the actual measured values and the corresponding values on the fitted regression line, it is expected that Min and Max have similar absolute values). Currently, $|-0.31| < |0.63|$, which is not considered a good fit, while 1Q and 3Q have similar absolute values $|-0.05| \sim |0.06|$, and this could be considered appropriate.
- Standard error for each coefficient: Low standard error for each coefficient indicates reliable conclusions. In this case, each coefficient's standard error is very low, approximately zero. Small values are expected for good functions.
- t-Value for each coefficient: If this value is large, it shows that the independent variable (predictor) affects more clearly the value of the dependent variable. In this case, PM10 has a value of approx. 34 and NO2 a value of 9, which leads to the conclusion that both predictors are considerable.
- Pr(>|t|) for each coefficient is compared with significance codes; PM10 and NO2 coefficients have approximately zero value, which aligns with values annotated with ***.
- Residual standard error: smaller values indicate a better a regression model fit to the dataset. In this model the value is 0.1163, which could be considered satisfactory. Multiple R-squared: High value (currently 0.88, close to 1) shows good fit of the function to the data.
- Adjusted R-squared: Very close value to the Multiple R-squared; a high value (currently 0.88, close to 1) shows good fit of the function to the data.
- The F-statistic is very large (907), which, in combination with the model *p*-value (very small, close to zero), brings to the conclusion that the overall regression model is highly evaluated. It means that both predictors contribute to the value of the dependent

variable (the null hypothesis of having one or both predictors insignificantly influence the value of the dependent variable is rejected).

- *p*-Value: Very small, near zero; small values are expected for good functions, and this is obtained in this case.

The created MLR prediction model that relates pollutant triplet values (in this example, it is CO = 0.0353465 + 0.0122536 PM_{10} + 0.0037720 NO_2) is used for creating predicted values. The R program (presented in Listing 3) was used with the predict function and the interval = 'confidence' parameter. The predicted values of CO were compared with the measured values obtained from measurement stations, and the mean absolute error (MAE) and root mean square error (RMSE) regarding the relation between measured and predicted (fit) CO values are presented (Table 11).

Table 11. Index-based comparison between measured and predicted values of CO [mg/m^3] (prediction interval = 'confidence').

No.	Measured Values Day	CO Measured [mg/m^3]	Prediction Input Day	CO Predicted (Fit) [mg/m^3]	CO Predicted (lwr) [mg/m^3]	CO Predicted (upr) [mg/m^3]
1	19 April 2024	0.5	21 March 2024	0.8165582	0.7884586	0.8446577
2	20 April 2024	0.49	22 March 2024	1.0028438	0.9656920	1.0399957
3	21 April 2024	0.31	23 March 2024	0.6585113	0.6156314	0.7013912
4	22 April 2024	0.32	24 March 2024	0.8816229	0.8543786	0.9088673
5	23 April 2024	0.33	25 March 2024	0.5439181	0.5235418	0.5642944
6	24 April 2024	0.4	26 March 2024	0.3999367	0.3797142	0.4201593
7	29 April 2024	0.34	27 March 2024	0.3497512	0.3329901	0.3665124
8	30 April 2024	0.32	28 March 2024	0.3058584	0.2849163	0.3268006
9	1 May 2024	0.27	29 March 2024	0.3672958	0.3415149	0.3930767
10	2 May 2024	0.3	30 March 2024	0.3736649	0.3450180	0.4023118
11	3 May 2024	0.34	31 March 2024	0.5738885	0.5555471	0.5922300
12	4 May 2024	0.38	1 April 2024	0.6747961	0.6582883	0.6913039
13	5 May 2024	0.44	2 April 2024	0.3746309	0.3558766	0.3933852
14	6 May 2024	0.46	3 April 2024	0.3957255	0.3659180	0.4255330
15	7 May 2024	0.43	4 April 2024	0.4209375	0.3931202	0.4487549
16	8 May 2024	0.38	5 April 2024	0.4811807	0.4443787	0.5179828
17	9 May 2024	0.34	6 April 2024	0.5234604	0.4855039	0.5614169
18	10 May 2024	0.34	7 April 2024	0.3573654	0.3334105	0.3813203
19	11 May 2024	0.38	8 April 2024	0.2698514	0.2500119	0.2896909
20	12 May 2024	0.37	9 April 2024	0.3058197	0.2818599	0.3297794
21	13 May 2024	0.37	10 April 2024	0.3519847	0.3309952	0.3729742
22	14 May 2024	0.36	11 April 2024	0.2862374	0.2661621	0.3063128
23	15 May 2024	0.27	12 April 2024	0.3438971	0.3188329	0.3689613
24	16 May 2024	0.26	13 April 2024	0.3711590	0.3417263	0.4005917
25	17 May 2024	0.29	14 April 2024	0.3878597	0.3452902	0.4304292
26	18 May 2024	0.4	15 April 2024	0.3920800	0.3510319	0.4331281
27	19 May 2024	0.35	16 April 2024	0.3609261	0.3380593	0.3837929
28	20 May 2024	0.34	17 April 2024	0.1933899	0.1700365	0.2167434
29	21 May 2024	0.33	18 April 2024	0.2732109	0.2496177	0.2968042
			MAE	0.134104697 mg/m^3		
			RMSE	0.197366426 mg/m^3		

Figure 16 presents the results of the predicted values of CO (presented with fit, lwr, and upr) compared (with an index-based approach) with the measured values from the in-sample set of 29 items (days) that belong to the annually collected data from monitoring station "Mostar" in the years 2023/2024.

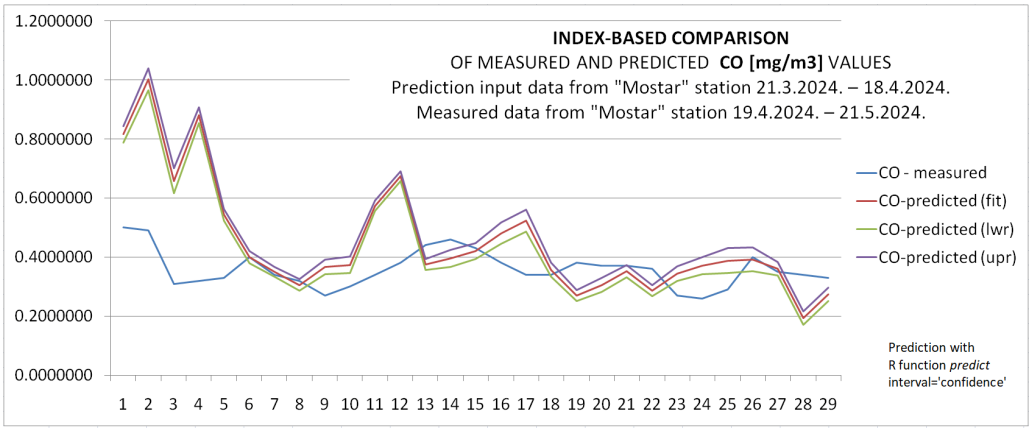


Figure 16. Index-based comparison of CO [mg/m^3] measured (19 April 2024–21 May 2024) and predicted values (based on 21 March 2024–18 April 2024 data); prediction with interval = "confidence" and MLR model created with cross-seasonal dataset obtained from "Mostar" station in the years 2023/24.

Table 11 and Figure 16 present the results related to CO pollutant concentration as measured values and predicted values. The predicted values are presented as interval-fit values (computed with the MLR function: CO = 0.0353465 + 0.0122536 PM_{10} + 0.0037720 NO_2) and upper and lower values, created with the *predict* command in the R program, with the prediction input dataset. Maximum predicted values are derived for the second day (~1.0), while the highest measured value was obtained on the first day (0.5) and second day (0.49).

The input data (values for PM_{10} and NO_2) for prediction were used from the prediction input dataset (21 March 2024–18 April 2024), and the comparison dataset consists of measured values of CO from the comparison dataset (19 April 2024–21 May 2024). The correspondence between prediction input data and comparison (measured) data was established with an "index-based" approach. For example, the first day's measured data of PM_{10} and NO_2 from the prediction input dataset were used to compute the predicted value of CO, which was compared with the first measured value of CO from the "comparison" dataset, etc. This "index-based" approach could not lead to appropriate conclusions regarding the MLR-based prediction model fit level, so additional and more precise in-sample fit analysis was performed with the "same date" approach, and the results are presented in Section 4.5.

4.5. In-Sample Fit Analysis of the MLR-Based Model Based on the "Same-Date" Approach

This section presents results of in-sample fit analysis for the MLR-based model CO~PM10+NO2 presented with statistical characteristics in Table 10. The "same-date" approach was applied in the prediction. The results of the CO prediction were computed with the same R program as in the previous "index-based" approach, but with the prediction input dataset used with the data that were the same as the comparison data; i.e., only one dataset was used from the period 19 April 2024–21 May 2024 for both purposes—as prediction input and comparison dataset.

The R program (presented in Listing 3) was used with one dataset (presented in Listing 4), the predict function and the interval = 'confidence' parameter. Table 12 and Figure 17 present the results of the predicted values and measured values of CO computed with the "same date" approach, whereby the prediction input and measured values for the comparison are acquired from the same dataset.

Table 12. "Same date"-based prediction and comparison of values of CO [mg/m^3] (prediction interval = 'confidence').

No.	Day	CO Measured [mg/m^3]	CO Predicted (Fit) [mg/m^3]	CO Predicted (lwr) [mg/m^3]	CO Predicted (upr) [mg/m^3]
1	19 April 2024	0.5	0.3306399	0.3069228	0.354357
2	20 April 2024	0.49	0.2945376	0.2697946	0.3192807
3	21 April 2024	0.31	0.2068756	0.1847618	0.2289895
4	22 April 2024	0.32	0.2274335	0.206178	0.248689
5	23 April 2024	0.33	0.2324108	0.2125768	0.2522448
6	24 April 2024	0.4	0.2061613	0.1855132	0.2268094
7	29 April 2024	0.34	0.1878582	0.1609138	0.2148026
8	30 April 2024	0.32	0.1988408	0.1771459	0.2205357
9	1 May 2024	0.27	0.1384354	0.1148366	0.1620341
10	2 May 2024	0.3	0.1361048	0.1130471	0.1591626
11	3 May 2024	0.34	0.178925	0.1561898	0.2016603
12	4 May 2024	0.38	0.1992614	0.1763819	0.2221409
13	5 May 2024	0.44	0.2186084	0.1952275	0.2419892
14	6 May 2024	0.46	0.2302872	0.2041399	0.2564345
15	7 May 2024	0.43	0.2554115	0.2245665	0.2862565
16	8 May 2024	0.38	0.1829763	0.1607356	0.205217
17	9 May 2024	0.34	0.1726872	0.1507162	0.1946583
18	10 May 2024	0.34	0.20429	0.1804036	0.2281764
19	11 May 2024	0.38	0.2443464	0.2167689	0.2719239
20	12 May 2024	0.37	0.2155022	0.1905685	0.240436
21	13 May 2024	0.37	0.3394409	0.3203858	0.3584959
22	14 May 2024	0.36	0.4312187	0.4092585	0.453179
23	15 May 2024	0.27	0.3052718	0.2876713	0.3228724
24	16 May 2024	0.26	0.3257578	0.307983	0.3435325
25	17 May 2024	0.29	0.4133072	0.3966286	0.4299858
26	18 May 2024	0.4	0.4450537	0.4278478	0.4622595
27	19 May 2024	0.35	0.3777592	0.3617792	0.3937392
28	20 May 2024	0.34	0.3275993	0.3103276	0.344871
29	21 May 2024	0.33	0.4536234	0.4373094	0.4699375
		MAE	0.128046831 mg/m^3		
		RMSE	0.141247829 mg/m^3		

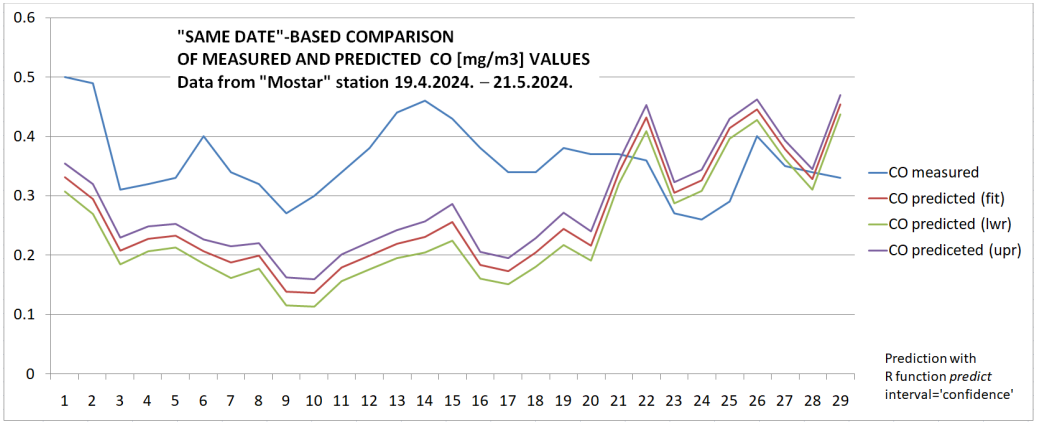


Figure 17. "Same date"-based comparison of CO [mg/m^3] predicted and comparison data (19 April 2024–21 May 2024); prediction with interval = "confidence" and MLR model created with cross-seasonal dataset obtained from "Mostar" station in the years 2023/24.

Listing 4. The R program for creating MLR prediction model and prediction processing.

```
datatablePredictionInput<- read.table
("MostarInSample_MeasuredForPredictionAndComparisonData.txt",
header = TRUE, fill = TRUE)
```

The R program (presented in Listing 3) was used with one dataset (presented in Listing 4), the predict function and the interval = 'prediction' parameter. Table 13 and Figure 18 present the results of the predicted values and measured values of CO, computed with the "same date" approach, whereby the prediction input and measured values for the comparison were acquired from the same dataset.

Table 13. "Same date"-based prediction and comparison of values of CO [mg/m^3] (prediction interval = 'prediction').

No.	Day	CO Measured [mg/m^3]	CO Predicted (Fit) [mg/m^3]	CO Predicted (lwr) [mg/m^3]	CO Predicted (upr) [mg/m^3]
1	19 April 2024	0.5	0.3306399	0.100420863	0.560859
2	20 April 2024	0.49	0.2945376	0.064210634	0.5248647
3	21 April 2024	0.31	0.2068756	−0.02318379	0.4369351
4	22 April 2024	0.32	0.2274335	−0.002544982	0.457412
5	23 April 2024	0.33	0.2324108	0.002559314	0.4622623
6	24 April 2024	0.4	0.2061613	−0.023761869	0.4360844
7	29 April 2024	0.34	0.1878582	−0.042715674	0.4184321
8	30 April 2024	0.32	0.1988408	−0.031178726	0.4288603
9	1 May 2024	0.27	0.1384354	−0.091771531	0.3686423
10	2 May 2024	0.3	0.1361048	−0.094047258	0.3662569
11	3 May 2024	0.34	0.178925	−0.051194953	0.409045
12	4 May 2024	0.38	0.1992614	−0.030872864	0.4293957
13	5 May 2024	0.44	0.2186084	−0.011576307	0.448793
14	6 May 2024	0.46	0.2302872	−0.000194869	0.4607693
15	7 May 2024	0.43	0.2554115	0.024349298	0.4864737
16	8 May 2024	0.38	0.1829763	−0.047095311	0.413048
17	9 May 2024	0.34	0.1726872	−0.057358503	0.402733
18	10 May 2024	0.34	0.20429	−0.025946586	0.4345265
19	11 May 2024	0.38	0.2443464	0.013697649	0.4749951
20	12 May 2024	0.37	0.2155022	−0.014845342	0.4458498
21	13 May 2024	0.37	0.3394409	0.109655296	0.5692264
22	14 May 2024	0.36	0.4312187	0.201174014	0.6612634
23	15 May 2024	0.27	0.3052718	0.075602311	0.5349414
24	16 May 2024	0.26	0.3257578	0.096074813	0.5554407
25	17 May 2024	0.29	0.4133072	0.183706481	0.6429079
26	18 May 2024	0.4	0.4450537	0.215414046	0.6746933
27	19 May 2024	0.35	0.3777592	0.148208163	0.6073102
28	20 May 2024	0.34	0.3275993	0.097954731	0.5572439
29	21 May 2024	0.33	0.4536234	0.224048919	0.683198
		MAE	0.128046831 mg/m^3		
		RMSE	0.141247829 mg/m^3		

For the purpose of the in-sample fit analysis, Table 14 shows the results of the measured CO values from the period of 19 April 2024–21 May 2024 compared with the predicted (fit) CO values (from Table 13) and persistence baseline prediction results. Persistence baseline predictions were created by using a simple principle—the next day's predicted value is the same as the measured value from the previous day. Finally, Table 14 contains mean

absolute error (MAE) and root mean square error (RMSE) for CO predicted (fit) values and persistence baseline prediction values.

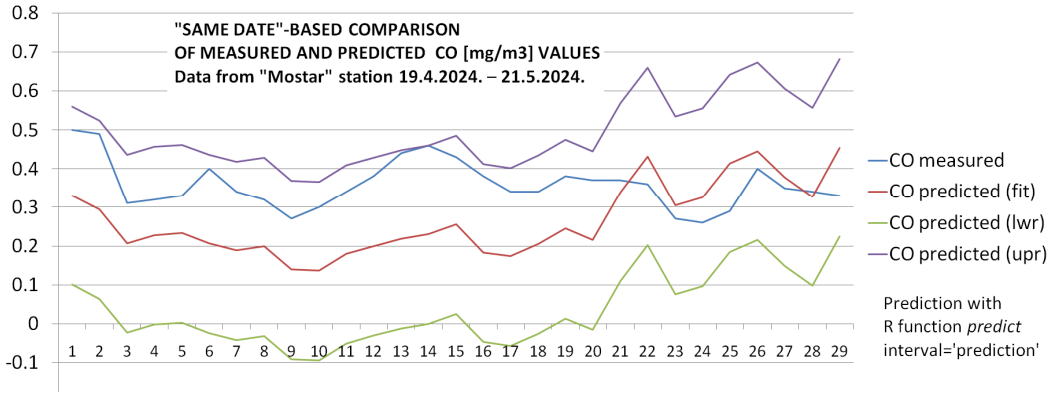


Figure 18. "Same date"-based comparison of CO [mg/m^3] predicted and comparison data (19 April 2024–21 May 2024); prediction with interval = "prediction" and MLR model created with cross-seasonal dataset obtained from "Mostar" station in the years 2023/24.

Table 14. Comparison of predicted values with MLR model and persistence baseline method.

No.	Day	CO Measured	CO Predicted (Fit) mg/m^3	Persistence Baseline Prediction mg/m^3
1	19 April 2024	0.5	0.3306399000	-
2	20 April 2024	0.49	0.2945376000	0.5
3	21 April 2024	0.31	0.2068756000	0.49
4	22 April 2024	0.32	0.2274335000	0.31
5	23 April 2024	0.33	0.2324108000	0.32
6	24 April 2024	0.4	0.2061613000	0.33
7	29 April 2024	0.34	0.1878582000	0.4
8	30 April 2024	0.32	0.1988408000	0.34
9	1 May 2024	0.27	0.1384354000	0.32
10	2 May 2024	0.3	0.1361048000	0.27
11	3 May 2024	0.34	0.1789250000	0.3
12	4 May 2024	0.38	0.1992614000	0.34
13	5 May 2024	0.44	0.2186084000	0.38
14	6 May 2024	0.46	0.2302872000	0.44
15	7 May 2024	0.43	0.2554115000	0.46
16	8 May 2024	0.38	0.1829763000	0.43
17	9 May 2024	0.34	0.1726872000	0.38
18	10 May 2024	0.34	0.2042900000	0.34
19	11 May 2024	0.38	0.2443464000	0.34
20	12 May 2024	0.37	0.2155022000	0.38
21	13 May 2024	0.37	0.3394409000	0.37
22	14 May 2024	0.36	0.4312187000	0.37
23	15 May 2024	0.27	0.3052718000	0.36
24	16 May 2024	0.26	0.3257578000	0.27
25	17 May 2024	0.29	0.4133072000	0.26
26	18 May 2024	0.4	0.4450537000	0.29
27	19 May 2024	0.35	0.3777592000	0.4
28	20 May 2024	0.34	0.3275993000	0.35
29	21 May 2024	0.33	0.4536234000	0.34
		MAE	0.128046831 mg/m^3	0.038928571 mg/m^3
		RMSE	0.141247829 mg/m^3	0.054412446 mg/m^3

Table 14 shows that an MAE of ~0.13 mg/m^3 and an RMSE of ~0.14 mg/m^3 were obtained for the predicted (fit) CO values, which were computed with the MLR-based

prediction model presented in Table 10. For the persistence baseline prediction, the MAE is ~0.04 mg/m^3 and the RMSE ~0.05 mg/m^3.

Finally, Table 15 presents a comparison of the MAE and RMSE values for all previously presented prediction types—"index-based" prediction (two different datasets for prediction input and comparison of predicted data with measured data), "same date -based" prediction (with one dataset for both purposes—prediction input and comparison) and persistence baseline prediction (having measured values of previous day as prediction of the next day).

Table 15. Comparison of MAE and RMSE for different types of predictions of CO [mg/m^3] based on MLR-based model CO~PM_{10} + NO_2.

Prediction Type	Table with Data	MAE [mg/m^3]	RMSE [mg/m^3]
"Index-based" prediction	Table 11	0.134104697	0.197366426
"Same date-based" prediction, interval 'confidence'	Table 12	0.128046831	0.141247829
"Same date-based" prediction, interval 'prediction'	Tables 13 and 14	0.128046831	0.141247829
Persistence baseline prediction	Table 14	0.038928571	0.054412446

Values of MAE and RMSE were calculated with formulas that include absolute values of differences between the measured and predicted (fit) values of CO. Table 5 shows MAE and RMSE value comparison for four types of predictions. The first three predictions are based on the MLR-based prediction model, while the fourth has a different prediction source.

The best results (the lowest values) were obtained for the persistence baseline prediction (MAE = 0.038928571 mg/m^3; RMSE = 0.054412446 mg/m^3). Among predictions that use the MLR-based prediction model, the "same date–based" prediction achieved better results, i.e., lower values for MAE = 0.128046831 and RMSE = 0.141247829, compared with the "index-based" predictions' MAE = 0.134104697 and RMSE = 0.197366426.

5. Discussion

5.1. Results' Relation to the Research Questions

RQ1 results are presented in Section 4.1, with a graphical representation in Figure 8. According to the 89 Serbian air-quality monitoring stations (status as of 23 April 2026), the most available sensors (at more than 50% of all monitoring stations) for air pollutants are related to PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO and O_3, so they were selected as Serbian-comparable. The Serbian-comparable list of air pollutants is also comparable with European and World lists. The European Air Quality Index [54] is based on presenting concentration values obtained from European air-quality monitoring stations, and they are related to five key pollutants: particulate matter (PM_{10}), fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), ozone (O_3) and sulphur dioxide (SO_2). The World Health Organization, in the global air-quality guidelines [61], provides recommendations regarding PM_{10}, $PM_{2.5}$, O_3, NO_2, SO_2, and CO. It can be concluded that the Serbian-comparable list includes all pollutants from the EAQI list and the WHO list.

RQ2 results are presented in Section 4.2, with tables for each season and diagrams for each comparable pollutant detectable at "Mostar" air-quality monitoring station in relation to meteorological parameters. The figures present comparisons between results of linear regression (single pollutant related to each single meteorological factor) and multiple linear regression (single pollutant related to all four meteorological factors). Two conclusions can be drawn: a) multiple linear regression R^2 values relating each pollutant to all four meteorological factors reach higher values compared with values computed in linear regression with single pollutants related to single atmospheric factors; b) comparable-

pollutant parameters reach the highest R-squared values in multiple linear regression relating all four meteorological factors in summer (PM_{10}, $PM_{2.5}$, and CO) and spring (SO_2 and NO_2).

RQ3 results are presented in Section 4.3. They are computed with cross-seasonal datasets, and they present MLR models and prediction results relating all comparable-pollutant triplets. The highest R-squared values (>0.7) is derived for 14 MLR models, and those among them that relate PM_{10} and $PM_{2.5}$ had the highest values for R^2—all above 0.95. Since PM_{10} and $PM_{2.5}$ usually have the same source and their appearance is simultaneous (in approx. 70% of cases) [60], MLR models relating both PM_{10} and $PM_{2.5}$ in the pollutant triplets were excluded from further analysis, and among other MLR models with $R^2 > 0.7$, the MLR model CO~PM_{10} + NO_2 is selected because of the highest value of the determination coefficient ($R^2 = 0.8776$).

Based on the results from RQ3, prediction with the R program was conducted with the MLR model CO~PM_{10} + NO_2, and the results are presented in Section 4.4. In this prediction, the R program utilized the *predict ()* function with interval = 'confidence'. Two different datasets with the same number of days were used for prediction and comparison with the "index-based" approach. In order to get a more precise evaluation of the MLR-based prediction function, in-sample fit analysis was conducted with the "same date-based" prediction, with one dataset being used as the prediction input and for the comparison of predicted and measured values. Values for MAE and RMSE were computed for four prediction types, and they are compared in Table 15. The following could be concluded:

- The interval parameter of the R function *predict* does not have any influence on the predicted fit value.
- Better results (the lowest value, i.e., the closest MAE and RMSE values to the values obtained for the persistence baseline prediction) were obtained with both "same date-based" prediction types, while the "index-based" prediction type gained higher values of MAE and RMSE.

5.2. Comparison with Previous Research

Our work has been positioned in the context of previous research, within the Related Work Section, which provides details about previous research results in relevant areas corresponding to our contribution aspects. In this section, we present a summary of the most important similarities and differences between previously published work and our contribution.

The most important similarities between our work and earlier research results are addressed in the following analysis: (a) Utilizing the multiple regression method with R programming has been presented in the work by Fisher [10], Seo and Lee [9], and Setiwan [40]. (b) Data clustering has been proposed to be related to seasons, similarly to [22], since Zhao et al. [15] concluded that seasonal prediction models were more accurate compared with annual data predictions. (c) Linear regression and multiple linear regression methods have been utilized in ecology research [17,23–27]. (d) All four seasons were included in this research study, similarly to [15,17].

Contribution differences between this research study and other research results are related to the following:

(a) Comparable list of pollutants. This research study created a list of comparable pollutants that are detected at multiple monitoring stations across Serbia. This way, a systematic approach was selected to gain comparable and comprehensive results. In previous research results, specific air pollution types or small groups of air pollutant types were addressed: CO_2 [23]; $PM_{2.5}$ [24]; NO_X and PM_{10} [25]; PM_{10}, $PM_{2.5}$ and

NO_2 [26]; O_3, CO, NO_2, NO_x, and benzene [27]; PM_{10}, NO, NO_2, and CO [17]; SO_2, NO_2, CO, and O_3 [15].

(b) Meteorological factors were separated from air pollutants in the multiple regression method used for creating prediction models. In this research study, (1) through impact analysis, all meteorological factors together (T, W, RH, and AP) were related to each comparable air pollutant with the multiple linear regression method applied upon seasonally clustered data; (2) cross-seasonal datasets were used for creating MLR models for prediction of pollutants compared with values of two other pollutants (no meteorological factors were mixed in the process of creating prediction models). In previous research, meteorological factors were used together with air pollutants in prediction models [18,22,41]. In [17], only some pollutants (PM_{10}, NO, NO_2, and CO) were related to meteorological factors by using seasonally clustered data. Multiple results from [17] show that each pollutant was related to each specific meteorological factor (wind, relative humidity, temperature, and rainfall) with seasonally clustered data by using linear regression, while only one summary table presents MLR results relating pollutants to meteorological factors generally, without referring to specific seasons.

(c) Selection of air-quality monitoring station location. This research study is focused on a heavy-traffic location and only vehicles' emissions as a single type of air pollution source. By focusing on one source of air pollution emissions, research results could be considered more precise. In previous research, spatial variations in traffic pollution have been addressed, with attention paid to the mixed ambient conditions of air-quality monitoring stations' locations, such as urban and industrial regions [17,29], urban non-motorized lanes and riverside [16].

5.3. Limitations

This research study has certain limitations, which will be elaborated on separately in detail.

RQ1 results' limitations:

- We report no consideration nor analysis of data regarding air-quality monitoring station locations, types or special purposes. These details remained out of scope for this research study.
- Air-quality monitoring stations change their status regarding sensor availability. The selected station, "Mostar", was previously able to detect and measure CO, but as of 23 April 2026, only PM_{10}, $PM_{2.5}$, SO2, and NO_2 values are measurable. This monitoring station has never supported O_3 measurements. However, this station is selected as the most appropriate, as it enables analysis from a single source of pollutant emissions.

RQ2 results' limitations:

- The conclusions that meteorological factor values are related to air pollutant values are based on Multiple R^2 values < 0.7. The highest value of 0.6906 was computed for $PM_{2.5}$~AP + T + W + RH with the summer dataset. It could be concluded that meteorological impact (in the seasonal context) on air pollutant values is supported by MLR-based prediction models with a moderate level of statistical fit, since all of the Multiple R^2 values are less than 0.7.

RQ3 limitations:

- Limited types of air pollutants were included in this research study. Not all comparable parameters were included, since "Mostar" station does not measure O_3.
- Number of independent variables: In this research study, pollutant triplets were related; i.e., one pollutant was related to two other pollutants. This approach considering pollutant triplets was selected because of the technical possibilities of R programming

for the visualization of their relation. Fortunately, the *ggplot2* library enables visualization of triplet relations with scatterplot diagrams. Another reason why only triplets were selected is the fact that the number of possible combinations of all comparable pollutants in their relations could be very large, considering that there could be one pollutant related to the values of three or more other pollutants.

- The results were obtained from cross-seasonal datasets, omitting the (moderate) impact of seasons in relation to meteorological factors on air pollutants.
- "Index-based" prediction was conducted "in-sample", where the annual dataset was divided in three separate (not overlapping) cross-seasonal segments. The size and seasonal referring of the prediction input is the same as that of the dataset for prediction comparison—both datasets have 29 items and are related to the same season, spring. However, large MAE and RMSE errors compared with the "same date–based" approach show the benefit of having one dataset used for the prediction and comparison of predicted values with measured values.
- The results are based on only a one-year period and related to only one air pollution monitoring station, located in a heavy-urban-traffic area.

General limitations:

- Fragmented annual dataset (52/366 days (14%) are missing, while 86% were properly collected), which leads to incomplete seasonally clustered datasets (spring has 4.44% of days missing, and the summer dataset has 53% of days missing), which has been reported in the description of the research sample characteristics and preprocessing. Therefore, the conclusions regarding meteorological parameters' relation to air pollutant values in spring and summer should be treated accordingly, with these incomplete-dataset facts in mind. However, the European Air Quality Index requires at least 75% of data available per year for a monitoring station to be evaluated as an eligible monitoring station in the EAQI network. According to the fact that this research study included 86% of valid data from "Mostar" monitoring station and that this monitoring station is registered as valid in the EAQI network, the results of this study could be considered approximately valid.
- The values obtained from the monitoring stations were daily average values, ignoring the fact that during each day, there are diverse values available. This aspect remains out of scope of this research study.
- The partial approach to the statistical analyses in this manuscript is related to the fact that the dominant role of the statistics used in this research study lies in the results having been obtained with R program computation (basic functions), which is addressed in the title, abstract and manuscript text. Research by Eren et al. [17] examined both linear and non-linear relations between air pollutants and meteorological factors and multiple statistical indicators. This research study, compared with [17], shows limited (only basic) statistical indicator results.

6. Conclusions

6.1. Summary of Findings and Contributions

This research study presents a systematic approach to the examination of the seasonal impact of meteorological factors on air pollution values in the heavy-urban-traffic area. The list of all air pollutants, comparable at Serbian, European and world levels, is used in their relation to four meteorological factors (atmospheric pressure, relative humidity, wind speed, and temperature) with seasonally clustered datasets. A comprehensive approach is utilized with all pollutants, meteorological factors and annually collected data, clustered by season, and both linear and multiple linear regression methods are used to compute prediction models for relating each pollutant to each meteorological factor, each pollutant

to all four meteorological factors and each pollutant to two other pollutants. Particular attention is given to the values of the R^2 indicator in the comparison and selection of prediction models. Compared with other similar research, we selected a location with a single pollution source (heavy traffic) and established one of the preconditions for more precise results.

Regardless of the fractional data having been obtained in a year-long period from the monitoring station (totally, 86% were collected properly in the years 2023/24), the results of this research study can be considered acceptable (the percentage of 86% of collected data is >75% of the EAQI standard data availability level). It can be concluded that season-related meteorological parameters establish a moderate ($R^2 < 0.7$) relation with air pollutant values. In the cross-seasonal analysis of pollutant triplet MLR relations, high ($R^2 > 0.7$) values were obtained in 14/29 cases. Particularly high values ($R^2 > 0.95$) were computed for triplets relating other pollutants to PM_{10} and $PM_{2.5}$ values. Because of the high PM_{10}~$PM_{2.5}$ correlation due to their similar sources in polluted areas [60], MLR models of pollutant triplets with both PM_{10} and $PM_{2.5}$ were separated from the list of the MLR models with the highest R^2 values.

This research study contributes to the literature with a methodology that shows comprehensiveness (all comparable pollutants, all seasons, all Serbian-measurable meteorological factors, and both linear and multiple linear regression prediction models) and preciseness in selecting a single source of pollution (heavy-urban-traffic location without industrial impact). The results of this research study contribute to research on the determination of the impact of meteorological factors on air pollution values, with seasonal considerations.

6.2. Implications

The proposed methodology is generally applicable to any region's data and type of pollutants. The developed R programs could be easily adapted to any dataset sources and their structure. This way, our research contributes to the literature with solutions that are tested, and the applicability is proved by the presented results.

Future research could be directed towards using and adapting the proposed methodology and the developed R program in more comprehensive studies with complete data or multiple years' datasets, wider lists of air pollutants and more monitoring stations, different types of locations, diverse monitoring station types (according to the purpose/location categorization) and more complete statistical indicator analysis.

The results of such studies could lead to organizational and technical solutions in traffic management and urban-life improvement, having considered meteorological factors in a seasonal context.

Author Contributions: Conceptualization, Z.K. and L.K.; methodology, Z.K. and L.K.; software, Z.K.; validation, Z.K., S.F. and L.K.; formal analysis, Z.K. and L.K.; investigation, Z.K. and L.K.; resources, Z.K. and L.K.; data curation, Z.K. and L.K.; writing—original draft preparation, Z.K., S.F. and L.K.; writing—review and editing, L.K., Z.K., and S.F.; visualization, Z.K.; supervision, Z.K., L.K. and S.F.; project administration, L.K.; funding acquisition, L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research study received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The R programming language GUI tool is free and open software that was used in this research study. It is available at https://www.r-project.org (accessed on 19 August 2026). The additional R module for diagrams GGPLOT2 is available at https://ggplot2.tidyverse.org/

(accessed on 19 August 2026). Data for this research study were obtained from https://sepa.gov.rs/ (accessed on 19 August 2026). The created R program is open-sourced, and the code is available at https://github.com/AirPolWRL/MulRegAirPolPredict (accessed on 19 August 2026).The datasets and computing results are available in the public repository: https://github.com/AirPolWRL/MulRegAirPolPredict (accessed on 19 August 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EAQI	European Air Quality Index
LR	Linear regression method
MAE	Mean absolute error
MLR	Multiple linear regression method
R	Programming language
R^2	R-squared statistical indicator
RGUI	R programming editor and console
RMSE	Root mean square error
SEPA	Serbian Environmental Protection Agency
WHO	World Health Organization

References

1. Miladinović, S.; Jaćimovski, S.; Nikač, Ž.; Kekić, D. The Influence of "Zelezara Smederevo" on the Quality of the Environment and its Ability to Improve Through the Monitoring System. *Tech. Gaz.* **2013**, *20*, 237–246.
2. Tusnio, N.; Fichna, J.; Nowakowski, P.; Tofilo, P. Air Pollution Associates with Cancer Incidences in Poland. *Appl. Sci.* **2020**, *10*, 7489. [CrossRef]
3. Brauer, M.; Hoek, G.; Smit, H.A.; De Jongste, J.C.; Gerritsen, J.; Postma, D.S.; Kerkhof, M.; Brunekreef, B. Air Pollution and Development of Asthma, Allergy and Infections in a Birth Cohort. *Eur. Respir. J.* **2007**, *5*, 879–888. [CrossRef] [PubMed]
4. Balogun, H.A.; Rantala, A.K.; Antikainen, H.; Siddika, N.; Amegah, A.K.; Ryti, N.R.I.; Kukkonen, J.; Sofiev, M.; Jaakkola, M.S.; Jaakkola, J.J.K. Effects of Air Pollution on the Risk of Low Birth Weight in a Cold Climate. *Appl. Sci.* **2020**, *10*, 6399. [CrossRef]
5. Parra, J.C.; Gómez, M.; Salas, H.D.; Botero, B.A.; Piñeros, J.G.; Tavera, J.; Velásquez, M.P. Linking Meteorological Variables and Particulate Matter PM2.5 in the Aburrá Valley, Colombia. *Sustainability* **2024**, *16*, 10250. [CrossRef]
6. Zeng, Y.; Yuanfei Cao, Y.; Qiao, X.; Seyler, B.C.; Tang, Y. Air pollution reduction in China: Recent success but great challenge for the future. *Sci. Total Environ.* **2019**, *663*, 329–337. [CrossRef] [PubMed]
7. Lu, H.; Song, J.; Di, T.; Kurdestany, J.M.; Wang, H. A Deep Believe Network Based Model for Urban Haze Prediction. *Tech. Gaz.* **2018**, *25*, 519–527. [CrossRef]
8. Pan, J.J.; Mahmoudi, M.R.; Baleanu, D.; Maleki, M. On Comparing and Classifying Several Independent Linear and Non-Linear Regression Models with Symmetric Errors. *Symmetry* **2019**, *11*, 820. [CrossRef]
9. Seo, J.Y.; Lee, H.M. A Study on Statistical Map of Air Pollution in Korea Using R. In Proceedings of the 4th International Conference on Computer Applications and Information Processing: Technology CAIPT, Kuta Bali, Indonesia, 8–10 August 2017. [CrossRef]
10. Fisher, R.; Wilson, S.K.; Sin, T.M.; Lee, A.C.; Langlois, T. A Simple Function for Full-subsets Multiple Regression in Ecology with R. *Ecol. Evol.* **2018**, *8*, 6104–6113. [CrossRef] [PubMed]
11. European Environment Agency. Monitoring Station Classifications and Criteria for Including Them in the EEA's Assessment Products. 2024. Available online: https://www.eea.europa.eu/en/topics/in-depth/air-pollution/monitoring-station-classifications-and-criteria (accessed on 22 April 2026).
12. Lishev, S.; Spasov, G.; Petrova, G. LoRaWAN IoT System for Measuring Air Parameters in a Traffic Monitoring Station. *Eng. Proc.* **2025**, *100*, 17. [CrossRef]
13. Stojanov, Z.; Brtka, V.; Jotanovic, G.; Jausevac, G.; Perakovic, D.; Dobrilovic, D. Graph Clustering-Based Model for Optimizing the Number of Air Pollution Measurement Stations in Urban Environments. *Mob. Netw. Appl.* **2025**, *30*, 1–11. [CrossRef]
14. Xie, P.; Zhang, C.; Wei, Y.; Zhu, R.; Chu, Y.; Chen, C.; Wu, Z.; Hu, J. Status of near-road air quality monitoring stations and data application. *Atmos. Environ. X* **2024**, *23*, 100292. [CrossRef]
15. Zhao, R.; Gu, X.; Xne, B.; Zhang, J.; Ren, W. Short Period PM2.5 Prediction Based on Multivariate Linear Regression Model. *PLoS ONE* **2018**, *13*, e0201011. [CrossRef] [PubMed]

16. Cao, R.; Luo, B.; Liu, K.; Wang, Z.; Cai, M.; Hu, X.; Xu, J.; Fan, Z. Characterizing and interpreting the spatial variation of traffic pollution in urban non-motorized lanes using mobile measurements. *Air Qual. Atmos. Health* **2023**, *16*, 1907–1929. [CrossRef]
17. Eren, B.; Serat, S.; Arifoglu, Y.D.; Ozdemir, S. Seasonal Analysis and Machine Learning-Based Prediction of Air Pollutants in Relation to Meteorological Parameters: A Case Study from Sakarya, Türkiye. *Appl. Sci.* **2025**, *15*, 4551. [CrossRef]
18. Choi, S.M.; Choi, H. Statistical Modeling for PM10, PM2.5 and PM1 at Gangneung Affected by Local Meteorological Variables and PM10 and PM2.5 at Beijing for Non and Dust Periods. *Appl. Sci.* **2021**, *11*, 11958. [CrossRef]
19. Žvirblis, T.; Vainorius, D.; Matijošius, J.; Kilikevičiene, K.; Rimkus, A.; Bereczky, Á.; Lukács, K.; Kilikevičius, A. Engine Vibration Data Increases Prognosis Accuracy on Emission Loads: A Novel Statistical Regressions Algorithm Approach for Vibration Analysis in Time Domain. *Symmetry* **2021**, *13*, 1234. [CrossRef]
20. Zheng, Y.; Xiuwen, Y.; Ming, L.; Ruiyan, L.; Zhangping, S.; Eric, C.; Tiannui, L. Forecasting Fine-Grained Air Quality Based on Big Data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2015, Sydney, Australia, 10–13 August 2015; pp. 2267–2276. [CrossRef]
21. Sharma, N.; Taneja, S.; Sagar, V.; Bhatt, A. Forecasting Air Pollution Load in Delhi Using Data Analysis Tools. *Procedia Comput. Sci.* **2018**, *132*, 1077–1085. [CrossRef]
22. Syafei, A.D.; Fujiwara, A.; Zhang, J. Prediction Model of Air Pollutant Levels Using Linear Model with Component Analysis. *Int. J. Environ. Sci. Dev.* **2015**, *6*, 519–525. [CrossRef]
23. Hussain, A.; Khan, F.; Albalawi, O. Modeling and Monitoring CO2 Emissions in G20 Countries: A Comparative Analysis of Multiple Statistical Models. *Sustainability* **2024**, *16*, 6114. [CrossRef]
24. Park, S.-Y.; Yoon, D.-K.; Park, S.-H.; Jeon, J.-I.; Lee, J.-M.; Yang, W.-H.; Cho, Y.-S.; Kwon, J.; Lee, C.-M. Proposal of a Methodology for Prediction of Indoor PM2.5 Concentration Using Sensor-Based Residential Environments Monitoring Data and Time-Divided Multiple Linear Regression Model. *Toxics* **2023**, *11*, 526. [CrossRef] [PubMed]
25. Vlachogianni, A.; Kassomenos, P.; Karppinen, A.; Karakitsios, S.; Kukkonen, J. Evaluation of a Multiple Regression Model for the Forecasting of the Concentrations of NOx and PM10 in Athens and Helsinki. *Sci. Total Environ.* **2011**, *409*, 1559–1571. [CrossRef] [PubMed]
26. Dun, M.; Xu, Z.; Chen, Y.; Wu, L. Short-Term Air Quality Prediction Based on Fractional Grey Linear Regression and Support Vector Machine. *Math. Probl. Eng.* **2020**, *8914501*, 13. [CrossRef]
27. Roy, S.S.; Paraschiv, N.; Popa, M.; Lile, R.; Naktode, I. Prediction of Air Pollutant Concentrations Using Hybrid Model of Regression and Genetic Algorithm. *J. Intell. Fuzzy Syst.* **2020**, *38*, 5909–5919. [CrossRef]
28. Ilijazi, V.; Jacimovski, S.; Milic, N.; Popovic, B. Software-supported Visualization of Mathematical Spatial-time Distribution Models of Air-pollutant Emissions. *J. Sci. Ind. Res.* **2021**, *80*, 915–923. [CrossRef]
29. Özdemir, U.; Taner, S. Impacts of Meteorological Factors on PM10: Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) Approaches. *Environ. Forensics* **2014**, *15*, 329–336. [CrossRef]
30. Basurko, E.; Berastegi, G.; Madariaga, I. Regression and Multilayer Perceptron-Based Models to Forecast Hourly O3 and NO2 Levels in the Bilbao Area. *Environ. Model. Softw.* **2006**, *21*, 430–460. [CrossRef]
31. Chan, B.K.; Chan, B.K. Data Analysis Using R Programming. In *Biostatistics for Human Genetic Epidemiology*; Springer: Cham, Switzerland, 2018; pp. 47–122. Available online: https://link.springer.com/chapter/10.1007/978-3-319-93791-5_2 (accessed on 19 August 2026).
32. Dutang, C.; Goulet, V.; Pigeon, M. Actuar—An R Package for Actuarial Science. *J. Stat. Softw.* **2008**, *25*, 1–37. [CrossRef]
33. Berea, A. R-Programming. In *Encyclopedia of Big Data*; Springer International Publishing: Cham, Switzerland, 2022; pp. 796–799. Available online: https://link.springer.com/referencework/10.1007/978-3-319-32010-6 (accessed on 19 August 2026).
34. R Core Team. *R: A Language and Environment for Statistical Computing*; Manual; R Foundation for Statistical Computing: Vienna, Austria, 2026. Available online: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf (accessed on 22 April 2026).
35. Kobal, M.; Ceglar, A.; Eler, K.; Medved-Cvikl, B.; Honzak, L.; Simonèiè, P.; Hladnik, D. On the Use of R Programming Language in the Analyses of Spatial Data. *Acta Silvae Ligni* **2013**, *102*, 55–62. [CrossRef]
36. Morandat, F.; Hill, B.; Osvald, L.; Vitek, J. Evaluating the Design of the R Language. In Proceedings of the European Conference on Object-Oriented Programming ECOOP, Beijing, China, 11–16 June 2012; pp. 104–131. [CrossRef]
37. Emura, T.; Matsumoto, K.; Uozumi, R.; Michimae, H.G. Ridge: An R Package for Generalized Ridge Regression for Sparse and High-Dimensional Linear Models. *Symmetry* **2024**, *16*, 223. [CrossRef]
38. Guenzi, D.; Fratianni, S.; Boraso, R.; Cremonini, R. CondMerg: An Open Source Implementation in R Language of Conditional Merging for Weather Radars and Rain Gauges Observations. *Earth Sci. Inform.* **2017**, *10*, 127–135. [CrossRef]
39. Frichot, E.; Francois, O. LEA: An R Package for Landscape and Ecological Association Studies. *Methods Ecol. Evol.* **2015**, *6*, 925–929. [CrossRef]
40. Setiawan, I. Time Series Air Quality Forecasting with R Language and R Studio. *J. Phys. Conf. Ser. Bristol* **2020**, *1450*, 1–8. [CrossRef]
41. Tularam, H.; Ramsay, L.F.; Muttoo, S.; Brunekreef, B.; Meliefste, K.; De Hoogh, K.; Naidoo, R.N. A Hybrid Air Pollution/Land Use Regression Model for Predicting Air Pollution Concentrations in Durban, South Africa. *Environ. Pollut.* **2021**, *274*, 116513. [CrossRef] [PubMed]

42. Hodgson, E.C.; Phillips, I.D. Seasonal variations in the synoptic climatology of air pollution in Birmingham, UK: Hodgson and Phillips. *Theor. Appl. Climatol.* **2021**, *146*, 1421–1439. [CrossRef]
43. Zeng, H.; Liu, Y.; Dai, H.; Zhao, X.; Tian, N. Mixed Forecast of Air Quality Index with a Bibranch Parallel Architecture Considering Seasonal Heterogeneity. *Entropy* **2026**, *28*, 419. [CrossRef] [PubMed]
44. RGUI Software Tool Download, The R Project for Statistical Computation. Available online: https://www.r-project.org/ (accessed on 7 August 2026).
45. GGPlot2 Library Download, Tidyverse—R Packages for Data Science. Available online: https://ggplot2.tidyverse.org/ (accessed on 18 June 2026).
46. Legal Regulations, Serbian Environmental Protection Agency, Ministry of Environmental Protection, Republic of Serbia. Available online: https://sepa.gov.rs/zakoni/ (accessed on 8 August 2026).
47. *SRPS ISO/IEC 17025:2017*; General Requirements for the Competence of Testing and Calibration Laboratories. Institute for Standardization of Serbia: Belgrade, Serbia, 2017. Available online: https://iss.rs/sr_Cyrl/project/show/iss:proj:60059 (accessed on 19 August 2026).
48. *SRPS EN 14212:2013*; Ambient Air—Standard Method for the Measurement of the Concentration of Sulphur Dioxide by Ultraviolet Fluorescence. Institute for Standardization of Serbia: Belgrade, Serbia, 2013. Available online: https://iss.rs/en/project/show/iss:proj:40989 (accessed on 19 August 2026).
49. *SRPS EN 12341:2023*; Ambient Air—Standard Gravimetric Measurement Method for the Determination of the PM_{10} or $PM_{2.5}$ Mass Concentration of Suspended Particulate Matter. Institute for Standardization of Serbia: Belgrade, Serbia, 2023. Available online: https://iss.rs/en/project/show/iss:proj:104555 (accessed on 19 August 2026).
50. *SRPS EN 14907:2008*; Ambient Air Quality—Standard Gravimetric Measurement Method for the Determination of the $PM_{2.5}$ Mass Fraction of Suspended Particulate Matter. Institute for Standardization of Serbia: Belgrade, Serbia, 2008. Available online: https://iss.rs/en/project/show/iss:proj:18668 (accessed on 19 August 2026).
51. *SRPS EN 14626:2013*; Ambient Air—Standard Method for the Measurement of the Concentration of Carbon Monoxide by Non-Dispersive Infrared Spectroscopy. Institute for Standardization of Serbia: Belgrade, Serbia, 2013. Available online: https://iss.rs/en/project/show/iss:proj:40993 (accessed on 19 August 2026).
52. *SRPS EN 14625:2025*; Ambient Air—Standard Method for the Measurement of the Concentration of Ozone by Ultraviolet Photometry. Institute for Standardization of Serbia: Belgrade, Serbia, 2025. Available online: https://iss.rs/en/project/show/iss:proj:105625 (accessed on 19 August 2026).
53. The National Network of Automatic Stations for Air Quality Monitoring—Map and Data, Serbian Environmental Protection Agency, Ministry of Environmental Protection, Republic of Serbia. Available online: https://vazduh.sepa.gov.rs/?lng=en (accessed on 8 August 2026).
54. European Air Quality Index—Map and Explanation (About). Available online: https://airindex.eea.europa.eu/AQI/index.html (accessed on 5 August 2026).
55. European Union. *Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on Ambient Air Quality and Cleaner Air for Europe (Recast), PE/88/2024/REV/1, OJ L, 2024/2881, 20.11.2024*; European Union: Brussels, Belgium, 2024. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202402881 (accessed on 8 August 2026).
56. Google Maps Satellite View of "Mostar" Air Quality Monitoring Station. Available online: https://www.google.com/maps/search/?api=1&query=44.798746,20.450184&zoom=15 (accessed on 19 August 2026).
57. Google Earth Presentation of "Mostar" Air Quality Monitoring Station. Available online: https://earth.google.com/web/@44.7990171,20.44656081,0a,1006.74926734d,35y,0h,0t,0r/data=CgRCAggCQgIIAEoICPbmsIYCEAA?authuser=1 (accessed on 5 August 2026).
58. Abdullah, S.; Napi, N.N.L.M.; Ahmed, A.N.; Mansor, W.N.W.; Abu Mansor, A.; Ismail, M.; Abdullah, A.M.; Ramly, Z.T.A. Development of Multiple Linear Regression forParticulate Matter (PM10) Forecasting during EpisodicTransboundary Haze Event in Malaysia. *Atmosphere* **2020**, *11*, 289. [CrossRef]
59. Xu, X.; Du, H.; Lian, Z. Discussion on regression analysis with small determination coefficient in human-environment researches. *Indoor Air* **2022**, *32*, e13117. [CrossRef] [PubMed]
60. Xue, H.; Guijian Liu, G.; Zhang, H.; Ruoyu Hu, R.; Xin Wang, X. Similarities and differences in PM10 and PM2.5 concentrations, chemical compositions and sources in Hefei City, China. *Chemosphere* **2019**, *220*, 760–765. [CrossRef] [PubMed]
61. World Health Organization—Global Air Quality Guidelines. Available online: https://iris.who.int/server/api/core/bitstreams/b729bbc4-7032-4799-898e-d112faa16f22/content (accessed on 11 August 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.