Scientific Papers. Series E. Land Reclamation, Earth Observation & Surveying, Environmental Engineering. Vol. XIV, 2025
Print ISSN 2285-6064, CD-ROM ISSN 2285-6072, Online ISSN 2393-5138, ISSN-L 2285-6064

MATHEMATICAL MODEL FOR PREDICTION OF NO_2 CONCENTRATION IN THE SOUTHEASTERN REGION OF ROMANIA, USING RECURSIVE LEAST SQUARES FILTER METHODS

Gabriel MURARIU[1], Sorin FRASINA[2] Adrian ROȘU[1], Corneliu DOROFTEI[3], Iulian RACOVIȚĂ[1], Bogdan ROȘU[1], Fetecău CĂTĂLIN[2], Mirela VOICULESCU[1], Cătălin NEGOIȚĂ[1]

[1]"Dunărea de Jos" University of Galați, Faculty of Sciences and Environment, 111 Domnească Street, 800201 Galați, Romania
[2]"Dunărea de Jos" University of Galați, Faculty of Engineering, 111 Domnească Street, 800201 Galați, Romania
[3]"Alexandru Ioan Cuza" University of Iași, Science Research Department, Institute of Interdisciplinary Research, Research Center in Environmental Sciences for the North-Eastern Romanian Region (CERNESIM), 11 Carol I Blvd, 700506, Iași, Romania

Corresponding author email: gabriel.murariu@ugal.ro

Abstract

Nitrogen dioxide is found in the atmosphere as a key ingredient in the photochemical formation of smog and acid rain, nitrogen dioxide is a poisonous gas that is formed during combustion. Toxic at high concentrations, it reacts with moisture in the air to form nitric acid, which is highly corrosive and dangerous to plants and animals. In this study, we present a predictive model for nitrogen dioxide concentrations measured between 2017 and 2024 at ground level in a national network of monitoring stations. The model is based on a statistical approach to measurements from 152 automatic measurement points, with an hourly resolution. The analysis carried out allowed the construction of a mathematical model in order to make an effective prediction. The algorithms used were of the Recursive least squares filter type. The application used was made possible by running a dedicated software in PyCharm. It was found that the model for daytime concentrations depends linearly on a series of parameters monitored by the national network.

Key words*: NO_2, algorithm, RLS, statistical analysis.*

INTRODUCTION

Nitrogen dioxide (NO_2) pollution has emerged as a major environmental and public health concern globally (Moreda-Piñeiro et al., 2021), particularly following efforts to reduce $PM_{2.5}$ concentrations to below regulatory thresholds. There is widespread scientific and regulatory interest in understanding the dynamics of ground-level NO_2, a key atmospheric pollutant, due to its well-documented harmful effects on human health. Research has also highlighted the detrimental impacts of NO_2 on vegetation, contributing to reduced plant growth and crop yields (Pietrogrande et al., 2021). According to the World Health Organization (WHO), strong evidence from both epidemiological and toxicological studies shows that elevated NO_2 concentrations are a primary contributor to adverse respiratory outcomes (Varga-Balogh et al., 2021). These effects range from decreased lung function and aggravated asthma symptoms to increased mortality, particularly among sensitive population groups such as children, the elderly, and individuals with pre-existing respiratory conditions.

Ground-level ozone is a secondary pollutant that results predominantly from the photochemical chain reactions involving nitrogen oxides (NO_X = NO + NO_2), carbon monoxide (CO) and volatile organic compounds (VOCs) using the catalysis of sunlight in the troposphere (Vîrghileanu et al., 2020).

Over the past few decades, greenhouse gas concentrations have increased around the world. With the rapid development of car traffic and the car fleet in particular, air pollution has become increasingly in South-Eastern Europe (Constantin et al., 2017). Since 2017, several legislative actions have been carried out, including the elimination of the environmental tax on vehicle registration. This

fact led to the increase in the level of air pollution in the South-East of Europe.
In this paper, the nitrogen dioxide monitoring data in the South-East of Europe are presented, taking into account a system of 7 automatic air quality monitoring stations within the national network. The data were used to analyse the characteristics of variation and the main causes of the concentration of nitrogen dioxide - NO_2 in the South-East area of Europe in combination with the relationship between different pollutants and meteorological factors.

MATERIALS AND METHODS

Since 2010, a national air quality monitoring network (https://www.calitateaer.ro/) has been established in South-Eastern Europe, which now includes 158 national automatic stations. Data for NO_2 were sourced from seven automatic air quality monitoring stations that are part of Romania's national monitoring network. These stations provided hourly measurements of NO_2 and associated pollutants (e.g., NO, NO_x, O_3, CO, PM), along with meteorological variables such as solar radiation, wind speed, temperature, and humidity, covering the years 2017 to 2024. This high-resolution dataset enabled the examination of NO_2 variability across multiple temporal scales.
To investigate temporal trends, the researchers employed classical statistical techniques, notably one-way Analysis of Variance (ANOVA) and the Kruskal-Wallis test. These methods were applied to assess seasonal, weekly, and diurnal fluctuations in NO_2 concentrations. The statistical tests consistently yielded p-values below 0.001, confirming significant variation in NO_2 levels across different time periods. Seasonal changes likely reflected atmospheric chemistry and heating or traffic patterns, while weekly and hourly patterns pointed to anthropogenic influences such as workweek traffic cycles and photochemical processes during daylight hours.
The data recorded between 2017 and 2025 were used together for data processing in this paper. The equipment of the automatic air quality monitoring stations collects automatically air samples and generates data reports every 30 minutes; then automatically uploads this data to the national database from local environmental protection departments. Table 1 shows the details of the parameters taken into the statistical analysis from each sampling stations, and Table 2 presents the location of the air quality stations used in this study based on their coordinates.

Table 1. Monitored parameters in the network of monitoring points

Parameters	Unit of measurement	Evaluation method	Notatia param.
O3	[µg/m³]	Hourly averaged	P1
CO.	[µg/m³]	Hourly averaged	P2
NO	[µg/m³]	Hourly averaged	P3
NO2	[µg/m³]	Hourly averaged	P4
NOx	[µg/m³]	Hourly averaged	P5
SO2	[µg/m³]	Hourly averaged	P6
Benzene	[µg/m³]	Hourly averaged	P7
Ethylbenzene	[µg/m³]	Hourly averaged	P8
m-Xilene	[µg/m³]	Hourly averaged	P9
o-Xilene	[µg/m³]	Hourly averaged	P10
p-Xilen	[µg/m³]	Hourly averaged	P11
Toluene	[µg/m³]	Hourly averaged	P12
SERIOUS. 10 - PM 10	[µg/m³]	Hourly averaged	P13
SERIOUS. 2.5 - PM 2.5	[µg/m³]	Hourly averaged	P14
LSPM10 - PM 10	[µg/m³]	Hourly averaged	P15
LSPM10 - PM 2.5	[µg/m³]	Hourly averaged	P16
Precipitation	[mm]	Hourly averaged	P17
Air pressure	[mbar]	Hourly averaged	P18
Solar radiation	[W/m²]	Hourly averaged	P19
Air temperature	[°C]	Hourly averaged	P20
Relative humidity	[%]	Hourly averaged	P21
Wind speed	[m/s]	Hourly averaged	P22
Wind direction	[grN]	Hourly averaged	P23

Table 2. Coordinates of monitoring stations

Area	Name	Wide	Long	Altitude
South-eastern part of Romania	GL5	45.82	27.44	31.00
	GL1	45.42	44.02	51.00
	GL4	45.41	44.05	38.00
	BR2	45.26	27.97	19.00
	BZ1	45.15	26.82	98.00
	FP2	45.18	28.77	35.00
	VN1	45.70	27.21	45.00

RESULTS AND DISCUSSIONS

In this section, we will present the results of statistical analysis and the results of numerical simulations using digital fitre (Voipan, Voipan, & Barbu, 2025).

Statistical analysis of the temporal variation of nitrogen dioxide concentration.

Seasonal variation analyses were performed for the concentration of NO_2 in the South-East area of Romania. Figure 1 shows, for example, box plot charts (Afshar-Mohajer et al., 2018) of NO_2 concentration in relation to the season in

question for the set of monitoring stations for 2018.

Obviously, seasonal variation was analyzed using ANOVA methods (Koziel, Pietrenko-Dabrowska, Wójcikowski, & Pankiewicz, 2025). In Table 3 are presented the results obtained for the seasonal variation presented above. It is noted that there is certainly a significant difference between the seasons considered between 2017 and 2024 (Table 3).

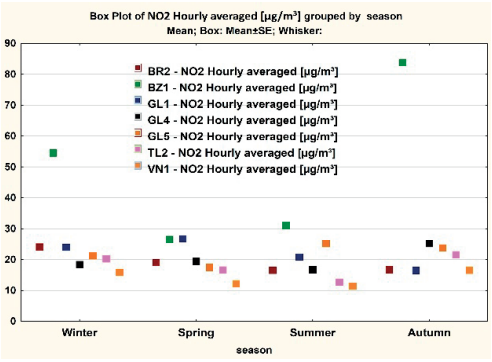

Figure 1. Seasonal variation of NO_2 during 2018

Thus, it is observed that in all cases submitted on NO_2 concentration, values of parameter p – confidence level, both for the analysis of using ANOVA methods one-way and for statistical analysis the Kruskal-Wallis test (KW) (Iticescu et al., 2019), are much smaller than 0.05 (Table 3).

Table 3. results obtained for the seasonal ANOVA method

Name	F value, p - value	H value, p- value
GL5	F(3.8751) = 223.6277, p = 0.0000	KW H(3.8755) = 861.4596, p = 0.0000
GL1	F(3.8751) = 1923.508, p = 0.0000	KW H(3.8755) = 2776.935, p = 0.0000
GL4	F(3.8751) = 186.0061, p = 0.0000	KW H(3.8755) = 995.1096, p = 0.0000
BR2	F(3.8751) = 72.9322, p = 0.0000	KW H(3.8755) = 179.6793, p = 0.0000
BZ1	F(3.8751) = 61.504, p = 0.0000	KW H(3.8755) = 173.5316, p = 0.0000
FP2	F(3.8751) = 133.792, p = 0.0000	KW H(3.8755) = 955.3999, p = 0.0000
VN1	F(3.8751) = 76.092, p = 0.0000	KW H(3.8755) = 424.889, p = 0.0000

Similarly, for all data sets, the weekly variation analysis for NO_2 concentration in the South-East area of Romania was performed.

Figure 2 shows, for example, the box plot graphs (NO_2 concentration in relation to the day of the week) for the set of monitoring stations for 2018 and Table 4 shows the values resulting from the ANOVA and Kruskal-Wallis test. Thus, it is observed that, in all cases, the values of the p parameter are much lower than 0.001, which means significant differences.

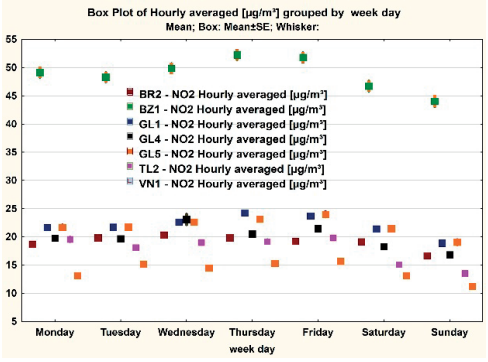

Figure 2. Weekdays variation of NO_2 during 2018

Table 4. Results obtained for weekly ANOVA (for a period of one week)

Name	F value, p – value	H value, p - value
GL5	F(6.8748) = 14.6618, p = 0.0000;	KW H(6.8755) = 118.6725, p = 0.0000
GL1	F(6.8748) = 7.906, p = 0.00000;	KW H(6.8755) = 133.5861, p = 0.0000
GL4	F(6.8748) = 16.3105, p = 0.0000;	KW H(6.8755) = 137.9648, p = 0.0000
BR2	F(6.8748) = 12.5073, p = 0.0000;	KW H(6.8755) = 162.6187, p = 0.0000
BZ1	F(6.8748) = 7.3831, p = 0.00000;	KW H(6.8755) = 239.9735, p = 0.0000
FP2	F(6.8748) = 27.9262, p = 0.0000;	KW H(6.8755) = 306.9294, p = 0.0000
VN1	F(6.8748) = 16.3649, p = 0.0000;	KW H(6.8755) = 138.5051, p = 0.0000

Lastly, we investigated the diurnal variability of NO_2 concentrations in the south-eastern region of Romania.

Figure 3 shows, for example, the box plot charts for the set of monitoring stations for 2018, and Table 5 shows the values for the ANOVA and Kruskal-Wallis test (NO_2 concentration in relation to the hour).

At this stage, the question arises whether a numerical model can be made that can integrate all these statistical data leading to the identification of some essential parameters - NO_x, NO, and solar radiation (Table 1).

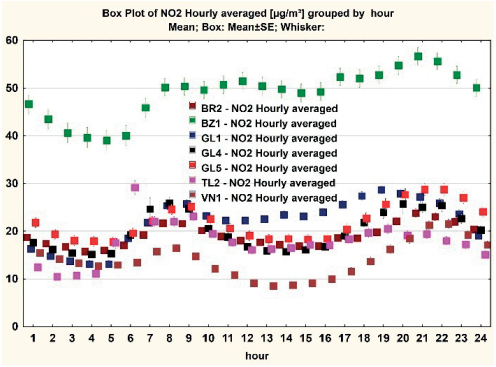

Figure 3. Diurnal variation of NO_2 during 2018

Table 5. Results obtained for daytime ANOVA

Name	F value, p - value	H value, p - value
GL5	F(23.8731) = 15.6522, p = 0.0000;	KW-H(23.8755) = 348.5298, p = 0.0000
GL1	F(23.8731) = 7.2156, p = 0.0000;	KW-H(23.8755) = 448.7269, p = 0.0000
GL4	F(23.8731) = 38.4603, p = 0.0000;	KW-H(23.8755) = 1896.5265, p = 0.0000
BR2	F(23.8731) = 13.8979, p = 0.0000;	KW-H(23.8755) = 680.6765, p = 0.0000
BZ1	F(23.8731) = 12.2314, p = 0.0000;	KW-H(23.8755) = 603.4816, p = 0.0000
FP2	F(23.8731) = 25.2004, p = 0.0000;	KW-H(23.8755) = 770.6485, p = 0.0000
VN1	F(23.8731) = 28.4301, p = 0.0000;	KW-H(23.8755) = 1139.2773, p = 0.0000

In this regard, a multidimensional model was developed to identify the set of physical variables influencing ground-level NO_2 concentration. Table 6 presents the results obtained for the BR2 station in 2017. The statistical model employed a factorial analysis approach, using the parameters listed in Table 1, with the sum of squares (SS) as a key metric.
It is important to note that, for assessing NO_2 conformity, the key parameters comprise NO concentration, NO_x concentration, and ground-level solar radiation values - as indicated in Table 6 (Iticescu et al., 2019).
At this stage, the reliability and robustness of the developed statistical model can be evaluated. Table 7 presents the results for the BR-2 station based on the 2017 dataset, with respect to the model's approximation accuracy. For all other stations considered, the models achieved an R^2 coefficient exceeding 0.90, indicating a high degree of explanatory power.

Table 6. Univariate Tests of Significance, Effect Sizes, and Powers for BR2 - NO_2 Hourly averaged [µg/m³]*

Effect	SS	F	p
Intercept	0.0	0	1.000000
O3 [µg/m³]	0.0	0	1.000000
CO [µg/m³]	0.0	0	1.000000
NO [µg/m³]	**129543.6**	**2854069**	**0.000000***
NOx [µg/m³]	**540546.5**	**11909170**	**0.000000***
SO2 [µg/m³]	0.0	0	1.000000
Benzene [µg/m³]	0.0	0	1.000000
Ethylbenzene [µg/m³]	0.0	0	1.000000
m-Xylene [µg/m³]	0.0	0	1.000000
o-Xylene [µg/m³]	0.0	0	1.000000
p-Xylene [µg/m³]	0.0	0	1.000000
Toluene [µg/m³]	0.0	0	1.000000
Precipitation [mm]	0.0	0	1.000000
Air pressure [mbar]	0.0	0	1.000000
Solar radiation [W/m²]	**0.3**	**6**	**0.012867***
Air temperature [°C]	0.0	0	1.000000
Relative humidity [%]	0.0	0	1.000000
Wind speed [m/s]	0.1	3	0.091114
Error	363.6		

*Sigma-restricted parameterization Effective hypothesis decomposition

Table 7. Test of SS Whole Model vs. SS Residual (BR2 Report 2017)

Dependent variable	Multiple R	MultipleR^2	Adjusted R^2	F	p
BR2 - NO_2 [µg/m³]	**0.999848**	**0.999695**	**0.999694**	**1545295**	**0.00**

Numerical simulation of the temporal variation of nitrogen dioxide concentration.
The specialized literature provides a variety of analysis and interpolation procedures, as well as algorithms, aimed at improving the accuracy of NO_2 estimation and prediction (Koziel et al., 2025). Several highly efficient algorithms implementing difference-based methods have been reported (Koziel et al., 2025). In the present study, we employed a filter with three input parameters and a single output parameter - NO_2 concentration (Figure 4) (Koziel et al., 2025). The algorithm follows the structure of a Recursive Least Squares (RLS) filter (Barbu, 2024). As a result of the iterative process, numerical simulations consistently converged in all cases, with the trace of the covariance matrix reaching its minimum value (Figure 5).

At the same time, the Nyquist diagram (Koziel et al., 2025) shows that the simulation process is stationary, being finally obtained closed trajectories in the right semiplane (Voipan et al., 2025). Finally, the response of the applied filter and the evolution of the estimation error exhibit a clear tendency toward convergence (Figure 6), indicating that the filter operates effectively.

Figure 4. Digital filter configuration

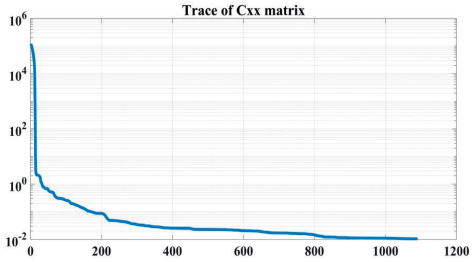

Figure 5. Digital filter covariance matrix trace and evolution towards convergence (matrix trace value versus simulation number)

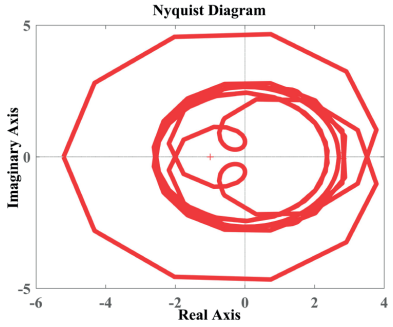

Figure 6. Nyquist filter response diagram

The final output of the filter is illustrated in Figure 7, where the amplification and refinement of the data following the filter's application are evident.

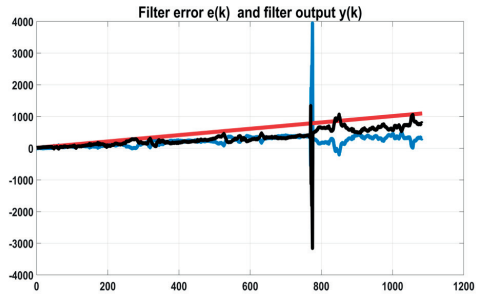

Figure 7. Digital filter response

CONCLUSIONS

Beyond exploratory statistics, the study advanced to multivariate modeling to identify key predictors influencing NO_2 concentrations. Using univariate tests of significance, our work determined that NO, NO_x, and solar radiation were the most impactful variables. These predictors were integrated into a factorial regression model, which showed exceptionally high accuracy, with R^2 values exceeding 0.99 for certain stations, such as BR2. This strong explanatory power highlighted the model's potential for accurate and localized air quality assessments.

The predictive component of the study was based on a Recursive Least Squares (RLS) filter, a numerical algorithm implemented in Python. This digital filter used the key predictors to estimate NO_2 concentrations in real-time, dynamically adjusting to new input data. The RLS model achieved convergence in all test cases, demonstrated by the minimization of the covariance matrix trace and stability confirmed through Nyquist diagram analysis. Model validation revealed high correlation with observed data and RMSE values under 3.5 $\mu g/m^3$, underscoring the model's effectiveness. These results indicate that the proposed approach not only enables reliable NO_2 prediction but also holds promise for cost-effective deployment in air quality monitoring and environmental management.

The corrective analysis and simulation process developed within the group exploited reference data and data provided by the national air quality monitoring network, collected in several locations, regarding the evaluation of the evolution and prediction of the NO_2 concentration. The rigorous verification indicates that the proposed correction technique achieves a very good accuracy of NO_2 monitoring, with a correlation coefficient exceeding 0.88, obtained for the reference data. Simultaneously, the RMSE error remains below 3.5 $\mu g/m^3$. Achieving such very high accuracy confirms the practicality and reliability of NO_2 detection using inexpensive detection devices. Further experiments involving alternative correction configurations emphasize the importance of the algorithmic tools developed in refining the correction scheme. Specifically, the inclusion of additional input variables, and the improvement of global data correlation together increase the accuracy of NO_2 detection.

AKNOWLEDGEMENTS

The work of Gabriel Murariu was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI - UEFISCDI, project number PN-IV-P8-8.1-PRE-HE-ORG-2024-0212, within PNCDI IV.

REFERENCES

Afshar-Mohajer, N., Zuidema, C., Sousan, S., Hallett, L., Tatum, M., Rule, A. M., & Koehler, K. (2018). Evaluation of low-cost electrochemical sensors for environmental monitoring of ozone, nitrogen dioxide, and carbon monoxide. *Journal of Occupational and Environmental Hygiene, 15*(2), 87–98. https://doi.org/10.1080/15459624.2017.1388918

Barbu, M. (2024). *Modeling, simulation, control and optimization in engineering with applications*. https://www.linkedin.com/posts/mathematics-mdpi_mathematics-activity-7004350949817274368-0UAW

Constantin, D.-E., Merlaud, A., Voiculescu, M., Van Roozendael, M., Arseni, M., Roşu, A., & Georgescu, L. (2017). NO_2 and SO_2 observations in Southeast Europe using mobile DOAS observations. *[Conference paper / report - details to be added if available]*.

Iticescu, C., Georgescu, L. P., Murariu, G., Topa, C., Timofti, M., Pintilie, V., & Arseni, M. (2019). Lower Danube water quality quantified through WQI and multivariate analysis. *Water, 11*(6), 1305. https://doi.org/10.3390/w11061305

Koziel, S., Pietrenko-Dąbrowska, A., Wójcikowski, M., & Pankiewicz, B. (2025). Nitrogen dioxide monitoring by means of a low-cost autonomous platform and sensor calibration via machine learning with global data correlation enhancement. *Sensors, 25*(8), 2352. https://doi.org/10.3390/s25082352

Moreda-Piñeiro, J., Sánchez-Piñero, J., Fernández-Amado, M., Costa-Tomé, P., Gallego-Fernández, N., Piñeiro-Iglesias, M., & Muniategui-Lorenzo, S. (2021). Evolution of gaseous and particulate pollutants in the air: What changed after five lockdown weeks at a Southwest Atlantic European region (Northwest of Spain) due to the SARS-CoV-2 pandemic? *Atmosphere, 12*(5), 562. https://doi.org/10.3390/atmos12050562

Pietrogrande, M. C., Casari, L., Demaria, G., & Russo, M. (2021). Indoor air quality in domestic environments during periods close to Italian COVID-19 lockdown. *International Journal of Environmental Research and Public Health, 18*(8), 4060. https://doi.org/10.3390/ijerph18084060

Varga-Balogh, A., Leelőssy, Á., & Mészáros, R. (2021). Effects of COVID-induced mobility restrictions and weather conditions on air quality in Hungary. *Atmosphere, 12*(5), 561. https://doi.org/10.3390/atmos12050561

Vîrghileanu, M., Săvulescu, I., Mihai, B.-A., Nistor, C., & Dobre, R. (2020). Nitrogen dioxide (NO_2) pollution monitoring with Sentinel-5P satellite imagery over Europe during the coronavirus pandemic outbreak. *Remote Sensing, 12*(21), 3575. https://doi.org/10.3390/rs12213575

Voipan, D., Voipan, A. E., & Barbu, M. (2025). Evaluating machine learning-based soft sensors for effluent quality prediction in wastewater treatment under variable weather conditions. *Sensors, 25*(6), 1692. https://doi.org/10.3390/s25061692