ASN, Vol. 12, No 3, Pages 1–16, 2025

Acta Scientifica Naturalis

Former Annual of Konstantin Preslavsky University of Shumen: Chemistry, Physics, Biology, Geography

Journal homepage: https://content.sciendo.com/view/journals/asn/asn-overview.xml

Application of artificial intelligence and GIS for integrated modeling of processes and factors related to the greening of urban environments

Radoslav Miltchev[1], Galin Milchev[2], Miroslava Nikolova[3]

[1] *Technical University, Faculty of Industrial Technology, Department of Theory of Mechanisms and Machines, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria*

[2] *University of Forestry, Faculty of Forest Industry, Department of Mathematics, Physics and Informatics, 10 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria*

[3] *Technical University, Faculty of Industrial Technology, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria*

Abstract: *As a tool for processing and analyzing multiple types of data in various fields, GIS is of significant importance and contributes greatly to the digitization, automation, and management of diverse processes. The functionalities of these systems can be enhanced with the capabilities of artificial intelligence (AI) for improving analysis algorithms. Their combined use, based on machine learning aimed at discovering features and generating solutions unnoticed by human factors, can significantly improve the quality of research and the results obtained. This is especially true in cases where large volumes of data in the form of time series need to be handled. This article analyzes the potential for the productive use of GIS and AI in the field of air pollution research and forecasting, based on available data on the distribution of different types of pollutants and climatic data for urbanized areas. The generated digital models through the functionalities of GIS are supplemented by using AI for time series analysis. The results of the study will be used in future developments aimed at using AI and GIS to study pollen pollution in urban conditions and reduce health risks.*

Keywords: AI, Urban environment, Ecologization, GIS

Introduction

Although artificial intelligence is not a new technology, it has undergone explosive development in recent years, both in terms of its scope and the attitudes toward its use and areas of application. In more and more fields,

Corresponding author: gmilchev@ltu.bg

Full Paper

DOI: 10.2478/asn-2025-0014

it is taking a major or a minor part, significantly supporting all other technologies. GIS, on the other hand, due to their long-term use in numerous ways in the fields of ecology and human health preservation, are of key importance in automating the processes of analyzing and modeling a number of related factors.

In recent years, the focus on their use has shifted from their application mainly as a means of acquiring, analyzing, and presenting spatial data based on cartographic principles, to the additional possibilities related to supporting well-founded decision-making in the sustainable management of ecological processes, including air pollution. As a tool for processing and analyzing multiple types of data in various fields, GIS is of significant importance and contributes greatly to the digitization, automation, and management of diverse processes.

Artificial intelligence enhances the functionalities of GIS by improving analysis algorithms. Machine learning focused on established computer systems, learning from data with included techniques, allows GIS to improve their performance over time. On the other hand, time series analysis is a useful tool in accounting for changes of certain characteristics over time, related to ecology, as well as comparing changes in other variables for the same period. The combined use of all these approaches and technologies finds its application in numerous studies.

Trends in the combined use of AI and GIS

As one of the main and valuable water resources, groundwater and its characteristics can be accurately mapped and classified based on hybrid models using artificial intelligence and GIS, combining modified RealAdaBoost, bagging, and rotation forest ensembles with functional tree [1]. The combined use of GIS and AI can be applied as neural networks together with machine learning to create predictions and interpretable scenarios in building explanatory models of demographic and social profiles of specific territorial areas, used for analysis and testing of pre-formulated hypotheses [2]. An integrated approach using remote sensing and GIS in the preservation of ecological zones associated with land use conflicts and environmental pressures [3] can be applied in building effective models for the optimal protection of ecological species.

The combined application of modern GIS and artificial intelligence is essential for facilitating sustainable agriculture worldwide by integrating them into conventional agricultural practices [4]. GIS and AI can be used in modeling ecosystems and landscape processes occurring on larger scales by extrapolating data or model predictions from fine scales to the larger landscape, requiring assumptions of spatial homogeneity or knowledge of resource heterogeneity [5]. With the introduction of autonomous GIS as an AI-based Geographic Information System (GIS), self-generation, self-organization, self-verification, self-execution, and self-expansion of the system are achieved, which improves spatial analysis and its accessibility to a wider audience [6]. The use of GIS in combination with time series can be employed for natural resource management and modeling the spatial behavior

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

of various natural sources through the collection of satellite images, their radiometric and atmospheric correction, and subsequent analysis [7].

Use of GIS and AI for Air Quality Monitoring

In recent years, the development of internet technologies and the creation of high-quality and relatively inexpensive sensors have deepened research in various fields, especially in ecology and environmental protection. Sensors can generate large volumes of data on an hourly or daily basis, which can serve as the basis for subsequent analyses and final decision-making. Combining air quality data with environmental factors can support the development of smart cities and more efficient urban management.

Air quality monitoring is important for forecasting its quality, a problem that is ubiquitous in large cities. In this regard, AI can be reliably used to create models based on hourly concentrations of NO_2 and CO from multisensor data and local meteorological data. For them to be reliable, properly trained models need to be developed that combine sensor outputs with weather data, considering that many factors can affect the accuracy of the models. Research in this area shows that combining multisensor data with advanced AI techniques offers a powerful tool, especially for modeling nonlinear processes like determining air quality [8].

The increasing construction in cities leads to serious consequences related to the circulation of air flows in cities, thereby deteriorating air quality. Many models related to pollution forecasting and mitigation consider air pollutants like particulate matter (PM) with the help of machine learning algorithms (MLA). There are also studies evaluating spatial hazard in terms of the Air Quality Index (AQI), which account for the crucial importance of PM in the AQI due to its easily inhalable microsize, which adversely affects ecology, the environment, and human health.

By developing machine learning models like eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbor (KNN), and other spatially processed data like slope and elevation, and subsequently integrating them into a GIS environment, it is possible to generate spatial data on air quality [9].

GIS and related remote sensing methods are technologies that enable real-time monitoring and mapping of changing phenomena, including air pollution. The evaluation of pollutants, obtained by combining data from satellites (Terra, Aqua, OMI, and Sentinel-5P) and data collected from air quality monitoring stations located in various parts of the areas of interest, confirms the possibility of using Sentinel-5P data for such purposes. The presence of strong and acceptable correlations between satellite data and on-site measurements indicates the ability of satellite imagery to monitor air pollution, especially over urban areas during fires [10]. Other studies complement the analyses related to remote sensing methods, sources of satellite data, and geospatial data analysis, considering aerosol optical depth (AOD) for air pollution monitoring. They highlight the important role of machine learning in correcting its values and filling gaps for missing aerosol values compared to ground sensors [11].

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

Developed as a tool for decision-makers, including city managers, air monitoring models in urban environments over the past decades have been divided into three main classes: statistical methods, machine learning methods, and hybrid methods. Combining these with GIS can improve forecasting methods. An analysis from a similar study [12] shows that forecasting with the help of machine learning and hybrid methods provides better results and also demonstrates that it is better to use emission inventories when choosing data, as they contain structured information from various sources of air pollution. Geographic Information Systems (GIS) offer potentially powerful tools for exposure assessment in support of both air pollution epidemiology and air quality policy. As noted in [13], most epidemiological applications rely on relatively simple techniques such as buffering and distance functions, while interpolation methods and dynamic modeling techniques, including time-activity pattern modeling with appropriate levels of spatial and temporal accuracy, can also be applied in such cases.

In some European projects funded by the EU related to urban air quality, GIS also finds application in the form of developed methodologies for mapping traffic-related pollution. In [14], a GIS containing data on observed air pollution levels, road network, traffic volume, land cover, elevation, and other locally determined characteristics is presented. Data from conducted monitoring based on predictive environmental variables and the composed regression equation are used to map air pollution in the study areas. As a tool related to the analysis and assessment of climate change and air quality, GIS combined with time series is used in the statistical and spatial assessment of investigated sequences of temperature and precipitation data from meteorological stations distributed over a certain territory. In [15], the analysis of trends, variability, and spatial patterns of temperature and precipitation is conducted based on linear regression, coefficient of variation, inverse distance weighting interpolation techniques, and geographic information systems. An autoregressive integrated moving average (ARIMA) model is used to simulate temperature and precipitation data.

Air pollution in urban environments can also be analyzed through long-term variations in the mean value (trend) of cyclic or periodic components using GIS and time series. In [16], statistically significant periodic components with higher frequency were observed. The developed model is based on a step-by-step approach to time series analysis applied to the mean daily concentrations of strong acidity (SA) and black smoke (BS) in the study area, with each step completed by correlation analysis of the residuals, allowing the identification of an optimal structure with residual white noise.

Studies related to air quality find broad application in analyses related to population morbidity. Time series analysis is a suitable solution here, which can be applied to a retrospective study aimed at determining whether the air quality index (AQI), its main pollutants (PM2.5, PM10, NO_2, and CO), and corresponding meteorological variables (daily temperature, highest temperature, lowest temperature, temperature difference, and sunshine duration) can increase morbidity [17]. The model is based on an initial descriptive analysis providing an overview

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

of morbidity and air quality over the study period, followed by a linear regression model related to each independent variable.

GIS and time series data analysis from sensor observation infrastructure have proven effective for controlling and managing air quality in megacities. Their combined use based on the SWE framework of OGC for integrating both sensors and their observations allows them to be used in spatial data infrastructure [18]. Developed in this way, it enables the collection, transfer, sharing, and processing of sensor observations, calculation of air quality status, and real-time reporting of critical conditions. By implementing it in the form of a four-layer architectural structure, including sensor, web service, logical, and presentation layers, and using defined routine procedures and subsystems, the system applies data from web-accessible sensors to a set of web services to generate warning information.

Another direction in the use of GIS in the field of ecology and urban air pollution is placing them at the core of methods for optimizing the air quality monitoring network (AQMN) through interpolation techniques and historical data. In a similar approach [19], existing air quality stations are systematically removed, and missing data are filled in using the most appropriate interpolation technique, followed by comparison of observed and interpolated data. Accuracy verification is conducted using predefined performance measurements such as root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation coefficient (r). Based on a developed algorithm, the process is simulated for several sets of variables, and overall, the technology allows determining the optimal number of stations without compromising concentration coverage in the study area.

Air quality and its monitoring through traditional methods are expensive and have limited spatiotemporal capability. The use of AI in combination with geospatial analyses, based on a specific controlled AI algorithm, represents a method that provides a good solution when supplemented with satellite imagery and GIS data to improve the accuracy and efficiency of air quality monitoring [20]. Based on the analysis of data collected from air quality sensors (public and private) through machine learning algorithms and geospatial data such as traffic density and land use, the results demonstrate the effectiveness of AI and geospatial analysis in predicting air quality parameters such as particulate matter and nitrogen oxides with high accuracy.

Material and Methods

The analysis shows that specialized neural networks such as RNNs and their variants (e.g., LSTM and GRU) are designed for processing sequential data and are highly effective for forecasting time series. They can capture long-term trends in the data, which makes them useful for predicting future values. At the same time, AI can detect anomalies or outliers in time series, which is important for detecting unusual events. Machine learning algorithms

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

can identify data points that significantly deviate from expected behavior. AI can analyze and separate trends and seasonal components in time series data.

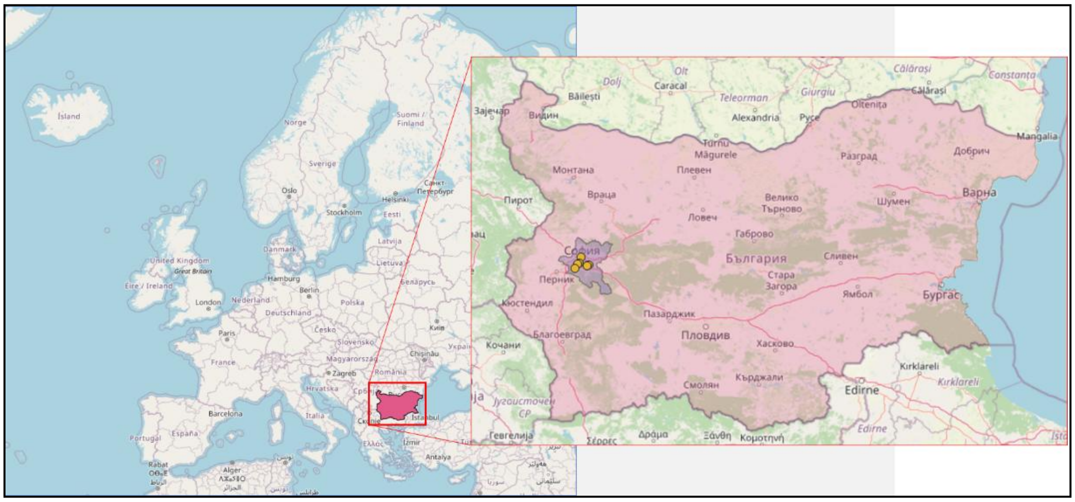

Figure 1. Studied geographic area – Republic of Bulgaria, Sofia District

The research hypothesis of the current study is that based on the combined use of time series analysis with AI and data interpolation with GIS, users can obtain a specialized tool that not only forecasts the development of specific phenomena in the context of their geographic characteristics but also enables the determination of critical values at a given point on the Earth's surface, regardless of the small number of measurement sensors. Such a tool can be crucial for efficiently managing various processes - for example, air pollution or assessing various measures to improve it. Figure 1 shows the study area, and Figure 2 shows the location of the Automatic Monitoring Stations (AMS) in the study area.

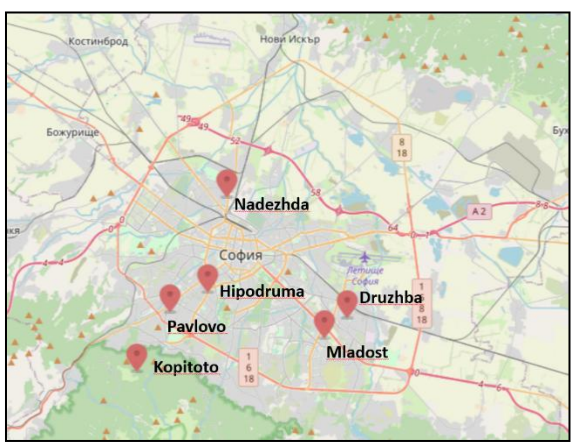

Figure 2. Location of automatic measuring stations, Sofia District

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

Primary Air Pollutants in Urban Environments

Air pollution is a deeply concerning issue that affects nearly every corner of the world, primarily due to human activities such as extensive industrialization, automotive activities, increasing fuel consumption, and all other urbanization, construction, electricity production, and essentially all other anthropogenic activities. Air pollution contributes to various health hazards and increases the risk of chronic and respiratory diseases. It is important to measure a range of pollutants to determine the dynamics of air pollution in a given area. Table 1 shows the primary pollutants and environmental factors used in the current study.

Table 1. Primary pollutants and environmental factors related to air pollution

Air pollutants	Environmental factors
• Fine particulate matter < 10um [PM10] • Ozone [O3] • Nitric oxide [NO] • Nitrogen dioxide [NO2] • Sulphur dioxide [SO2] • Benzene [Benzene]	• Temperature [AirTemp] • Wind direction [WD] • Wind speed [WS] • Relative humidity [UMR] • Atmospheric pressure [Press] • Solar radiation [GSR]

According to California Air Resources Board (https://ww2.arb.ca.gov/resources):

• Particulate matter (PM) - airborne particulate matter (PM) is a complex mixture of various chemical species, including small droplets of liquid, dry solid fragments, and solid cores with liquid coatings. For air quality regulatory purposes, particles are classified by their diameter. Specifically, particles with a diameter of 10 microns or less (PM10) are considered inhalable into the lungs and can cause adverse health effects.

• Sulphur dioxide (SO_2) - Sulfur oxides (SOx), including sulfur dioxide (SO_2), are emitted when sulfur-containing fuels are burned. Sources of such emissions in general, include motor vehicles, locomotives, ships, and off-road diesel equipment using high-sulfur fuels. Additionally, SO_2 and other SOx could be released from industrial processes such as natural gas and petroleum extraction, oil refining, and even metal processing.

• Nitrogen dioxide (NO_2) - although nitrogen dioxide (NO_2) can be directly emitted from combustion sources, much of it in ambient air is formed through photochemical reactions between nitric oxide (NO) and other air pollutants in the presence of sunlight. NO_2 levels in the air vary based on direct emission levels and changing atmospheric conditions, particularly the amount of sunlight.

• Ozone (O_3) - Ozone is formed in the atmosphere through chemical reactions between pollutants emitted from vehicles, factories, industrial sources, fossil fuel combustion, consumer products, and etc. Hydrocarbons and

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

ASN, Vol. 12, No 3, Pages 1–16, 2025

nitrogen oxide gases react in the presence of sunlight to create ozone, with hot, sunny, and calm weather conditions promoting its formation.

Creating a dataset

Problems with the Source Data - during the course of the study, it was found that the data publicly available was not in a suitable format either for artificial intelligence analysis methods or for visualization in GIS. Figure 3 displays O_3 data with added georeferenced data before their processing.

	Date	Kopitoto	Pavlovo	Hipodruma	Nadezhda	Mladost	Druzhba	Temp	Hum
0	2021-01-01	78.72	19.37	17.61	24.20	NaN	22.80	5.1	48.3
1	2021-01-02	80.02	11.70	15.30	17.50	NaN	23.07	3.0	48.6
2	2021-01-03	73.24	9.76	23.51	16.04	NaN	20.29	5.7	77.4
3	2021-01-04	57.12	10.88	13.61	8.36	NaN	12.95	6.7	88.4
4	2021-01-05	73.19	38.82	35.32	32.06	NaN	27.88	6.7	NaN
5	2021-01-06	79.70	11.74	12.16	15.05	NaN	20.57	5.5	55.9
6	2021-01-07	76.44	33.80	30.05	29.97	NaN	20.71	4.8	81.2
7	2021-01-08	67.39	14.45	17.35	26.09	NaN	18.15	4.4	92.1
8	2021-01-09	57.41	28.68	30.71	31.57	NaN	27.75	0.3	94.3
9	2021-01-10	43.81	16.02	24.16	27.48	NaN	29.00	1.0	93.4
•••									
1236	2024-05-21	82.42	57.30	42.44	57.07	NaN	63.85	21.5	69.8
1237	2024-05-22	100.88	50.58	30.15	44.45	NaN	46.55	19.8	81.4
1238	2024-05-23	116.22	71.91	40.93	56.90	NaN	56.07	20.0	74.5
1239	2024-05-24	113.57	84.00	50.51	65.63	NaN	67.29	21.1	67.9
1240	2024-05-25	120.76	73.98	42.79	53.22	NaN	62.27	21.9	61.7

```
RangeIndex: 1241 entries, 0 to 1240
Data columns (total 9 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Date       1241 non-null   object
 1   Kopitoto   1241 non-null   float64
 2   Pavlovo    1241 non-null   float64
 3   Hipodruma  1241 non-null   float64
 4   Nadezhda   1241 non-null   float64
 5   Mladost    0 non-null      float64
 6   Druzhba    1241 non-null   float64
 7   Temp       1220 non-null   float64
 8   Hum        1095 non-null   float64
```

Figure 3. Data before processing for O_3

Preliminary data analysis showed significant gaps in measurements—missing data for certain periods and pollutants; stations do not have an equal set of sensors—some stations lack sensors for temperature and humidity or for ozone; data is not in a format that allows for automatic importation—there are differences in types even for the same parameter; data written in Cyrillic requires transformation. This necessitates data preprocessing and subsequent transformation to achieve the objectives set forth in this study.

Formation of an Adequate Dataset

To form a suitable dataset, scripts were developed that automatically extract data from online repositories, verify data types, and convert them into the correct format for GIS and AI analysis. Additionally, data interpolation at the daily average level was used, excluding periods with missing sensor data spanning three to six months to avoid significantly distorting forecasts. The study data covers the period from January 1, 2021, to May 25, 2024, on a daily basis.

Methods Used for Time Series Analysis

For the purposes of this study, two methods for time series analysis and forecasting were tested based on their characteristics:

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

• Method using the "Prophet" model – time series forecasting tool developed by Facebook (now Meta, https://facebook.github.io/prophet/) designed for easy and accurate forecasting at scale. It is particularly effective for time series that have strong seasonal effects. The following advantages can be noted when using it: quick setup - the model can be quickly set up and applied to time series data, making it accessible for fast prototyping and implementation; missing data – designed to handle missing data gracefully, which is a common issue in real-world time series datasets; automatic seasonality detection - automatically detects and models daily, weekly, and yearly seasonality in the data; custom seasonalities - users can add custom seasonal components to capture specific periodic patterns that are not covered by default; holiday effects - allows users to specify holidays and special events that might affect the time series, and these effects are automatically included in the model; outliers - the model is robust to outliers, reducing their impact on the overall forecasting accuracy; trend changepoints - automatically detects changepoints in the data where the trend changes. This is crucial for time series with sudden shifts in trends due to external factors.

• Method using the "LSTM" model [21] - Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that is particularly well-suited for modeling and forecasting time series data. Unlike traditional RNNs, LSTMs can capture long-term dependencies in the data, making them ideal for sequences where relationships between distant time steps are important. The advantages of its use can be summarized as follows: ability to Learn Long-Term Dependencies - LSTM networks can capture long-range dependencies in time series data, which is crucial for accurate forecasting; robust to Noisy Data - LSTM models are robust to noise and can handle irregular time series data effectively; flexibility - LSTM networks can be used for a variety of tasks, including classification, regression, and anomaly detection.

GIS Methods Used for Analysis and Visualization

From the perspective of analyses and visualizations executed in the GIS environment during the study, Spatial Analysis and IDW Interpolation were applied. Spatial analysis is the process of manipulating spatial information to extract new information and meaning from the original data. Spatial interpolation is the process of using points with known values to estimate values at other unknown points. Spatial interpolation can estimate the concentration of pollutant at locations without recorded data by using known temperature readings at nearby weather stations. In the IDW interpolation method, the sample points are weighted during interpolation such that the influence of one point relative to another declines with distance from the unknown point in which we want to calculate missing value. IDW interpolation method also has some disadvantages: the quality of the interpolation result can decrease, if the distribution of sample data points is uneven. In GIS, interpolation results are usually shown as a two-dimensional raster layer.

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

Results

Initial conditions during the study and related to the forecasting process, the dataset was divided into training and testing sets in an 80:20 ratio. Forecasting was conducted for pollutants: Fine particulate matter < 10um [PM10], Ozone [O_3], Nitrogen dioxide [NO_2], Sulphur dioxide [SO_2]. Forecasting was performed separately for each monitoring station. IDW interpolation has been used in the QGIS environment to illustrate concentration maps derived from real and forecasted data. Subsequent figures and tables display results for O_3 studied similarly to analyses conducted for other air pollution parameters, the subject of this article. Figure 4 shows O_3 data from all stations for the period 2021 – 2024.

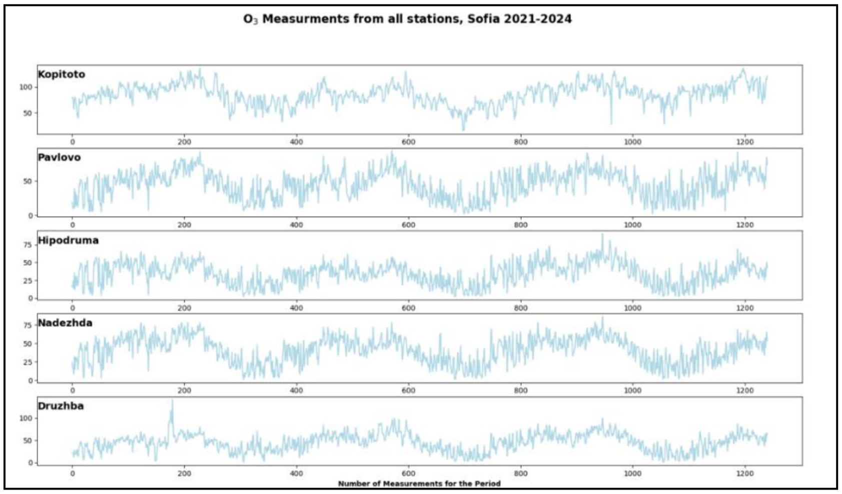

Figure 4. O_3 Measurements 2021-2024, Sofia

Figure 5 shows the O_3 data with temperature and humidity included before interpolation for applying the "Prophet" model.

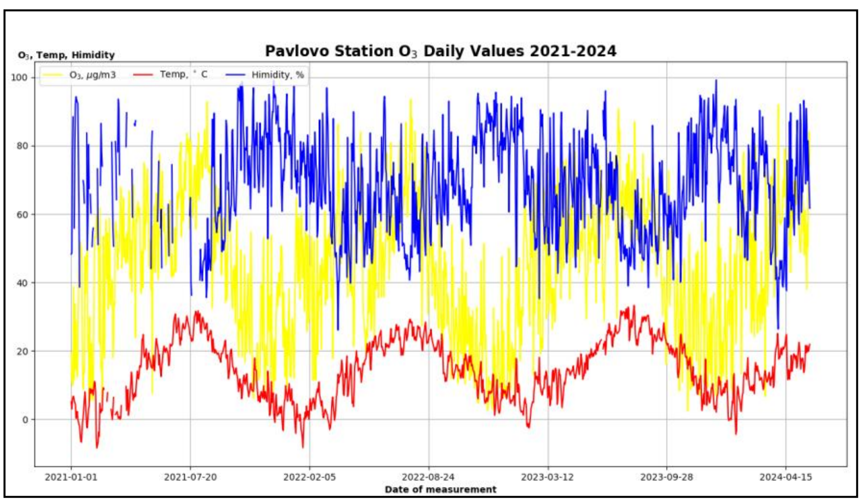

Figure 5. O_3 Before interpolation for temperature and humidity

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

Figure 6 displays the individual components of the "Prophet" model, while Figure 7 demonstrates the model's predictability compared to observations and measurements, utilizing an 80% Uncertainty Interval.

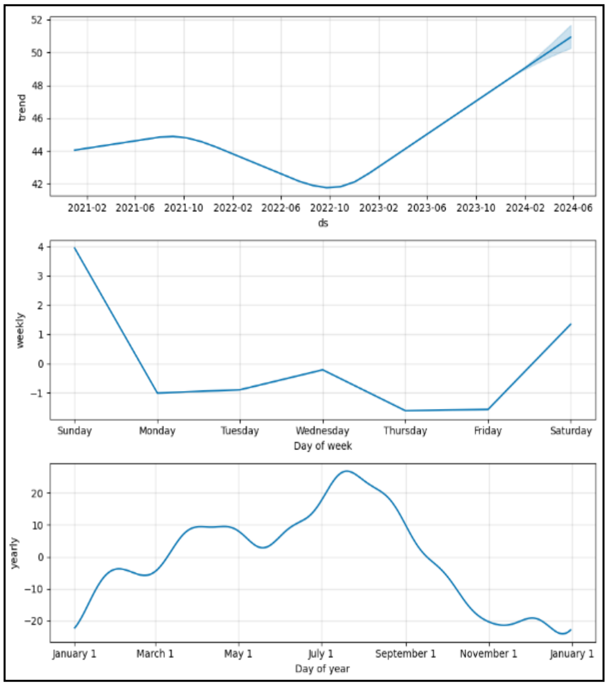

Figure 6. Prophet model – components

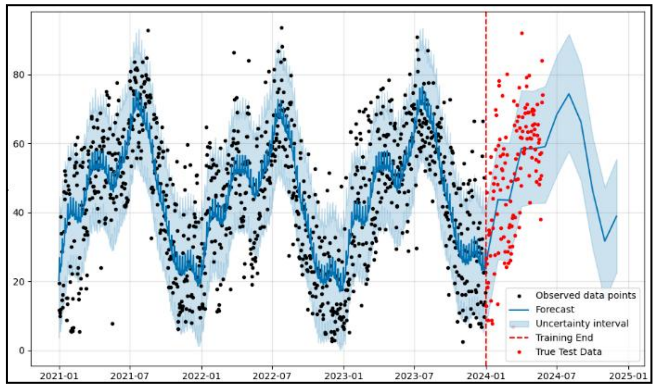

Figure 7. Prophet model – forecast and observation

Figures 8 and 9 illustrate data from training and results from using the LSTM method during the study.

```
Epoch 1/50 13/13 - 1s - 110ms/step - loss: 0.0266 - val_loss: 0.0075
Epoch 2/50 13/13 - 0s - 4ms/step - loss: 0.0121 - val_loss: 0.0096
....
Epoch 48/50 13/13 - 0s - 4ms/step - loss: 0.0072 - val_loss: 0.0043
Epoch 49/50 13/13 - 0s - 4ms/step - loss: 0.0072 - val_loss: 0.0043
Epoch 50/50 13/13 - 0s - 4ms/step - loss: 0.0071 - val_loss: 0.0043
```

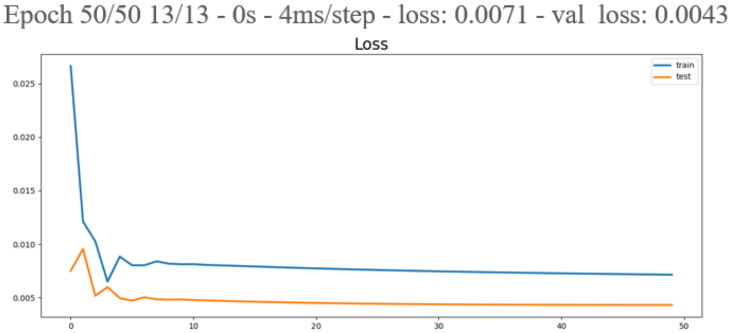

Figure 8. LSTM Training

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

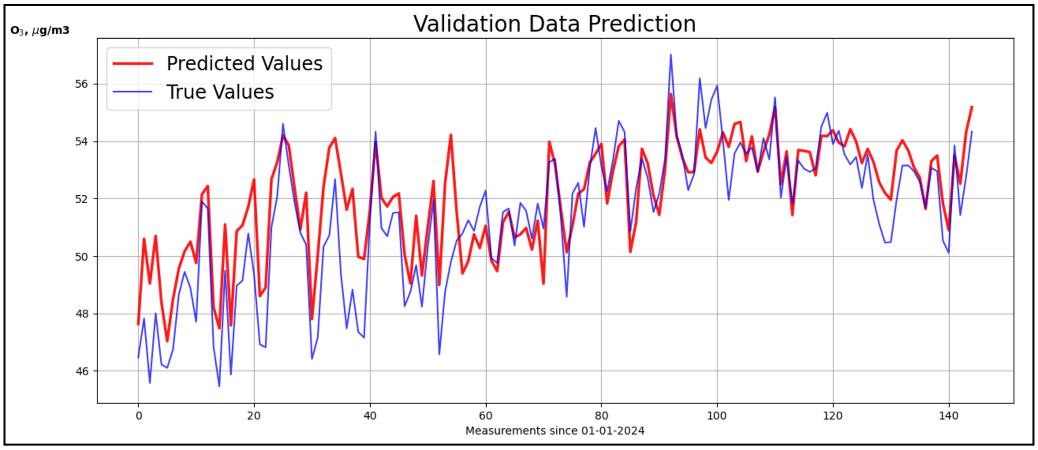

Figure 9. LTSM results

Table 2, as well as Figure 10, presents a comparative analysis and results when using both methods.

Table 2. Comparative analysis of the results of using different methods

Station	Measured	Prophet model	LSTM
Kopitoto	**120.76**	**72.46**	**96.61**
Pavlovo	**73.98**	**44.39**	**59.18**
Hipodruma	**42.79**	**25.71**	**34.23**
Nadezhda	**53.22**	**31.93**	**42.58**
Druzhba	**62.27**	**37.36**	**49.81**

The values shown in the table, obtained from both research methods, represent the midpoint of the range formed by maximum and minimum forecasted values, indicating that LSTM closely approximates the real measured results to a much greater extent.

Corresponding author: gmilchev@ltu.bg Full Paper
DOI: 10.2478/asn-2025-0014

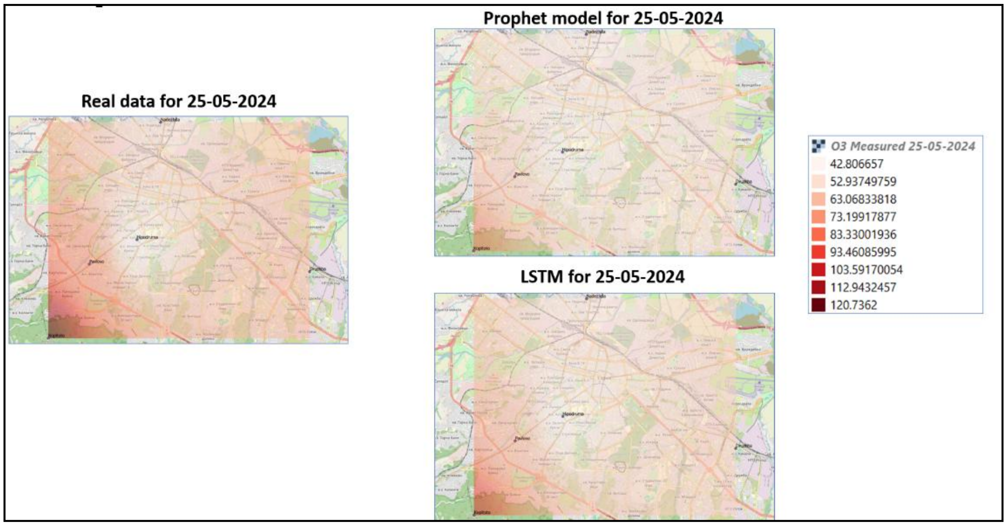

Figure 10. Comparison between different methods

Discussion

Currently, there are various sources of pollution-related data, but their formats require additional processing to be used in GIS or for artificial intelligence analysis. Preliminary analysis and screening are necessary due to occasional missing single measurements or periods of varying duration, which could complicate the fine-tuning of analyses both in GIS and AI. The quantity of data strongly influences time series analyses and forecast preparation, necessitating subsequent study of parameters for fine-tuning models. Challenges in using the proposed models include:

Related to the Prophet model and Computational Resources - while "Prophet" is relatively efficient, large datasets or complex models with many seasonalities and holidays can become computationally expensive; customization limits - while users can add custom seasonalities and changepoints, other aspects of the model are less customizable compared to more flexible machine learning models; stationarity assumptions - prophet assumes that the time series data is approximately stationary or can be made stationary with transformations; non-stationary data may require additional pre-processing; trend saturation - the logistic growth model in prophet assumes a saturation point for trends, which may not be appropriate for all datasets; historical data requirement -the model's performance heavily relies on having several seasons of historical data to accurately detect and model seasonality and trends; limited short-term forecasting - prophet may not perform as well for short-term forecasting if there is insufficient historical data; changepoint sensitivity - the default settings for changepoint detection may not always be optimal, and tuning these parameters can be challenging.

Regarding LSTM and Computational Complexity - training LSTM networks can be computationally intensive, especially for large datasets; hyperparameter tuning - selecting appropriate hyperparameters (e.g.,

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

number of layers, number of neurons) can be challenging and often requires extensive experimentation; overfitting - LSTMs can be prone to overfitting, particularly with small datasets. Regularization techniques and proper cross-validation are essential to mitigate this risk. As an alternative option for similar studies related to air pollution research, the ARIMA method can be used. (Autoregressive Integrated Moving Average) proposed by [22]. The predictions obtained from the application of this method would converge the mean of the time series for a further period if a stationarity is ensured [23]. The ARIMA method gives consistent results for short-term predictions. It is not preferable for long-term predictions, especially in the case of periodicity in time series. Long-term predictions, e.g., annual, might be more useful for city authorities to define long-term policies and measures. Thus, the periodicity of the time series has to be considered in order to provide reasoning based on the hidden data cycles. The knowledge about the underlying structure can be used as the measurement of effectiveness of the measures taken by policy makers [24].

Conclusions

The achieved results demonstrate good forecasting capabilities and their integration with GIS. The research results can be used to both make informed final decisions for implementing management practices, and to monitor the long-term effects of these practices. Integrating the results into GIS provides an easy and convenient tool not only for visualizing the results but also for obtaining values resulting from interpolation and at points other than the measurement stations.

Acknowledgments

• The authors would like to thank the Artificial Intelligence and CAD Systems Laboratory at Research & Development & Innovation Consortium at Sofia Tech Park for supporting of the presented research.

• This study was conducted in the framework of the Project NIS-B-1405 Study of the capabilities of artificial intelligence in assessing the soil-vegetation system as a factor in the greening of cities funded by the Scientific sector of the University of Forestry.

References

[1]. Phong, T.V., Pham, B.T., Trinh, P.T., Ly, H.-B., Vu, Q.H., Ho, L.S., Le, H.V., Phong, L.H., Avand, M. and Prakash, I., Groundwater Potential Mapping Using GIS-Based Hybrid Artificial Intelligence Methods. *Groundwater*, **2021**, *59*, 745-760, https://doi.org/10.1111/gwat.13094.

[2]. Abarca-Alvarez, Francisco Javier; Campos-Sanchez, Francisco Sergio, Reinoso-Bellido, Rafael. Methodology of Decision Support through GIS and Artificial Intelligence: Implementation for Demographic

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

Characterization of Andalusia based on Dwelling. *Estoa* [online]. **2017**, *6(11)*, 39-64, https://doi.org/10.18537/est.v006.n011.a03.

[3]. Liu, X.; Lao, C.; Li, X. et al. An integrated approach of remote sensing, GIS and swarm intelligence for zoning protected ecological areas. *Landscape Ecol*, **2012**, *27*, 447–463, https://doi.org/10.1007/s10980-011-9684-1.

[4]. Denis, V.; Hazlett, R.; Stevens, R.; Bornmalm, L., Sustainable Agriculture, GIS and Artificial Intelligence, *International Multidisciplinary Scientific Geo Conference: SGEM; Sofia*, **2022**, Vol. 22, Issue 5.1, DOI:10.5593/sgem2022/5.1/s20.056.

[5]. Pastor, J.; Johnston, C.A., Using Simulation Models and Geographic Information Systems to Integrate Ecosystem and Landscape Ecology. In: Naiman, R.J. (eds) Watershed Management. Springer, New York, NY, **1992**, https://doi.org/10.1007/978-1-4612-4382-3_11.

[6]. Li, Z., & Ning, H., Autonomous GIS: the next-generation AI-powered GIS. *International Journal of Digital Earth*, **2023**, *16(2)*, 4668–4686, https://doi.org/10.1080/17538947.2023.2278895.

[7]. El-Hallaq, M. & Habboub, M., Using GIS for Time Series Analysis of the Dead Sea from Remotely Sensing Data. *Open Journal of Civil Engineering*, **2014**, *04*, 386-396, 10.4236/ojce.2014.44033.

[8]. Ly, H.-B.; Le, L.M.; Phi, L.V.; Phan, V.-H.; Tran, V.Q.; Pham, B.T.; Le, T.-T.; Derrible, S., Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data. *Sensors*, **2019**, *19*, 4941, https://doi.org/10.3390/s19224941.

[9]. Tella, A.; Balogun, AL., GIS-based air quality modelling: spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms. *Environmental Science and Pollution Research*, **2022**, *29*, 86109–86125, https://doi.org/10.1007/s11356-021-16150-0.

[10]. Soleimany, A.; Grubliauskas, R.; Šerevičienė, V., Application of satellite data and GIS services for studying air pollutants in Lithuania (case study: Kaunas city). *Air Qual Atmos Health*, **2021**, *14*, 411–429, https://doi.org/10.1007/s11869-020-00946-z.

[11]. Goswami, T.; Sarma, H., Intelligent Computing for Air Pollution Monitoring Using GIS, Remote Sensing and Machine Learning. In: Hitendra Sarma, T., Sankar, V., Shaik, R. (eds) Emerging Trends in Electrical, Communications, and Information Technologies. Lecture Notes in Electrical Engineering, **2020**, vol 569, Springer, Singapore, https://doi.org/10.1007/978-981-13-8942-9_12.

[12]. Zhalehdoost, A., & Taleai, M., A Review of the Application of Machine Learning and Geospatial Analysis Methods in Air Pollution Prediction. *Pollution*, **2022**, *8(3)*, 904-933, doi: 10.22059/poll.2022.336044.1300.

[13]. Briggs, D., The Role of Gis: Coping with Space (And Time) in Air Pollution Exposure Assessment. *Journal of Toxicology and Environmental Health*, Part A, **2005**, *68(13–14)*, 1243–1261. https://doi.org/10.1080/15287390590936094.

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014

[14]. Briggs, D. J.; Collins, S.; Elliott, P.; Fischer, P.; Kingham, S.; Lebret, E.; … Van der Veen, A., Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science*, **1997**, *11(7)*, 699–718, https://doi.org/10.1080/136588197242158.

[15]. Rahman, M.R.; Lateh, H., Climate change in Bangladesh: a spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model. Theoretical and Applied Climatology, **2017**, *128*, 27–41, https://doi.org/10.1007/s00704-015-1688-3.

[16]. Salcedo, R.L.R.; Alvim Ferraz, M.C.M.; Alves, C.A.; Martins, F.G., Time-series analysis of air pollution data, *Atmospheric Environment*, **1999**, 33(15), 2361-2372, https://doi.org/10.1016/S1352-2310(99)80001-6.

[17]. He Li, Xiao-Long Xu, Da-Wei Dai, Zhen-Yu Huang, Zhuang Ma, Yan-Jun Guan, Air pollution and temperature are associated with increased COVID-19 incidence: A time series study, *International Journal of Infectious Diseases*, **2020**, *97*, 278-282, https://doi.org/10.1016/j.ijid.2020.05.076.

[18]. Akbari, M.; Zahmatkesh, H.; Eftekhari, M., A GIS-Based System for Real-Time Air Pollution Monitoring and Alerting Based on OGC Sensors Web Enablement Standards, *Pollution*, **2021**, *7*, 25-41. https://doi.org/10.22059/poll.2020.296938.741.

[19]. Shareef, M.; Husain, T.; Alharbi, B., Optimization of Air Quality Monitoring Network Using GIS Based Interpolation Techniques. *Journal of Environmental Protection*, **2016**, *7*, 895-911. https://doi.org/10.4236/jep.2016.76080.

[20. Andrei, Nistor & Ioanid, Alexandra. Potential use of artificial intelligence and geospatial analysis in environmental monitoring: Air quality in a large city. *International Conference of Management and Industrial Engineering*, **2023**, *11*, 369-376, https://doi.org/10.56177/11icmie2023.31.

[21]. Hochreiter, S.; Schmidhuber, J., Long short-term memory. *Neural Computation*, **1997**, *9(8)*, 1735–1780.

[22]. Marinov, E.; Petrova-Antonova, D.; Malinov, S., Time Series Forecasting of Air Quality: A Case Study of Sofia City, *Atmosphere*, **2022**, *13(5)*, 788, https://doi.org/10.3390/atmos13050788.

[23]. Liu, P.W.G., Simulation of the daily average PM10 concentrations at Ta-Liao with Box–Jenkins time series models and multivariate analysis. *Atmospheric Environment*, **2009**, *43*, 2104–2113.

[24]. Zhang, L.; Lin, J.; Qiu, R.; Hu, X.; Zhang, H.; Chen, Q.; Wang, J., Trend analysis and forecast of PM2. 5 in Fuzhou, China using the ARIMA model, *Ecological Indicators*, **2018**, *95*, 702–710.

Corresponding author: gmilchev@ltu.bg
Full Paper
DOI: 10.2478/asn-2025-0014