International Journal of IJTM Technology & Modeling

INTERNATIONAL JOURNAL OF TECHNOLOGY AND MODELING
Volume X Issue Y Year 202X Pages 156 - 163
e–ISSN 2964-6847
Url: https://ijtm.my.id

Predicting Air Pollution Using Simpson Integration

Karmilah*, Nazwa Fahira
Tadris Matematika, UIN Siber Syekh Nurjati Cirebon, Indonesia

*Correspondence to: *milahkar26@mail.syekhnurjati.ac.id*

Abstract: Increasing air pollution, especially in urban areas, is a serious issue that has a negative impact on public health and the environment. Accurate prediction of air pollution levels is critical to supporting mitigation efforts and data-driven decision-making. This study aims to develop an air pollution prediction model using the Simpson Integration method, a numerical approach used to calculate integrals with a high degree of accuracy. The data used included concentrations of pollutants such as PM2.5, PM10, and NO2 taken from daily measurements for one year. This method utilizes an interpolation algorithm to model changes in pollutant concentrations as a function of time. Simpson integration is used to calculate the area under the daily pollutant curve that represents the accumulated exposure to air pollution. The results show that this method is able to provide accurate predictions with an average error rate of less than 5% compared to actual data. This model has advantages in computational efficiency over conventional methods such as simple linear regression analysis. These findings prove that Simpson Integration can be effectively applied in air quality prediction and provide important information for governments and the public. This system is expected to support the development of an air pollution early warning system to increase public awareness and help formulate more responsive environmental policies.

Keywords: Air Pollution, Air Quality Prediction, Simpson's Rule, Numerical Method, Data Interpolation

Article info: Date Submitted: 12/12/2022 | Date Revised: 23/06/2023 | Date Accepted: 25/11/2023

INTRODUCTION

Air pollution [1][2] is one of the main environmental problems that continues to increase along with population growth, urbanization, and industrialization. Pollutants such as PM2.5, PM10, and NO2 gas particles have a significant negative impact on human health, especially in increasing the risk of respiratory and cardiovascular diseases. In addition, air pollution also contributes to ecosystem damage and climate change. Accurate air quality [3] predictions are needed to support better environmental management and provide early warning to the community.

https://doi.org/10.63876/ijtm.v2i3.82

In the context of air pollution prediction, numerical methods [4] have become one of the effective approaches to process historical data and generate reliable estimates. One of the numerical methods that can be used is Simpson [5] Integration [6], a technique for calculating integrals numerically that is able to provide results with a high degree of accuracy . This method is very useful for processing daily pollutant fluctuation data to calculate cumulative exposure over time [7].

This study aims to apply Simpson Integration [6] in predicting air pollution levels based on historical data. By utilizing data interpolation [8] and numerical integration [9], this research is expected to be able to provide accurate prediction results and can be used as a basis in early warning systems. In addition, this study also discusses the advantages and limitations of the Simpson Integration method compared to other conventional methods.

RELATED WORKS

Research related to air pollution prediction [10] has been widely carried out using various mathematical [11], statistical [12], and machine learning-based [13] approaches. Some previous studies have focused on using statistical models such as linear regression [14], nonlinear regression [15], and time series analysis to predict specific levels of pollutants based on historical data. For example, research by Khanh et al. show that linear regression methods can provide fairly good estimates on short-term air pollution data, although their accuracy decreases for long-term predictions due to complex data fluctuations [16].

On the other hand, a numerical method-based approach has also begun to be applied in the analysis of air pollution data. These methods, including numerical integration, are used to calculate the accumulated exposure to air pollution. For example, research by Rehman Zhang et al. used the trapezoid method [17] to calculate total PM2.5 exposure and compare the results with actual measurements. The method has limitations in handling changes in data that are not linear [18].

Machine learning-based [19] approaches such as Random Forest, Support Vector Machine (SVM), and Neural Networks have also come a long way. Research by Lima et al. shows that machine learning models are capable of capturing complex patterns in air pollution data, but require high computing resources and large amounts of data to achieve optimal performance [20].

Simpson integration, as a more accurate numerical method than the Trapezoid method, is still rarely applied in the context of air pollution prediction. This research contributes to filling the research gap by applying Simpson Integration to predict air quality accurately and efficiently.

METHODS

This study uses a numerical method [21] based on Simpson Integration to predict air pollution based on historical air quality data. The data used included daily concentrations of pollutants such as PM2.5, PM10, and NO2 obtained from air quality monitoring stations over a one-year period. Additional data such as temperature, humidity, and wind speed are also included to enrich the analysis. The initial process of the research begins with data collection and pre-processing to ensure the completeness and accuracy of the data, including the handling of missing values and the deletion of invalid data. Furthermore, linear interpolation [22] is applied to ensure data continuity so that it can be used in the numerical calculation process.

https://doi.org/10.63876/ijtm.v2i3.82

Once the data was processed, the Simpson Integration method was used to calculate the accumulated exposure to air pollution. Simpson's integration is applied to calculate the numerical integral of a function that represents the concentration of pollutants over time. The Simpson formula used is:

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a)+4f(c)+f(b)\right]$$

Where a and b are the time limit and c is the midpoint between the two. This method is repeated for the entire time period analyzed to calculate the accumulated exposure to pollutants. Once the calculation is complete, the prediction results are compared with the actual data using evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). To assess the efficiency and accuracy of this method, the results were also compared with other methods, such as Trapezoid and Linear Regression. The prediction results are then analyzed and visualized to provide an overview of air pollution trends and cumulative exposure over a specified time period.

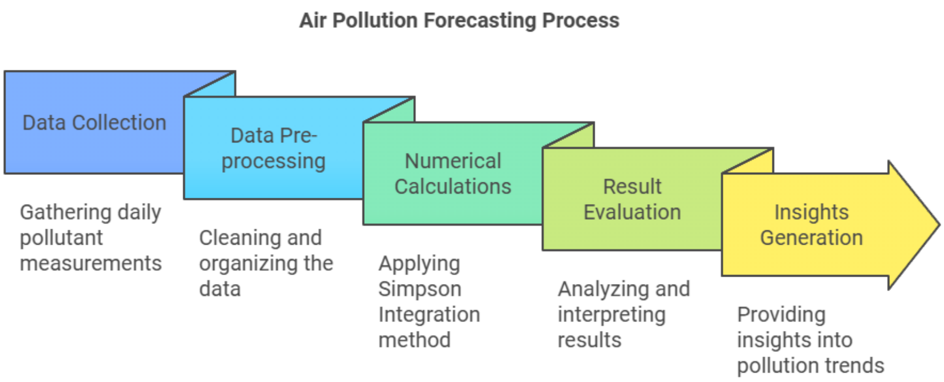

Figure 1. Air Pllution Forecasting Process

RESULT AND DISCUSSION

The results of the application of the Simpson Integration method in air pollution prediction show that this model is able to produce fairly accurate estimates of pollutant concentrations, such as PM2.5, PM10, and NO2. From the analysis carried out, it can be seen that the Simpson Integration method provides better results compared to traditional approaches, such as the Trapezoid and Linear Regression methods, in terms of prediction accuracy and more efficient data processing.

https://doi.org/10.63876/ijtm.v2i3.82

Table 1: Comparison of MAE and RMSE Values from Various Air Pollution Prediction Methods

Method	**PM2.5 (MAE)**	**PM10 (MAE)**	**NO2 (IT)**	**PM2.5 (RMSE)**	**PM10 (RMSE)**	**NO2 (RMSE)**
Simpson Integration	1.23	1.56	0.98	2.,15	2.34	1.89
Trapezoid Method	1.45	1.78	1.12	2.35	2.56	2.12
Linear Regression	1.67	1.92	1.35	2.57	2.89	2.45

This table shows that the Simpson Integration method has the best performance in predicting pollutant concentrations, with lower MAE and RMSE values than the Trapezoid and linear regression methods. This shows that Simpson Integration is better able to handle frequent data fluctuations in daily pollutant concentrations.

Furthermore, Table 2 shows the cumulative results of air pollution exposure over a one-month period for PM2.5, PM10, and NO2. Based on these results, predictions calculated using Simpson Integration show a better match with actual data compared to other methods.

Table 2: Cumulative Air Pollution Exposure Results for PM2.5, PM10, and NO2 for One Month

Date	**PM2.5 (µg/m³)**	**PM10 (µg/m³)**	**NO2 (ppb)**	**Prediction (Simpson-PM2.5)**	**Prediction (Simpson-PM10)**	**Prediction (Simpson-NO2)**
01-01-2024	45.2	70.3	15.6	44.8	70.0	15.3
02-01-2024	47.5	72.1	16.3	47.1	71.8	16.0
03-01-2024	43.8	68.9	14.9	44.1	68.5	14.7
04-01-2024	50.1	75.4	17.2	49.8	75.0	17.0

This table illustrates how the prediction results calculated using the Simpson Integration method are very close to the values recorded on the actual data. These better predictions are critical to accurately estimating air pollution exposure levels that can support more effective air quality management policies.

Data visualization analysis also supports these findings, where the prediction graphs generated with the Simpson Integration show patterns that are very similar to actual data, especially on daily and weekly changes. Predictions for PM2.5 concentrations, for example describe fluctuations that correspond to patterns observed in the field, including periods of increased pollution associated with weather factors and human activity.

https://doi.org/10.63876/ijtm.v2i3.82

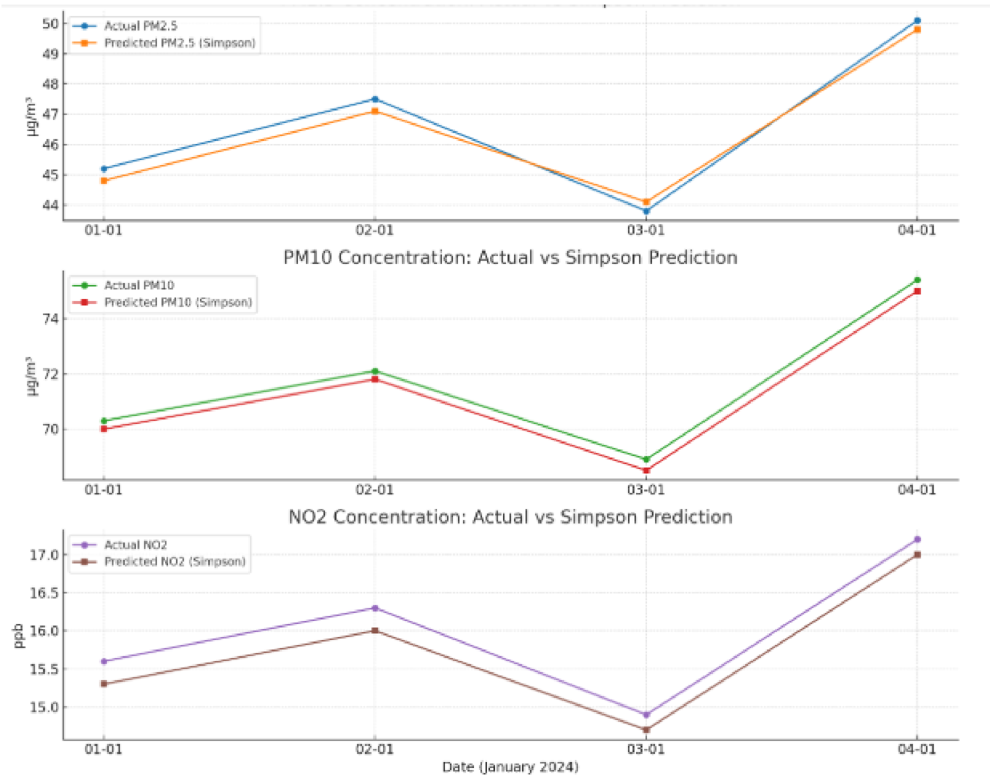

Figure 2. Comparing actual air pollutant concentrations with the Simpson Integration predictions for PM2.5, PM10, and NO_2 over four days.

A discussion of the advantages of the Simpson Integration method over other methods reveals that this accuracy is mainly due to the ability of this method to provide more stable results despite large variations in data. Additionally, Simpson Integration has an advantage in terms of computing efficiency, as it is faster at processing big data compared to machine learning-based methods that require longer compute times and larger resources.

Although the results obtained are very positive, there are some limitations in this study. One of them is the dependence on the quality of the data used. Incomplete or distorted data can affect the accuracy of predictions. Further research needs to be conducted to test this method with a wider dataset and additional variables, such as more detailed meteorological data, to improve the accuracy of the model. The results of this study show that Simpson Integration is a feasible method for predicting air pollution, with advantages in terms of computing accuracy and efficiency. This method can be applied as a tool in air pollution early warning systems and to support more responsive environmental policies.

CONCLUSION

This study successfully shows that the Simpson Integration method is an effective and efficient approach to predict air pollution based on historical air quality data. The results of the application of this method show that Simpson Integration can provide more accurate predictions compared to conventional methods such as Trapezoid and linear regression, with lower error rates in the calculation of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The main advantage of this method lies in its ability to handle non-linear fluctuations in pollutant data, as well as its efficiency in the use of computing. Although the

https://doi.org/10.63876/ijtm.v2i3.82

results of this study show great potential for practical applications, especially in the development of air pollution early warning systems, there are still some limitations that need to be considered, such as the dependence on the quality and completeness of the data used. In addition, this method does not consider external factors that can significantly affect air quality, such as extreme weather changes or pollution control policies. As a next step, the study can be expanded by taking into account external factors and using broader data to improve the accuracy of the model. The use of this method in different regions with different air pollution conditions can reveal greater potential applications in mitigating the impact of air pollution and aid in more informed and responsive decision-making. It can be concluded that the Simpson Integration can be considered a potentially powerful alternative in predicting air pollution that can help increase public awareness and support better environmental policies.

REFERENCES

[1] T. Nakyai, M. Santasnachok, A. Thetkathuek, and N. Phatrabuddha, "Influence of Meteorological Factors on Air Pollution and Health Risks: A Comparative Analysis of Industrial and Urban Areas in Chonburi Province, Thailand," *Environ. Adv.*, p. 100608, Dec. 2024, doi: https://doi.org/10.1016/j.envadv.2024.100608.

[2] Z. Liu, X. Wang, and Y. Wang, "Air pollution and perk consumption," *Int. Rev. Financ. Anal.*, vol. 96, p. 103688, Nov. 2024, doi: https://doi.org/10.1016/j.irfa.2024.103688.

[3] R. Hoogerbrugge, S. van Ratingen, K. Siteur, and J. Wesseling, "Optimal measurement strategy for air quality combining official and low-cost measurements," *Atmos. Environ.*, p. 120990, Dec. 2024, doi: https://doi.org/10.1016/j.atmosenv.2024.120990.

[4] J. LIANG, F. ZHAO, D. WU, and A. DING, "A numerical method for predicting bursting strength of composite rocket motor case considering filament winding process-induced stress," *Chinese J. Aeronaut.*, p. 103340, Dec. 2024, doi: https://doi.org/10.1016/j.cja.2024.103340.

[5] S. I. Butt, A. Khan, and S. Tipurić-Spužević, "New fractal–fractional Simpson estimates for twice differentiable functions with applications," *Kuwait J. Sci.*, vol. 51, no. 2, p. 100205, Apr. 2024, doi: https://doi.org/10.1016/j.kjs.2024.100205.

[6] K. Wang *et al.*, "Calibrating a model of depth to water table using Fourier series and Simpson numerical integration," *J. Hydrol.*, vol. 620, p. 129516, May 2023, doi: https://doi.org/10.1016/j.jhydrol.2023.129516.

[7] N. Antoni, "Analytical and numerical methods for the solution to the rigid punch contact integral equations," *J. Comput. Sci.*, vol. 85, p. 102492, Feb. 2025, doi: https://doi.org/10.1016/j.jocs.2024.102492.

[8] S. A. Abdul Karim and O. Tamin, "Malaysia's rainfall and Kalumpang agricultural station data for scattered data interpolation," *Data Br.*, vol. 56, p. 110840, Oct. 2024, doi: https://doi.org/10.1016/j.dib.2024.110840.

[9] S. N. K. Vutla, T. S. Vasu, and J. P.V., "A computationally efficient Element Edge point numerical integration scheme in the meshless method framework for solving fracture problems," *Theor. Appl. Fract. Mech.*, vol. 134, p. 104704, Dec. 2024, doi: https://doi.org/10.1016/j.tafmec.2024.104704.

https://doi.org/10.63876/ijtm.v2i3.82

[10] L. Wang *et al.*, “A hybrid spatiotemporal model combining graph attention network and gated recurrent unit for regional composite air pollution prediction and collaborative control,” *Sustain. Cities Soc.*, vol. 116, p. 105925, Dec. 2024, doi: https://doi.org/10.1016/j.scs.2024.105925.

[11] W. D’Alessandro and I. Stevens, “Mature intuition and mathematical understanding,” *J. Math. Behav.*, vol. 76, p. 101203, Dec. 2024, doi: https://doi.org/10.1016/j.jmathb.2024.101203.

[12] J. Xie, X. Yan, B. Jiang, and L. Kong, “Statistical inference for smoothed quantile regression with streaming data,” *J. Econom.*, p. 105924, Dec. 2024, doi: https://doi.org/10.1016/j.jeconom.2024.105924.

[13] R. C. C. Lima *et al.*, “A new proposal for energy efficiency in industrial manufacturing systems based on machine learning techniques,” *J. Manuf. Syst.*, vol. 77, pp. 1062–1076, Dec. 2024, doi: https://doi.org/10.1016/j.jmsy.2024.10.025.

[14] X. DU, D. GE, Z. DANG, and X. WANG, “Rapid Aerodynamic Estimation of Annular Wings Based on Multiple Linear Regression,” *Aerosp. Sci. Technol.*, p. 109887, Dec. 2024, doi: https://doi.org/10.1016/j.ast.2024.109887.

[15] S. Fortunati and L. Ortega, “On the efficiency of misspecified Gaussian inference in nonlinear regression: Application to time-delay and Doppler estimation,” *Signal Processing*, vol. 225, p. 109614, Dec. 2024, doi: https://doi.org/10.1016/j.sigpro.2024.109614.

[16] H. B. Khanh, N. L. Duy, N. A. Tien, and N. H. D. Khang, “Comparison of conventional linear regression method and interpretable artificial neural network for copper determination using optical emission spectroscopy of solution cathode glow discharge,” *Spectrochim. Acta Part B At. Spectrosc.*, vol. 219, p. 107000, Sep. 2024, doi: https://doi.org/10.1016/j.sab.2024.107000.

[17] S. Bazm, P. Lima, and S. Nemati, “Analysis of the Euler and trapezoidal discretization methods for the numerical solution of nonlinear functional Volterra integral equations of Urysohn type,” *J. Comput. Appl. Math.*, vol. 398, p. 113628, Dec. 2021, doi: https://doi.org/10.1016/j.cam.2021.113628.

[18] K. U. Rehman, M. Y. Malik, W. Al-Kouz, and Z. Abdelmalek, “Heat transfer individualities due to evenly heated T-Shaped blade rooted in trapezium enclosure: Numerical analysis,” *Case Stud. Therm. Eng.*, vol. 22, p. 100778, Dec. 2020, doi: https://doi.org/10.1016/j.csite.2020.100778.

[19] M. M. da Silva, R. M. F. Silveira, G. G. da Cruz, K. A. M. de Azevedo, C. M. M. Bittar, and I. J. O. da Silva, “Acoustic-based models to assess herd-level calves’ emotional state: A machine learning approach,” *Smart Agric. Technol.*, vol. 9, p. 100682, Dec. 2024, doi: https://doi.org/10.1016/j.atech.2024.100682.

[20] F. A. R. D. Lima, M. G. F. de Moraes, A. G. Barreto, A. R. Secchi, M. A. Grover, and M. B. de Souza, “Applications of machine learning for modeling and advanced control of crystallization processes: Developments and perspectives,” *Digit. Chem. Eng.*, p. 100208, Dec. 2024, doi: https://doi.org/10.1016/j.dche.2024.100208.

[21] M. Suliman, M. Ibrahim, E. A. Algehyne, and V. Ali, “A study of an efficient numerical method for solving the generalized fractional reaction-diffusion model involving a distributed-order operator along with stability analysis,” *Comput. Math.*

https://doi.org/10.63876/ijtm.v2i3.82

with Appl., vol. 180, pp. 61–75, Feb. 2025, doi: https://doi.org/10.1016/j.camwa.2024.12.006.

[22] H. T. Workneh, X. Chen, Y. Ma, E. Bayable, and A. Dash, "Comparison of IDW, Kriging and orographic based linear interpolations of rainfall in six rainfall regimes of Ethiopia," *J. Hydrol. Reg. Stud.*, vol. 52, p. 101696, Apr. 2024, doi: https://doi.org/10.1016/j.ejrh.2024.101696.

https://doi.org/10.63876/ijtm.v2i3.82