Tikrit Journal of Pure Science Vol. 31 (1) 2026

DOI:https://doi.org/10.25130/tjps.v31i1.1933

IRAQI Academic Scientific Journals

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: https://tjpsj.org/

Predicting the Percentage of Air Pollution Gases Using the Particle Swarm Optimization Algorithm

Salwa Shakir Baawi, Zuhal Adil Madlool

Computer science and information technology, Al-Qadisiyah University, Al-Qadisiyah, Iraq

Received: 8 Mar. 2025 **Received in revised forum:** 18 Apr. 2025 **Accepted:** 20 Apr. 2025

Final Proofreading: 27 Jul. 2025 **Available online:** 25 Feb. 2026

ABSTRACT

Caring for the environment is a major priority in many countries, as environmental pollution poses a serious threat to natural resources, including water, air, and land. Pollution levels have reached alarming rates, prompting researchers across various scientific disciplines to focus on studies aimed at reducing and controlling these pollutants within permissible limits.

Air and atmospheric pollution are among the most dangerous forms of pollution affecting human health and the environment. They contribute to global warming and ozone layer depletion by emitting harmful gases, especially nitrogen dioxide (NO_2). When the concentration of NO_2 in the air reaches 0.07%, it transforms into nitric acid, a lethal gas that can cause death within half an hour. These oxides react with hemoglobin in the blood, hindering oxygen transport to cells, making children particularly vulnerable. Symptoms such as blue lips are common signs of this type of poisoning. In industrial regions like the United States, nitrogen oxides are major contributors to acid rain.

With significant advances in digital data recording technologies, environmental data is now captured as a time series. This allows the application of mathematical models to analyze pollutant behavior for control and prediction. In this research, the Particle Swarm Optimization (PSO) algorithm was applied to analyze nitrogen dioxide (NO_2) levels in Baghdad from 2015 to 2017, using weekly averages across 157 observations. The model achieved a prediction accuracy rate of approximately 94%.

Keywords: Big data, Machine learning, deep learning, Optimization, Prediction

Name: **Zuhal Adil Madlool** **E-mail:** zuhal.adil@qu.edu.iq

Tikrit Journal of Pure Science Vol. 31 (1) 2026

DOI:https://doi.org/10.25130/tjps.v31i1.1933

التنبؤ بنسبة الغازات الملوثة للهواء باستخدام خوارزمية تحسين سرب الجسيمات

سلوى شاكر بعيوي، زحل عادل مدلول

كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة القادسية، العراق

الملخص

تُعد العناية بالبيئة من الأولويات المهمة في مختلف دول العالم، حيث يُشكل التلوث البيئي تهديدًا كبيرًا لمكونات البيئة، بما في ذلك الماء، والهواء، واليابسة. وقد وصلت مستويات التلوث إلى حدود خطيرة، مما استدعى اهتمام الباحثين في مختلف التخصصات العلمية لإيجاد طرق للحد من هذه الملوثات والسيطرة عليها ضمن الحدود المسموح بها.

يُعد تلوث الهواء والغلاف الجوي من أخطر أنواع التلوث على صحة الإنسان والبيئة، إذ يؤدي إلى ظواهر مثل الاحتباس الحراري واستنزاف طبقة الأوزون، نتيجة لانبعاث الغازات الضارة، وعلى رأسها ثاني أكسيد النيتروجين (NO_2). فعند وصول تركيز هذا الغاز في الهواء إلى 0.07%، يتحول إلى حمض النيتريك، وهو غاز قاتل يمكن أن يسبب الوفاة خلال نصف ساعة. وتتسبب هذه الأكاسيد بتثبيط نقل الأكسجين في الدم، مما يجعل الأطفال والفئات الضعيفة أكثر عرضة للتسمم، ويؤدي إلى ظواهر مثل زرقة الشفاه. كما أن هذه الغازات تساهم في تكوين الأمطار الحمضية، خاصة في المناطق الصناعية مثل الولايات المتحدة.

ونظرًا للتطور الكبير في أجهزة تسجيل البيانات الرقمية، أصبحت البيانات تُسجّل على شكل سلاسل زمنية، مما أتاح إمكانية استخدام نماذج رياضية لدراسة سلوك الملوثات. وفي هذا البحث، تم استخدام خوارزمية تحسين أسراب الجسيمات (PSO) لتحليل بيانات ثاني أكسيد النيتروجين في مدينة بغداد للفترة 2015–2017، باستخدام متوسطات أسبوعية لـ157 مشاهدة، وبلغت دقة النتائج حوالي 94%.

INTRODUCTION

Since ancient times, man has been interested in anticipating the future, predicting its changes, and predicting its values. At the institutional and national levels, reading the future is a very important process because of its close, intimate connection with various planning and development initiatives. (1, 2).

Time series analysis using the particle swarm optimization algorithm is one of the most important scientific methods for predicting air pollutant concentrations. The research problem boils down to how to build an optimization model for predicting and how to use it in time series data. (1, 3).

With a growing population comes a greater complexity in studying the factors that contribute to pollution, such as toxic gases, pollutants, vehicle and industrial exhausts, wind, etc. As a result, longer-term forecasting has become more complex and more accurate projections are needed. The interplay between polluting gases and extraneous factors is complex and nonlinear, significantly limiting the predictive power of standard models. This led to the current line of inquiry, which seeks to optimize the study and prediction of environmental phenomena to better inform decision-makers about both the here and now and the future.(1, 4).

RELATED WORK

Haidong Kan, Renjie Chen, Shilu Tong (5) This article discusses the relationships among China's population health, climate change, and ambient air pollution. Using particle swarm optimization, our research predicts the proportions of air-polluting gases with an accuracy of up to 94%.

Jansakoo, T., Sekizawa, S., Fujimori, S. et al. (6). Our research presents a particle swarm optimization algorithm to predict the percentage of air-polluting gases, and this research utilized a chemical transport model and an integrated assessment model to determine changes in fine particulate matter. It also suggested that human health could benefit from improved air quality resulting from climate change mitigation efforts, particularly through dietary changes and reduced food waste.

Huda Kadem Alwan, Ghaidaa A. AL-Sultany, Saba Mohammed Hussein [7] This research uses machine learning and neural networks to predict air pollution. This paper provides a survey of recent air quality prediction techniques and highlights the essentials.

Lingling Lv., Peng Wei., Juan Li., Jingnan Hu. [8] To enhance numerical simulation predictions of PM2.5 and chemical components, this research applies machine learning methods. Three common approaches were used to calibrate the discrepancies. We used important meteorological factors as inputs, along with simulated and observed pollution concentrations.

Anıl UTKU, Umit CAN [9] The findings showed that the advanced model proposed in this study outperformed the other 11 models used to predict PM2.5 in terms of MSE, RMSE, MAE, and R2.

Jhayron S. Pérez-Carrasquilla, Paola A. Montoya, Juan Manuel Sánchez, K. Santiago Hernández, and Mauricio Ramírez [10] This research forecasts the average PM2.5 concentration over 24 hours in the Aburrá Valley using tree-based ML models, global projections, and satellite data.

Jianlin Hu, Bart Ostro, Hongliang Zhang, Qi Ying and Michael J. Kleeman [11] This work improved the exposure assessment of PM2.5 components by using chemical transport models in California and using their predictions.

MAIN CONCEPT

Big data

Given the diversity of sources that might provide data, this term has been widely used. Because it requires big storage systems and might be chaotic, dealing with this data may be difficult. In 2001, Douglas Laney proposed the concept of big data, based on the three Vs: volume, velocity, and variety. Since then, many have used this notion. Some, nevertheless, have contemplated increasing the number of Vs by four, five, six, or even eleven. Another approach to explaining big data is to view it as an application, emphasizing its many uses depending on the kinds of data, in contrast to Barry Devlin's explanation, which centers on process-mediated data, human-sourced information, and machine-generated data. By examining data on contacts, observations, and transactions, Shaun Connolly hoped to gain insights using big data technology. [12].

Machine learning

An approach that builds systems capable of learning from their own mistakes and improving their performance over time, acquiring valuable skills in areas like optimization and prediction through algorithms that interpret data. Particularly in the realm of new technology, this is crucial. [13]. Computer programs that learn from their mistakes and improve their performance are the focus of machine learning, which also seeks to understand the theoretical, computational, and informational principles that govern the functioning of learning systems. [13]Figure 1 shows the type of machine learning.

In [14] Three ways of looking at machine learning are possible:

- Supervised learning: which involves training a model using previously labelled data and known inputs/outputs to make predictions about new data. Experts train algorithms to achieve desired outcomes using this most commonly utilised strategy.

- Unsupervised learning: Unlike supervised approaches, which create preset outputs, this methodology examines, analyses, and consolidates unclassified incoming data. It finds insights and patterns in your data by free-training on it. Machines may learn previously unseen patterns and procedures with little to no human input using this strategy, which is based on self-learning [15].

One such method of learning is reinforcement learning, which entails training models to make sequential choices in response to incentives for good performance. This approach allows one to learn how to do things in uncertain circumstances by rewarding each successful step towards an aim in a complicated environment. While this method is conceptually similar to direct feedback systems, it differs significantly in that it relies

DOI:https://doi.org/10.25130/tjps.v31i1.1933

on trial-and-error rather than predefined data for learning. [16].

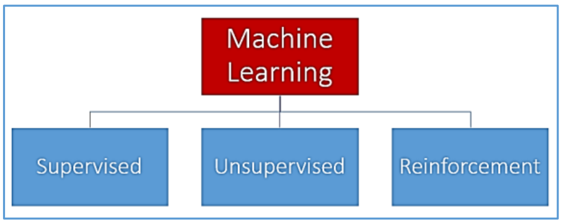

Fig. 1: Types of Machine Learning

Prediction

The purpose of data analysis in prediction is to show what the future holds for attributes whose precise nature is currently unclear. Two parts make up the whole: a numerical prediction work for predicting ordered or continuous values, and a classification job for predicting class labels. [17]. The nature of the target characteristic determines whether a task involves numerical prediction (continuous values) or classification (binary values). Several statistical approaches have been used for this type of numerical prediction, with regression analysis being the most commonly used. [16, 18].

Air Pollution

The scientific community, governments, and businesses have all taken notice of the persistent problem of air pollution. Fine particulate matter is a form of air pollution that has garnered significant attention. anything that contaminates the air on a large scale, including both solid and liquid particles suspended in the atmosphere. Because air pollution is a problem all across the world, regardless of where you are, finding a solution will need expertise from many different fields [1, 6].

Predicting when dangerous levels of air pollutants will be present is an important step in safeguarding public health.[2].

Statement of the problem

The problem lies in creating an integrated system that can predict the percentage of air-polluting gases. Most researchers have studied the problem of air pollution for specific gases, for example, in [8, 9] They addressed the problem of predicting PM2.5 levels and methods for improving prediction using various algorithms. In contrast, we focused on predicting NO2 levels because of its serious impact on air pollution, despite most researchers' lack of interest in this gas. Our research is therefore somewhat unique among the rest.

Contribution

High prediction accuracy is achieved for 10-week forecasts of air-polluting gas percentages using the particle swarm optimization method. The accuracy value for the number of weeks given above is used in conjunction with the error measures MSR and RMSR. This approach handles missing values and standardizes the dataset to improve solution performance.

Strategy of Evaluation

We employed particle swarm optimization to forecast the percentage of air-polluting gases over 10 weeks to assess the efficacy of the proposed system upgrade. The three most basic metrics for evaluating unsupervised performance are accuracy, RMSR, and MSR.

METHODOLOGY OF PARTICLE SWARM OPTIMIZATION ALGORITHM (PSO)

A "swarm" is a collection of components upon which an algorithm relies; these components are then randomly spread throughout a constrained region to find the best solution for this problem. To get things off, we will list and briefly describe some basic things about PSO:

Grouping is what a population does, and statistical methods are the backbone of this grouping.

In turn, the assembly is a collection of parts, or particles, that stand in for a collection of solutions in any system that employs the algorithm.

Drs created the 1995 model. The movement patterns of bird flocks or fish schools influenced Eberhart and Kennedy.

Several evolutionary computation approaches are comparable to PSO. Take "Genetic Algorithm GA" as an example; it mimics the way birds forage for food to find the best possible solution to a problem. Initially, each system using this approach will be constructed from a random pool of solutions. Then, updating generations will hunt for the best solution within this pool.

By following the best particles at any given moment, the PSO algorithm always seeks the optimal solution, much as bees and ants do.

Here are the main points of the representation of the search algorithm:

- Data that has been divided into training and test sets must be entered first.
- Part two involves finding the matrix that extracts specifications.
- After that, you may start optimizing the training data using the particle swarm technique.
- Step four involves setting the starting points for the positions and velocities, with the velocities initially set to zero.
- Step five in calculating inertia [19]:

$$\theta = \theta_{max} - \frac{\theta_{max} - \theta_{min}}{i_{max}} i \quad \ldots\ldots(1)$$

- Step six: Use the following equation to get the updated location and velocity for each particle:

▪ The rate of updates [19]:

$$u_i = \theta\, u_{i-1} + c_1 r_1 [x^* - x_{i-1}] + c_2 r_2 [g^* - x_{i-1}] \quad \ldots (2)$$

▪ Review current role [19]:

$$x_i = x_{i-1} + u_i \quad \ldots\ldots (3)$$

- In the seventh step, determine the new location based on f using the formula (check/Update: X*, g*).
- Verify convergence in step eight.

Suggested Technological Framework

1. Read the data set.
2. Managing data that is missing.
3. Data normalization.
4. Using the Particle Swarm Optimization (PSO) technique to forecast the amount of gases that pollute the air.
5. Determining the Mean Squared Error and ultimate error rate up to 10 weeks.

The proposed system's technical steps are shown in the figure below:

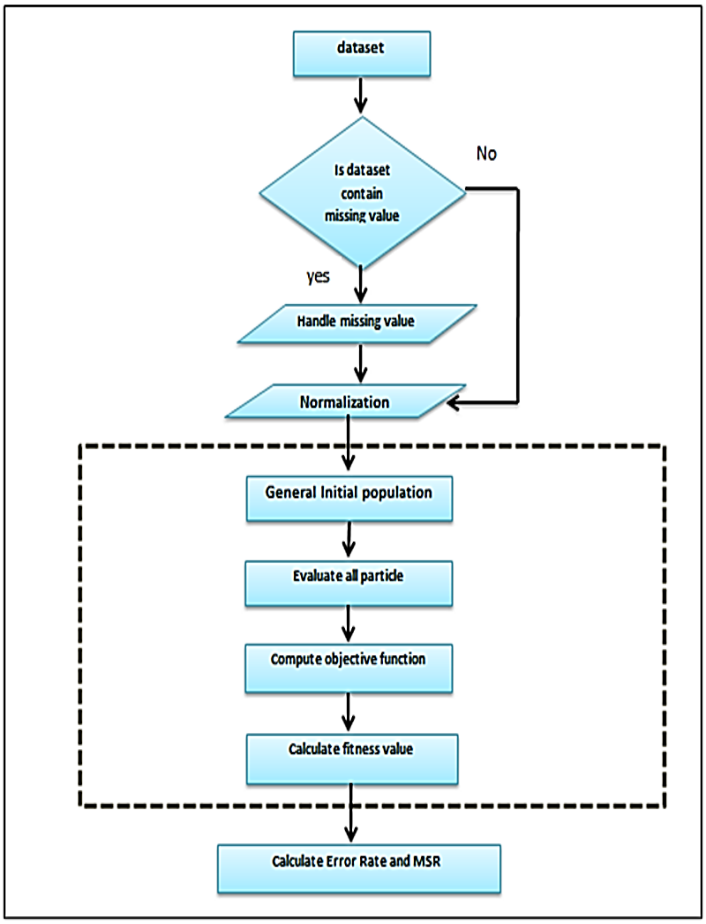

Fig. 2: The proposed system

DOI:https://doi.org/10.25130/tjps.v31i1.1933

Algorithm 1. Pseudo Code of Particle Swarm Optimization (PSO)

Input: N, x1, x2, c1, c2, imax, f
output: N-vector swarm S
Initialize Fix S to an initial value and then create X, the location of each particle relative to S, at random. A goal function's x1, xu limits;
Initialize Put all u-velocities to zero at the beginning;
Initialize Establish the starting points for the optimal locations x* (and their corresponding values) of the particles and determine g*;
For r1 and r2, pick two numbers at random from the interval [0,1];
Repetition i=0;
Set θmin and θmax to their initial values.
while i<imax do
  Determine the moment of inertia using equation (1);
  In iteration i, the values for all particles in S are:
   1. Recalculate the speed using the formula (2).
   2. Revise location in accordance with equation (3);
   3. Determine the new position's value using f:
   4. verify/update: X*, g*
  (Optional) verify convergence;
  Recursive optimization: i=i+1;
11. End
12. Return S

Valuation stage

After applying the particle swarm optimization algorithm to polluted nitrogen dioxide (NO2) gas in the city of Baghdad for the period (2015-2017), with a weekly average, we obtained predictive values for 10 weeks with an accuracy of 94%, as shown in Table 1. The absolute or relative error between the least and the best values found determines solution quality. In most cases, when looking for the minimum, the method produces a function value of zero, which is considered the optimal result. As the value decreases, the method's efficiency improves.

Table 1: Percentage of predictive values for NO2

Forecasts (predicting)	Standard Error	Upper	Lower
0.036501	0.1582275	0.0724417	0.018718644
0.03582	0.1659085	0.082695	0.017878281
0.03541	0.1668578	0.0826424	0.017344356
0.035021	0.1687117	0.0824484	0.017088834
0.034624	0.1719966	0.0822141	0.016745571
0.034407	0.1726444	0.0819857	0.016575517
0.034223	0.1729199	0.0817808	0.016454414
0.0341	0.1733749	0.0815984	0.016366845
0.034187	0.1734528	0.08146540	0.16302775
0.034128	0.1823417	0.0813498	0.016255440

The table above shows the predictive values for air pollution over 10 weeks, with an accuracy of 94%, where the air pollution prediction depends on the upper and lower bounds of NO2 levels.

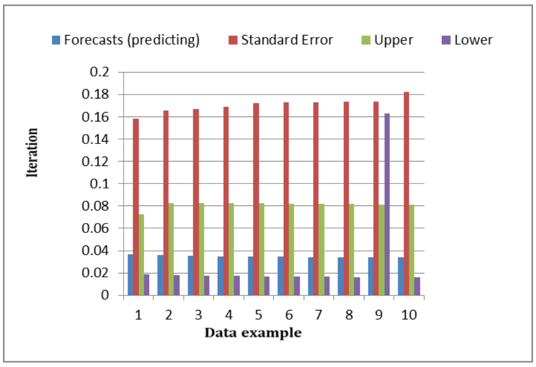

Fig. 3: Results of the PSO Algorithm

The figure above shows the results from Table 1 after applying the particle swarm optimization algorithm to predict NO2 gas concentrations.

Table 2: The values of MSR and RMSR

MSR	RMSR
0.02767	0.158352

To assess how well the predictions in this study performed, the following error measures were used, as shown in the table above. The likelihood of achieving the best outcome, defined here as the discrepancy between the expected and observed values, is proportional to the accuracy measure. The resultant squared error is the mean squared error, which is calculated by summing all the values used in the computation and dividing by the total number of data points.

CONCLUSION AND FUTURE WORK

Predicting the amount of air pollutants is done using the particle swarm optimization technique. The air pollutant data contained missing values, and readings were not recorded on official holidays, so they needed to be processed before analysis. These data also showed instability in the variance, which required processing this problem. The study recommends the necessity of building a central and integrated dataset (holiday registration) to record pollutant data in the Ministry of Health and Environment to obtain data reports and tables that are consistent with the needs of researchers, and to apply the models used in this research to pollutant data in other governorates on the integrated central dataset.

REFERENCES

1. Mannucci PM, Franchini M. Health effects of ambient air pollution in developing countries. International journal of environmental research and public health. 2017;14(9):1048. https://doi.org/10.3390/ijerph14091048

2. Moore FC. Climate change and air pollution: exploring the synergies and potential for mitigation in industrializing countries. Sustainability. 2009;1(1):43-54. https://doi.org/10.3390/su1010043

3. Eze IC, Schaffner E, Fischer E, Schikowski T, Adam M, Imboden M, et al. Long-term air pollution exposure and diabetes in a population-based Swiss cohort. Environment International. 2014;70:95-105. https://doi.org/10.1016/j.envint.2014.05.014

4. Lee KK, Bing R, Kiang J, Bashir S, Spath N, Stelzle D, et al. Adverse health effects associated with household air pollution: a systematic review, meta-analysis, and burden estimation study. The Lancet Global Health. 2020;8(11):e1427-e34. https://doi.org/10.1016/s2214-109x(20)30343-0

5. Kan H, Chen R, Tong S. Ambient air pollution, climate change, and population health in China. Environment International. 2012;42:10-9. https://doi.org/10.1016/j.envint.2011.03.003

6. Jansakoo T, Sekizawa S, Fujimori S, Hasegawa T, Oshiro K. Benefits of air quality for human health resulting from climate change mitigation through dietary change and food loss prevention policy. Sustainability Science. 2024;19(4):1391-407. https://doi.org/10.1007/s11625-024-01490-w

7. Yadav D. Machine learning: Trends, perspective, and prospects. Science. 2020;349:255-60. https://doi.org/10.1126/science.aaa8415

8. Lv L, Wei P, Li J, Hu J. Application of machine learning algorithms to improve numerical simulation prediction of PM2. 5 and chemical components. Atmospheric Pollution Research. 2021;12(11):101211. https://doi.org/10.1016/j.apr.2021.101211

9. Utku A, Can Ü. Deep learning based air quality prediction: A case study for London. Türk Doğa ve Fen Dergisi. 2022;11(4):126-34. https://doi.org/10.46810/tdfd.1201415

10. Pérez-Carrasquilla JS, Montoya PA, Sánchez JM, Hernández KS, Ramírez M. Forecasting 24 h averaged PM2.5 concentration in the Aburrá Valley using tree-based machine learning models, global forecasts, and satellite information. Adv Stat Clim Meteorol Oceanogr. 2023;9(2):121-35. https://doi.org/10.5194/ascmo-9-121-2023

11. Hu J, Ostro B, Zhang H, Ying Q, Kleeman MJ. Using chemical transport model predictions to improve exposure assessment of PM2. 5 constituents. Environmental Science & Technology Letters. 2019;6(8):456-61. https://doi.org/10.1021/acs.estlett.9b00396

12. Mehta P, Bukov M, Wang C-H, Day AG, Richardson C, Fisher CK, et al. A high-bias, low-variance introduction to machine learning for physicists. Physics reports. 2019;810:1-124. https://doi.org/10.1016/j.physrep.2019.03.001

13. Kothandaraman D, Praveena N, Varadarajkumar K, Madhav Rao B, Dhabliya D, Satla S, et al. Intelligent forecasting of air quality and pollution prediction using machine learning. Adsorption Science & Technology. 2022;2022:5086622. https://doi.org/10.1155/2022/5086622

14. Mohammed M.S. Talib H. A. Using machine learning algorithms in intrusion detection systems: A review. Tikrit Journal of Pure Science. 2024;29:3. https://doi.org/10.25130/tjps.v29i3.1553

15. Maaeda Mohsin R. K-means Clustering, Unsupervised Classification, K-NN, Euclidean Distance, Genetic Algorithm. Tikrit Journal of Pure Science. 2023;22(9):113-7. https://doi.org/10.25130/tjps.v22i9.884

16. Sun S, Cao Z, Zhu H, Zhao J. A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics. 2019;50(8):3668-81. https://arxiv.org/pdf/1906.06821

17. Zubair ZRS. A deep learning-based optimization model for the computer interface of wheelchair directional control. Tikrit Journal of Pure Science. 2021;26(1):108-12. https://doi.org/10.25130/tjps.v26i1.107

18. Chaabene WB, Flah M, Nehdi ML. Machine learning prediction of mechanical properties of concrete: Critical review. Construction and Building Materials. 2020;260:119889. https://doi.org/10.1016/j.conbuildmat.2020.119889

19. Abed SN. Predicting the Optimal Treatment for Diseases Using the Genetic Method by the Developed (PSO) Optimization Technique. Journal of Al-Qadisiyah for Computer Science and Mathematics. 2024;16(2).