Forecasting Atmospheric Air Pollution in Tehran Using Random Forest Model

Nilufar Makky[1,], Zahra Eghrari[2]*

[1] RSGIS Department., University of Tabriz, Tabriz, Iran, rsgis.nilufarmakki73@gmail.com
[2] GIS Department, College of Eng, University of Tehran, Tehran, Iran, zahraeqrari@ut.ac.ir
* corresponding author
doi: 10.5281/zenodo.11584441

Abstract: Air pollution poses a significant threat to human health, particularly for those with respiratory issues. Since 2018, the Sentinel-5 TROPOMI has been identified as the most effective tool for monitoring air pollution. We conducted an assessment of environmental factors such as precipitation, NDBI (Normalized Difference Built-up Index), temperature, and population to understand the impact of air pollution in megacities, especially in Tehran, from 2020 to 2024. Our study focused on atmospheric pollutants including CO, O_3, Aerosol, NO_2, and SO_2. Using the Random Forest model, we predicted pollutant concentrations and achieved an R2 value exceeding 0.90 for all parameters. Recognizing and understanding these environmental factors is crucial for effective government crisis management.

Keywords: Sentinel-5; air pollution; precipitation; temperature; NDBI; population growth; random forest model.

1 Introduction

Urban populations surged from 751 million in 1950 to 4.2 billion in 2018, dramatically altering city landscapes and exacerbating atmospheric air pollution (Jonson et al. 2020, Glencross et al. 2020). Air pollution's detrimental effects on human health and the environment highlight the urgency of addressing this issue (Glencross et al. 2020, Tainio et al. 2021). Major contributors include emissions from automobiles, aviation, industrial activities, mining, fossil fuel use, and agriculture (Colvile et al. 2001, Ciais et al. 2020). Urban expansion, land use changes, increased traffic, and industrial emissions significantly impact air quality (Kastratović 2019). Air pollution causes millions of premature deaths annually, underscoring the need for effective management strategies (Giles-Corti et al. 2016). Rising temperatures and natural disasters further emphasize the importance of addressing climate challenges (Boer and Hendrix 2000, Kaufmann et al. 2011, Eghrari et al. 2023, Kamran et al. 2023, Makky et al. 2023, Anggraini et al. 2024). Nitrogen dioxide (NO_2) is a critical pollutant causing significant health and ecological damage (Damtoft et al. 2008). Satellite observations, such as TROPOMI, provide essential insights into NO_2 levels and highlight the pandemic's effects on air quality (Duprè et al. 2010). Fossil fuel emissions and pollutants like ozone and aerosols exacerbate urban air pollution and climate change (Li et al. 2021). Cities face significant challenges related to energy consumption and air pollution, with transportation being a major factor. Measures like low emission zones have had mixed results in reducing NO_2 levels (Mojtehedzadeh 2019, Eghrari et al. 2023). In Tehran, addressing transportation and industrial impacts on air pollution is crucial, with research focusing on alternative transportation modes and satellite data to predict pollutant patterns (Fernandez-Moran et al. 2021). In this study, we aim to predict the concentrations of five air pollution parameters to identify the regions where each type of pollution poses the greatest risk to human health.

2 Materials and methods

Using the Google Earth Engine cloud-based platform streamlines the analysis of the intricate interplay between environmental factors and the Earth's surface. This study delves into the dynamic relationship between key environmental factors precipitation, temperature, population growth, and Normalized difference built-up index, additionally pollutants such as CO, NO_2, SO_2, Aerosol, and O_3 within Tehran city. The European Space Agency's Sentinel-5 Precursor (Sentinel-5P) is a crucial Earth Observation satellite equipped with the innovative TROPOMI sensor, designed to bridge the gap between the Envisat mission and the upcoming launch of Sentinel-5 (UN). The TROPOMI sensor, a nadir-viewing imaging spectrometer, plays a vital role in monitoring air pollution by measuring electromagnetic

spectral wavelengths from UV to SWIR spectrum. With a

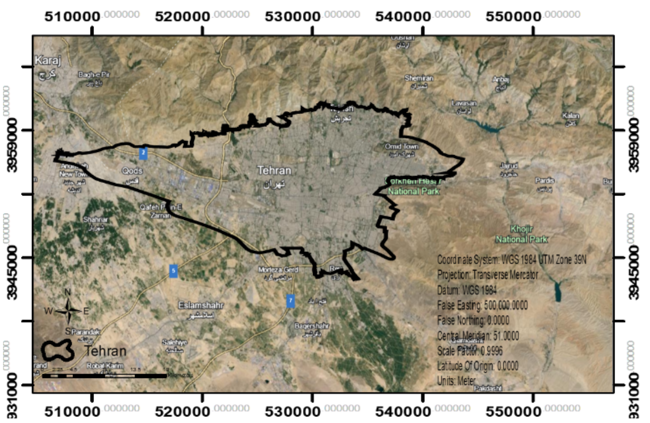

Figure 1. Overview of the study area.

push-broom configuration and a wide swath of 2600 km on the ground, the sensor ensures thorough coverage, albeit with varying pixel dimensions for different spectral bands. The sensor's Level-1B and Level-2 data products have been widely acclaimed for their utility in atmospheric pollution mapping and monitoring (Tabunschik et al. 2023). In our study, we used 60 images per month from 2020 to 2024 for each air pollution parameter. In addition, we obtained 48 images for each environmental factor. We then averaged these images to create one composite image per year for each environmental factor. Utilizing a robust Random Forest model, we precisely determined the concentrations of Atmospheric Pollutant Parameters (APP) in Tehran. This city Located at coordinates 51°19’ E, 35°41’ N, Tehran, the capital of Iran, boasts a unique geographical setting. Nestled amidst the Alborz Mountains in the north and bordered by a central desert in the south, the city exhibits distinct elevation and climate differences between its northern and southern regions (Zargari et al. 2024). Figure 1 and Figure 2 indicates the study area and methodology.

This approach allowed for the collection of diverse and accurate data to understand the complex relationship between environmental factors and pollution dynamics. Sentinel-5 data was beneficial for assessing pollutants and predicting air pollution in high-risk areas, such as those with vulnerable populations (Trenchev et al. 2023). Air and environmental pollution, resulting from natural and human activities, are impacted by seasonal variations affecting pollutant levels (Bekkar et al. 2021).

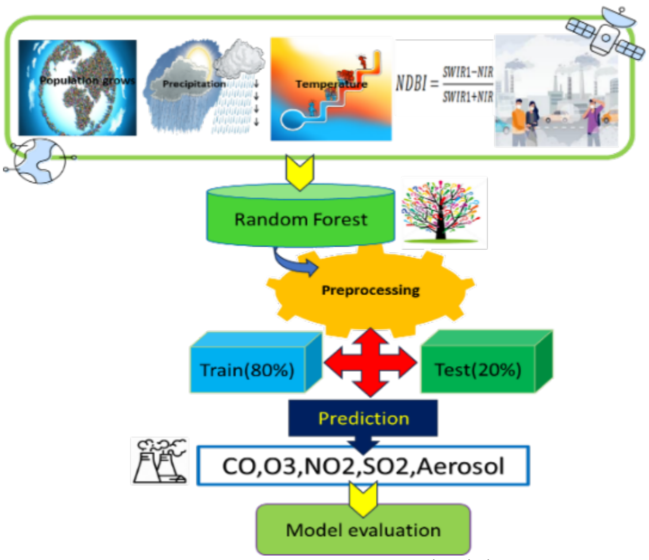

Figure 2. Data processing methodology.

3 Results

Environmental factors such as Precipitation, NDBI, Temperature, and population were presented in Figures 3–6. Furthermore, the Figures 7–11 indicate the most significant parts of the city facing various pollution parameters. Additionally, Figure 12 illustrates the correlation coefficient between pollution parameters and environmental factors.

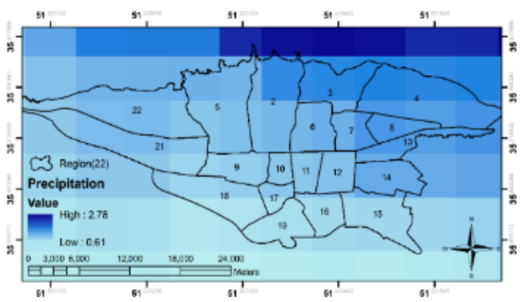

Figure 3. Precipitation values.

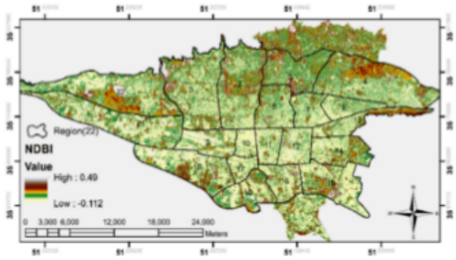

Figure 4. NDBI values.

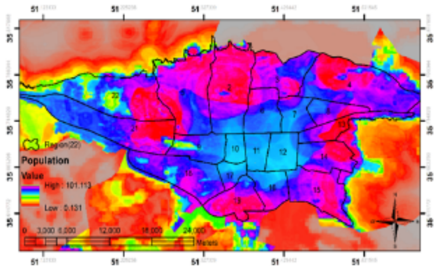

Figure 5. Population values.

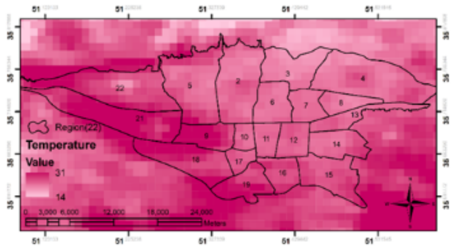

Figure 6. Temperature values.

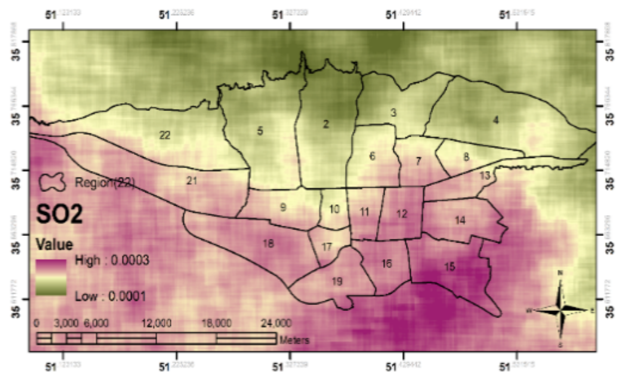

Figure 7. Illustrating predicted concentration of SO_2 in Tehran based on RF.

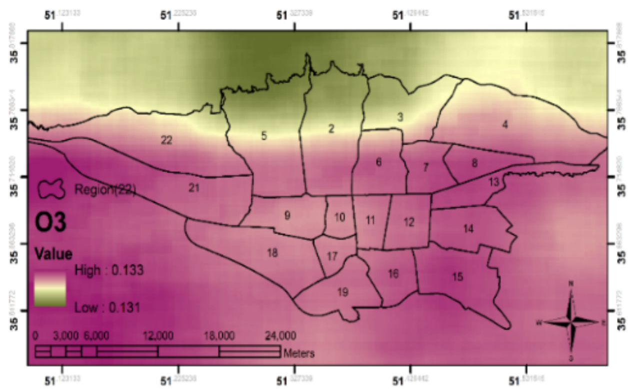

Figure 8. Illustrating predicted concentration of O_3 in Tehran based on RF.

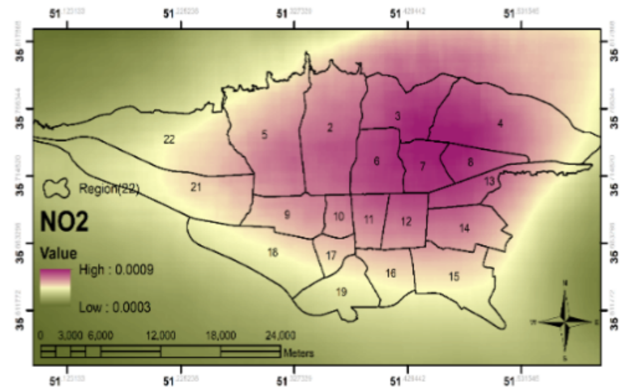

Figure 9. Illustrating predicted concentration of NO_2 in Tehran based on RF.

The Random Forest model (Khajavi and Rastgoo 2023) is applied to predict the Atmospheric Pollutant Parameters in Tehran (Table 1), all accuracy assessments are indicated.

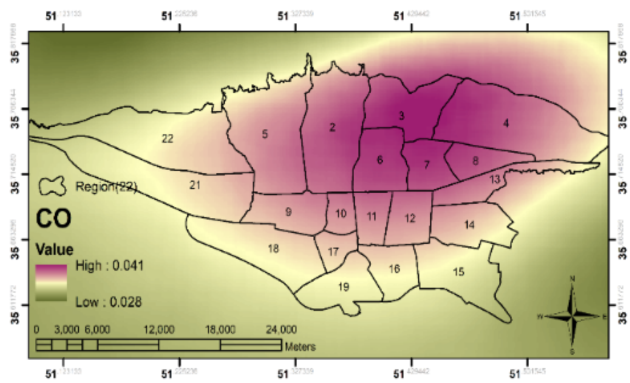

Figure 10. Illustrating predicted concentration of CO in Tehran based on RF.

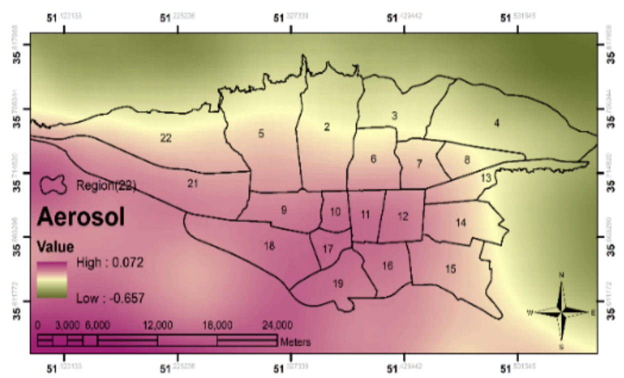

Figure 11. Illustrating predicted concentration of Aerosol in Tehran based on RF.

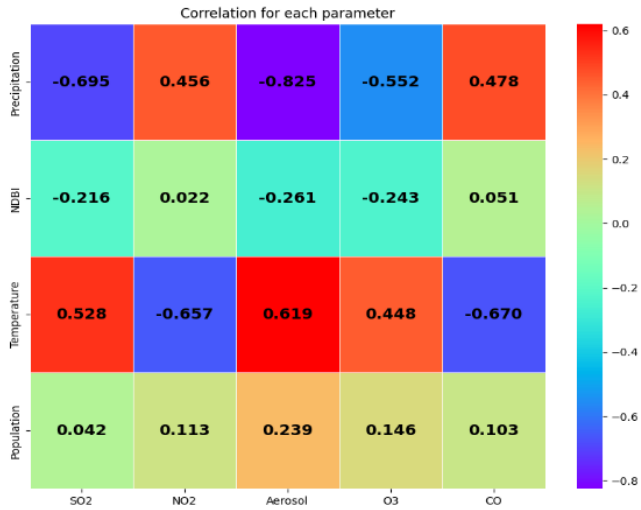

Figure 12. Cross-correlation between environmental factors and Atmospheric Pollutant Parameters.

4 Discussion

This study showed that the Random Forest model accurately predicts atmospheric pollutant concentrations in Tehran and revealed complex relationships between pollutants and environmental factors. Interestingly, precipitation was negatively correlated with aerosol and O_3 levels, indicating a cleansing effect. Positive correlations were found between temperature and pollutants, as well as population and pollutants, highlighting the impact of these factors on air quality.

5 Conclusions

The RF model was highly accurate in predicting APP concentrations in Tehran city using environmental factors and time series pollutant data from 2020 to 2024, with an R^2 value exceeding 0.90 for All, Train, and Test data. Correlation analysis revealed relationships between atmospheric pollutants (SO_2, O_3, NO_2, CO, aerosol) and environmental factors (Precipitation, NDBI, Temperature, Population). Precipitation showed a strong negative correlation with aerosol (−0.8253) and O_3 (−0.5524), suggesting a mitigating effect on these pollutants. Temperature had positive correlations with SO_2 (0.5282) and aerosol (0.6194), and negative correlations with NO_2 (−0.6567) and CO (−0.6698). Population positively correlated with all pollutants, indicating a link to urbanization and emissions. These findings enhance understanding of pollutant-environment relationships and emphasize the importance of considering multiple factors in air quality management.

Table 1. Random Forest (RF) Model Performance Metrics for Air Quality Predictions.

Parameter	All (R^2)	Train (R^2)	Test (R^2)	All (NRMSE)	Train (NRMSE)	Test (NRMSE)
CO	0.9569	0.9428	0.9276	0.0485	0.0685	0.0673
NO_2	0.9515	0.9510	0.9422	0.0245	0.0340	0.0662
SO_2	0.9330	0.9194	0.9177	0.0605	0.0886	0.0832
O_3	0.9648	0.9573	0.9443	0.0522	0.0851	0.0819
Aerosol	0.9033	0.9094	0.9059	0.0811	0.0999	0.1196

6 References

Anggraini, T.S., Irie, H., Sakti, A.D., Wikantika, K., 2024. Machine learning-based global air quality index development using remote sensing and ground-based stations. Environmental Advances 15, 100456.

Boer, E.P.J., Hendrix, E.M., 2000. Global optimization problems in optimal design of experiments in regression models. Journal of Global Optimization 18 (4), 385-398.

Ciais, P., Wang, Y., Andrew, R., Bréon, F.-M., Chevallier, F., Broquet, G., Nabuurs, G.-J., Peters, G., McGrath, M., Meng, W., ... 2020. Biofuel burning and human respiration bias on satellite estimates of fossil fuel CO2 emissions. Environmental Research Letters 15, 074036.

Colvile, R.N., Hutchinson, E.J., Mindell, J.S., Warren, R.F., 2001. The transport sector as a source of air pollution. Atmospheric environment 35, 1537-1565.

Damtoft, J.S., Lukasik, J., Herfort, D., Sorrentino, D., Gartner, E.M., 2008. Sustainable development and climate change initiatives. Cement and Concrete Research 38 (2), 115-127.

Duprè, C., Stevens, C.J., Ranke, T., Bleeker, A., Peppler-Lisbach, C.O.R.D., Gowing, D.J., Dise, N.D., Dorland, E., Bobbkink, R., Diekmann, M., 2010. Changes in species richness and composition in European acidic grasslands over the past 70 years: The contribution of cumulative atmospheric nitrogen deposition. Global Change Biology 16, 344-357.

Eghrari, Z., Delavar, M.R., Zare, M., Beitollahi, A., Nazari, B., 2023. Land subsidence susceptibility mapping using machine learning algorithms. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-4/W1-2022, 129-136.

Eghrari, Z., Delavar, M.R., Zare, M., Mousavi, M., Nazari, B. and Ghaffarian, S., 2023. Groundwater level prediction using deep recurrent neural networks and uncertainty assessment. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. X-1/W1-2023. 493-500.

Fernandez-Moran, R., Gómez-Chova, L., Alonso, L., Mateo-García, G., López-Puigdollers, D., 2021. Towards a novel approach for Sentinel-3 synergistic OLCI/SLSTR cloud and cloud shadow detection based on stereo cloud-top height estimation. ISPRS Journal of Photogrammetry and Remote Sensing 181, 238-253.

Giles-Corti, B., Vernez-Moudon, A., Reis, R., Turrell, G., Dannenberg, A.L., Badland, H., Foster, S., Lowe, M., Sallis, J.F., Stevenson, M., Owen, N., 2016. City planning and population health: a global challenge. The Lancet 388 (10062), 2912-2924.

Glencross, D.A., Ho, T.R., Camina, N., Hawrylowicz, C.M., Pfeffer, P.E., 2020. Air pollution and its effects on the immune system. Free Radical Biology and Medicine 151, 56-68.

Jonson, J.E., Gauss, M., Schulz, M., Jalkanen, J.-P., Fagerli, H., 2020. Effects of global ship emissions on European air pollution levels. Atmospheric Chemistry and Physics 20, 11399-11422.

Kamran, K.V., Makky, N., Charandabi, N.K., 2023. Investigating the flooded area of Bangladesh by Sentinel_1 and CHIRPS images in the GEE system. Intercontinental Geoinformation Days (IGD) 6, 83-88, Baku, Azerbaijan.

Kastratović, R., 2019. Impact of foreign direct investment on greenhouse gas emissions in agriculture of developing countries. Australian Journal of Agricultural and Resource Economics 63, 620-642.

Kaufmann, R.K., Kauppi, H., Mann, M.L., Stock, J.H., 2011. Reconciling anthropogenic climate change with observed temperature 1998–2008. Proceedings of the National Academy of Sciences 108 (29), 11790-11793.

Li, M., Liu, W., Bi, M., Cui, Z., 2021. Dynamic substance flow analysis of lead in the fossil fuel system of China from 1980 to 2018. Journal of Cleaner Production 313, 127918.

Makky, N., Valizadeh Kamran, K., Karimzadeh, S., 2023. Impact of Global Warming on water height using machine learning algorithms. Proceedings of the 5th International Electronic Conference on Remote Sensing, 7-21 November 2023, MDPI: Basel. Switzerland.

Mojtehedzadeh, M., 2019. Assessment of Urban Transport System in Tehran. Suti Report; United Nations Economic and Social Commission for Asia and Pacific (ESCAP): Bangkok, Thailand. Available online: https://www.unescap.org/ sites/default/files/Tehran.pdf (Accessed on 17 July 2019).

Tainio, M., Andersen, Z.J., Nieuwenhuijsen, M.J., Hu, L., de Nazelle, A., An, R., Garcia, L.M., Goenka, S., Zapata-Diomedi, B., Bull, F., ... 2021. Air pollution, physical activity and health: A mapping review of the evidence. Environment international 147, 105954.

Zargari, M., Mofidi, A., Entezari, A., Baaghideh, M., 2024. Climatic comparison of surface urban heat island using satellite remote sensing in Tehran and suburbs. Scientific Reports 14 (1), 643.