ISSN: 2959-6386 (Online), Volume 3, Issue 2, June 2024

Journal of Knowledge Learning and Science Technology

Journal homepage: https://jklst.org/index.php/home

Assessing the Efficacy of Air Pollution Mitigation in Beijing: Insights from GIS and Machine Learning Analyses

Yue Wang (1), **Jingbo Zhang** (2), **Yueshu Zhao** (3), **and Xinrui Li** (4)

[1] *Independent Researcher, Mexico City, Mexico 06500*
[2] *Independent Researcher, Arlington, VA 22202*
[3] *Independent Researcher, Washington D.C., USA 20431*
[4] *Independent Researcher, New York, USA 10065*

Abstract

This paper examines Beijing air pollution treatment through examining the Aerosol Optical Depth from 2013 to 2019, and pollutants density change in the region of Beijing city, Tianjin city, and Hebei province, using remote sensing products, i.e., aerosol optical depth from MODIS and the Sentinel- 5P SO2 and NO2 density products in Google Earth Engine and QGIS. The study finds that the whole region experienced an AOD decrease. And the patterns suggest that Beijing didn't shift its pollution to the surrounding area. However, Beijing city's SO2 and NO2 density still remain high in the region. Additionally, a machine learning model is introduced to further understand the dynamics of PM2.5 in Beijing.

Keywords: Remote sensing, Sentinel-5P, MODIS, Aerosol Optical Depth, Air pollution, Beijing, Machine Learning, Logistic Model

Article Information:
Received: 20-Janc-24 *Accepted:* 27-March-24 *Online:* 08-April-24 *Published:* 25-June-24
DOI: https://doi.org/10.60087/jklst.vol3.n2.p96

[i] **Correspondence author:** Yuanzhou Wei **Email:** joewei011@gmail.com

Introduction

Beijing's air quality has improved in recent years. Quote from a UN environment report in 2019, "In the five years from 2013 to 2017, fine particulate pollution (PM2.5) in Beijing fell by 35% and by 25% in surrounding regions." The data and effects are convincing because the Winter Olympics 2022 will take place in the region (People's Daily Online 2014). But the method remains questionable. Beijing has a history of shutting about 10,000 factories in the surrounding area in order to have 11-days of blue-sky during the APEC conference in 2014 (Larson 2014). So here comes the question, did Beijing shift the pollution to other areas? Machine learning has attained paramount importance in generating robust research outcomes and finds application across diverse domains, including malware detection (Li et al., 2024). To complement the examination of remote sensing data, this study incorporates a machine learning model to gain deeper insights into the dynamics of PM2.5 levels in Beijing.

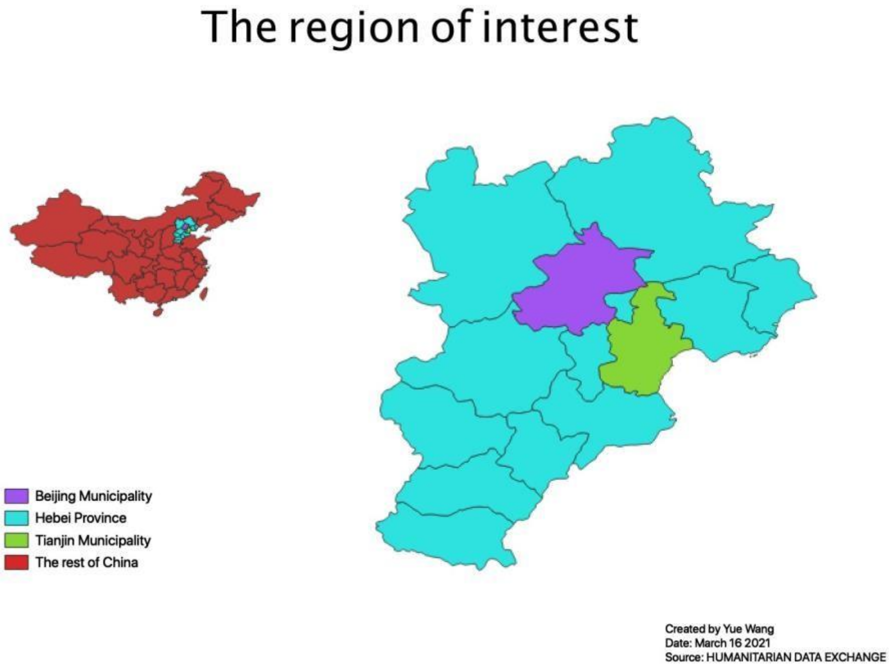


Figure 1 My region of interest (ROI) consists of Beijing Municipality, Hebei Province, Tianjin Municipality. The winter Olympics 2022 will take place in the circled area.

Conceptual Framework

Based on this question, I come up with two hypotheses.

***H0**: If Beijing treated the pollution effectively, we should observe that decreased the pollution density across years in surrounding areas.*

***H1**: If Beijing just shifted pollution, we should observe a higher pollutant density in surrounding areas.*

I attempt to test these hypotheses through processing remote sensing data in Google Earth Engine and visualizing them in QGIS.

Data

The main pollutions of Beijing are sulfur dioxide (SO2), nitrogen dioxide (NO2), and particulate matter (e.g., PM2.5). To assess the pollutant density change, I have chosen MODIS (known as Moderate Resolution Imaging Spectroradiometer) and Sentinel-5P products. MODIS is a multifunctional satellite platform of NASA since 1999. It offers a satellite product called Aerosol Optical Depth (AOD). It captures the amount of direct sunlight that is an obstacle from reaching the ground by aerosol particles. Many researchers have used AOD to predict PM2.5 concentration because it correlated with PM2.5 (Filonchyk 2019). Xu et al. discovered that Beijing city has a "southeast high and northwest low" PM2.5 distribution using AOD and multiple other indexes (2020). Sentinel-5P is a European satellite product issued in October 2017. It offers air quality-related remote sensing data, e.g., sulfur dioxide, nitrogen dioxide data.

In our machine learning analysis, we utilized satellite-derived PM2.5 data provided by the Atmospheric Composition Analysis Group at Washington University in St. Louis.

(*Satellitederived PM2.5 | Atmospheric Composition Analysis Group | Washington University in St. Louis*, n.d.)

Methods

Due to the time limitation of the Sentinel-5P, I use “MODIS Terra & Aqua MAIAC Land Aerosol Optical Depth Daily 1 km” for my main analysis. Later, I use “Sentinel-5P OFFL NO2: Offline Nitrogen Dioxide” and “Sentinel-5P OFFL SO2: Offline Sulphur Dioxide” for a complimentary analysis.

Main analysis method

To assess the PM2.5 change over a long period, I chose 2013 as my starting year and 2019 as the ending year for a sharp contrast. To obtain this dataset, I created an image collection variable by importing MODIS AOD data and selecting the “OD_055” band (Green band (0.55 μm) aerosol optical depth overland). Then, I use date filters to create the 2013 and 2019 image collection variables. To remove the noise caused by the cloud, I applied a mean reducer and obtained two images. Finally, I subtract the 2013 image from the 2019 image and exported the difference as a raster layer through my ROI geometry[1].

In QGIS, I first imported and selected my ROI administration boundary shape fine. Then, I imported the 2019-2013 difference raster layer from GEE. Finally, I used the zonal statistic tool to obtain the mean of the change at city levels and area levels.

Complementary analysis method

Regarding the Sentinel-5P NO2 and SO2 data, I first created annual column density image collections separately for the years 2018, 2019, 2020. Using these image collections, I plotted the UI day-by-year series chart. Finally, I computed the image for each year using a mean reducer and exported the image for zonal statistical analysis in QGSI.

Machine Learning analysis method

In addition to the remote sensing data analysis, a machine learning approach was applied to predict and understand the dynamics of PM2.5 levels in Beijing. A logistic model was chosen due to the observed S-shaped trend in Beijing's PM2.5 levels from 2013 to 2022.

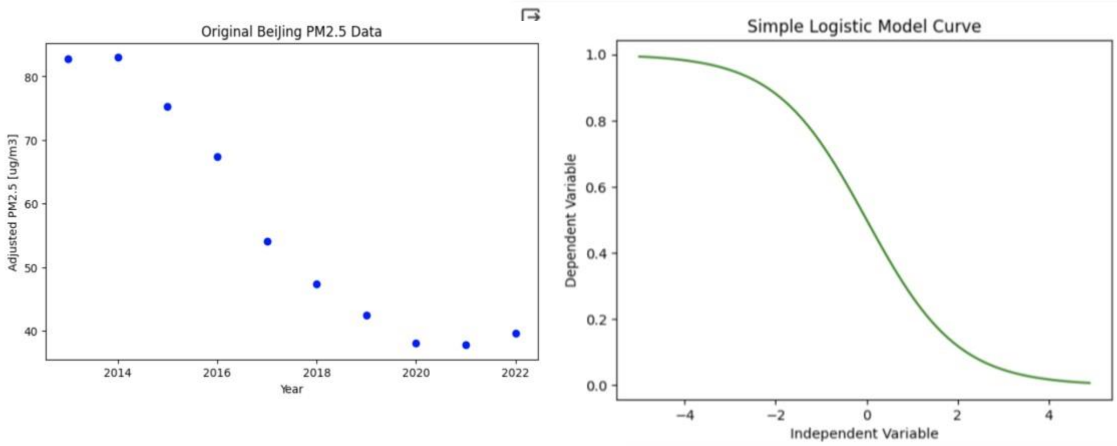


Figure 2. Plotting Beijing PM2.5 Data with Yea.

1 My ROI geometry layer is made from China administration boundary shapefile.

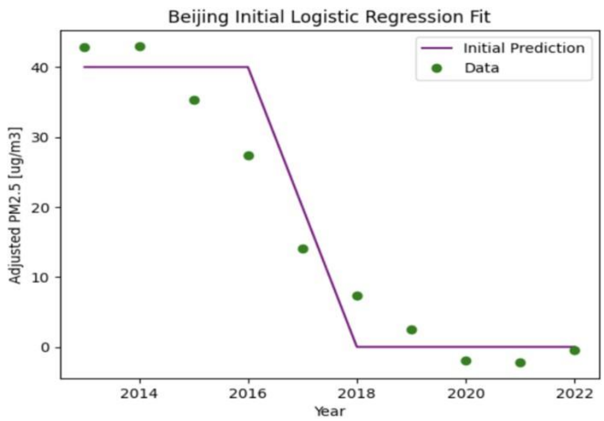
There is a clear S curve. More specifically, Beijing experienced severe air pollution during 2013 and 2014, and then PM2.5 dropped significantly during 2015 to 2020. After 2020, the level of PM2.5 in Beijing seemed the logistic model, chosen for its similarity to the observed "S" shape in PM2.5 data, estimates the probability of the dependent variable based on influencing independent variables (Cui & Zhao, 2019). The formula for the simple logistic model is as follows:

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

where μ is a location parameter and s is a scale parameter. And its visualization is the picture below (L. Liu et al., 2010).

Initially, we define a custom sigmoid function based on the logistic regression function. Estimation of the location and scale parameters is essential before inputting the function into machine learning libraries. A rough estimation is obtained through visualization. Next, we normalize the data by dividing both the Year and adjusted PM2.5 by their respective maximum values, ensuring their ranges fall within 0 and 1.

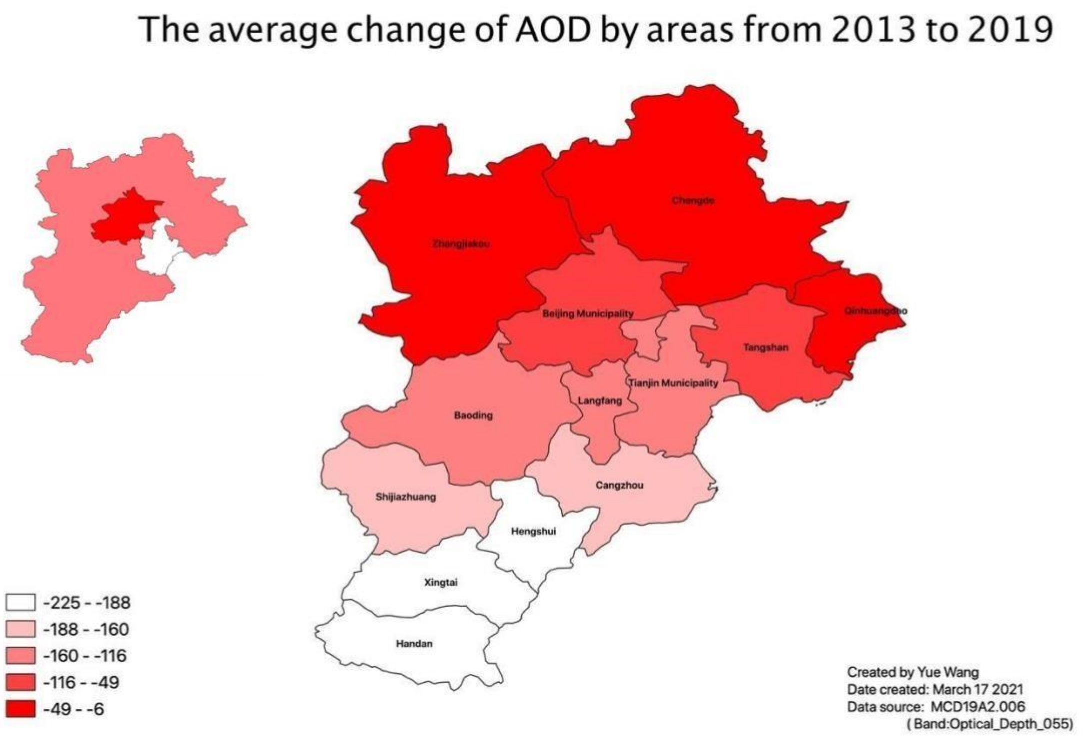
Normalization is a widely acknowledged practice to enhance the performance and stability of machine learning models (Cabello-Solorzano et al., 2023). Thirdly, we use *scipy.optimize.curve_fit* to fit our defined sigmoid function using non-linear least squares for best location parameter and scale parameter.[2]



Results

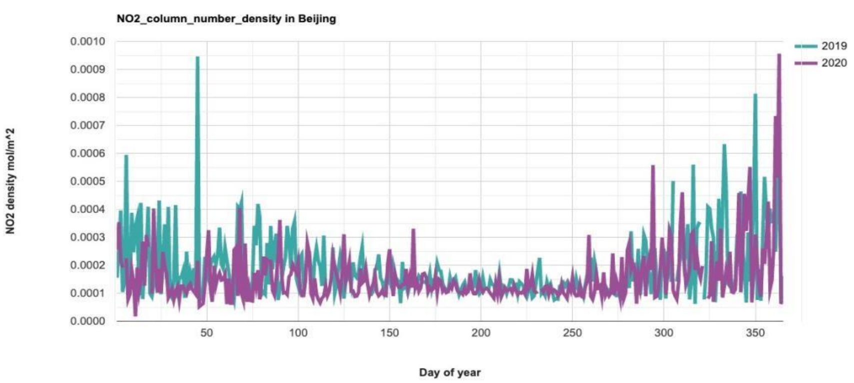
Main analysis result

2 scipy.optimize.curve_fit. Python. function.
https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

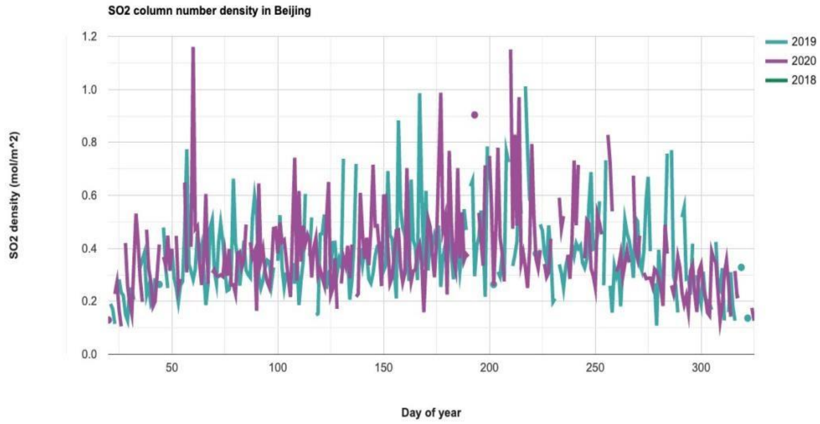


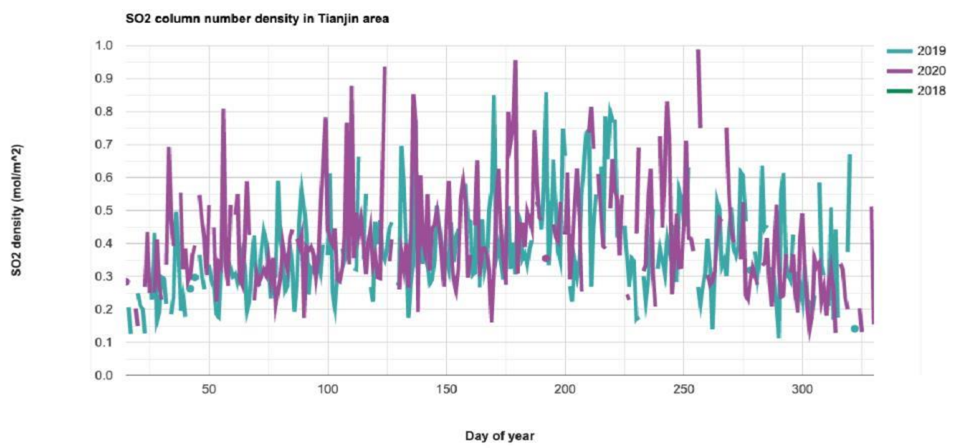
In general, the graph shows that the AOD decreased in the region because all changes are negative. The most significant decrease happened in Xingtai, Hengshui, Handan areas, which belong to Hebei Province. And these areas are far from Beijing city. Beijing's ADO decrease is consistent with the AOD decreasing in its surrounding areas, which is the opposite as the H1 suggested. Instead, there could be a positive spillover from the South to the North.

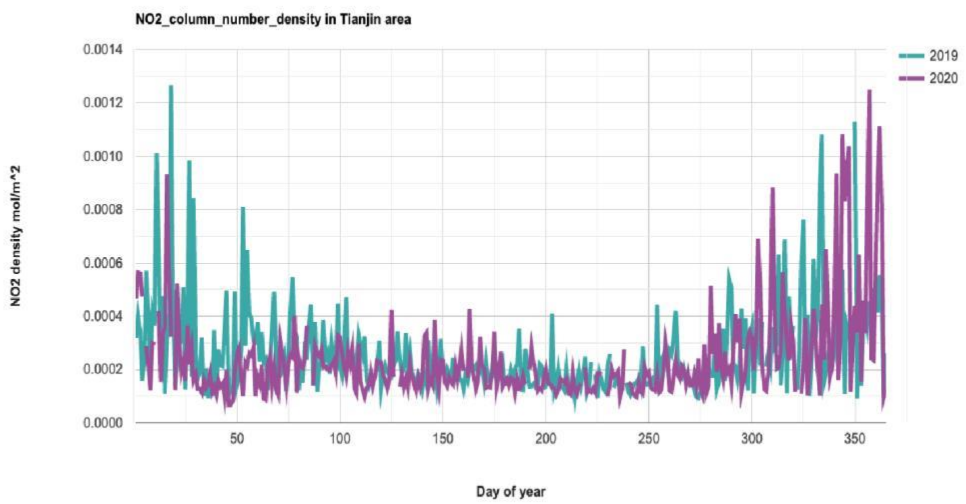
Complementary analysis result[3]



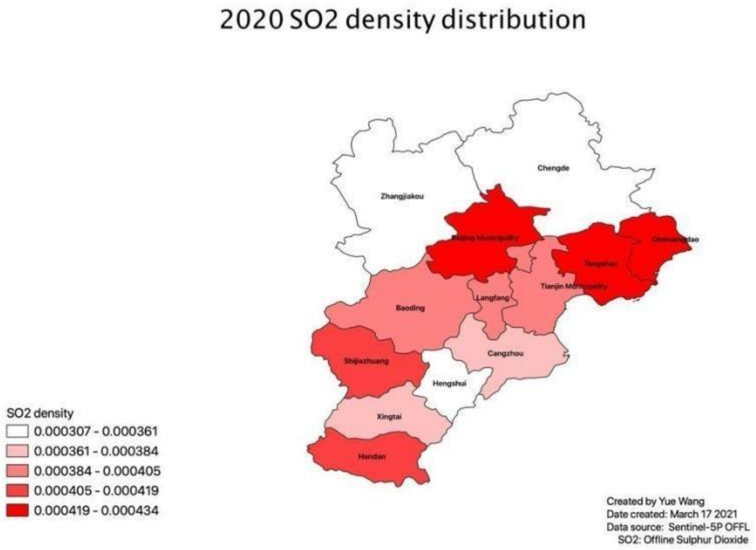
3 Hebei province is omitted because Google Earth Engine refused to compute, but the chart code worked for Tianjin and Beijing city.

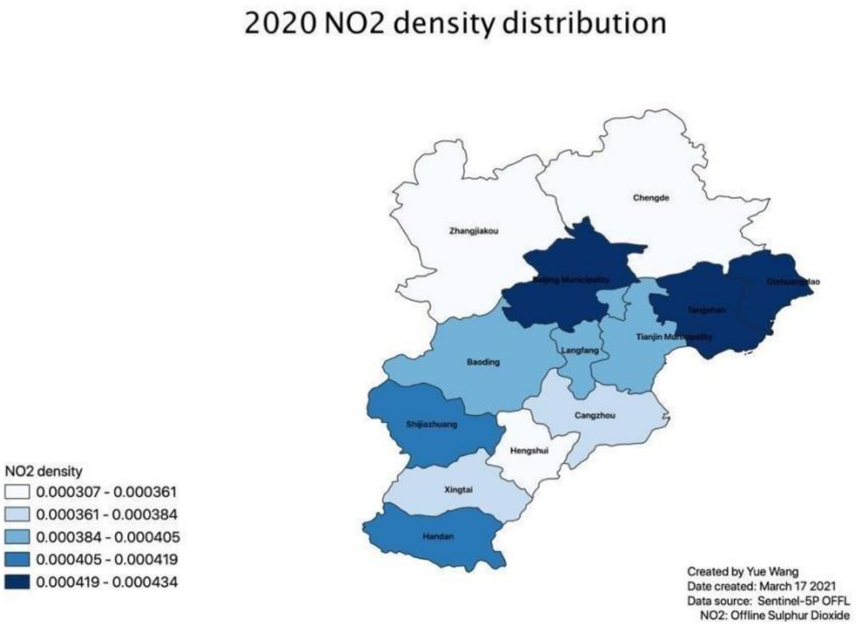






Based on the UI chart, NO2 and SO2 density didn't have significant changes from 2019 to 2020. And 2018 data is missing. Hence, I used 2020 data for visualization to present the current pollutants density level using QGIS. Even though the change across the year is not significant, it is worth mentioning that the NO2 density in the first six months in 2020 is observable lower than the same period in 2019. It could be caused by the COVID-19 lockdown when people used less transportation. Based on SO2 graphs, I think that 2020 has a slightly average higher SO2 density than 2019, it could be caused by people using more electricity at home during lockdown since SO2 is from burning fusil fuel and burning fusil fuel is the main source of generating electricity.





Beijing and several towns in Hebei provinces have relatively high SO2 and NO2 densities. For example, the Handan area in Hebei province has relatively high SO2, NO2 densities, although it had a huge decrease in AOD from 2013 to 2019. These distributions are possibly related to local industries and urbanization levels.

Machine learning analysis results

The obtained R2 values for Beijing and Henan are 0.99 and 0.96, indicating that the model can account for 99% and 96% of the variability in PM2.5 values for Beijing and Henan, respectively.

Discussions and Conclusions

In conclusion, the integration of remote sensing data and machine learning models provides a comprehensive understanding of Beijing's air pollution dynamics. The combined evidence suggests that Beijing has not shifted its pollution to surrounding areas, but challenges persist with relatively high SO2 and NO2 density. The machine learning model offers valuable predictions for future PM2.5 levels and underscores the ongoing need for effective environmental policies.

Remote sensing conclusion:

While the results suggest that Beijing did not transfer its pollution to surrounding areas, they do not sufficiently support the efficiency of the pollution treatment policy, as evidenced by the persistently high SO2 and NO2 densities in Beijing and certain regions. Additionally, the study acknowledges limitations, such as the absence of an extensive data processing regimen. The use of primary AOD and Sentinel-5P SO2 and NO2 data with a mean reducer for cloud removal lacks the sophistication required for accurate analysis, including the consideration of factors like seasonality. Despite these limitations, the study offers valuable insights, prompting potential future research avenues. The examination of pollution shifting at the provincial level, encompassing more regions for counterfactual analysis, and investigating the drivers behind the observed decrease in AOD in the Southern part of the region emerge as promising areas for further investigation.

Machine Learning Analysis

The analysis of PM2.5 data spanning from 2013 to 2022 reveals that Beijing did not transfer pollution to Hebei, as indicated by the absence of intensified air pollution in Hebei since

2013. Moreover, the predictive capabilities of the logistic model, specifically tailored for Beijing, anticipate a stabilization of PM2.5 levels around 40 [ug/m3] in the absence of new external influences, such as environmental policies. Notably, this projected level surpasses the annual PM2.5 standard set by the US EPA at 9.0 [ug/m3], posing potential long-term respiratory health risks (National Ambient Air Quality Standards (NAAQS) for PM | US EPA, 2024).

Future Research Directions

The considerable impact of transportation on air pollution necessitates a comprehensive exploration of transportation-related variables employing statistical methodologies (X. Ma et al., 2020). It is imperative to incorporate traffic volume as a co-factor, with its measurement facilitated through the estimation of traffic queues using onramp metering techniques, as explicated by Liu et al. (2022). Moreover, a pivotal aspect of this inquiry involves capturing temporal variations in traffic volume, as advocated by X. Ma et al. (2023).

In the context of environmental analysis, the application of machine learning, particularly leveraging transformer, attention-related technologies (Lyu, Zheng, et al., 2022b)., and Graph Convolutional Network (Wang et al., 2023), proves instrumental in unveiling intricate relationships within satellite imagery This approach offers a sophisticated computational methodology, such as deep neural networks, for discerning nuanced patterns and associations inherent in the satellite data (Wang, Jin, et al., 2023).

Future research should delve into integrating machine learning techniques to augment the accuracy and predictive capabilities of air pollution models. This exploration can extend to investigating the interplay between environmental policies and pollution levels, utilizing advanced analytical tools such as graph convolutional networks and deep learning (Wei et al., 2023). Additionally, in the machine learning domain, the potential deployment of deep learning becomes feasible with a more granular and extensive dataset. Deep learning, wellestablished in various applications like recommendations (Wu & Chi, 2024), excels in its intelligence and capacity to discern intricate patterns. A larger volume of data would facilitate a more nuanced exploration, thereby enhancing the model's ability to extract meaningful insights.

Another salient factor warranting examination in the context of environmental dynamics pertains to the perceptions of residents. Employing sentiment analysis proves instrumental in discerning the nuanced perspectives of netizens regarding fluctuations in air quality across temporal intervals (Wu et al., 2024). Predominantly, such sentiment analysis methodologies leverage sophisticated computational models, notably Large Language Models (Wu, Xiang, et al., 2024), GenAI models (Xiang et al., 2024), BERT models (Pang et al., 2019), and

Multimodal Transformer (Lyu, Dong, et al., 2022b).

References

Cabello-Solorzano, K., De Araujo, I. O., Peña, M. A., Correia, L., & Tallón-Ballesteros, A. J. (2023). The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis. In *Lecture notes in networks and systems* (pp. 344–353). https://doi.org/10.1007/978-3-031-42536-3_33

Christina Larson. November 18, 2014. How did Beijing achieve "APEC blue"?. *Bloomberg*. Retrieved 17 March 2015

Cui, X., & Zhao, W. (2019). Study of the Modified Logistic Model of Chinese Electricity Consumption Based on the Change of the GDP Growth Rate under the Economic New Normal. *Mathematical Problems in Engineering, 2019*, 1–13. https://doi.org/10.1155/2019/3901821

Filonchyk, M., Yan, H., Shareef, T.M.E. et al. (2019). Aerosol contamination survey during dust storm process in Northwestern China using ground, satellite observations and atmospheric modeling data. *Theor Appl Climatol, 135*, 119–133. https://doi.org/10.1007/s00704-017-2362-8

Filonchyk, M., Yan, H., Zhang, Z. et al. (2019). Combined use of satellite and surface observations to study aerosol optical depth in different regions of China. *Sci Rep, 9*, 6174. https://doi.org/10.1038/s41598-019-42466-6

Li, Z., Zhu, H., Liu, H., Song, J. J., & Qian, C. (2024). Comprehensive evaluation of MalAPI-2019 dataset by machine learning in malware detection. *International Journal of Computer Science and Information Technology, 2*(1), 1–9. https://doi.org/10.62051/ijcsit.v2n1.01

Liu, L., Sankarasubramanian, A., & Ranjithan, S. R. (2010). Logistic regression analysis to estimate contaminant sources in water distribution systems. *Journal of Hydroinformatics, 13*(3), 545–557. https://doi.org/10.2166/hydro.2010.106

Lyu, W., Dong, X., Wong, R., Zheng, S., Abell-Hart, K., Wang, F., & Chen, C. (2022b). A Multimodal Transformer: Fusing Clinical Notes with Structured EHR Data for Interpretable In-Hospital Mortality Prediction. *arXiv (Cornell University)*, 2022, 719– 728. https://arxiv.org/abs/2208.10240

Lyu, W., Zheng, S., Ma, T., & Chen, C. (2022b). A study of the attention abnormality in trojaned BERTs. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* https://doi.org/10.18653/v1/2022.naacl-main.348

Ma, Xiaobo & Karimpour, Abolfazl & Wu, Yao-Jan. (2020). Statistical evaluation of data requirement for ramp metering performance assessment. *Transportation Research Part A: Policy and Practice, 141*(C), 248-261.

Ma, X., Karimpour, A., & Wu, Y. (2023). Eliminating the impacts of traffic volume variation on before and after studies: a causal inference approach. *Journal of Intelligent Transportation Systems*, 1–15. https://doi.org/10.1080/15472450.2023.2245327

Pang, N., Li, Q., Lyu, W., & Yang, J. (2019). Transfer Learning for Scientific Data Chain Extraction in Small Chemical Corpus with BERT-CRF Model. *arXiv (Cornell University)*. https://arxiv.org/pdf/1905.05615

Satellite-derived PM2.5 | Atmospheric Composition Analysis Group | Washington University in St. Louis. (n.d.). https://sites.wustl.edu/acag/datasets/surface-pm2-5/

UN Environment (2019). A Review of 20 Years' Air Pollution Control in Beijing. *United Nations Environment Programme*, Nairobi, Kenya.

National Ambient Air Quality Standards (NAAQS) for PM | US EPA. (2024, February 15). *US EPA*. https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqspm

Wang, Y., Zhao, Y., Zhang, Y., & Derr, T. (2023). Collaboration-Aware Graph Convolutional Network for Recommender Systems. *arXiv*. https://doi.org/10.1145/3543507.3583229

Wang, Y., Jin, X. & Castro, C. (2023). Accelerating the characterization of dynamic DNA origami devices with deep neural networks. *Sci Rep, 13*, 15196.

Wei, Y., Gao, M., Xiao, J., Liu, C., Tian, Y., & He, Y. (2023). Research and implementation of cancer gene data classification based on deep learning. *Journal of Software Engineering and Applications, 16*(06), 155–169. https://doi.org/10.4236/jsea.2023.166009

Wu, K., & Chi, K. (2024). Enhanced e-commerce customer Engagement: A Comprehensive Three-Tiered Recommendation System. *Journal of Knowledge Learning and Science Technology, 2*(3), 348–359. https://doi.org/10.60087/jklst.vol2.n2.p359

Wu, Y., Jin, Z., Chenxi, S., Liang, P., & Zhan, T. (2024). Research on the application of deep learning-based BERT model in sentiment analysis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.08217

Wu, Y., Xiang, Y., Huo, S., Gong, Y., & Liang, P. (2024). LORA-SP: Streamlined Partial Parameter Adaptation for Resource-Efficient Fine-Tuning of Large Language Models. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.08822

Yafei , X., Wu, Y. ., Song, J. ., Gong, Y., & Lianga, P. . (2024). Generative AI in Industrial Revolution: A Comprehensive Research on Transformations, Challenges, and Future Directions. *Journal of Knowledge Learning and Science Technology*, 3(2), 11-20. https://doi.org/10.60087/jklst.vol.3n2.p20