Yuchen Ji[1], Xiaonan Zhang[1], and Yueqian Cao[1]

[1]School of Transportation and Civil Engineering, Nantong University

August 05, 2024

Abstract

The impact of the COVID-19 lockdown on air quality in seven major Chinese cities was investigated by utilizing long-term datasets of air pollutants and meteorological conditions from 2016 to 2021. Generalized additive model (GAM) was developed to predict air quality during the lockdown period. The model accounting for weather conditions demonstrated high accuracy with predictions compared against measurements during the lockdown. Significant reductions in NO2, CO, and PM10 concentrations were observed primarily due to decreased vehicular traffic and industrial activities. Notable reductions were particularly evident in cities with high traffic volumes and industrial emissions prior to the lockdown. The study also employed transfer learning to enhance the accuracy of lockdown model with limited data. Despite occasional anomalies caused by specific events like fireworks and agricultural burning, the findings suggest that extended training periods and advanced modeling techniques can significantly improve air quality predictions. This research highlights the potential long-term benefits of sustained reductions in human activities and provides valuable insights for future air quality management and policy-making.

1. Introduction

Air quality is crucial for human health, e.g., PM_{10} (particulate matter smaller than 10 microns) can penetrate deep into the lungs, causing respiratory issues, blood disorders, and neurodevelopmental problems such as autism, attention deficit disorders, and cognitive delays (Neogi, 2023). Besides, air pollution negatively impacts cognitive function in the elderly and is linked to higher mortality rates (Lin et al., 2024). The economic costs of air pollution are substantial, underscoring the importance of reducing pollution (Lu et al., 2016), especially in densely populated urban areas. During the COVID-19 pandemic, long-term exposure to PM_{10} and nitrogen dioxide (NO_2) significantly increased human's susceptibility to the virus (Copat et al., 2020) and led to higher fatality rates among infected individuals (Zhang et al., 2023). Additionally, emerging evidence suggests that the virus can be detected in outdoor particulate matter, raising concerns about the transmission routes (Comunian et al., 2020). While further investigations are necessary to fully understand these dynamics, it is increasingly clear that air pollution played a crucial role in both the transmission and severity of COVID-19.

During the winter and spring of 2020, most cities in China implemented lockdown measures of various durations and strictness levels to combat the spread of the virus, which has been proved effective in significantly slowing the transmission of COVID-19 (Ahmed et al., 2024; Murphy et al., 2023). Concurrently, the widespread adoption of reduced mobility prompted researchers to examine the potential benefits of pollution reduction, which is particularly pertinent in the context of global warming, as these shifts in behavior and operations might offer a blueprint for reducing pollution while maintaining the essential functions of cities and nations. The pandemic inadvertently provided a unique opportunity to rethink and reimagine urban planning and environmental policies for a more sustainable future (Afrin et al., 2021; Sharifi and Khavarian-Garmsir, 2020).

Despite multiple studies have estimated pollution reductions during lockdowns across different countries (Briz-Redón et al., 2021; Fu et al., 2020; Ghahremanloo et al., 2021; Venter et al., 2021, 2020), these results often aggregate differences to various baselines without modeling the relationships between air pollution, local weather conditions, time variations, and land-use patterns (Tsilimigkas et al., 2016). On the other hand, the short duration of lockdowns resulted in a scarcity of representative data, complicating detailed analysis. Moreover, the strict initial lockdown measures in many countries were only implemented for a few weeks, making it difficult to establish effective models for the lockdown period. Addressing those challenge will help accurately measure the reduction in local pollution during the lockdown and understand its spatiotemporal variations, as well as predicting how pollution patterns might change if lockdowns occur in different seasons or are extended in duration.

In this paper, a long-term predictive model for the lockdown period (LD model) was developed to address the challenges associated with analyzing air quality during the lockdown period. Historical weather data preceding the lockdown were utilized to train interpretable pre-LD model based upon the generalized additive model (GAM), even with limited records available during the lockdown. Model predictions were then compared with actual measurements during the lockdown. Adjustments for environmental factors were made to enhance accuracy, and transfer learning was employed to refine parameters related to land use and daytime activities. The research approach and findings offer valuable insights into the potential impacts of lockdown measures on air quality in China.

2. Study area and datasets

2.1. Study area

The study area encompasses seven Chinese cities: Beijing, Changchun, Chongqing, Guangzhou, Hangzhou, Wuhan, and Xiamen. Boasting substantial populations, high levels of urbanization and robust economic prowess, these metropolises represent major urban centers across different geographical regions and experience significant daily traffic volumes leading to motor vehicle emissions that impact local air quality. Notably, Changchun is a prominent industrial hub in China where industrial emissions significantly contribute to urban air pollution. The onset of stringent lockdown measures in early 2020 resulted in a marked reduction in traffic flow and industrial emissions across various cities in China — providing an opportunity for our research endeavors. Consequently, these seven cities were strategically selected as representatives for studying the epidemic-induced lockdowns' influences on Chinese urban air quality.

2.2. Datasets

Training pre-LD and LD models utilized hourly datasets of air pollutants and meteorology conditions from January 1, 2016 to December 31, 2021 across the aforementioned seven Chinese cities with one site designated per city. The lockdown period is from January 23 to April 7, 2020. Air pollutant data, encompassing NO_2 concentration ($??/?^3$), CO concentration ($??/?^3$), and PM_{10} concentration ($??/?^3$), as well as meteorological records, including temperature (), pressure (hPa), dew point (), wind direction (deg), and wind speed (m/s), were both sourced from the China Meteorological Data Center.

3. Methods

To estimate air pollution levels in Chinese urbans unaffected by pandemic-related lockdown measures, the pre-LD and LD models should be with high accuracy and robust predictive power for data forecasting during the lockdown period, while also maintaining a high level of interpretability. A more interpretable model facilitates better comprehension of the impact of traffic patterns on air quality, enabling effective measures to mitigate air pollution through adjustments in relevant traffic flows. Previous studies (Dominici, 2002; John et al., 2011; Khaiwal et al., 2019) have demonstrated the success of GAM in predicting air pollution. Therefore, as illustrated by **Fig. 1**, this work adopted the GAM as the pre-lockdown module. Subsequently, a transfer learning mechanism was employed for training the LD model. Given variations in pollutant concentrations across different regions, certain parameters were adjusted and optimized to enhance prediction accuracy.

3.1. Pre-LD Model

Hastie and Tibshirani (1986) expanded the application of additive models (Stone, 1985) to include the Generalized Additive Model (GAM) as a versatile and flexible statistical tool for identifying non-linear regression effects:

where and represents a non-parametric smooth function such as a smooth spline function, kernel function, or local regression smooth function. The distribution of air pollutant concentration closely follows a lognormal distribution (Limpert et al., 2001). The non-parametric nature of GAM provides significant flexibility to the model and facilitates the elucidation of nonlinear effects arising from derived variables.

In the context of model selection, this study employs the forward selection method, which has been utilized in related fields of environmental science with promising outcomes (Ghasemi and Amanollahi, 2019). The model incorporates two key indicators: the Akaike Information Criterion (AIC) (Bozdogan, 1987) and the Variance Inflation Factor (VIF) (Akaike, 1974).

AIC serves as a standard for assessing the adequacy of statistical model fitting, which is defined as follows:

where is the number of model parameters, and denotes the maximum value of the model likelihood. A small indicates a parsimonious model, whereas a large signifies a precise model. AIC emphasizes the significance of data fitting while endeavoring to mitigate overfitting. Consequently, the preferred model for consideration is the one with the lowest AIC value.

VIF is a number that characterizes the degree of complex collinearity between observations of the independent variable as follows:

where marks the coefficient of determination for the regression analysis of the ?-th variable with all other explanatory variables. Multicollinearity is a linear or approximate linear relationship between regression variables. The general criteria of VIF are: 0 < VIF [?] 5, no multicollinearity; 5 < VIF [?] 10, weak multicollinearity; 10 < VIF [?] 100, moderate or strong multicollinearity; VIF > 100, severe multicollinearity. The VIF threshold is typically set at 2.5 for processing meteorological data (Senaviratna and A. Cooray, 2019). Thus, this study excluded variables with VIF > 2.5.

For each explanatory variable, a GAM containing only one variable is fitted, and the model with the lowest AIC was chosen. Subsequently, an iterative process was employed to identify the next optimal variable to be added to the existing model. In order to accommodate the weekly variations in pollutants resulting from the pandemic lockdown, the explanatory variable 'weekday' was artificially incorporated into cities where it had not been selected.

3.2 LD Model

The pre-LD model (GAM) was trained through the pre-LD data. Tied to the limited duration of the lockdown, it is unfeasible to employ the same GAM for training during this period; hence, transfer learning (Pan and Yang, 2010) was employed. Given the consistent relationship between weather conditions and air pollution, insights gained from studying how weather impacts air pollution prior to the lockdown can be leveraged during this period. In conducting experiments, it becomes imperative to adjust model variables, particularly with regard to the 'weekday' variable that signifies fluctuations in traffic intensity — a pivotal factor influencing air pollution.

4. Results and analysis

4.1 The impact of lockdown on air pollutants and weather

As depicted by **Figs. 2-4**, the concentrations of NO_2, CO, and PM_{10} in the seven cities during the 2020 lockdown were consistently lower than those recorded during the same period in 2019 (Kumari and Toshniwal, 2020; He et al., 2020). This reduction highlights the significant impact of decreased human activities, such as reduced vehicular traffic and industrial operations, on mitigating air pollution in Chinese urban areas. Particularly, there was a significant reduction in NO_2 levels in Beijing, Changchun, and Wuhan during the lockdown (**Fig. 3**). NO_2 primarily originates from automobile exhaust (Wang et al., 2022) and industrial emissions, and Beijing experienced prolonged and widespread traffic congestion (Bian et al., 2016). The substantial decrease in traffic volume during the lockdown resulted in a marked decline of NO_2 levels. Changchun's extensive industrial emissions were notably reduced tied to the lockdown measures, leading to a direct decrease in nitrogen dioxide emissions. In comparison to other cities, Wuhan's stringent lockdown measures significantly lowered its pollutant concentration. However, despite the overall decline in pollutant levels, there were notable exceptions. For instance, a marked increase in pollutant concentrations was observed in Beijing in 2020 mid-February. This anomaly can be attributed to the extensive use of fireworks during the Spring Festival, which significantly contributed to the surge in air pollutants (Zhang et al., 2017). Furthermore, low wind speeds during this period exacerbated the situation by trapping pollutants close to the ground, further intensifying air quality issues. Similarly, a significant rise in pollutant levels occurred in Changchun in 2020 early April was attributed to the burning of agricultural straw in rural areas surrounding the cities (Liu et al., 2020), which released substantial quantities of particulate matter and other pollutants into the atmosphere. The reduction in pollution levels during the lockdown provides a compelling argument for the potential long-term benefits of sustained reductions in human activities. Nevertheless, the occasional spikes in pollutant concentrations during specific events highlight the need for targeted air quality management strategies that address both routine and exceptional situations of pollution.

The comparison in **Fig. 5** illustrates the variation in weather elements between the pandemic lockdown period in 2020 and the corresponding period in 2019. Overall, a significant majority of cities experienced cooler temperatures during the lockdown in 2020 compared to the same period in 2019, attributed to a notable reduction in human activities that weakened the urban heat island effect (Liu et al., 2022). However, it is worth noting that these lower temperatures, slower wind speeds, and higher air pressure observed in 2020 are not conducive to pollutant diffusion.

4.2 Model predictions

The pre-LD model's predictions utilized data from training periods of 3, 6, 9, 12, 18, and 24 months prior to the lockdown, with subsequent testing against the next month's data. As shown by **Fig. 6**, the model's performance improves with longer training durations, achieving optimal results when trained on 24 months of data. This indicates that a longer training period allows the model to better capture complex temporal patterns, enhancing predictive accuracy. However, the increased RMSE of PM_{10} in Beijing, despite extended training, can be attributed to the frequent occurrence of sand and haze weather, leading to extreme PM_{10} values prior to the lockdown (Liu et al., 2023). The inclusion of these extreme values during the extended training period influenced the model's accuracy. Overall, training on 24 months of data generally provided better outcomes, as it offered a richer context for predicting future conditions. This is consistent with the statement that extended training periods promote model performance by incorporating more comprehensive historical data (Eklund and Kapetanios, 2008).

Due to the scarcity of data during the lockdown period, only three consecutive days of data were used as the test set in the experiment, and the remaining data during the lockdown were used to train the LD model. **Table 1** provides a summarized performance of pre-LD and LD models in predicting air pollutant concentrations (NO_2, CO, and PM_{10}) across seven cities. Among the seven cities, Xiamen's coastal location, high temperatures, strong wind speeds, and open terrain promote the dispersion of air pollutants. The city also experiences low industrial emissions and maintains smooth traffic year-round. Consequently, Xiamen's air quality remains generally high on non-lockdown days (Lin et al., 2023). As a result, the level of air pollution blocked by the epidemic decreased less than usual, and the predictions of the pre-LD and LD models are more accurate than that of the other six cities. Generally, the RMSE of CO exhibits a significantly lower value against the ones of other two pollutants and there was an average reduction of 14% for CO, 34% for NO_2, and 28% for PM_{10} across the seven cities, after comparing observations from the 2020 lockdown with those during the corresponding period in 2019. In contrast to the pre-LD model, the LD one estimated a reduction of 44%, 32%, and 30% in the average RMSE for NO_2, CO, and PM_{10} predictions in the seven cities, respectively.

Similarly, **Figs. 7-9** support this behavior, proving that the LD model predicted more accurately than the pre-LD model did. The discrepancy between predicted and actual values is notably reduced with the LD model, indicating enhanced predictive accuracy. This improved performance can be attributed to the transfer learning that is crucial for enhancing prediction accuracy. However, **Fig. 7** reveals a significant increase in CO concentrations during the 2020 winter in Beijing and Changchun, leading to considerable deviations between the LD model's predictions and actual observed values. This discrepancy is likely because of higher CO emissions resulted from increased heating activities in northern cities during the winter (Fan et al., 2020; Liu et al., 2022). Moreover, pollution levels were generally higher in 2019 than 2020, which can be explained by the higher emissions of air pollutants resulted from much larger traffic volumes prior to the lockdown in the seven cities. In contrast, the lockdown period, with its restrictions on movement and reduced industrial activity, contributed to a significant promotion in overall air quality.

Table 2 presents the agreement between the LD model-predicted NO_2 reduction and the measurements from satellite images. The model's predictions fall within a reasonable range, with those for Beijing and Wuhan aligning closely with observations from both satellites. Discrepancies in observations explain Chongqing and Guangzhou predictions being consistent with one of the satellites, which is understandable. This comparison highlights the LD model's predictive accuracy.

5. Conclusions

This study investigated the impact of the COVID-19 lockdown on air quality in seven major Chinese cities representing diverse geographic regions with high urbanization and economic activities to understand the broader effects of reduced human activity on air pollution levels. Using long-term datasets of air pollutants and meteorological conditions from January 1, 2016 to December 31, 2021, generalized additive models (GAM) was employed to develop predictive models for the pre-lockdown (pre-LD) and lockdown (LD) periods. The GAM showed high accuracy in predicting air pollutant levels via incorporating weather conditions and capturing complex temporal patterns. The pre-LD model was trained by extensive historical data, while the LD models utilized transfer learning to adjust for the limited data during the lockdown.

The results showed significant reductions in NO2, CO, and PM10 concentrations during the lockdown compared to the same period in 2019. This reduction was primarily attributed to decreased vehicular traffic and industrial activities (e.g., Beijing and Changchun). In contrast, Xiamen, with its coastal location and lower industrial emissions, showed less pronounced improvements, indicating the diverse impact of lockdown measures across different urban settings.

The study also highlighted the importance of extended training periods in enhancing model performance. Models trained on 24 months of data generally provided better predictive accuracy by offering a richer context for understanding future conditions. Specific anomalies (e.g., high pollution levels in 2020 caused by extensive use of fireworks and agricultural straw burning in Beijing and Changchun, respectively) were observed, which underscores the need for targeted air quality management strategies addressing both routine and exceptional pollution sources. The application of transfer learning was crucial in refining the LD model since allowing them to leverage pre-lockdown insights despite the limited lockdown data. This approach significantly enhanced the model's predictive accuracy, as evidenced by the reduced discrepancies between predicted and actual values. The LD model was more accurate than the pre-LD one, which validates the efficacy of transfer learning in adapting to new, constrained datasets.

The COVID-19 lockdown provided a unique natural experiment to explore the impact of reduced human activity on urban air quality. The findings underscore the potential long-term benefits of sustained reductions in human activities for air quality promotion. Additionally, the study demonstrates the feasibility of combining model interpretability with advanced techniques like transfer learning to enhance the reliability and accuracy of air quality predictions. These insights can inform future policy-making and air quality management strategies to mitigate urban pollution more effectively.

CRediT authorship contribution statement

Yuchen Ji: Software, Data Curation, Writing - original draft. Xiaonan Zhang: Software, Visualization, Writing - original draft. Yueqian Cao: Conceptualization, Methodology, Writing – review & editing, Funding acquisition.

Acknowledgments

This research was jointly supported by the National Natural Science Foundation of China (Grant No. 42301428). The authors appreciate the China Meteorological Data Center for providing the meteorological data and air pollutants records.

Declaration of Competing Interest

The authors declare no conflict of interest.

References

Afrin, S., Chowdhury, F.J., Rahman, M.M., 2021. COVID-19 Pandemic: Rethinking Strategies for Resilient Urban Design, Perceptions, and Planning. Front. Sustain. Cities 3.

Ahmed, F., Shafer, L., Malla, P., Hopkins, R., Moreland, S., Zviedrite, N., Uzicanin, A., 2024. Systematic review of empiric studies on lockdowns, workplace closures, and other non-pharmaceutical interventions in non-healthcare workplaces during the initial year of the COVID-19 pandemic: benefits and selected unintended consequences. BMC Public Health 24, 884.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 52, 345–370.

Briz-Redón, Á., Belenguer-Sapiña, C., Serrano-Aroca, Á., 2021. Changes in air pollution during COVID-19 lockdown in Spain: a multi-city study. Journal of Environmental Sciences 101, 16–26.

Comunian, S., Dongo, D., Milani, C., Palestini, P., 2020. Air Pollution and COVID-19: The Role of Particulate Matter in the Spread and Increase of COVID-19's Morbidity and Mortality. Int J Environ Res Public Health 17, 4487.

Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S.S., Conti, G.O., Ferrante, M., 2020. The role of air pollution (PM and NO2) in COVID-19 spread and lethality: A systematic review. Environ Res 191, 110129.

Dominici, F., 2002. On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health. American Journal of Epidemiology 156, 193–203.

Fan, C., Li, Y., Guang, J., Li, Z., Elnashar, A., Allam, M., de Leeuw, G., 2020. The Impact of the Control Measures during the COVID-19 Outbreak on Air Pollution in China. Remote Sensing 12, 1613.

Fu, F., Purvis-Roberts, K.L., Williams, B., 2020. Impact of the COVID-19 Pandemic Lockdown on Air Pollution in 20 Major Cities around the World. Atmosphere 11, 1189.

Ghahremanloo, M., Lops, Y., Choi, Y., Mousavinezhad, S., 2021. Impact of the COVID-19 outbreak on air pollution levels in East Asia. Science of the Total Environment 754, 142226.

Ghasemi, A., Amanollahi, J., 2019. Integration of ANFIS model and forward selection method for air quality forecasting. Air Qual Atmos Health 12, 59–72.

Hastie, T., Tibshirani, R., 1986. Generalized Additive Models. Statistical Science 1, 297–310.

John L. Pearce, Jason Beringer, Neville Nicholls, Rob J. Hyndman, Nigel J. Tapper, 2011. Quantifying the influence of local meteorology on air quality using generalized additive models. Atmospheric Environment 45, 1328–1336.

Khaiwal Ravindra, Preety Rattan, Suman Mor, Ashutosh Nath Aggarwal, 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. Environment International 132, 104987.

Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. BioScience 51, 341–352.

Lin, Y.-C., Fan, K.-C., Wu, C.-D., Pan, W.-C., Chen, J.-C., Chao, Y.-P., Lai, Y.-J., Chiu, Y.-L., Chuang, Y.-F., 2024. Yearly change in air pollution and brain aging among older adults: A community-based study in Taiwan. Environment International 190, 108876.

Liu, Z., Lai, J., Zhan, W., Bechtel, B., Voogt, J., Quan, J., Hu, L., Fu, P., Huang, F., Li, L., Guo, Z., Li, J., 2022. Urban Heat Islands Significantly Reduced by COVID-19 Lockdown. Geophysical Research Letters Volume 49.

Lu, X., Yao, T., Fung, J.C.H., Lin, C., 2016. Estimation of health and economic costs of air pollution over the Pearl River Delta region in China. Science of The Total Environment 566–567, 134–143.

Murphy, C., Lim, W.W., Mills, C., Wong, J.Y., Chen, D., Xie, Y., Li, M., Gould, S., Xin, H., Cheung, J.K., Bhatt, S., Cowling, B.J., Donnelly, C.A., 2023. Effectiveness of social distancing measures and lockdowns for reducing transmission of COVID-19 in non-healthcare, community-based settings. Philos Trans A Math Phys Eng Sci 381, 20230132.

Neogi, S., 2023. EFFECTS OF TRAFFIC-RELATED AIR POLLUTION ON COGNITIVE FUNCTION AND ENVIRONMENTAL AWARENESS OF ADOLESCENTS. GLOBAL JOURNAL FOR RESEARCH ANALYSIS 97–99.

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22, 1345–1359.

Senaviratna, N.A.M.R., A. Cooray, T.M.J., 2019. Diagnosing Multicollinearity of Logistic Regression Model. AJPAS 1–9.

Sharifi, A., Khavarian-Garmsir, A.R., 2020. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. Science of The Total Environment 749, 142391.

Stone, C.J., 1985. Additive Regression and Other Nonparametric Models. The Annals of Statistics 13, 689–705.

Tsilimigkas, G., Stathakis, D., Pafi, M., 2016. Evaluating the land use patterns of medium-sized Hellenic cities. Urban Research & Practice.

Venter, Z.S., Aunan, K., Chowdhury, S., Lelieveld, J., 2021. Air pollution declines during COVID-19 lockdowns mitigate the global health burden. Environmental research 192, 110403.

Venter, Z.S., Aunan, K., Chowdhury, S., Lelieveld, J., 2020. COVID-19 lockdowns cause global air pollution declines. Proceedings of the National Academy of Sciences 117, 18984–18990.

Zhang, J., Lim, Y.-H., So, R., Jorgensen, J.T., Mortensen, L.H., Napolitano, G.M., Cole-Hunter, T., Loft, S., Bhatt, S., Hoek, G., Brunekreef, B., Westendorp, R., Ketzel, M., Brandt, J., Lange, T., Kolsen-Fisher, T., Andersen, Z.J., 2023. Long-term exposure to air pollution and risk of SARS-CoV-2 infection and COVID-19 hospitalisation or death: Danish nationwide cohort study. European Respiratory Journal 62.