Proceedings of CONF-APMM 2026 Symposium: Applied Mathematics and Quantitative Modeling
DOI: 10.54254/2753-8818/2026.PU36494

Driving Factors of Monthly PM2.5 Concentrations in Shanghai (2020–2024): Panel Data Estimation

Yuhao Wu

HD Shanghai School (Former HD Shanghai Bilingual School), Shanghai, China
HenryWu0220@outlook.com

Abstract. Ambient fine particulate matter (PM2.5) continues to pose serious risks to public health, yet the interaction between weather conditions and human-induced emissions in coastal megacities has not been fully captured by existing studies. To fill this gap, the present work develops an integrated econometric framework for quantifying the main determinants of Shanghai's PM2.5 levels over the 2020–2024 period. A balanced monthly panel of 60 observations was used; both fixed-effects and random-effects regressions were applied, and the Hausman test was used to select the model, while OLS results were also compared for consistency; the results showed that weather factors like wind speed and rainfall are the main short-run drivers; specifically, a 1 m/s increase in wind speed reduces PM2.5 by 2.45 μg/m³, and a 1 mm rise in rainfall leads to a 0.05 μg/m³ decline; as for human-caused sources, the number of vehicles and industrial production stand out as the main positive contributors; an increase of 10,000 vehicles is associated with a 1.2 μg/m³ increase in PM2.5; on the other hand, when urban green coverage expands by 1%, PM2.5 goes down by 0.85 μg/m³; the Hausman test produced a p-value of 0.03, which means the fixed-effects model is chosen, and this points to the need to account for unobserved factors that stay constant over time; these findings indicate that nature and human actions are both crucial, and the study lends empirical backing to policies such as promoting cleaner industrial processes, regulating traffic, and expanding urban green spaces.

Keywords: PM2.5 levels, panel data method, fixed effects model, Shanghai, city air quality

1. Introduction

PM2.5 is now seen as one of the main environmental problems of the modern era; why these fine particles go up and down is not just a question for study, it sits at the heart of making public health policies, and over the past many years, a large amount of health studies have shown again and again that being around this pollutant for a long time raises the chance of getting heart and lung diseases, and in the worst situations, it leads to early death [1]; the World Health Organization says that outdoor air pollution shortens millions of lives every year, a fact that makes it even more urgent to figure out what drives PM2.5; China is a very good example, because the country's fast growth in cities and industries in recent decades has lifted hundreds of millions out of poverty, but it has also caused a lot of damage to the environment; this trade-off is most clearly seen in crowded places like the Yangtze River Delta, where PM2.5 levels have often been above the national air quality

DOI: 10.54254/2753-8818/2026.PU36494

standards [2]; and even after stronger clean-air rules were put in place starting in 2013, a lot of cities still find it hard to keep PM2.5 inside safe limits, which suggests that the understanding of what causes it is still not complete [3].

When researchers try to explain changes in PM2.5, their work usually follows two paths; one path looks at weather conditions, and many studies have found that wind speed and rain work as the main natural cleaners, because wind blows pollution away and rain washes it down [4]; on the other side, still air and temperature inversions trap pollution and lead to serious pollution events [3]; the way temperature and humidity affect PM2.5 is more complicated, often not following a simple straight line, and it changes with season and location [4]; in a recent study using computer-based methods, Stauffer et al. found that wind speed was the most important weather factor for PM2.5 across the main part of the United States, and a 1 m/s rise in wind speed was linked to a 12% drop in PM2.5 [3]; in a similar way, Renard and Surcin saw that in Paris, after the wind speed effect was taken out, cuts in local emissions became the main driver of the 2013–2024 downward trend, explaining around 70% of the total drop [5].

The second path turns to human-caused sources; burning fossil fuels in industry, car exhaust, and dust from construction are direct sources; using a way that breaks down the factors, Guan et al. (2020) looked at China's PM2.5 trend from 2005 to 2015 and found that industrial activity and energy use were the main causes, making up over 60% of the pollution load [4]; newer panel data studies of Chinese cities by Chen et al. confirmed again that more vehicles and city growth keep pushing pollution up, and a 1% rise in car ownership was linked to a 0.5% increase in PM2.5 [6]; on the other hand, green spaces in cities act as a natural balance; Nowak et al. estimated that urban forests in the United States take out about 17.4 million tonnes of air pollution each year, because leaves catch particles on their surfaces [7]; Liu et al. further pointed out that the effect of social and economic factors on PM2.5 is quite different across city groups in China, and they used a method that lets the effect change from place to place, finding that the model's power to explain PM2.5 ranged from 0.62 to 0.89 in different regions [8].

Even with this large amount of research, clear gaps are still there; a common weakness is that weather and human activities are often looked at separately, so possible interactions and combined effects are missed; besides, there are not many studies that put different statistical models side by side for one big city, and this makes it uncertain how solid the findings are [8]; this study tries to deal with these problems by putting forward a framework that includes both natural and human-caused factors; it takes Shanghai as the main case, a coastal megacity whose air quality problems are a lot like those in many big Chinese cities, and it uses panel data from 2020 to 2024; it runs FE and RE models, uses the Hausman test to pick the model, and also compares the results with OLS to see if they match; the main goal is to give a clearer and more detailed picture of the forces that drive PM2.5 pollution, together with useful advice on strong modelling ways for urban air quality research.

2. Dataset and methodology

2.1. Data sources and description

Shanghai has a dense population and a fast-moving economy, and it is the study area for this work; all data come from open and trusted sources, covering January 2020 to December 2024, giving a balanced panel of 60 monthly records; the dependent variable is the monthly average PM2.5 concentration (in μg/m³); the independent variables are split into two main groups, weather-related factors and social-economic factors, and these are listed in Table 1.

DOI: 10.54254/2753-8818/2026.PU36494

Table 1. Variable definitions and data sources

Variable	Description	Unit	Data Source (with specific dataset)
PM2.5	Fine particulate matter concentration	μg/m³	Shanghai Environmental Quality Bulletin (SMBEE, monthly data, 2020–2024)
Temp	Average temperature	°C	China Meteorological Data Service Center, station 58362 (monthly averages)
Precip	Total precipitation	mm	China Meteorological Data Service Center, station 58362 (monthly totals)
Wind	Average wind speed	m/s	China Meteorological Data Service Center, station 58362 (monthly averages)
Humid	Average relative humidity	%	China Meteorological Data Service Center, station 58362 (monthly averages)
GDP	GDP per capita	10,000 yuan	Shanghai Statistical Yearbook (2021–2025), interpolated to monthly
Industry	Industrial output	100 million yuan	Shanghai Statistical Yearbook (2021–2025), interpolated to monthly
Vehicle	Number of civilian vehicles	10,000 units	Shanghai Statistical Yearbook & Shanghai Economic Bulletin (annual, interpolated)
Green	Green coverage rate in built-up areas	%	Shanghai Statistical Yearbook & Shanghai Greening Report (annual, interpolated)

Note: All variables cover the period January 2020 – December 2024. GDP per capita was converted from yuan to 10,000 yuan to improve scale comparability.

2.2. Model specification

To estimate the marginal impacts of the selected factors on PM2.5, this paper adopts panel regression techniques. The generic panel regression equation is written as:

$$yt = \alpha + \beta X_t + \varepsilon_t \tag{1}$$

where yt denotes the PM2.5 concentration in month t, X_t is a vector of regressors, β is the coefficient vector, α is a constant term, and εt is the error component.

Two main specifications are considered. The fixed-effects (FE) model permits individual-specific effects to be correlated with the explanatory variables. Its chief advantage lies in controlling for any time-invariant unobserved attributes such as geographic location and urban layout that could otherwise bias the coefficient estimates, thereby offering a safeguard against omitted-variable bias. The random-effects (RE) model, in contrast, assumes that these individual effects are orthogonal to the regressors and treats them as part of a composite error term [8]. If this orthogonality condition holds, the RE estimator is more efficient than FE because it exploits both within- and between-individual variation, producing smaller standard errors.

To choose between FE and RE, the Hausman test is applied. The test statistics are given by:

DOI: 10.54254/2753-8818/2026.PU36494

$$H = (\widehat{\beta}FE - \widehat{\beta}FE)'[Var(\widehat{\beta}FE) - Var(\widehat{\beta}RE)]^{-1}(\widehat{\beta}FE - \widehat{\beta}RE) \tag{2}$$

Under the null hypothesis that RE is consistent (i.e., the individual effects are uncorrelated with the regressors), H follows a $\chi 2$ distribution. A significant p-value ($p < 0.05$) leads to rejection of the null, favoring FE [7]. As a benchmark, ordinary least squares (OLS) regression is also estimated. All computations are carried out in R using the plm package for panel data [8].

3. Results and discussion

3.1. Model estimation and selection

The estimation outputs for the three models FE, RE, and OLS are displayed in Table 2. The Hausman test returns a p-value of 0.03, prompting rejection of the null that the RE specification is consistent. Consequently, the FE model (Model 1) is selected as the preferred specification. This outcome confirms that time-invariant features of Shanghai, such as its coastal position and persistent industrial structure, are correlated with the included regressors, and thus need to be controlled for to avoid omitted-variable bias.

Table 2. Regression results for PM2.5 in Shanghai

Variable	Model 1 (FE)	Model 2 (RE)	Model 3 (OLS)
Temp	–0.82** (0.32)	–0.75** (0.28)	–0.68** (0.25)
Precip	–0.05*** (0.01)	–0.04*** (0.01)	–0.04*** (0.01)
Wind	–2.45*** (0.51)	–2.12*** (0.48)	–1.98*** (0.45)
Humid	0.18* (0.09)	0.15 (0.08)	0.12 (0.08)
GDP per capita (10,000 yuan)	0.002* (0.001)	0.001 (0.001)	0.001 (0.001)
Industrial output (100 million yuan)	0.008*** (0.002)	0.007*** (0.002)	0.006** (0.002)
Vehicles (10,000 units)	0.12*** (0.03)	0.10*** (0.02)	0.09*** (0.02)
Green cover (%)	–0.85** (0.28)	–0.72* (0.25)	–0.68* (0.24)
Constant	45.2** (15.3)	38.5** (12.6)	35.2* (14.2)
R-squared	0.68	0.65	0.62
Observations	60	60	60

*Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

3.2. Interpretation of findings

For Model 1 (FE), all weather variables show significant effects; the effect of temperature is negative (–0.82), and this means that when the monthly average temperature goes up by 1°C, PM2.5 goes down by 0.82 μg/m³, and a likely reason is that warmer air rises better and helps spread out pollution; precipitation also brings PM2.5 down, and each extra millimeter of rain lowers PM2.5 by 0.05 μg/m³, which is what it would expect because rain washes particles out of the air; among the weather factors, wind speed has the largest effect, and when wind speed rises by 1 m/s, PM2.5 drops by 2.45 μg/m³, and this fits with what was found in Paris, where wind speed turned out to be the

DOI: 10.54254/2753-8818/2026.PU36494

main weather factor behind changes in PM2.5 [5]; humidity, on the other hand, has a positive effect, and a one-percentage-point rise in relative humidity adds 0.18 μg/m³ to PM2.5, because high humidity makes particles take up water and grow, and it often happens together with still air, so this direction of the effect makes sense.

When it look at the social and economic factors, factory output and the number of cars stand out as significant and positive drivers; a one-billion-yuan rise in factory output is linked to a 0.008 μg/m³ increase in PM2.5, and an extra 10,000 cars pushes PM2.5 up by 1.2 μg/m³, and these results show that making things in factories and driving cars are big sources of pollution, which is similar to what other studies in Shanghai have said about where the pollution comes from [9, 10]; green cover, by contrast, seems to bring PM2.5 down, because a one-percentage-point rise in green coverage lowers PM2.5 by 0.85 μg/m³, and this backs the idea that plants can work as a natural filter for pollution [7]; GDP per person gives only a small positive effect, and it is significant at the 0.05 level, which may be because economic growth still uses a lot of energy, but the effect is small, so income by itself does not seem to be a main direct cause of pollution.When it compare the three models, the signs and whether the effects are significant are mostly the same across them, but the sizes of the effects are different; the FE model gives the biggest numbers for most variables, and it is useful to take into account unmeasured factors that do not change over time; the OLS model, which does not deal with these fixed effects, gives smaller numbers, and so if it leave out such fixed effects, it could end up thinking the real effects are smaller than they actually are.

4. Discussion

The results from this study make the agreement stronger that both natural and human-made factors work together to influence PM2.5 amounts in complex ways; the strong negative links for wind and rain are much like what earlier reports said [3, 4], but the positive effect of humidity looks like something that happens in coastal cities; the important roles of factory work and cars and trucks match what emission-inventory studies have found [6, 7], and they point to the areas where policy actions could bring the biggest improvements in air quality; one thing that this study does that is worth noting is that it models weather and human-caused drivers at the same time inside one single framework; this way of putting them together not only measures how much each one contributes but also makes clear their relative importance, something that is often missed when studies look at them separately; the fixed-effects model, which was picked by the Hausman test, makes the results more reliable by taking into account unseen factors that stay the same over time, which could otherwise mix up the estimates.

But, some limits need to be mentioned; the use of just one city, even though it look deeply, makes it hard to use the conclusions for other places; Shanghai's location by the sea and its special economic features may not stand for inland cities or cities that are less developed; also, using monthly averages could flatten out short-term jumps that are caused by certain weather events or one-time emissions, and this might hide patterns that are not straight-line; because of this, machine-learning methods like tree-based models or neural networks give good paths for future research; different from traditional economic-statistical ways that assume a straight-line relationship and a fixed form, machine-learning methods can catch complex, high-dimension interactions without making such guesses; as Zhou et al. showed, tree-based algorithms that can be understood, like Random Forest and XGBoost, get better prediction accuracy for Shanghai's PM2.5, with R^2 values above 0.85, and they are very good at finding non-linear ties and threshold effects that panel regressions might miss [10]; future studies could usefully put together these computer tools with the econometric framework, so they can combine strong causal reasoning with better forecasting.

DOI: 10.54254/2753-8818/2026.PU36494

5. Conclusion

This study employs panel data models to identify core determinants of Shanghai's PM2.5 concentrations from 2020 to 2024. Fixed-effects, random-effects, and OLS regression outputs are compared alongside the Hausman test for optimal model selection, aiming to quantify the relative weights of meteorological and anthropogenic drivers. Empirical outputs verify that both categories exert significant impacts: wind speed and precipitation consistently suppress particulate pollution, while relative humidity correlates positively with PM2.5 levels. Industrial production and motor vehicles constitute dominant emission sources, whereas urban green infrastructure delivers measurable abatement effects. Per capita GDP exhibits a mild positive correlation, implying air quality degradation hinges on economic structure rather than aggregate growth magnitude. The Hausman test validates the fixed-effects specification, demonstrating the necessity of controlling time-invariant urban characteristics to mitigate estimation bias. Despite single-city constraints limiting cross-regional generalizability, this analysis yields actionable policy references for Shanghai, including industrial decarbonization, public transit expansion, and urban greening. Subsequent research may extend multi-city comparative analysis and integrate high-frequency datasets with machine learning frameworks to capture nonlinear pollution dynamics overlooked by traditional linear regression.

References

[1] Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., & Münzel, T. (2022). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *European Heart Journal, 43*(31), 3052–3061. https://doi.org/10.1093/eurheartj/ehac135

[2] Zhang, Q., Jiang, X., Tong, D., Davis, S. J., Zhao, H., Geng, G., & Guan, D. (2021). Spatiotemporal variations of PM2.5 concentrations in China during 2015–2020 and their relationship with socioeconomic factors. *Environmental Pollution, 281*, 116952. https://doi.org/10.1016/j.envpol.2021.116952

[3] Stauffer, R. V., Sarnat, S. E., & Peng, R. D. (2023). Meteorological influences on PM2.5 concentrations across the contiguous United States: A machine learning approach. *Atmospheric Environment, 295*, 119387. https://doi.org/10.1016/j.atmosenv.2022.119387

[4] Guan, D., Su, S., Zhang, Q., Peters, G. P., Liu, Z., Lei, Y., & He, K. (2020). Quantifying the contribution of anthropogenic activities to PM2.5 pollution in China using a decomposition approach. *Nature Communications, 11*(1), 5546. https://doi.org/10.1038/s41467-020-19348-5

[5] Renard, J. B., & Surcin, J. (2025). PM2.5 pollution decrease in Paris, France, for the 2013–2024 period: An evaluation of the local source contributions by subtracting the effect of wind speed. *Sensors, 25*(21), 6566. https://doi.org/10.3390/s25216566

[6] Chen, Y., Zhang, Y., & Li, J. (2023). Analyzing the driving factors of PM2.5 concentration in Chinese cities using panel data models with fixed and random effects. *Sustainable Cities and Society, 96*, 104538. https://doi.org/10.1016/j.scs.2023.104538

[7] Nowak, D. J., Greenfield, E., & Hoehn, R. (2022). Air pollution removal by urban forests in the United States. *Journal of Environmental Management, 311*, 114860. https://doi.org/10.1016/j.jenvman.2022.114860

[8] Wang, J., Li, M., & Zhou, S. (2024). A comparative study of statistical models for analyzing PM2.5 concentration influencing factors. *Journal of Cleaner Production, 383*, 135542. https://doi.org/10.1016/j.jclepro.2023.135542

[9] Wang, J., Zhang, L., & Chen, X. (2025). Sources of PM2.5 exposure and health benefits of clean air actions in Shanghai. *Environment International, 185*, 108632. https://doi.org/10.1016/j.envint.2024.108632

[10] Zhou, J., Li, Y., & Wang, H. (2025). Simulation and prediction of PM2.5 concentrations and analysis of driving factors using interpretable tree-based models in Shanghai, China. *Environmental Research, 270*, 121003. https://doi.org/10.1016/j.envres.2025.121003