Environmental Science

AIMS Environmental Science, 12(6): 1031–1058.
DOI: 10.3934/environsci.2025045
Received: 14 July 2025
Revised: 25 November 2025
Accepted: 15 December 2025
Published: 17 December 2025

https://www.aimspress.com/journal/environmental

Research article

Key driving factors and model predictions of the characteristics of air pollution in the urban area of Changchun, a typical city on the Northeast China Plain

Yujia Song*

College of Geographic Science and Tourism, Jilin Normal University, Siping 136000, China

* **Correspondence:** Email: jlsongyujia@126.com; Tel: 8613596171347.

Abstract: In this study, we used monitoring data on six representative pollutants, meteorological data, and socioeconomic data from January 2018 to December 2023 to clarify the spatiotemporal characteristics of air pollution in Changchun, Jilin Province, China, and to establish a short-term prediction model. First, correlation analysis (CA) was used to identify the key factors affecting pollution characteristics. Then, wavelet analysis (WA) was applied to capture the major periodic coupling characteristics of the time series, thereby analyzing the variation law of air quality. The results showed that the six air pollutants in Changchun exhibited significant seasonal differences. The peak concentration of O_3 occurred in summer, while the peaks of other pollutants were concentrated in winter and spring. The average air quality index (AQI) exhibited a downward, overall fluctuating trend, with $PM_{2.5}$ and PM_{10} as the major pollutants. Urbanization rate and green space coverage area were identified as key social driving factors, while temperature, relative humidity, and wind speed were identified as key meteorological driving factors. Furthermore, WA revealed a coupling phenomenon between the major periods of the AQI sequence and the $PM_{2.5}$ and PM_{10} sequences. Additionally, the low-frequency wavelet signals showed an overall downward trend, confirming that composite pollution is a typical feature of air pollution in Changchun, while the regional air quality is gradually improving. A short-term AQI forecasting equation based on ordinary least squares (OLS) was constructed. Unlike previous AQI prediction models that used only meteorological factors as independent variables, this model included "AQI data of the same period in the previous year" as a key independent variable. Verification showed that the model's fit outperformed that of traditional meteorologically driven models, yielding more accurate short-term AQI predictions for Changchun, driven by the following two major innovations. First, combined wavelet and CA comprehensively determined the periodic coupling characteristics between AQI and major pollutants, providing an efficient tool for analyzing the causes of composite pollution. Second, incorporating historical AQI

data into the OLS prediction model mitigated the limitation of traditional models that rely solely on meteorological factors.

Keywords: urban air pollution; short-term air quality index; socioeconomic factors; meteorological factors; Northeast China Plain

1. Introduction

Fossil fuels, including coal, oil, and natural gas, have been crucial to promoting economic development worldwide, though air pollution has steadily worsened. Burning fossil fuels, biomass, and minerals is the primary cause of air pollutants. The pollutants emitted by vehicles mostly include hydrocarbons, nitric oxides, and carbon monoxide. Coal burning results in the emissions of hazardous substances, such as SO_2, NO_x, and CO_2. Regional air pollution characteristics are not only influenced by pollution sources but also by several other factors. Identifying the driving factors of regional air pollution characteristics is the first step toward a more accurate prediction of air quality changes.

The driving factors of temporal and spatial air pollution characteristics on different scales have been extensively studied in recent years [1–3]. These driving factors are roughly divided into two categories, namely socioeconomic and meteorological. Socioeconomic factors, including population aggregation [4], urbanization level [5], and urban green coverage [6], determine the total amount of pollutants emitted into the atmosphere.

Fang et al. reported that urbanization harms China's air quality through factors such as population size, urbanization rate, vehicle density, and the proportion of secondary industries [7]. Fernando et al. performed a spectral analysis that identified the 36- to 77-day period of air quality in Southern Brazil, with 18- to 28-day oscillations. Moreover, the anthropogenic emissions exhibited distinct periodic characteristics [8]. Lei et al. discovered temporal variability in anthropogenic emissions of air pollutants, as manifested in "morning/evening peak hours", "day-of-the-week effect", and "holiday effect" [9]. Due to the periodic changes in weather systems and anthropogenic emissions, air pollutant concentrations also display marked periodicity. Research has examined the characteristics of air pollution and their influence on various environmental factors during smog weather. It has been generally accepted that industrial production and automotive exhaust emissions cause smog, deterioration of urban air quality, and reduced atmospheric visibility. Investigations have also been conducted into the optical characteristics and local distribution of aerosols during a haze outbreak [10–12]. Revealing the variation characteristics and the driving factors of regional air quality provides scientific evidence and support for environmental management.

China underwent rapid urbanization as the country shifted to an industrial economy following the 1980s reform and "opening up" policies. Ten million people a year migrated from rural areas to China's large cities during this period. This unprecedented scale and accelerated pace of China's urbanization, linked to the country's energy consumption, have led to serious resource, energy, and environmental crises, as well as significant increases in air pollutants and carbon dioxide emissions over the past three decades. Given that China's urbanization trend will likely continue for another 30 years, the conflict between urbanization and the atmosphere will persist into the foreseeable future.

Natural factors, including wind velocity [13], air temperature [14], and precipitation [15], usually affect the duration of air pollutants and pollutant concentrations.

Tan analyzed air quality monitoring data from Sichuan and Chongqing, China, from 2015 to 2021, and performed WA. They found that the air quality index (AQI) had a significant positive correlation

with air temperature and sunshine index in the short term, but a significantly negative correlation with wind velocity and precipitation [16]. The formation of air pollutants is also closely related to meteorological elements, including relative atmospheric humidity, temperature, and wind velocity. Meteorological elements influence the diffusion and transport of air pollutants [17]. Decreased atmospheric diffusion capacity will cause aerosols to accumulate near the ground, aggravating near-surface air pollution. Moreover, the meteorological elements directly influence the formation of secondary aerosols. For example, increased relative atmospheric humidity during smog weather promotes the heterogeneous reaction of liquid-phase SO_2 to produce sulfates. Strong photochemical reactions in summer facilitate the conversion of NO_2 in the atmosphere into particulate matter, that is, the conversion of gaseous precursors into particulate matter [18]. Li et al. identified Rossby waves with an oscillatory period of about 7 days in the mid- and high-latitude regions of the Northern Hemisphere, driven by the combined effects of emissions, meteorological diffusion, and chemical reaction processes. The low-frequency oscillations mostly consisted of quasi-biweekly oscillations (10-20 d) and intraseasonal oscillations (30-60 d) [19]. Li et al. employed the reverse trajectory method to analyze the major features and sources of air pollutants in Urumchi [20]. Xu et al. used a generalized additive model (GAM) to quantify the contributions of each meteorological factor and its gaseous precursors to $PM_{2.5}$ and ozone. They found that meteorological factors primarily influenced daily changes in ozone concentration [21].

Atmospheric relative humidity (RH) can regulate air quality, with its influence mechanisms involving the conversion of gaseous pollutants into secondary aerosols and the hygroscopic growth of particulate matter. High RH accelerates sulfate and nitrate formation by promoting liquid-phase oxidation of sulfur dioxide (SO_2) and nitrogen oxides (NO_x), thereby increasing $PM_{2.5}$ concentrations [22]. For instance, when RH exceeds 70%, the sulfur oxidation rate (SOR) and nitrogen oxidation rate (NOR) increase by 40% and 19%, significantly exacerbating fine particulate pollution [23].

In addition, water clusters (WCs) catalyze the formation of nitroaromatic compounds (NACs) in the gas phase. As important components of secondary organic aerosols (SOA), the formation of NACs is more prominent in high-humidity environments. Moreover, humidity exerts a dual effect on the hygroscopic properties of particulate matter: under low humidity, hygroscopic growth is limited; under high humidity, however, hydrophilic particulates increase in size after absorbing moisture, further reducing visibility and aggravating haze. These studies reveal the complex interaction between atmospheric relative humidity and aerosol chemistry.

Geographically weighted regression (GWR) [24] and geographic detectors are favored methods for determining pollution's temporal and spatial features at different levels [25]. However, these methods can detect only small-scale changes in pollutant concentrations and are hardly adequate for analyzing periodic changes in pollutants. WA is a time-frequency analysis method that emerged in the past two decades and provides a reliable tool for identifying the evolution law of nonstationary time series at different time scales [26]. Offering time and frequency resolution, the WA enables multi-scale refinement of discrete data while eliminating the influence of accidental factors on the variation law [27]. The wavelet transform enables us to analyze nonstationary time series effectively and to identify different frequency components and their periodic changes. When studying the characteristics of air pollution over a specific period, the rationality of WA can be fully demonstrated by its unique multi-scale time-frequency resolution, which far surpasses the limitations of traditional time-series methods to single-dimensional analysis. Unlike the Fourier transform, which captures only overall trends, the wavelet transform can simultaneously localize the local characteristics of pollution signals in both the time and frequency domains; this is crucial for analyzing nonstationary pollutant

concentration sequences with multi-scale fluctuations.

For example, Kumar et al. successfully identified the 330-day main cycle of $PM_{2.5}$ concentration and the daily fluctuation peak at 19:00–20:00 using Morlet wavelets, revealing the coupling impact mechanism of seasonal emissions and human activities. Moreover, Zhang et al. [28] reported that wavelet decomposition can decompose complex pollution sequences into approximate trend and detailed fluctuation components [29]. Combining this method with the ARMA model improved the prediction accuracy of heavy pollution events by 27% compared with a single model, effectively addressing the bottleneck of traditional methods in analyzing transient pollution processes. In addition, the robustness of the wavelet transform to noise enables it to perform excellently in the integrated analysis of multi-site monitoring data, providing a more reliable quantitative tool for pollution source tracing and spatiotemporal characteristic mapping. Analyzing periodic changes in air pollutants and understanding the multiple mechanisms underlying their formation, based on a long time series, is paramount for determining and predicting heavy pollution.

In summary, WA can identify the temporal characteristics of air pollution across temporal scales, which distinguishes it from traditional research methods. However, this method is often applied to study the characteristics of a single pollutant or a category of pollutants. It is rarely used to explore the variation relationship between air quality and its influencing factors. Therefore, based on the relevant socioeconomic data, AQI data, and monitoring data of major pollutants in Changchun, the geographical center of the Northeast China Plain, from January 2018 to December 2023, combined with the surface meteorological data of Changchun, we employed wavelet denoising analysis to investigate the variation characteristics of air quality in Changchun and its influencing factors, thereby addressing this research gap.

In addition, researchers who used prediction models to forecast air quality changes relied solely on meteorological conditions. In contrast, the impact of historical AQI data on future air quality was rarely considered. This practice ignores the cumulative effect of pollutants. Therefore, in constructing the air quality prediction model for Changchun, we incorporated historical AQI data to improve model accuracy and enrich the content of model prediction research.

2. Materials and methods

2.1 Study area

Changchun City (43°10′- 45°15 N, 124°01′- 127°05′ E) is in the mid-latitude temperate zone in the Northern Hemisphere, in the middle of Northeast China. As the political, economic, and cultural center of Jilin Province, Changchun City covers an area of 24,744 square kilometers. Its main urban area is on the Yitong River tableland within the Songliao Plain, where low mountains and hills dominate the landscape. Lying in the southeast of Changchun city, the main urban area belongs to the Big Black Mountain. Changchun extends from northeast to southwest, with an altitude of 250 m to 350 m and a relative height of 50 m to 100 m. Changchun has a four-season, monsoon-influenced, humid continental climate. The spring is windy in Changchun, with a maximum wind velocity of 30 m/s. The southeast wind prevails in summer, with an average temperature of 21.9C°. Despite considerable temperature differences, the weather is sunny and warm in autumn over several consecutive days. Besides, the wind velocity is usually smaller in autumn than in spring. The weather is sunny and dry in winter, with light winds and an average temperature of -12 °C. As for the industrial structure, the secondary sector is the dominant sector in Changchun, followed by the tertiary sector. The added value of the primary sector is the lowest. The location of the study area is shown in Figure 1.

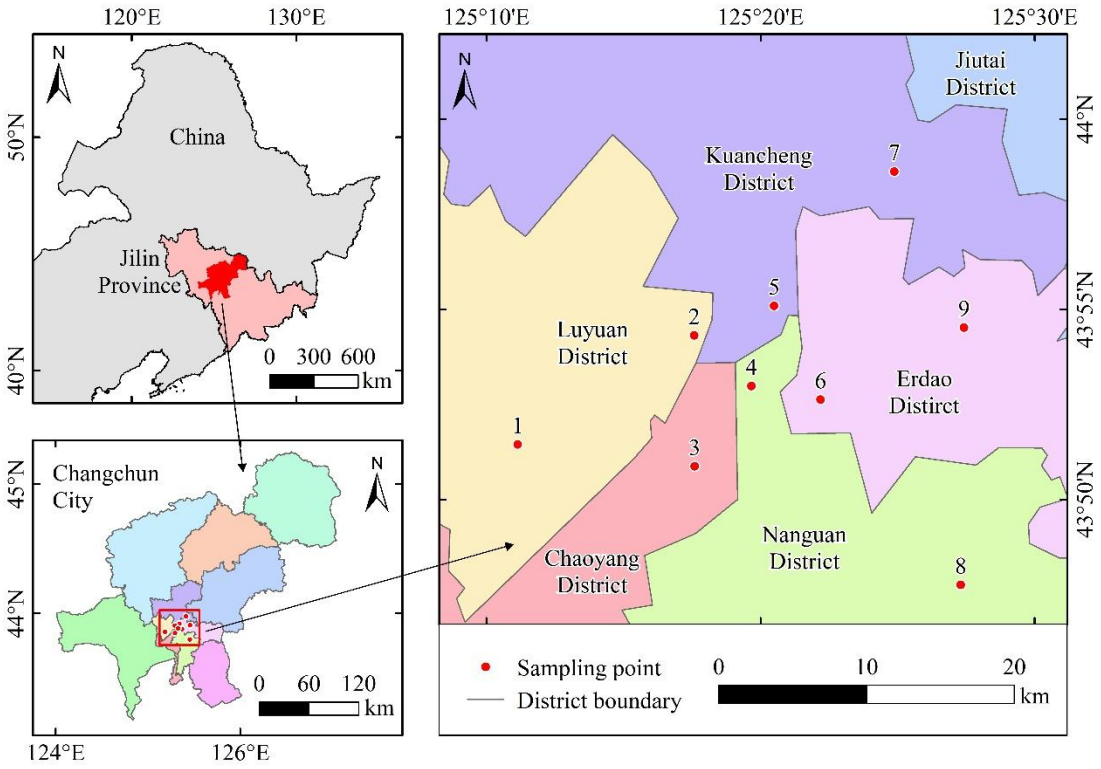

Figure 1. Study area and locations of atmospheric pollutant sampling points

2.2. Data sources and processing

In accordance with China's Regulations on the Delimitation of Ambient Air Quality Functional Zones in the Changchun Planning Area, the Changchun Municipal Bureau of Ecology and Environment has established a total of 9 automatic monitoring stations in the urban area. These stations cover the urban built-up area and suburban regions, and their site selection complies with the Technical Specifications for Layout of Ambient Air Quality Monitoring Sites (HJ664—2013), fully covering Class I areas (ecological protection zones) and Class II areas (residential, industrial, and traffic zones). They provide reliable data support for urban air quality assessment, pollution control, and public health protection.

We also used monitoring data from the nine stations listed above. This ensured that the monitoring data covered all types of urban areas and were highly representative. The pollutants monitored were PM_{10}, $PM_{2.5}$, CO, O_3, NO_2, and SO_2. The automatic air quality monitoring systems operated 365 days a year for at least 18 hours daily. The locations of the sampling points are shown in Figure 1.

According to the Technical Regulations on Ambient Air Quality Index (in Chinese) (HJ633-2012), the six major pollutants detected for air quality assessment are fine particulate matter, inhalable particulate matter, sulfur dioxide, nitrogen dioxide, ozone, and carbon monoxide.

The selected meteorological factors were air temperature, atmospheric pressure, wind velocity, and relative humidity. Data on meteorological factors from January 1, 2018, to December 31, 2023, were obtained from the China Meteorological Data Network (https://data.cma.cn/).

The socioeconomic factors considered were Changchun's GDP, urbanization rate, urban green coverage, and industrial electricity consumption from 2018 to 2023. Data on socioeconomic factors were from the Changchun Statistical Yearbook 2023 [30].

In China, the urbanization rate is defined as the percentage of the permanent urban population in the total permanent population of a city and serves as a core indicator of urbanization development. The permanent urban population generally refers to residents who have lived in urban areas (including urban downtown areas and seats of established towns) for more than half a year, covering registered residents and migrant populations that meet statistical standards. The total population refers to all permanent residents within the administrative region. The urbanization rate is a key foundational indicator in research fields such as regional development planning, ecological environment governance, and public policy formulation.

The urban green coverage area is the sum of the vertical projection areas of a city's vegetation (including arbors, shrubs, and lawns).

Industrial electricity consumption is the total electricity consumed by all industrial legal entities within a city's administrative region during production and operation activities in a reporting period (usually one year). It is one of the core energy indicators reflecting the vitality of a city's industrial economy and energy use efficiency.

Gross domestic product (GDP) is defined as the total value of final goods and services produced by all resident units within a city's administrative region during a reporting period (usually one year). It is a core macroeconomic indicator for measuring a city's economic volume, development scale, and growth rate.

2.3. Methodology

2.3.1. Wavelet coefficient analysis

The wavelet transform can characterize signal features in both time and frequency domains and has been widely used to analyze nonstationary, discontinuous time series in various fields [31], including hydrology [32], geology [33], and meteorology [34,35]. Continuous wavelet transform of any time series function is represented as follows:

$$w_f(a,b) = a^{-1/n}\int_{-\infty}^{\infty} f(t)\varphi\left(\frac{t-b}{a}\right)d_t \qquad a \in \mathrm{R}^+, b \in \mathrm{R} \tag{1}$$

where a is the scaling factor for stretching or compressing the wavelet; b is the translation factor for shifting the wavelet in time; $w_f(a,b)$ is the wavelet coefficient. According to Torrence, the scale is related to the period of the wavelet in the following way [36]:

$$T \approx 1.033a \tag{2}$$

where T is the period.

2.3.2. Wavelet variance

Integrating the square of the wavelet transform coefficient in the time domain gives wavelet variance:

$$Var(a) = \int_{-\infty}^{\infty} |W_f(a,b)|^2 db \tag{3}$$

where $Var(a)$ is the wavelet variance. The plot of wavelet variances by scale shows the variation in

wavelet variance with scale, reflecting the scale-based energy distribution. Wavelet variance can be used to determine the relative intensity and dominant time scale of signal disturbances at different scales, the latter known as the dominant period. We performed a wavelet analysis of variance using MATLAB.

2.3.3. Wavelet denoising

In this study, the AQI values under investigation were non-measured raw data whose trends were not readily apparent to the naked eye. These AQI values constituted a highly irregular time series, which was decomposed and reconstructed by wavelet transform to reveal the hidden trend. Among the commonly used wavelet families, we chose the Daubechies wavelets (hereafter referred to as db wavelets) to decompose and reconstruct the variation characteristics of air pollution. The Daubechies wavelet is a family of binary wavelets proposed by Daubechies in 1988 [37]. Daubechies wavelets construct orthogonal and localized bases using discrete filters, which are represented by dbN in MATLAB. Here, N is the wavelet's serial number, taking values 2, 3, ..., 9, 10. We further studied the periodic variation characteristics of air pollution in Changchun. The db3 wavelets were applied to AQI and meteorological factors. Wavelet denoising is essential when using wavelet analysis to study air pollution characteristics during a specific period. The purpose of wavelet denoising is to separate and eliminate irrelevant interference information from noisy time-series data of air pollutant concentrations while retaining valid signals that reflect the true variation patterns of pollution. This lays a high-quality data foundation for subsequent pollution characteristic analysis, source apportionment, and prediction modeling.

Specifically, monitoring data of air pollutant concentrations are susceptible to various types of noise interference. On the one hand, there are systematic errors, such as electronic noise from the monitoring equipment itself and environmental electromagnetic interference. On the other hand, there are irregular accidental interferences, such as transient traffic exhaust pulses and local temporary pollution sources (e.g., sudden construction dust). Such noise can obscure the true trends of pollutant concentrations (e.g., seasonal and diurnal variation patterns) and key mutation information (e.g., the start and end nodes of heavy pollution events), making wavelet decomposition-based denoising analysis necessary.

The db3 wavelet, a member of the Daubechies wavelet family, is widely used in air pollution research. The primary reason lies in its high compatibility between its characteristics and the features of air pollution data, which can be summarized as follows:

The db3 wavelet has a compact support property (short support length), which makes it highly capable of capturing local mutations in signals (e.g., sudden increases or decreases in concentration during heavy pollution events).

As a low-order wavelet (order 3), db3 balances a certain degree of smoothness and anti-interference capability. For "nonstationary, high-frequency noise-containing" data such as air pollution data, low-order db3 filters can effectively remove high-frequency noise without excessive smoothing that would lead to the loss of valid details (e.g., short-term fluctuation patterns of pollutant concentrations).

The decomposition and reconstruction of low-order Daubechies wavelets require less computational effort. This can significantly improve analysis efficiency and reduce computational costs for long-term time-series air pollution data (e.g., multi-year daily/hourly data). The Daubechies wavelets were applied to signals in MATLAB.

2.3.4. Correlation analysis

The Pearson Correlation Coefficient (PCC) of air pollutant concentrations with socioeconomic and meteorological factors was performed using GraphPad Prism 8.0.

We used Pearson's correlation analysis to examine the pairwise associations between two categories of factors and air pollutants. The first category included associations between 4 meteorological conditions (atmospheric relative humidity, average wind speed, atmospheric pressure, and air temperature) and six pollutants (PM_{10}, $PM_{2.5}$, SO_2, CO, NO_2, and O_3). The second category covered associations between 4 socioeconomic factors (GDP, industrial electricity consumption, urban green coverage area, and urbanization rate) and six pollutants. A total of (4×6) + (4×6) = 48 independent tests were conducted.

To control the risk of false-positive results from multiple tests and ensure the rigor of statistical conclusions, the FDR correction method was applied to adjust all original P-values, with the adjusted significance level set at $\alpha = 0.05$.

2.3.5. Constructing the predictive models

A variety of natural factors influence short-term air pollution. We employed OLS to estimate the coefficient that minimizes the difference between the predicted and actual values, and the optimal linear model was established accordingly. The OLS model assumes a linear relationship between dependent and independent variables and finds the optimal regression coefficients by minimizing the residual sum of squares. Minimization is usually achieved by differentiation and the extreme value theorem, and the chosen regression coefficient minimizes the residual sum of squares. The mathematical expression of the OLS regression is the function of the sum of squares of the residuals between the predicted and true values during minimization. The concept of best subset selection was employed for OLS regression, and the optimal equation for short-term air quality prediction was established with and without superimposing the previous year's AQI, respectively.

When constructing the prediction models, we used SPSSAU software to perform ANOVA calculations for the two models separately, obtained the sum of squared residuals, and further derived R^2 and adjusted R^2. We also calculated the models' RMSE and Durbin-Watson (DW) values, as shown in Table 1. The R^2 values of the two models were 0.908 and 0.861, respectively, and the adjusted R^2 values were 0.832 and 0.782, respectively.

Table 1. Model validation parameters.

AQI prediction model	R^2	Adjusted R^2	RMSE	DW
$Y_{1(AQI)}$	0.861	0.782	8.119	2.107
$Y_{2(AQI)}$	0.908	0.832	6.602	2.057

Note：$Y_{1\ (AQI)}$ model did not incorporate the AQI values from the previous year; $Y_{2\ (AQI)}$ model incorporated the AQI values from the previous year

In air pollution research, dependent variables (e.g., AQI) are often influenced by multiple meteorological factors, including atmospheric temperature, relative humidity, atmospheric pressure, and average wind speed. While these factors were briefly discussed in the introduction, we also proposed adding a fifth variable, AQI, and verifying whether it is a confounding factor. The judgment criterion is whether the change range of the regression coefficient of the core independent variable exceeds 10%–20% before and after introducing the new variable. Since air pollution data often exhibit natural fluctuations, the threshold for confounding factors can be flexibly relaxed to 20%.

Among the variables in the prediction models, the atmospheric relative humidity has the smallest

P-value, so it is the core independent variable. The change range of its regression coefficient after introducing the AQI variable is less than 10%, indicating that the AQI variable is not a confounding factor. In addition, we used the DW test to rule out residual autocorrelation in the models.

3. Results and discussion

3.1. Temporal variation characteristics of AQI and pollutant concentrations

The day-to-day variations of the concentrations of the six air pollutants (SO_2, NO_2, O_3, CO, $PM_{2.5}$, and PM_{10}) in Changchun from 2018 to 2023 are shown in Figure 2.

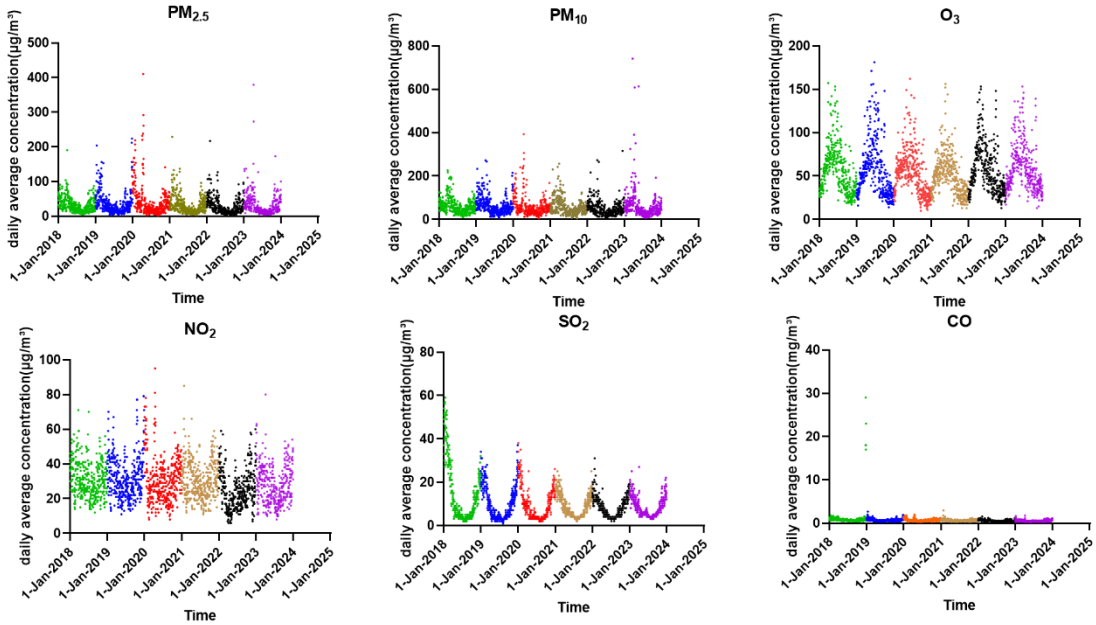

Figure 2. Daily variation trends of concentrations of six air pollutants (SO_2, NO_2, O_3, CO, $PM_{2.5}$, and PM_{10}) from 2018 to 2023.

All six pollutants displayed strong seasonal variations but no distinct interannual variation. Two distribution patterns were identified for the daily mean concentrations of the six pollutants throughout the year, namely unimodal and bimodal. Except for the daily mean O_3 concentration, which showed a unimodal distribution pattern, the daily mean concentrations of all other pollutants peaked in winter and spring, followed by a decrease afterward. This is because O_3 pollution is more common in summer, when high temperatures and strong sunlight more readily mediate photochemical reactions that produce O3. During the winter heating season, SO_2 and other pollutant emissions increase dramatically, and pollutant dispersion conditions are worse in winter than in other seasons due to unfavorable meteorological conditions. For the above reasons, the peak season of ozone pollution was not synchronized with those of other pollutants.

Given the daily limits of pollutants for second-grade air quality specified in the Ambient Air Quality Standard, the number of days when $PM_{2.5}$ and PM_{10} concentrations exceeded the second-grade daily limits was greater than that of other pollutants. $PM_{2.5}$ and PM_{10} pollutants were found in all four seasons, indicating severe atmospheric particulate matter pollution in Changchun. The time series of $PM_{2.5}$ and PM_{10} concentrations showed trends similar to those of AQI values, indicating that particulate matter was the primary pollutant in Changchun. As shown by the annual mean and extreme

concentrations of the other four pollutants, there were insignificant changes in the mean pollutant concentrations and the annual numbers and percentages of days with good and moderate air quality from 2018 to 2023. Besides, the concentrations of these four pollutants did not exceed the daily mean limit for second-grade air quality most of the time. As analyzed above, these four pollutants were not the major air pollutants in Changchun.

According to annual AQI statistics, the mean AQI first increased and then decreased, fluctuating from 2018 to 2023. The annual number of air pollution days was 34 in 2018, 59 in 2019, and 61 in 2020, indicating that the annual number peaked in 2020. After that, the annual number of air pollution days was 35 in 2021, 29 in 2022, and 51 in 2023.

3.2. CA of air pollutants, socioeconomic factors, and meteorological factors

3.2.1. Correlation between air pollutants and socioeconomic factors

A PCC was performed between the annual mean concentrations of six air pollutants (SO_2, NO_2, O_3, CO, $PM_{2.5}$, and PM_{10}) and socioeconomic data for Changchun in the same periods. The results of the correlation analysis are presented in Figure 3.

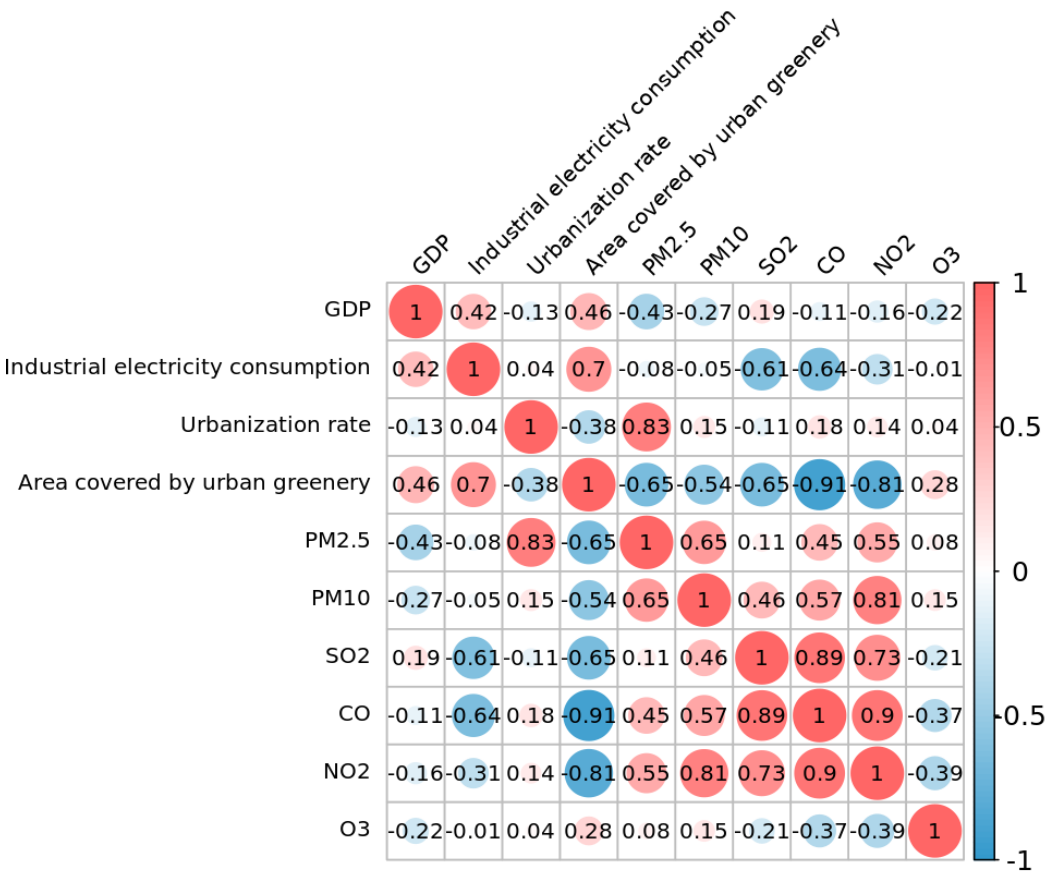

Figure 3. Correlation between concentration of air pollutants and social factors.

During this period, the urban green coverage, urbanization rate, and industrial electricity consumption in Changchun increased. The urbanization rate significantly influenced $PM_{2.5}$ concentration, which positively correlated with it (P<0.05). The higher the urbanization rate, the higher

the $PM_{2.5}$ concentration, and increasing urbanization led to a growing population. Additionally, the higher the population, the higher the energy consumption, thereby indirectly increasing atmospheric $PM_{2.5}$ concentrations. Urban green coverage showed a significant negative correlation with NO_2 and CO concentrations (P<0.05). Plants absorb NO_2 through stomata on their leaves, converting it into nitrates, a nutrient necessary for plant survival. NO_2 absorption varies across plant species, with daily values of 1.9 mg for cherry and 0.6 mg for Chinese maple [38]. Moreover, plants absorb carbon dioxide through photosynthesis and release oxygen into the atmosphere, helping regulate CO_2 levels. While CO_2 and CO share a carbon–oxygen bond, their atmospheric behavior differs. CO levels are mostly affected by combustion processes and atmospheric chemical reactions, not by photosynthesis. However, long-term changes in CO_2 concentration and climate can indirectly influence CO levels by altering temperature and the abundance of oxidizing agents such as hydroxyl radicals (OH). In a word, the urbanization rate and the urban green coverage were important socioeconomic driving factors of changes in air pollutant concentrations in Changchun. Furthermore, GDP and industrial electricity consumption exerted only a weak influence.

The purification effect of green spaces on pollutants may exhibit significant "lag effects" or seasonal variations across different seasons. For instance, vegetation is lush in summer, and its absorption and deposition effects on pollutants are far stronger than in winter.

Therefore, to capture the relationship between socioeconomic factors and air quality from a macro and long-term perspective, we conducted a correlation analysis using annual average pollutant concentrations (2018-2023, 6 years) and the corresponding annual average values of socioeconomic indicators over the same period. The inherent logic of this analysis method, based on multi-year averages, is to smooth and offset short-term (e.g., monthly, seasonal) fluctuations and lag effects to the greatest extent possible. Specifically, compressing 6-year data into an annual average effectively averages out differences in purification efficiency caused by seasonal changes (such as variation in the ecological activity of green spaces between January and June) each year.

Thus, the significant negative correlation we identified between green space area and concentrations of NO_2 and CO does not reflect an instantaneous effect in a specific season, but rather a long-term, stable macro-statistical association that transcends seasonal fluctuations. This helps us determine whether green space area is a significant social driving factor over a longer time scale. This method effectively reduces the interference of short-term lag effects on the study conclusions.

We consider the spatial heterogeneity of green space distribution to be a key factor affecting local air quality. Our original objective of this study was to treat Changchun as an integrated system to explore the relationship between its macro socioeconomic indicators and overall ambient air quality. Therefore, at the data analysis stage, we conducted correlation analysis using the total green coverage area of the entire Changchun administrative region and the municipal-level annual average pollutant concentrations (which represent the overall air quality of Changchun). This approach revealed the overall regularity at the city level. From a city-wide perspective, an increase in the total amount of green space resources is associated with a decrease in background concentrations of NO_2 and CO in the urban atmosphere. This conclusion provides a reference for urban top-level planners to formulate macro ecological construction strategies.

Certainly, we are well aware that to reveal the micro-mechanisms by which green spaces deeply affect pollutants, it is necessary to introduce more refined data (e.g., zoned green space data and monitoring station data) and analytical methods (e.g., Geographically Weighted Regression) in subsequent studies. We have also stated in the "Discussion and Prospects" section of the paper that the spatial heterogeneity pattern is a research direction warranting further in-depth exploration.

In summary, within our framework, we initially aimed to identify key socioeconomic driving

factors from a long-term, overall perspective by adopting macro-correlation analysis based on multi-year averages. Although this method has limitations in exploring mechanisms, it can effectively avoid interference from short-term lags and spatial heterogeneity when judging macro trends, thereby providing robust decision-making support for urban-level environmental management.

3.2.2. CA between air pollutants and meteorological factors

The results of PCC between air pollutant concentrations and meteorological factors in Changchun from January 2018 to December 2023 are shown in Figure 4.

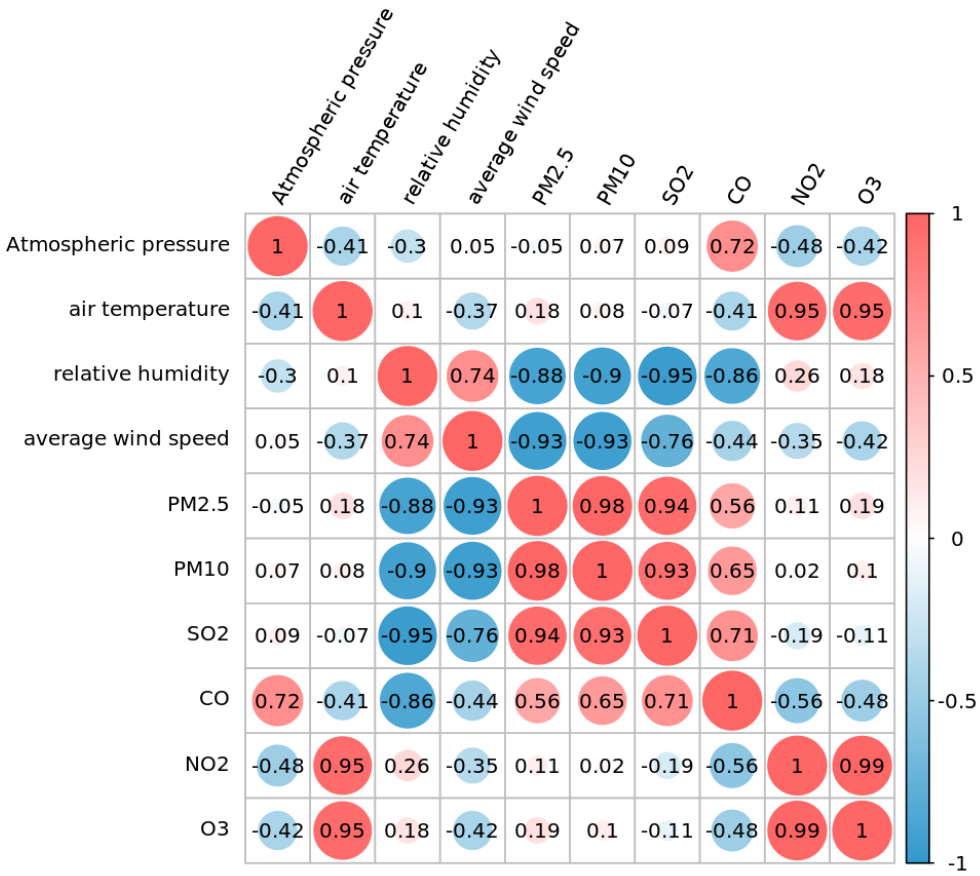

Figure 4. Correlation between concentration of air pollutants and meteorological factors.

Some pollutant concentrations showed strong correlations with meteorological factors. The O_3 and NO_2 concentrations had a strong positive correlation (P<0.01). The higher the temperature, the higher the O_3 and NO_2 concentrations would be, though a lower atmospheric pressure was not conducive to pollutant dispersion in winter. As a primary air pollutant, O_3 is not directly emitted into the atmosphere; it is produced through photochemical reactions. To produce ozone, volatile organic compounds (VOCs) and nitrogen oxides (NOx) undergo photochemical reactions at high temperatures and under strong illumination. Therefore, the photochemical reactions are more pronounced in summer for the above reasons, and summer is the ozone season. Relative atmospheric humidity showed significant negative correlations with $PM_{2.5}$ and PM10 concentrations (P<0.05) and with SO_2 concentration (P<0.01). The results indicated that higher atmospheric humidity was associated with lower concentrations of the three pollutants. At higher relative atmospheric humidity, water molecules

bind to particulate matter, increasing its weight and causing it to settle. $PM_{2.5}$ and PM_{10} can absorb water vapor from the air, which can cause them to expand and settle. When relative humidity exceeds 80%, particulate matter settles out of the air, reducing atmospheric concentrations. The correlation between the relative atmospheric humidity and SO_2 concentration mainly manifests as the former promoting the conversion of SO_2 into the sulfuric acid mist. SO_2 usually exists as a gas at a lower relative atmospheric humidity. As the relative atmospheric humidity increases, SO_2 is more likely to bind to water vapor to produce sulfuric acid and settle out of the air. Wind velocity showed a negative correlation with $PM_{2.5}$ and PM_{10} concentrations. The higher the wind velocity, the more dispersed $PM_{2.5}$ and PM_{10} pollutants are, improving air quality [39].

3.3. Periodic features of air pollutant concentrations based on wavelet transform

Morlet wavelet transform was applied to the AQI values and daily mean concentrations of six pollutants in Changchun from 2018 to 2023, and the real part of the wavelet coefficient was obtained, as shown in Fig.5. The intensity of the oscillatory signals is represented as follows: The higher the wavelet coefficient, the darker the color, and the higher the pollutant concentration, and vice versa.

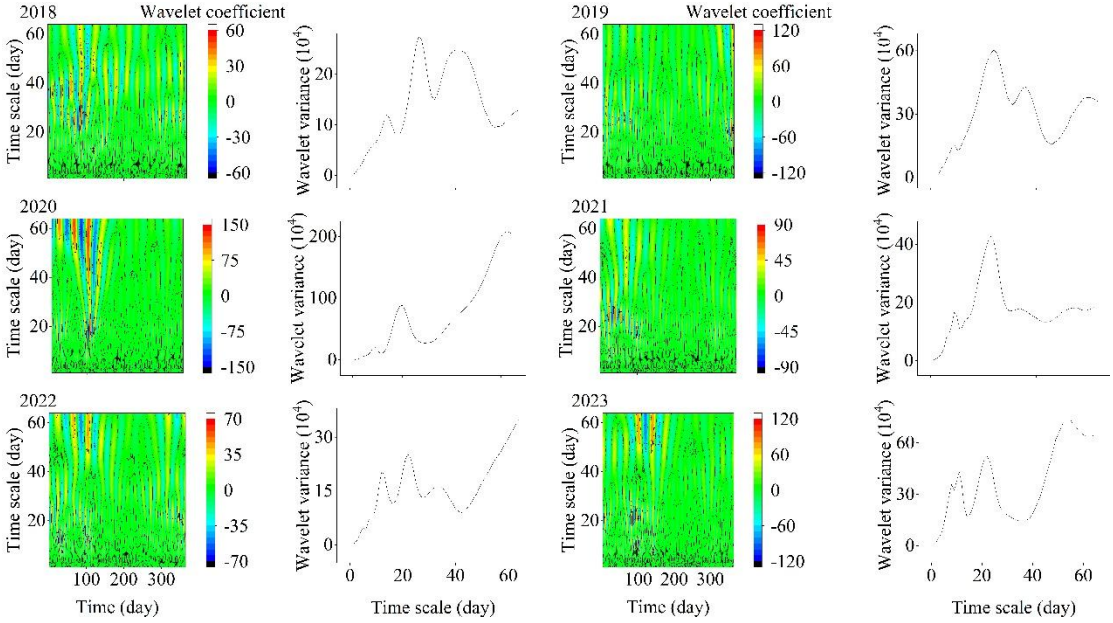

Figure 5. Wavelet Coefficients and Wavelet Variance of AQI during 2018–2023.

Figure 5. shows that in 2018, 2019, 2021, and 2023, the first dominant period of the AQI time series was 25 days; in 2020 and 2022, it was 20 days. From 2018 to 2023, the second dominant period of the time series was 16, 40, 10, 10, 12, and 12 days, respectively.

The real parts of the wavelet coefficient and wavelet variance for the time series of AQI values and concentrations of six pollutants ($PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO) in Changchun are plotted in Figs. 6-11, respectively.

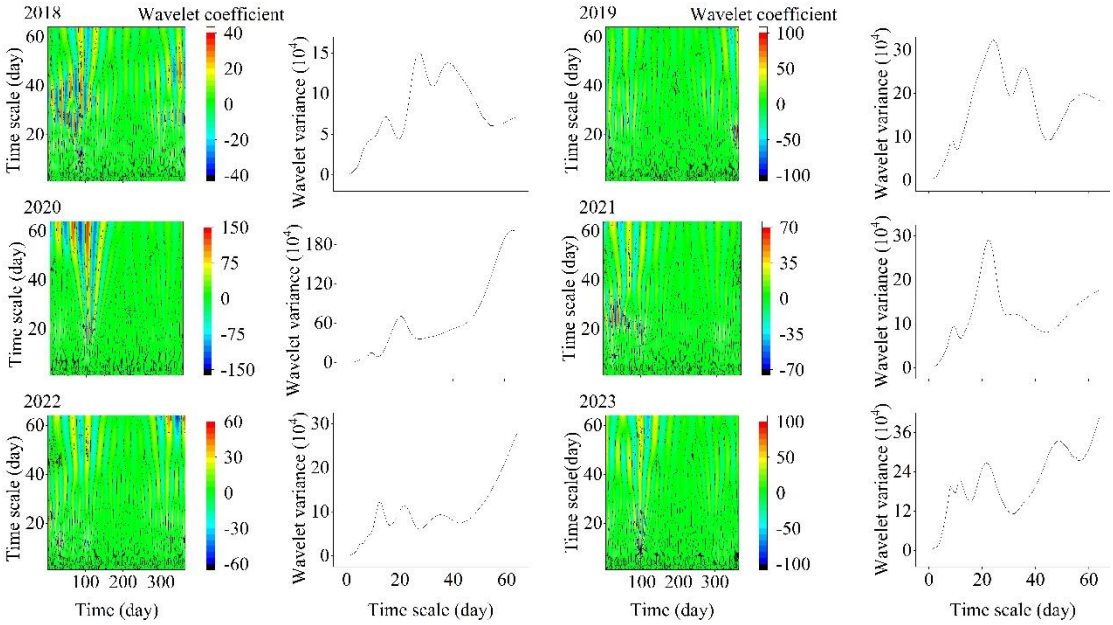

Figure 6. Wavelet Coefficients and Wavelet Variance of $PM_{2.5}$ during 2018–2023

Figure 6 shows that in 2018, 2019, 2021, and 2023, the first dominant period of the time series of $PM_{2.5}$ concentrations was 26 days; in 2020 and 2022, the first dominant period of the time series of $PM_{2.5}$ concentration was 20 and 15 days, respectively. From 2018 to 2023, the second dominant period of the time series of $PM_{2.5}$ concentrations was 15, 40, 10, 10, 20, and 14 days, respectively.

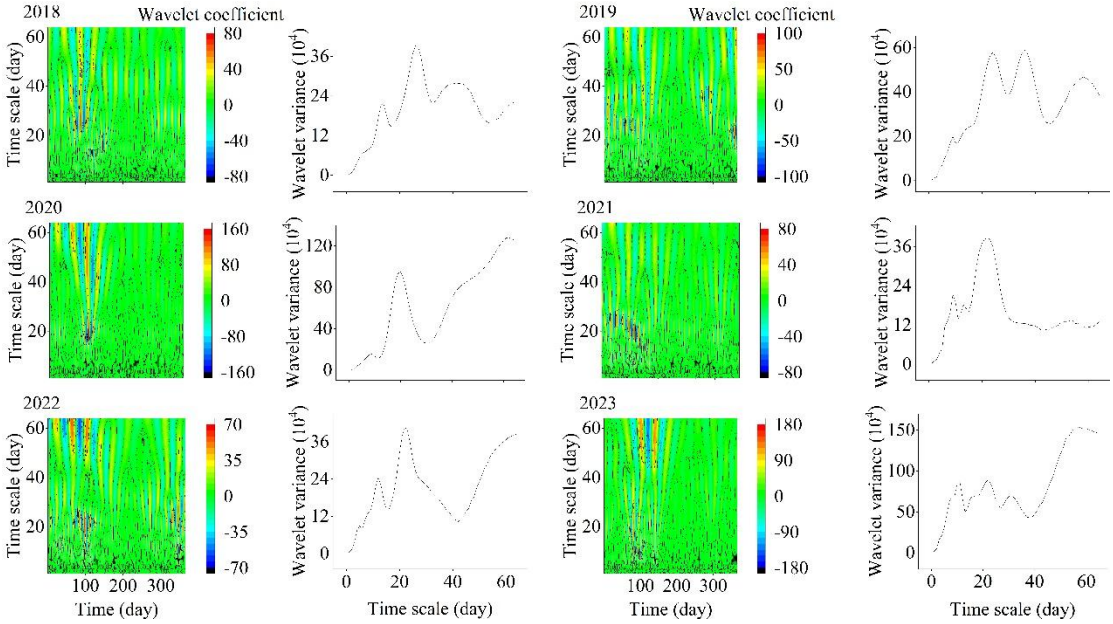

Figure 7. Wavelet coefficients and wavelet variance of PM_{10} during 2018–2023.

It can be inferred from Figure 7 that in 2018, 2021, and 2023, the first dominant period of the PM_{10} concentrations time series was 26 days. In 2019, the first dominant period of the time series of the PM_{10} concentrations was 40 days. In 2020 and 2022, the first dominant period of the PM10 concentrations time series was 20 days; from 2018 to 2023, the second dominant period was 17, 25, 10, 10, 10, and 10 days, respectively.

As shown in Figure 8, the first dominant period in the time series of NO_2 concentrations was 40 days in 2018. In 2019, the first dominant period of the time series of the NO_2 concentrations was 60 days. In 2020 and 2021, the first dominant period of the NO_2 concentrations time series was 20 days; in 2022, 34 days; and in 2023, 50 days. From 2018 to 2023, the second most dominant period in the time series of NO_2 concentrations was 24, 40, 10, 10, 24, and 24 days, respectively.

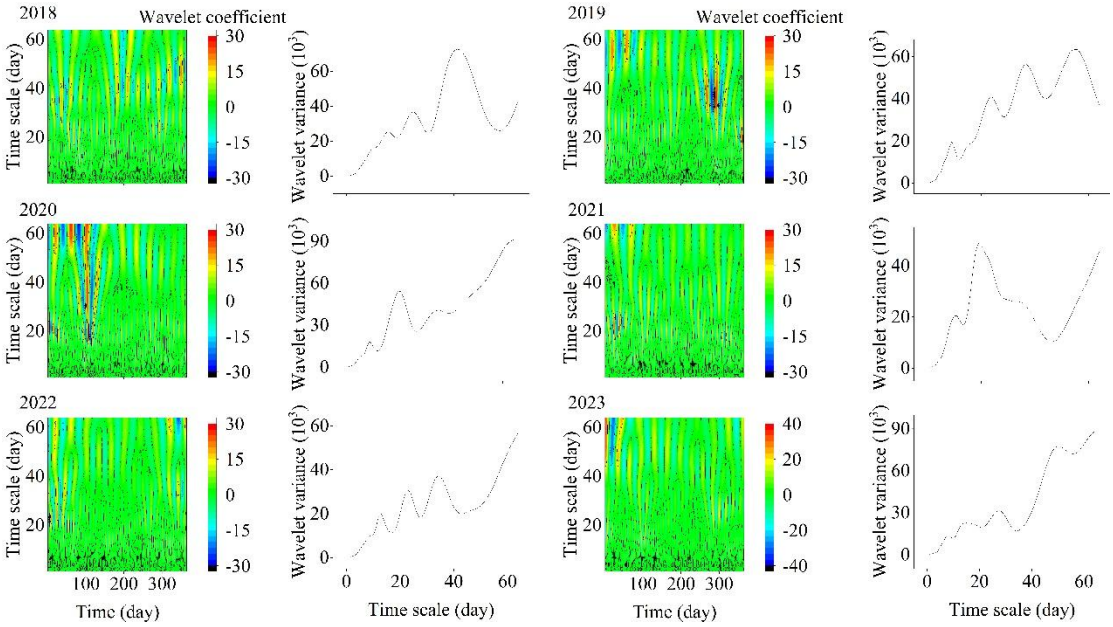

Figure 8. Wavelet coefficients and wavelet variance of NO_2 during 2018–2023

The analysis in Figure 9 shows that the first dominant periods of the time series of the O_3 concentrations were 35 days in 2018 and 2022, and 20 days in 2020 and 2021. The first dominant periods of the time series of the O_3 concentrations were 16 days in 2023 and 50 days in 2019. From 2018 to 2023, the second dominant periods of the time series were 17, 30, 10, 30, 24, and 0 days, respectively.

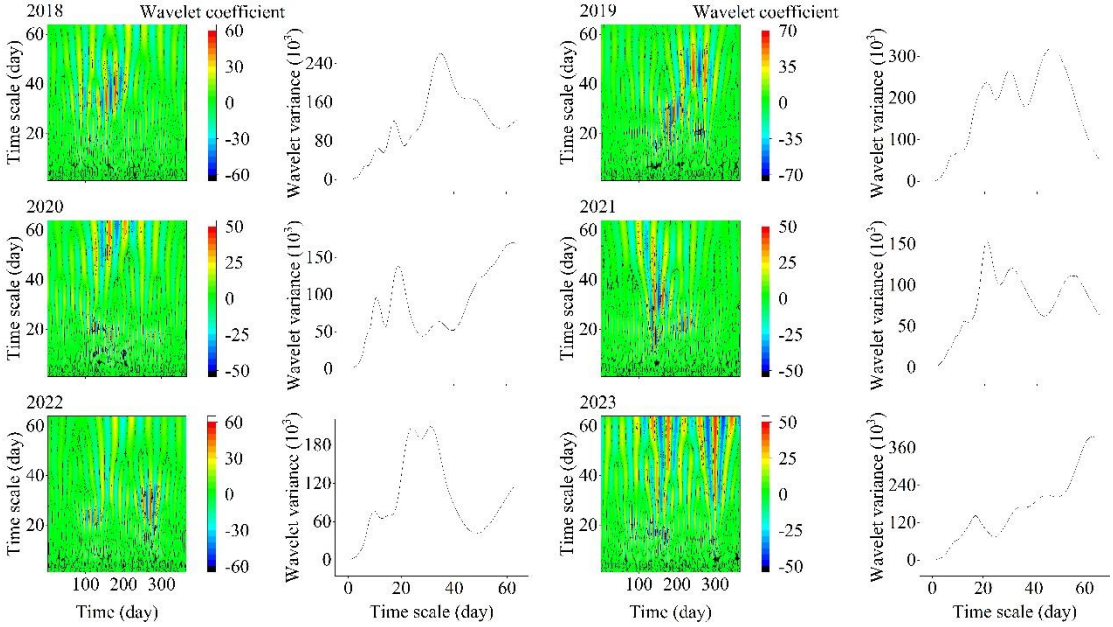

Figure 9. Wavelet coefficients and wavelet variance of O_3 during 2018–2023.

In Figure 10, the first dominant period of the time series of SO_2 concentrations was 20 days in 2018 and 2022, versus 25 days in 2019. The first dominant period of SO_2 concentrations was 55 days in 2021, 33 days in 2022, and 47 days in 2023. From 2018 to 2023, the second-most dominant periods in the time series of SO_2 concentrations were 16, 17, 8, 40, 16, and 10 days, respectively.

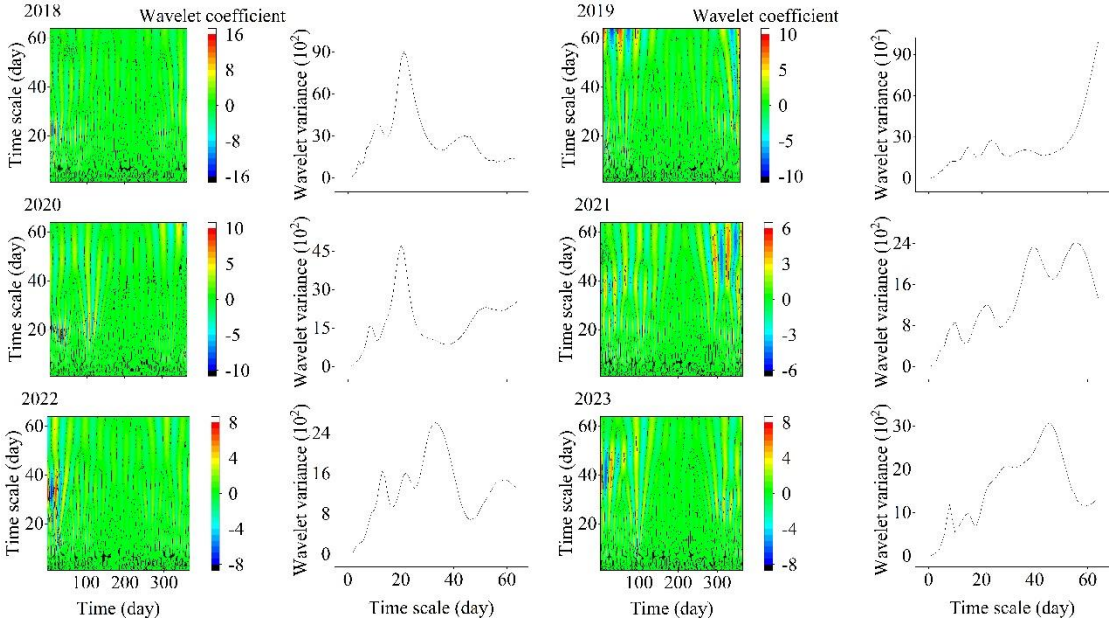

Figure 10. Wavelet coefficients and wavelet variance of SO_2 during 2018–2023.

As shown in Figure 11, the first dominant periods of CO concentrations in Changchun were 20 days in 2018 and 2020, and 40 days in 2019. The first dominant periods of the time series of the CO concentrations were 24 days in 2021, 24 days in 2022, and 47 days in 2023. From 2018 to 2023, the second-most dominant periods in the time series of CO concentrations were 7, 60, 10, 26, 30, and 20 days, respectively.

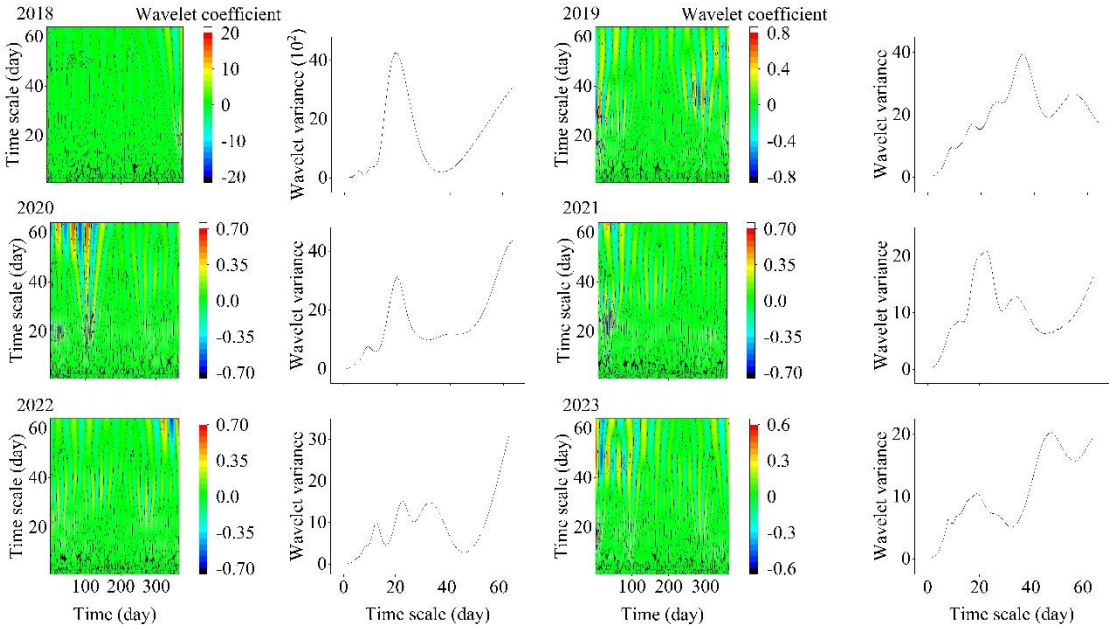

Figure 11. Wavelet coefficients and wavelet variance of CO during 2018–2023.

The dominant periods of the time series of AQI values and pollutant concentrations were compared, revealing that the AQI's first and second dominant periods were very similar to those of $PM_{2.5}$ and PM_{10} concentrations. A main periodic coupling phenomenon was observed, indicating that combined pollution caused by $PM_{2.5}$ and P_{10} persisted in Changchun during this period. However, the pollution situation in 2020 was an exception. The main periodic coupling phenomenon was observed between the first dominant periods of the AQI time series and the concentrations of six pollutants. This finding indicated poor air quality in Changchun in 2020, when pollution from all six pollutants lingered.

For the first main cycle of 25–26 days (e.g., AQI, $PM_{2.5}$, and PM_{10} in 2018 and 2019), an analysis of Changchun's meteorological data shows that this cycle is highly consistent with the local "intra-monthly cold air activity cycle". When cold air passes through, wind speed increases and diffusion conditions improve, decreasing pollutant concentrations. During cold-air intervals, stable weather occurs more frequently, enabling pollution to accumulate, thus forming a roughly 25-day fluctuation cycle.

In 2023, Changchun experienced more precipitation in summer and higher winter wind speeds, resulting in better meteorological diffusion conditions than in previous years. This slowed the rate of pollutant accumulation and extended the fluctuation cycle. Moreover, the growth in local motor vehicle ownership and adjustments to the industrial emission structure in 2023 also altered the emission cycles of gaseous pollutants such as NO_2 and CO compared to previous years. These factors manifested as a longer main cycle.

In 2020, the annual average concentration of $PM_{2.5}$ in Changchun was 42 μg/m³, the highest value of this indicator during the 2018–2023 period. Moreover, the number of days with good air quality stood at 305, accounting for 83.3% of the total monitoring days; lower than the 83.8% in 2019 and the levels in 2021–2023. The following three groups of factors primarily control this trend:

1. *Meteorological factors*. In 2020, meteorological conditions unfavorable to pollutant dispersion (e.g., frequent stable weather) led to easy accumulation of pollutants over the city and hindered their dispersion. For example, during winter, persistent temperature inversions occurred several times. These inversions blocked vertical air convection, making it difficult for pollutants to rise and increasing the concentration of ground-level pollutants.

2. *Socioeconomic factors*. Although industrial production and traffic decreased in 2020 due to the COVID-19 pandemic, pollutant emissions were not effectively controlled by other factors. For instance, some enterprises relaxed their environmental protection measures to compensate for production losses during the pandemic. Additionally, residents spent more time at home, increasing domestic energy consumption and, in turn, the levels of coal- and gas-combustion pollutants. Furthermore, reductions in nitrogen oxides (NO_x, mainly NO_2) from vehicle exhaust disrupted the original photochemical balance, increasing ozone concentrations (rather than decreasing them) in certain periods and regions.

3. *Pollution control factors*. While Changchun had been advancing air pollution control efforts, the effectiveness of some control measures in 2020 had not yet been fully realized. For example, in terms of dust pollution control, the progress of urban construction projects led to increased dust emissions. Still, the corresponding dust prevention measures failed to keep pace promptly.

The combined effect of these adverse factors led to an abnormal decline in air quality in 2020.

Based on the discovery of the 20–26-day main cycle of major air pollutants, this study provides a scientific basis for the government to formulate precise air pollution prevention and control strategies. First, this cycle reveals the high-frequency fluctuations in PM_{10} and $PM_{2.5}$ concentrations, indicating that a complete cycle of pollution accumulation and diffusion occurs approximately every 26 days. The government can establish a short-term pollution early warning mechanism based on this and

launch emergency emission-reduction measures in advance during high-concentration periods (e.g., dry, cold winter weather with low wind speed), such as restricting the movement of high-emission vehicles and strengthening control of construction dust.

Second, considering the positive correlation between meteorological factors (e.g., low wind speed and high humidity, which are prone to inversion layer formation) and urbanization rate, it is suggested to optimize the layout of green spaces in urban planning to improve humidity regulation capacity and reduce local pollution source emissions through industrial structure adjustment. In addition, sensitive populations can be reminded in advance to reduce outdoor activities; this can shift from a passive emergency mode that relies solely on current pollution levels to an active, pre-regulatory approach to pollution.

Finally, this periodicity can guide environmental supervision departments in optimizing monitoring frequency and law enforcement intensity. For example, strengthening inspections of pollution sources 3–5 days before the periodic peak can transform management from "passive response" to "active prevention and control." The systematic implementation of these measures will significantly improve the timeliness and accuracy of air quality improvements and support the city's sustainable development.

3.4. Recognition of air quality variation trend based on wavelet decomposition

We further studied the periodic variation characteristics of air pollution in Changchun. The *db3* wavelets were applied to AQI and meteorological factors. Figure 12 shows the plots after the three-layer wavelet decomposition using *db3* for AQI, atmospheric pressure, wind velocity, outdoor humidity, and outdoor temperature.

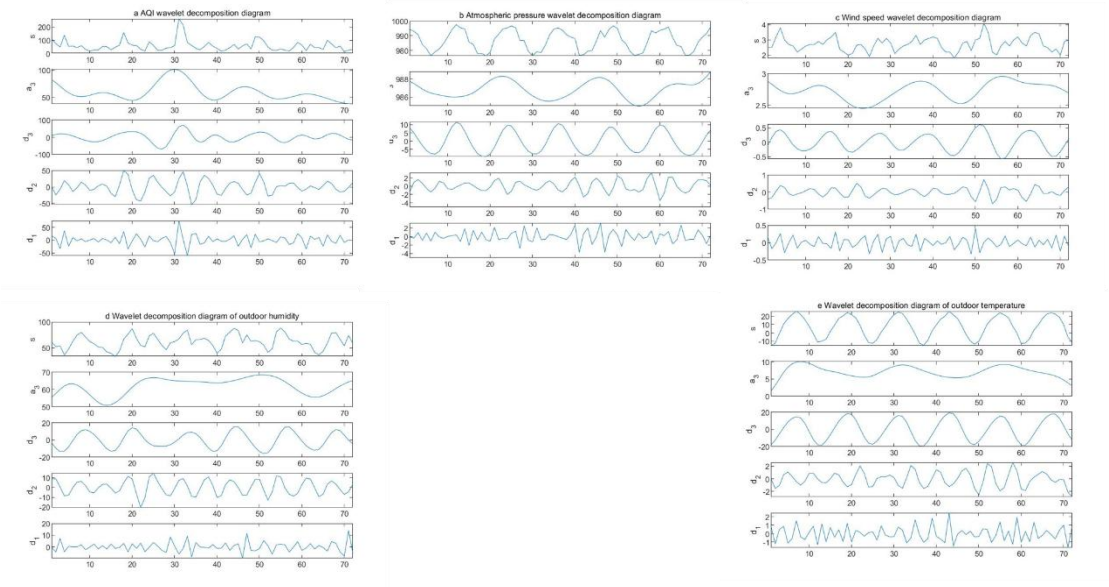

Figure 12. Three-layer wavelet decomposition diagrams: (a) AQI values; (b) atmospheric pressure; (c) wind velocity; (d) outdoor humidity; and (e) outdoor temperature.

The *x*-axis is time (Mth). A variety of factors influence air quality variations. The db wavelet decomposes a complex time series consisting of components of different frequencies into subseries of different frequencies, including a low-frequency component, a_3, and three high-frequency components, d_1, d_2, and d_3. The low-frequency component a_3 captures the major features of the signals, namely

stability and the dominant air quality trend over time.

In signal processing based on the db3 wavelet, each symbol represents the decomposition result at a different scale, with specific definitions as follows:

1. s: Approximation Coefficients

"s" typically denotes the original, undecomposed input signal, which reflects only the overall variation pattern of data over long time scales. In this study, "s" refers to the original monthly AQI series or meteorological indicators (atmospheric pressure, humidity, wind speed, temperature). It is the sum of all decomposed components (d1, d2, d3, a3), i.e., s = a3 + d3 + d2 + d1.

2. a3: Approximation Coefficients at Level 3

It filters out all high-frequency and medium-frequency fluctuations represented by d1, d2, and d3. a3 mainly illustrates the overall evolution trend of AQI or meteorological indicators throughout the entire study period (2018–2023), such as whether AQI shows a continuous decrease, remains stable, or increases.

3. d1, d2, d3: Detail Coefficients

These correspond to the high-frequency components of the signal, reflecting short-term fluctuations at different scales in the 3-level decomposition. The higher the level (d3 > d2 > d1), the longer the fluctuation period:

d1: Detail Coefficients at Level 1, corresponding to high-frequency fluctuations with the shortest period. It reflects the random changes in data over a month, such as small fluctuations in AQI caused by sudden pollution events (e.g., straw burning, short-term unfavorable weather).

d2: Detail Coefficients at Level 2, corresponding to fluctuations with a medium period. It reflects the seasonal variation pattern of data, such as quarterly fluctuations in AQI caused by seasonal changes (e.g., winter heating pollution, summer rainfall purification).

d3: Detail Coefficients at Level 3, corresponding to fluctuations with a longer period. It reflects changes in data over half a year, such as phased fluctuations in AQI caused by annual meteorological anomalies (e.g., drought or heavy rainfall).

By observing the a3 component of AQI, we can determine the change trend of Changchun's overall air quality over the 6-year time scale. For example, if the AQI a3 curve shows a downward fluctuation, it indicates the effectiveness of air pollution control policies in Changchun during this period. We can identify the impact scales of different meteorological factors by comparing the consistency of fluctuations between the high-frequency components (d1, d2, and d3) of meteorological indicators and AQI. This more accurately reveals the main meteorological drivers of AQI changes at different time scales, avoiding confusion between short-term fluctuations and long-term trends.

The AQI values of Changchun were higher in winter and lower in summer across the years from January 2018 to December 2023. There were two peak seasons of air pollution within a year, one in spring and the other in winter. Air pollution is more serious in winter than in spring. This phenomenon can be attributed to Changchun's distinctive pollution-emission characteristics and the influence of meteorological factors during this period. Changchun's heating period lasts from winter to spring the next year, and coal burning results in massive production of suspended particulate matter, reducing air quality. The above finding coincided with the correlation analysis between air pollutants and meteorological factors. As for interannual variations, the a_3 component generally showed a decreasing trend, indicating that Changchun's air quality was much better in 2023 than in other years. Changchun has improved in urban air quality in recent years.

3.5. AQI prediction model elaboration

The AQI values were predicted for Changchun using best-subset regression, which selects the most effective subset of independent variables from the set influencing the dependent variable, following specific criteria. A regression equation was established using the selected independent variables, and the model with the highest coefficient of determination was considered the best subset regression model [40].

We found that if air pollutant emissions were very high the previous year, air quality the following year was more likely to worsen without effective emissions-reduction measures. Since meteorological factors significantly impacted short-term pollutant concentrations and their interannual changes, the distributions and concentrations of the pollutants in the following year would be affected by meteorological factors, as determined by the concentrations of these pollutants in the preceding year [41–44]. The predictive models for AQI, influenced by two meteorological factors, were built using OLS regression. The first model used data on current-year meteorological factors to predict future AQI values; the second model used current-year meteorological factors and the previous year's AQI values to predict future AQI values. The coefficient of determination assessed the goodness-of-fit of the two models, and the optimal model was determined. The core purpose of including the AQI of the same period in the previous year as an independent variable in this study is to proactively capture and control the "intertemporal dependence effect of air quality," rather than enabling temporal autocorrelation to persist.

Logically, short-term air quality (AQI) exhibits a significant "path dependence" characteristic. Specifically, the pollutant stock from the previous year and diffusion foundations (e.g., the cumulative effect of pollutants under a long-term, stable meteorological background, which exerts intertemporal impacts) are important antecedent factors influencing the current year's AQI. This belongs to an "observable and explicit temporal correlation," rather than "unobserved and implicit temporal autocorrelation omitted by the model."

From the perspective of model design, if this variable is not included, the "impact of the previous year’s AQI on the current year’s AQI" will be "omitted" and transferred to the residuals, which would instead lead to "residual autocorrelation" (i.e., the residuals implicitly contain unexplained temporal dependence). By explicitly incorporating it as an independent variable, we essentially convert this "implicit temporal correlation" into an "explicit and quantifiable explanatory variable," which is fundamentally aimed at "eliminating potential temporal autocorrelation in the residuals" rather than "introducing" it.

To verify that "no residual autocorrelation occurs in the model after adding the previous year’s AQI," we conducted the Durbin-Watson (D-W) test. For the "basic model containing only meteorological variables," the D-W value is 2.107; for the "extended model with the previous year’s AQI added," the D-W value increases to 2.057 (close to the ideal value of 2). The above results indicate that adding the previous year's AQI not only does not introduce temporal autocorrelation but also improves the model's rationality and prediction accuracy by "making the temporal dependence explicit." The results using the two Regression models are listed in Tables 2 and 3.

Table 2. The regression results of predicting future AQI values using current-year meteorological factors (monthly data).

	Unstandardized coefficients		Standardized coefficients	*t*	*p*	Collinearity diagnosis	
	B	Standard error	*Beta*			VIF	Tolerance
Constant	534.329	553.196	-	0.966	0.366	-	-
Atmospheric pressure	-0.480	0.563	-0.133	-0.852	0.422	1.221	0.819
Air temperature	0.142	0.248	0.090	0.571	0.586	1.252	0.799
Relative atmospheric humidity	-0.907	0.310	-0.559	-2.927	0.022*	1.838	0.544
Mean wind velocity	22.124	9.063	0.471	2.441	0.045*	1.880	0.532
R^2	0.861						
Adjusted R^2	0.782						
F	$F(4,7)=10.865$, $p=0.004$						
D-W Statistic	2.107						
CI	95%						

Note：Dependent variable = AQI

* $P<0.05$ ** $P<0.01$

In Table 2, atmospheric pressure (hPa), air temperature, relative atmospheric humidity, and mean wind velocity were treated as independent variables, and AQI was the dependent variable. A regression analysis was performed, correspondingly, and the model was built as follows:

$$Y_{1(AQI)} = 534.329\text{-}0.480\,X_1 + 0.142X_2\text{-}0.907X_3 + 22.124\,X_4 \quad (4)$$

where $Y_{1(AQI)}$ is the AQI value to be predicted for the target year; X_1 is atmospheric pressure((hPa)); X_2 is air temperature(℃); X_3 is relative atmospheric humidity (%); and X_4 is mean wind velocity(m/s).

The model's R^2 was 0.861, indicating that atmospheric pressure (hPa), air temperature, relative atmospheric humidity, and mean wind velocity explained 86.1% of the variation in AQI. The model passed the F-test (F=10.865, p=0.004<0.05). The mean wind velocity significantly positively influenced AQI, and relative atmospheric humidity significantly negatively influenced AQI.

With a superposition of the AQI values of the previous year, that is, by treating the AQI of the previous year, atmospheric pressure (hPa), air temperature, relative atmospheric humidity, and mean wind velocity of the previous year as independent variables. The results of the regression analysis are listed in Table 3.

Table 3. The regression results of predicting future AQI values using current-year meteorological factors and previous-year AQI values (monthly data).

	Unstandardized coefficients		Standardized coefficients	*t*	*p*	Collinearity diagnosis	
	B	Standard error	*Beta*			VIF	Tolerance
Constant	471.809	516.688	-	0.913	0.396	-	-
AQI_0	0.555	0.248	0.439	2.235	0.067	2.523	0.396
Atmospheric pressure	-0.401	0.527	-0.120	-0.761	0.476	1.632	0.613
Air temperature	0.738	0.289	0.447	2.551	0.043*	2.005	0.499
Relative atmospheric humidity	-0.941	0.420	-0.737	-3.671	0.010*	2.640	0.379
Mean wind velocity	13.258	7.551	0.273	1.756	0.130	1.587	0.630
R^2	0.908						
Adjusted R^2	0.832						
F	$F(5,6)=11.885$, $p=0.005$						
D-W Statistic	2.057						
CI	95%						

Note: Dependent variable = AQI; AQI_0 is the AQI value of the year preceding the target year

* $P<0.05$ ** $P<0.01$

The constructed model can be expressed as follows:

$$Y_{2(AQI)}=471.809 + 0.555\,X_5 - 0.401\,X_1 + 0.738\,X_2 - 0.941\,X_3 + 13.258\,X_4 \quad (5)$$

where $Y_{2(AQI)}$ is the AQI value prediction for the target year; X_1 is atmospheric pressure (hPa); X_2 is air temperature (℃); X_3 is relative atmospheric humidity (%); X_4 is mean wind velocity (m/s); and X_5 is the AQI value of the year preceding the target year.

The R^2 of the model was 0.908, indicating that AQI of the previous year, atmospheric pressure (hPa), air temperature, relative atmospheric humidity, and mean wind velocity explained 90.8% of the variations of AQI. The model passed the F-test ($F=11.885$, $p=0.005<0.05$). This indicated that air temperature significantly positively influenced AQI, and relative atmospheric humidity significantly negatively influenced AQI.

The 1/VIF ratio was consistently above 0.2 for both models, indicating the absence of multicollinearity. As shown in Figure 13, the residuals from the two models were randomly distributed around the line at zero on the *Y*-axis.

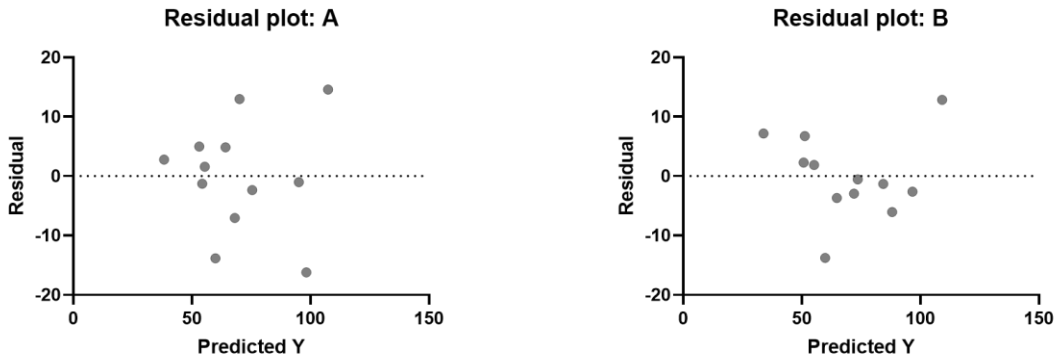

Figure13. Random distribution of residual values via different models: (a) $Y_{1(AQI}$; and (b) $Y_{2(AQI)}$

As shown in the quantile-quantile plot in Figure 14, the residuals approximately followed a normal distribution, satisfying the assumption of normal errors in the multiple linear regression model. Thus, the proposed model was reliable.

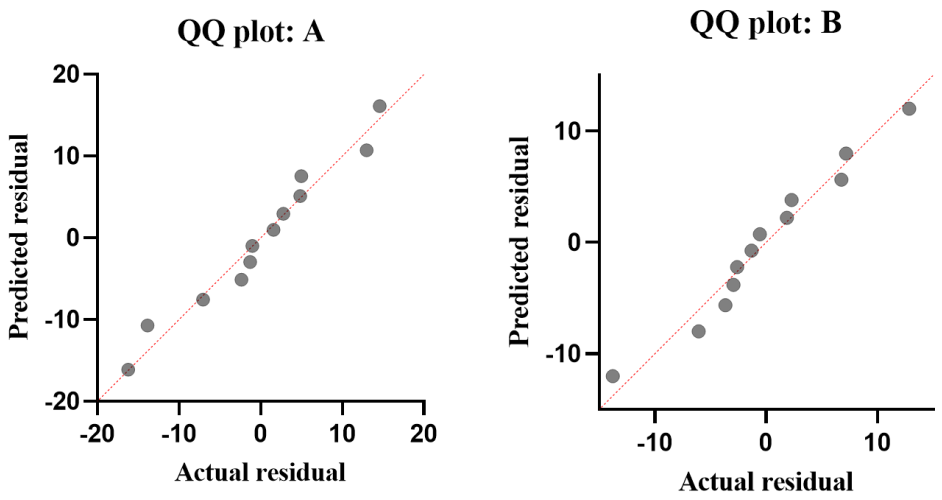

Figure 14. Normal distribution of residual values via different models: (a) $Y_{1(AQI}$; (b) $Y_{2(AQI)}$.

Based on the above analysis, the predictive equation's R2 correlation coefficient increased after superimposing AQI values from previous years. This predictive equation was more accurate in the short-term AQI prediction in Changchun.

4. Conclusions and limitations

4.1. Conclusions

The results obtained in this study made it possible to draw the following conclusions:

1. Concentrations of six air pollutants ($PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO) and short-term AQI in nine administrative regions of Changchun, China, displayed distinct seasonal patterns from 2018 to 2023. Except for the daily mean O_3 concentration, which showed a unimodal distribution, the daily mean concentrations of all other pollutants peaked in winter and spring, then decreased afterward.

2. The correlation analysis revealed that some socioeconomic and meteorological factors significantly correlated with pollutant concentrations. The urbanization rate showed a significant positive correlation with $PM_{2.5}$ concentration; urban green coverage showed a significant negative correlation with NO_2 and CO concentrations. Air temperature had a significant positive correlation with both O_3 and NO_2 concentrations. Relative atmospheric humidity significantly negatively correlated with $PM_{2.5}$, PM_{10}, and SO_2 concentrations. Wind velocity had a significant negative correlation with $PM_{2.5}$ and PM_{10} concentrations. The above driving factors controlled the variation characteristics of air pollution in the area under study.

3. The wavelet transform analysis revealed the main periodic coupling phenomenon between the first and the second dominant periods of the time series of AQI value and $PM_{2.5}$ and PM_{10} concentrations, respectively. The main periodic coupling phenomenon was identified between the time series of AQI values and those of six pollutants in 2020, indicating that combined pollution was a major feature of atmospheric pollution in Changchun. The low-frequency components of wavelet signals were attenuated over time, indicating a significant improvement in air quality in Changchun. The main periodic coupling indicates that the same type of periodic driving factors regulates multiple air pollutants. This characteristic can directly guide coordinated control of pollution sources, avoiding a focus on single-pollutant control while ignoring coordinated emissions from other pollutants, thereby improving control efficiency. Moreover, the coordinated early warning mechanism for multiple pollutants can be optimized based on the consistent rhythm of the coupling period.

The regression analysis results indicate that the regression equation incorporating the AQI value of the previous year helps optimize the short-term air quality prediction results.

4.2. Limitations

The above conclusions are based on the air quality analysis of Changchun from 2018 to 2023. However, this study has the following three major limitations.

(1) Limitation in research time scope: The data of this study spans 6 years (2018–2023), which can reflect the variation characteristics of air quality and short-term prediction rules during this period. However, it does not cover the impacts of long-term driving factors, such as climate cycles over longer time scales (e.g., 10 years or more) and in-depth adjustments to the industrial structure. The long-term applicability of the conclusions needs further verification.

(2) Lack of emission inventory data: We do not include Changchun's pollution source emission inventory data and only conducts analysis based on monitored concentration data, meteorological data, and socioeconomic data. It is difficult to accurately quantify the contribution ratio of different pollution sources (such as industry, transportation, and dust) to air quality, which limits the depth of analysis on the causes of pollution.

(3) Lack of verification with external datasets: The short-term AQI prediction equation constructed in this study is only fitted and verified based on internal data from 9 monitoring stations in Changchun. It has not been cross-validated using independent external datasets (e.g., monitoring data from surrounding cities and air quality data from third-party platforms). The equation's external applicability and prediction stability need further improvement.

The follow-up study envisages mitigating these limitations as follows:

The research period will be extended, and data with longer time series can be included to reveal the long-term evolution rules of air quality and its correlation with macro-climatic driving factors. Furthermore, high-resolution spatiotemporal emission inventories (e.g., the MEIC inventory) can be integrated, and environmental concentration data can be combined with source emission data to more

accurately assess the contribution rates of various pollution sources, thereby providing a direct scientific basis for source-specific control.

In the future, researchers can further explore the nonlinear, synergistic, and antagonistic effects among pollutants, especially in the context of composite pollution. In addition, advanced data mining technologies, such as machine learning (e.g., random forest, neural networks), can be used to handle complex nonlinear relationships among variables, or numerical simulations can be conducted using chemical transport models. These efforts aim to build a more accurate prediction and source-tracking system and, eventually, shift from statistical correlation to mechanism-based cognition.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This project was supported by the Doctoral Research Initiation Fund of Jilin Normal University (0420002).

Conflict of interest

The authors declare no conflict of interest.

References

1. Zhang L, Ji A, Liu M, et al. (2020) Spatiotemporal variations and influencing factors of $PM_{2.5}$ concentrations in Beijing, China. *Environ Pollut* 262: 114276. https://doi.org/10.1016/j.envpol.2020.114276
2. Maji KJ, Sarkar C (2020) Spatio-temporal variations and trends of major air pollutants in China during 2015-2018. *Environ. Sci Pollut R* 27: 33792-33808. https://doi.org/10.1007/s11356-020-09646-8
3. Bai X, Tian H, Liu X, et al. (2021) Spatial-temporal variation characteristics of air pollution and apportionment of contributions by different sources in Shanxi Province of China. *Atmos Environ* 244: 117926. https://doi.org/10.1016/j.atmosenv.2020.117926
4. Chen J, Wang B, Huang S, et al. (2020) The influence of increased population density in China on air pollution. *Sci Total Environ* 735: 139456. https://doi.org/10.1016/j.scitotenv.2020.139456
5. Han L, Zhou W, Li W, et al. (2014) Impact of urbanization level on urban air quality: a case of fine particles ($PM_{2.5}$) in Chinese cities. *Environ Pollut* 194: 163-170. https://doi.org/10.1016/j.envpol.2014.07.022
6. Wang S, Cheng S, Qi X, (2020) Effect of urban greening on incremental $PM_{2.5}$ concentration during peak hours. *Front Public Health* 8: 551300. https://doi.org/10.3389/fpubh.2020.551300
7. Fang C, Liu H, Li G, et al. (2015) Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability-Basel* 7: 15570-15592. https://doi.org/10.3390/su71115570

8. Fernando DC, Jaci MBS, Nisia K (2003) Periodicity of atmospheric phenomena occurring in the extreme South of Brazil. *Atmos Sci Lett* 5: 65–76. https://doi.org/10.1016/j.atmoscilet.2003.12.003
9. Lei Y, Zhang X, Tang Y, et al. (2015) Holiday effects on $PM_{2.5}$ and other major pollutants in Beijing. *Acta Scientiae Circumstantiae* 35: 1520-1528.
10. Schichtel BA, Husar RB, Falke SR, et al. (2001) Haze trends over the United States，1980-1995. *Atmos Environ* 35: 5205-5210. https://doi.org/10.1016/S1352-2310(01)00317-X
11. Rokjin JP, Daniel JJ, Naresh K, et al. (2006) Regional visibility statistics in the United States ： Natural and transboundary pollution influence，and implications for the regional haze rule. *Atmos Environ* 40: 5405-5423. https://doi.org/10.1016/j.atmosenv.2006.04.059
12. Yao Y, Zou Q, Chen C, et al. (2014) The Analysis of Chemical Component of $PM_{2.5}$ on Haze Formation in Suzhou City. *Environ Monitor China* 30: 62-68.
13. Xie J, Sun T, Liu C, et al. (2022) Quantitative evaluation of impacts of the steadiness and duration of urban surface wind patterns on air quality. *Sci Total Environ* 850: 157957. https://doi.org/10.1016/j.scitotenv.2022.157957
14. Maria I, Yan Z, Liu Y, et al. (2016) Assessing the possible impacts of temperature change on air quality and public health in Beijing, 2008–2012. *Nat Hazards* 84: 153-165. https://doi.org/10.1007/s11069-015-2061-7
15. Khaiwal R, Tanbir S, Akash B, et al. (2021) Impact of COVID-19 lockdown on ambient air quality in megacities of India and implication for air pollution control strategies. *Environ Sci Pollut R* 28: 21621-21632. https://doi.org/10.1007/s11356-020-11808-7
16. Tan S, Xie D, Ni C, et al. (2023) Spatiotemporal characteristics of air pollution in Chengdu-Chongqing urban agglomeration (CCUA) in Southwest, China:2015-2021. *J Environ Manage* 325: 116503. https://doi.org/10.1016/j.jenvman.2022.116503
17. Quan J, Xu X, Jia X, et al. (2020) Multi-scale processes in severe haze events in China and their interactions with aerosols: Mechanisms and progresses. *Chin Sci Bull* 65: 810-824. https://doi.org/10.1360/TB-2019-0197
18. Zamora, ML, Peng J, Hu M, et. al. (2019) Wintertime aerosol properties in Beijing. *Atmos Chem Phys* 19: 14329–14338. https://doi.org/10.5194/acp-19-14329-2019
19. Li C (1990) The intraseasonal oscillation in the atmosphere. *Scientia Atmospherica Sinica* 14: 32-45.
20. Li Z, Wang Y, Xu Z, et al. (2021) Characteristics and sources of atmospheric pollutants in typical inland cities in arid regions of central Asia: A case study of Urumqi city. *PLoS One* 16: e0249563. https://doi.org/10.1371/journal.pone.0249563
21. Xu T, Zhang C, Liu C, et al. (2023) Variability of $PM_{2.5}$ and O_3 concentrations and their driving forces over Chinese megacities during 2018-2020. *J Environ Sci* 124: 1-10. https://doi.org/10.1016/j.jes.2021.10.014
22. Duan J, Huang R J, Wang Y, et al. (2024) Measurement Report: Size-resolved secondary organic aerosol formation modulated by aerosol water uptake in wintertime haze. *Atmos Chem Phys* 24: 7687–7698. https://doi.org/10.5194/acp-24-7687-2024
23. Xiong H, Liu X, Sun C, et al. (2025) Atmospheric water cluster–catalyzed formation of nitroaromatics as a secondary aerosol source. *Sci adv* 11: eadv7805. https://doi.org/10.1126/sciadv.adv7805
24. Fotheringham AS, Yue H, Li Z (2019) Examining the influences of air quality in China′s cities using multi-scale geographically weighted regression. *T GIS* 23: 1444-1464. https://doi.org/10.1111/tgis.12580

25. Wang J, Xu C (2017) Geodetector: Principle and prospective. *Acta Geographica Sinica* 72: 116-134.
26. Oshan TM, Li Z, Kang W, et al. (2019) MGWR: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS Int J Geo-Inf* 8: 269. https://doi.org/10.3390/ijgi8060269
27. Cazelles, B, Chavez M, Berteaux D, et al. (2008) Wavelet analysis of ecological time series. *Oecologia* 156: 287-304. https://doi.org/10.1007/s00442-008-0993-2
28. Kumar A (2017) Spectral Analysis of Particulate Matter in the Atmosphere using wavelet transforms. *International Journal of Advance Research, Ideas and Innovations in Technology* 3: 155-159.
29. Zhang H, Zhang S, Wang P, et al. (2017) Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. *J Air Waste Manage.* 67: 776-788. https://doi.org/10.1080/10962247.2017.1292968
30. Changchun survey team of National Bureau of Statistics, (2023) Changchun Statistical Yearbook 2023, Peking: China Statistics Press, 49-61.
31. Stoy PC, Katul GG, Siqueira MBS, et al. (2005) Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis. *Tree Physiol* 25: 887-902. https://doi.org/10.1093/treephys/25.7.887
32. Wang W, Ding J, Xiang H (2002) Application and prospect of wavelet analysis in hydrology. *Advances in Water Science* 13: 515- 520.
33. Gao J, Jian M, Man W, et al. (2006) On the denoising method of prestack seismic data in wavelet domain. *Chin J Geop* 49: 1155-1163.
34. Jiang X, Liu S, Ma M, et al. (2009) A wavelet analysis of the precipitation time series in Northeast China during the last 100 years. *Geog Res* 28: 354-362.
35. He W, Bu R, Xiong Z, et al. (2013) Characteristics of temperature and precipitation in Northeastern China from 1961 to 2005. *Acta Ecologica Sinica* 33: 519-531. https://doi.org/10.5846/stxb201111241799
36. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *B Am Meteorol Soc* 79: 61-78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2
37. Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Commun Pur Appl Math* 41: 909-996. https://doi.org/10.1002/cpa.3160410705
38. Yang H (1980) The effectiveness of sunflowers in addressing air pollution lies in their ability to absorb a substantial amount of nitrogen dioxide (NO_2). *Environ Sustain Dev* 21: 10-11.
39. Quan J, Gao Y, Zhang Q, et al. (2013) Evolution of planetary boundary layer under different weather conditions, and its impact on aerosol concentrations. *Particuology* 11: 34–40. https://doi.org/10.1016/j.partic.2012.04.005
40. Yang Y (2015) Comparative research of regression model selection criteria. *J Kunming Univ Sci Technol* 40: 134-138.
41. Zhong Q, Ma J, Shen G, et al. (2018) Distinguishing emission-associated ambient air $PM_{2.5}$ concentrations and meteorological factor-induced fluctuations. *Environ Sci Technol* 52: 10416–10425. https://doi.org/10.1021/acs.est.8b02685
42. Zhang Q, Zheng Y, Tong D, et al. (2019) Drivers of improved PM2.5 air quality in China from 2013 to 2017. *Proc Natl Acad Sci USA* 116: 24463–24469. https://doi.org/10.1073/pnas.1907956116

43. Xiao Q, Zheng Y, Geng G, et al. (2021) Separating emission and meteorological contributions to long-term PM2.5 trends over eastern China during 2000–2018. *Atmos Chem Phys* 21: 9475–9496. https://doi.org/10.5194/acp-21-9475-2021
44. Chang Y, Liu W, Zhou X (2024) Characteristics of atmospheric volatile organic compounds and their relationship with ozone concentration in Lanzhou based superstation observation. *Environ Chem* 43: 1025-1034.

AIMS Press