Environment, Development and Sustainability
https://doi.org/10.1007/s10668-025-06113-7

Spatial–temporal effects of industrial pollutants and economic development on depopulation in the mega urban agglomeration

Xinyue Gu[1] · Zijian Miao[2] · Yan Mao[3] · Xintao Liu[1]

Received: 14 October 2024 / Accepted: 28 February 2025

Abstract
The negative effects of industrial pollutants on the whole environment have long been a contentious issue in urbanization. While studies have analyzed the impacts of the economy and industry on depopulation separately, few focus on their combined and interactive effects at the grid level. This study investigates the complex impacts of depopulation from 2010 to 2020 in China's Beijing-Tianjin-Hebei agglomeration. Using remote-sensing and geoinformatics data, we calculated the composite economic index and the composite pollution index to observe the spatial–temporal effects on depopulation. Subsequently, the study employs the interpretable ensemble model to establish the relations between industrial pollutants, economic variables, and population density and obtained nine distinct change types. Findings reveal that many areas still confront population decline due to industrial pollution despite GDP remaining an attraction to the population. The proportion of this type of area in the nine types of changes over 10 years has ranked second steady, reaching 27.175%, 26.964%, and 29.569% in 2015–2010, 2010–2015, and 2010–2020, respectively. In addition, combined with the visualization map of spatial distribution, it is observed that the problem of population loss is particularly serious in lagging cities. Furthermore, the top three pollutants with the most prominent adverse effects are O_3, $PM_{2.5}$, and PM_{10}. This study offers important insights into sustainable urban development.

Keywords Industrial pollution · Economic development · Population decline · Combined effects · Sustainable development

1 Introduction

The Industrial Revolution has facilitated rapid social and economic development worldwide. However, it has also presented numerous challenges, including energy and resource crises and ecological and environmental crises (Agboola et al., 2024; Alotaibi et al., 2024). Consequently, a range of air pollution issues has emerged, which has gained significant attention from various sectors of society (Dong et al., 2020). Research has shown that industrial pollution has gradually become a critical factor constraining sustainable

Extended author information available on the last page of the article

Published online: 20 March 2025

economic growth and social development (Miao & Gu, 2024), which deserves special attention.

With the transformation of an era heavily reliant on heavy industry development, research on industrialization gradually shifted towards environmental effects in the twenty-first century. Many studies have found a positive correlation between industrial and economic development (Gu et al., 2024a, 2024b). However, other research also found that industrial city populations have become more vulnerable to environmental risk factors and changes (Dvir et al., 2024). This vulnerability has led individuals to leave industrial cities for better living conditions (Kovats, 2012).

Research on the relationship between air pollutants and population mainly focuses on the following two aspects: Firstly, studying population outflow caused by natural environmental disasters. These studies often explore the long-term effects of environmental disasters on individual willingness to relocate based on national or regional scales (Brottrager et al., 2023; Grace et al., 2018). Secondly, research on the impact of air pollution on population outflow in the city scales (Chen et al., 2022; Liu & Yu, 2020). However, limited by the low precision of panel data, research on environmental quality data at a large spatial scale often focuses on describing overall characteristics, inevitably overlooking local variations within the scale.

Significant regional differences in industrial development exist between different regions in China and within cities, and the spatial imbalance of pollution emissions is also prominent (Shen et al., 2017). Industry's impact on the environment varies within the same province or city. Hence, the effects of industrial pollution on the microscale of the population still need to be confirmed. Additionally, previous studies often limited their duration to within 5 years (Germani et al., 2021; Liu & Yu, 2020; Zhao et al., 2021), since population changes occur over a long period, the relationship between population, industrial pollution, and economic levels over a longer period need to be explored further.

Regarding the methods, traditional studies mainly rely on various regression models employing linear regression analysis, such as Multivariate Regression models (Tella et al., 2021) or Land Use Regression models (Son et al., 2018) to understand the relationship between industrial pollution, population loss, and urban development indicators using. In general, linear regression models are relatively simple in computation and implementation, but non-linear problems and low-precision performance limit them (Gu et al., 2024a, 2024b; Ma et al., 2020).To better understand the complicated effects of economic or pollution and population, researchers have started using non-linear spatial econometric models (Huang et al., 2020; Li & Wu, 2017).

Machine learning (ML) is the scientific and statistical models that construct a mathematical model based on sample data, also known as training data (Ma & Cheng, 2016). In general, the causes of air pollutants and their spatial impact mechanisms are complex, and pollutants exhibit high dynamics, volatility, and spatiotemporal heterogeneity (Liu et al., 2022). ML can utilize massive amounts of observational data for training and testing, effectively capturing patterns and temporal periodicity in pollutant concentration changes. Therefore, based on previous studies (Balogun et al., 2021), ML is used to elucidate the complex non-linear effects of economic development and industrial pollution on urban depopulation.

However, studying the combined effects of economic and industrial development still requires further verification. Although previous research on using machine learning has achieved good results in modeling nonlinear problems, most models are still black-box models that lack interpretability and cannot evaluate the importance of variables nonlinearly. Therefore, Shapley Additive exPlanations (SHAP) proposed by Roth (1988) was used

as an interpretation method combined with ML, which can calculate the contribution of each variable in the overall relationship, namely the importance of variables (Gu et al., 2025).

Furthermore, previous studies have rarely delved into the interaction effects between economic development and industrial pollution variables influencing the population. Interaction effects are the simultaneous effects of two or more independent variables on the dependent variable. These reveal how economic development and industrial pollution jointly influence urban population changes, identifying which variable combinations are most sensitive to these changes. Hence, it has become imperative to investigate the synergistic effects among various variables.

Based on these backgrounds, the Beijing-Tianjin-Hebei (BTH) urban agglomeration is an excellent testbed. From 2000 to 2021, the urbanization rate of BTH has increased from 38.92 to 69.35%, according to the National Bureau of Statistics of the People's Republic of China, 2020. Meanwhile, industrialization issues in BTH are particularly prominent, and the expansion of heavy industries has led to severe air pollutants (Lu et al., 2022). According to data released by the Ministry of Environmental Protection of China, 8 out of the 20 cities with the most severe pollution in 2022 are in BTH (Wang & Ge, 2023). Furthermore, some research has indicated that the BTH has reached a mature stage of development, characterized by a declining population and shrinking cities (Jiang & Lv, 2022; Lyu & Jiang, 2022). Hence, the complex impact of economic and industrial development on the internal population needs to be further explored and confirmed.

In this context, the main objective of this study is to examine population inequality across various regions under the combined influences of economic and pollution factors. To achieve this, the methodology entails utilizing high-resolution urban data (1 km*1 km) from multiple sources to calculate the CEI and CPI for each region. Subsequently, an interpretable LightGBM and SHAP model was employed to analyze the nonlinear effects of economic and industrial pollution factors on the population. The main contributions of this study can be summarized as follows: (1) Observing the combined impact of composite economic development and industrial pollution index. (2) Describing the inequal combined effects of population decline on the micro-scale within cities. (3) Explaining the detailed impact of different factors on the population using interpretable machine learning models. This article brings new perspectives to the field, providing valuable insights for researchers and policymakers regarding economic development and pollution control in the megaregion.

2 Research materials

2.1 Research area

This study selects the Beijing-Tianjin-Hebei (BTH) urban agglomeration as the study area. The total area is 218,000 square kilometers, with a population of 113 million in 2019 and a GDP of 8.46 billion yuan. Located in northern China, BTH is one of the largest and most economically vibrant urban agglomerations. However, the heavy industry in this region used to be highly developed (Lyu & Jiang, 2022), with multiple cities included in the "National Plan for the Adjustment and Transformation of Old Industrial Bases (2013–2022)". Therefore, the contradiction between urban development and air pollutants in BTH has become increasingly prominent (Liang et al., 2019).

As shown in Fig. 1, BTH includes 14 cities, namely Beijing and Tianjin, 11 cities in Hebei Province, and Anyang City in Henan Province. China's capital, Beijing, sits at the center of the BTH and is the central city of this urban agglomeration. Due to the gradual population decline in this area since 2010 (Meng & Long, 2022), we select three-time points, 2010, 2015, and 2020, to explore the impact of economic development and industrial pollution on population outflows in the BTH. In addition, to better understand the impact at a meso-micro scale, the study area is divided into 379,228 grids using a 1×1 km fishnet based on GIS Pro.

2.2 Data collection and processing

All the data collected and used in the study are presented in Table 1. The dependent variable is the population density of each grid. Multiple high-resolution urban datasets were

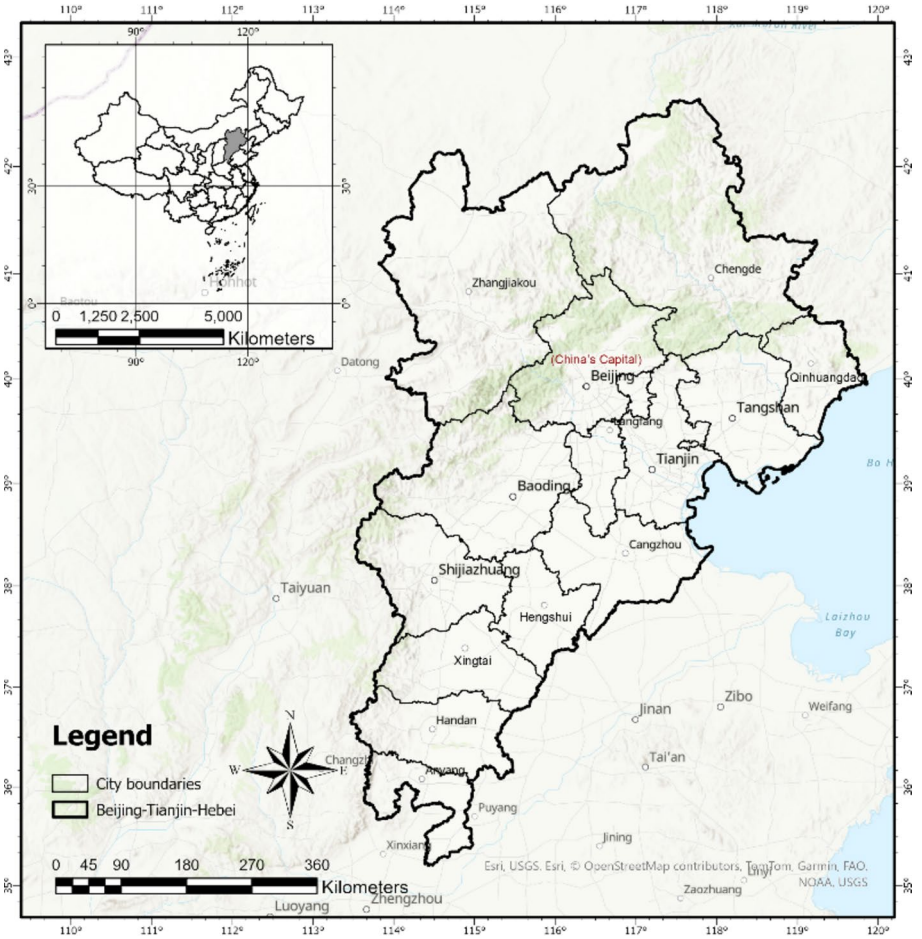

Fig. 1 The geographical location of BTH urban agglomeration in China

Table 1 Descriptions of dependent, explanatory, and control variables

Name	Description	Resolution	Data Source	References
Dependent variable				
PD	The number of people per square kilometer is based on the country's totals adjusted to match the official United Nations population estimates	1 km	WorldPop (WorldPop, & Bondarenko, 2020)	Yu et al. (2023)
Independent variables of economic development (explanatory)				
GDP	It is estimated with nighttime imagery and economic survey data	1 km	Resource and Environmental Science Data Platform	Yan et al. (2022)
RD	This data is calculated by dividing the number of roads by the grid area	vector data	OSM	Laouzai and Ouafi (2022) Shuhua et al. (2023)
Independent variables of industrial pollution (explanatory)				
NO_2	It contains estimated global NO_2 surface values derived using a Land Use Regression (LUR) model	0.83	Earth data (Ananberg, 2023)	Cui et al. (2019), Liu and Yu (2020)
$PM_{2.5}$	This is a yearly ground-level $PM_{2.5}$ in China to describe the particles with a size of 2.5 microns	1 km	ChinaHighPM$_{2.5}$ (Wei et al., 2021a, 2021b)	Chen et al. (2022), Liu and Yu (2020)
PM_{10}	This is a yearly ground-level PM_{10} in China to describe the particles with a size of 10 microns	1 km	ChinaHighPM$_{10}$ (Wei et al., 2021a, 2021b)	Liu and Yu (2020)
O_3	It is a yearly spatial distribution of O_3 in China	1 km	ChinaHighAirPollutants, CHAP	Liu and Yu (2020)
SO_2	It is a yearly spatial distribution of SO_2 in China	1 km		Hao et al. (2015), Jiang et al. (2020)
Independent variables of demographic structure (control)				
Old	It reflects the number of people (60 years +) per grid	1 km	WorldPop	Fan et al. (2022)

selected for 2010 and 2020 as explanatory variables to explore the impact of economic development and industrial pollution on population density (PD).

First, we collected the China population density data (1 km resolution) in 2010, 2015, and 2020, which were sourced from WorldPop and extracted and aggregated to each grid. The rate of change in population data reflects the population trend over a decade. The unit of the dataset is population per square kilometer, based on the corresponding official United Nations population estimates, adjusted for country totals (WorldPop, & Bondarenko, 2020).

Secondly, a series of ambient air pollutants remote sensing datasets in 2010, 2015, and 2020 are collected and integrated. These pollutants include SO_2, NO_2, $PM_{2.5}$, PM_{10}, and O_3. These substances are all associated with exhaust emissions generated during industrial production, significantly impacting the environment. These are commonly used variables in previous studies. The selected reference, source, and resolution of each dataset can be seen in Table 1. It is worth noting that there is a lack of data for 2010 as the high-precision SO_2 remote sensing dataset is only available up to 2013. Therefore, this study obtained data for the corresponding year through the downscaling process in Sect. 3.1.

Finally, the study also collected some control socio-economic data, GDP, and road density, to present the economic development of cities. Many studies have shown that GDP is significantly related to the economic level of cities and a two-way causal relationship between environmental pollution and economic growth (Yan et al., 2022). Efficient transportation systems have a positive impact on urbanization, and the higher the urbanization level of a city, the greater the density of its road network, and the higher the economic level of a city. Road density is chosen as an indicator to represent road traffic volume, which has a strong correlation with the economic development of cities (Liu et al., 2020). The road vector data is derived from the OSM historical dataset, but its dataset only covers the years 2014–2020. Based on the fact that the construction of roads in cities is relatively regular and constantly evolving and improving, in this study, the data for all these years were aggregated using the grid where the study area is located, and then the data for the year 2010 were fitted by a linear regression model. Additionally, to eliminate the influence of depopulation caused by demographic structure, this study incorporated one other control variable (elderly population) that reflects the demographic situation. This 1 km grid-level dataset for the research year is also attained from Worldpop.

2.3 Data accuracy verification

Since the integrated and estimated data in this study are based on remote sensing data, there is some uncertainty. Numerous studies have shown that too small a study cell magnifies the inaccuracy of the data, while aggregation to larger blocks means loss of information. In contrast, the use of a 1 km*1 km grid as the unit of analysis provides relatively good statistics of satellite data from different sources and can provide more reliable data estimates.

To estimate the accuracy of the data, we counted the most important PD and various pollutant data (2020) at the municipal unit of BTH and analyzed them against the official government-released demographic yearbook data and pollutant monitoring station data averages in spatial distribution. As shown in Fig. 2, when the estimated data coming from remote sensing data and the observed data officially released by the government are normalized by the maximum and minimum, the distribution and variability of the data on the municipal units maintain a high degree of consistency. Since the focus of this study is on

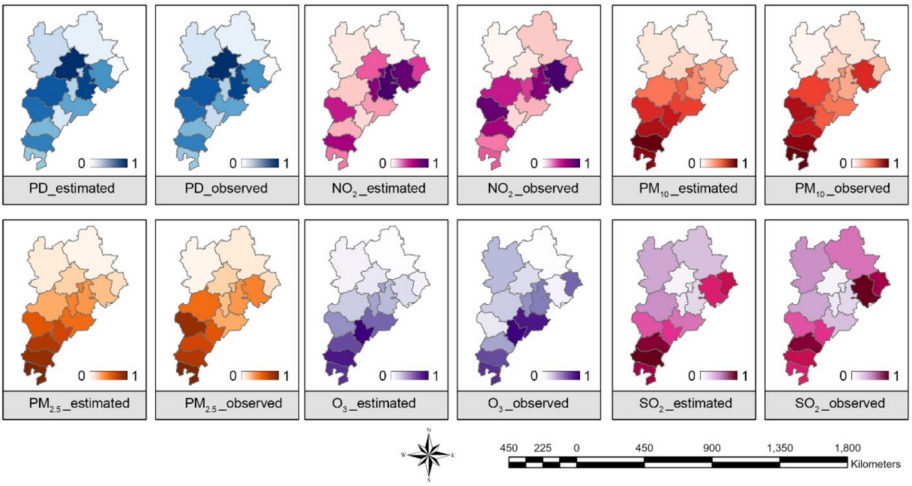

Fig. 2 Comparing the estimated grid-level data and observed data by the government

analyzing the spatial patterns of relative changes between different regions and the differential distribution of population contraction, a small amount of error in the data is acceptable. This type of high-precision remote sensing data has also been used extensively in related urban science and environment studies because of the breadth and relative accuracy of the data, resulting in more detailed results of urban unit characterization.

Overall, all explanatory and control variables were extracted and linked to the grids using the multi-value-to-point tool in GIS Pro. The point location values were interpolated using the bilinear interpolation method. After data normalization, cleaning, and collinearity testing, out of the initial 379,228 grids in the study area, 375,944 grids were included in the subsequent model construction and data analysis.

3 Methodology

3.1 Downscaling of ambient air pollutant

SO_2 is a non-negligible pollutant in air pollution, but due to the lack of 1 km SO_2 data monitoring products before 2013, this study adopts downscaling to realize the generation of 1 km SO_2 data from 2010. The technical core of downscaling is to construct a data relationship with the 2015 1 km SO_2 values by finding data available for both 2010 as well as 2015, and then mapping this relationship to the 2010 data to calculate the 2010 1 km SO_2 values. Details of the implementation are given below:

The first step is to collect variables related to the spatial distribution of SO_2. According to previous studies, there are many factors closely related to the spatial distribution of SO_2, one is the atmospheric factor, and the other is the terrestrial factor. For atmospheric factors, CO2 emissions, aerosols, and six commonly used near-surface atmospheric moisture indices were selected in this study, all of which were satellite remote sensing image data. For terrestrial factors, NDVI was selected to characterize the vegetation cover and unnatural surface cover of the land. In addition, SO_2 in different spatial and temporal dimensions will

be correlated with the spatial distribution of the current year. Therefore, we additionally acquired the SO_2 distribution dataset with an accuracy of 0.25degree for the same year, the SO_2 emissions dataset with an accuracy of 0.1°, and the SO_2 distribution dataset for 2013 with an accuracy of 10 km. Due to the spatial autocorrelation of SO_2 itself, the study further refers to digital image processing by using the tool Neighborhood Summary Statistics in ArcGIS Pro to calculate the mean values of the nearest 4-neighborhood and 8-neighborhood pixels of each pixel, which are collectively included as variables in the subsequent downscaling model. Full details of the variables in this step and the data sources are shown in Table 2.

Subsequently, the full satellite data values were INTEGRATED into a 1 km grid in ArcGIS Pro, nulls and outliers were removed, and the data were converted into pure dimensionless values using Min–Max normalization to facilitate the ability to compare metrics of different units and magnitudes. We then used the machine learning model LightGBM (see Sect. 3.3.2 for details) to construct a model of the relationship between the values of 1 km SO_2 in 2015 and the variables in Table 2 and to predict the values of 1 km SO_2 in 2010 based on this relational model. Here, the data of 2015 is divided into training and testing sets according to the ratio of 70% and 30%, and the parameters of the model after hyper-parameter tuning are learning rate = 0.35, max depth = 11, and estimators = 500. The predicted performance of the model is good, as shown in Table 3.

Finally, to verify the reliability of the results, we also carried out a pre-experimental model, first constructing a model with 2020 data to generate the 1 km SO_2 values for 2015 and measuring the accuracy of the model predictions against the values from the original 2015 dataset, which yielded ideal predictions.

Table 2 Descriptions of impact factors for SO_2 downscaling model

Name	Resolution	Data Source
Atmospheric factors		
CO_2 emission	0.1°	Emissions Database for Global Atmospheric Research
Aerosol Optical Depth (AOD)	1 km	MCD19A2.061 from GEE
actual vapor pressure (AVP)	1 km	HiMIC-Monthly
vapor pressure deficit (VPD)		
dew point temperature (DPT)		
mixing ratio (MR)		
specific humidity (SH)		
Land factor		
Normalized Difference Vegetation Index (NDVI)	1 km	China Resource and Environmental Science Data Platform
SO_2 factors		
SO_2	0.25°	OMI/Aura Sulphur Dioxide (SO_2) Total Column
SO_2 emission	0.1°	Emissions Database for Global Atmospheric Research
SO_2 emission (N4)		
SO_2 emission (N8)		
SO_2_2013	1 km	Multiple air pollution dataset (MuAP)
SO_2_2013 (N4)		
SO_2_2013 (N8)		

Table 3 The performance evaluation metrics of the LightGBM regression model

	R-square	MAE	MSE	RMSE
2020_Train model	0.8932	0.0259	0.0013	0.0366
2020_Test model	0.8623	0.0283	0.0017	0.0418
2015_Train model	0.9173	0.0201	0.0088	0.0296
2015_Test model	0.9007	0.0216	0.0011	0.0325

3.2 Calculation of Composite Index: Entropy Weight Method (EWM)

Referring to previous studies (Wan et al., 2015), this study adopted the EWM to determine the weights of the variables mentioned in Table 1, which include four economic and ten types of pollutant data. These data were normalized and weighed using Python, resulting in the calculation of the composite economic index (CEI) and composite pollution index (CPI) for 2010, 2015, and 2020. The main formulas involved in the measures are as follows:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln\left(p_{ij}\right), j = 1, \ldots, m \tag{1}$$

$$s_i = \sum_{j=1}^{m} \frac{1 - e_j}{\sum_{j=1}^{m} 1 - e_j} x_{ij}, i = 1, \ldots, n \tag{2}$$

where, e_j means the entropy value of the j indicator, and s_i represents the overall score of the system. x_{ij} is the normalized data of each indicator and p_{ij} means the probability of the occurrence. Note that the indicators under the facilities system are all positive, whereas, under the perceptions system, openness as well as greenness are positive while enclosed is negative.

3.3 Analysis of impact factors: LightGBM model and SHAP

3.3.1 Noncollinearity test

To ensure the reliability of our analysis, it is essential to test for noncollinearity among the independent variables. In this study, we employ the Ordinary Least Squares (OLS) regression model to detect multicollinearity. The primary purpose of using the OLS model is to examine the Variance Inflation Factor (VIF) values. VIF values exceeding 10 indicate significant multicollinearity issues, which can distort the regression results. By identifying and addressing these issues, we ensure that our LightGBM model will provide more accurate results.

3.3.2 Regression analysis

To analyze and explain the combined effects of the variables on the urban population, this study combines the advanced LightGBM model with SHAP. LightGBM regression model, an integrated ML algorithm based on decision trees and gradient boosting, has

demonstrated superior performance to artificial neural networks and traditional linear statistical models regarding accuracy and generalization (Ke et al., 2017). Due to this study's large amount of data and diverse variables, we utilized ML regression models to construct the nonlinear relationships between the change rates of the eight independent variables and the dependent variable (population density) from 2010 to 2020.

Additionally, hyperparameter tuning was conducted to optimize the performance of the LightGBM model. This process involved adjusting parameters such as the number of leaves, learning rate, max depth, and feature fraction to minimize the loss function and enhance the model's predictive accuracy. Grid search and cross-validation techniques were employed to ensure robust model selection and prevent overfitting.

3.3.3 Interactive effects interpretation

Subsequently, the SHAP model was employed to decouple the nonlinear relationships further and observe the impact of individual variables and the combined effect of multiple variables on the dependent variable. SHAP is an effective eXplainable AI (XAI) technique that can model, interpret, and visualize complex processes (Kuhn & Tucker, 2016). SHAP builds upon this concept and quantifies the contribution value of each feature in the model, considering the dimensions of individual observations (Štrumbelj & Kononenko, 2014). Therefore, this study applied the SHAP to explain the LightGBM model and decouple the importance of each variable in detail. All data computations, cross-validation, training, testing, and model tuning were performed using Python.

To provide a comprehensive analysis, we also explored the interactive effects of variables using SHAP interaction values. This method allows us to identify and visualize how pairs of variables jointly influence the population density. By analyzing these interactions, we gain deeper insights into the complex interdependencies and nonlinear dynamics present in the data. All data computations, cross-validation, training, testing, and model tuning were performed using Python, leveraging libraries such as LightGBM, SHAP, and Scikit-learn for implementation.

4 Results

4.1 Spatial–temporal distribution of economic, industrial pollution and population

This study utilizes the rate of change in the CEI to represent economic fluctuations, the rate of change in the CPI to characterize industrial pollution changes, and the rate of change in PD to signify population density changes between 2010 and 2020. The economic, industrial pollution, and population dynamics variations are categorized and discussed, yielding nine distinct change types, as outlined in Table 4.

Subsequently, we conducted a spatial distribution visualization analysis of these nine types of regions, as shown in Fig. 3, where the numbers in the legend represent the types of regions. These types are not uniformly distributed across various cities within the urban agglomeration; instead, significant type clustering effects are observed among different cities.

It is observed that in the two periods from 2010 to 2015 and from 2015 to 2020, the largest proportion is accounted for by the seventh category (economic decline, pollution increase, and population increase), comprising 48.930% and 49.111% respectively. This

Table 4 The nine types of changes are economic, pollution, and population

Type	Economic change	Pollution change	Population change	2015 to 2020	2010 to 2015	2010 to 2020
0	None of the following types	4184	4185	3942		
1	Increase	Increase	Decline	19,624	19,355	20
2	Increase	Decline	Decline	2528	2527	14
3	Increase	Increase	Increase	37,805	37,595	32
4	Increase	Decline	Increase	8150	8150	193
5	Decline	Increase	Decline	102,162	101,369	111,164
6	Decline	Decline	Decline	5889	6844	29,937
7	Decline	Increase	Increase	184,629	183,949	180,814
8	Decline	Decline	Increase	10,973	11,970	49,828

The data on the table represents the number of grids

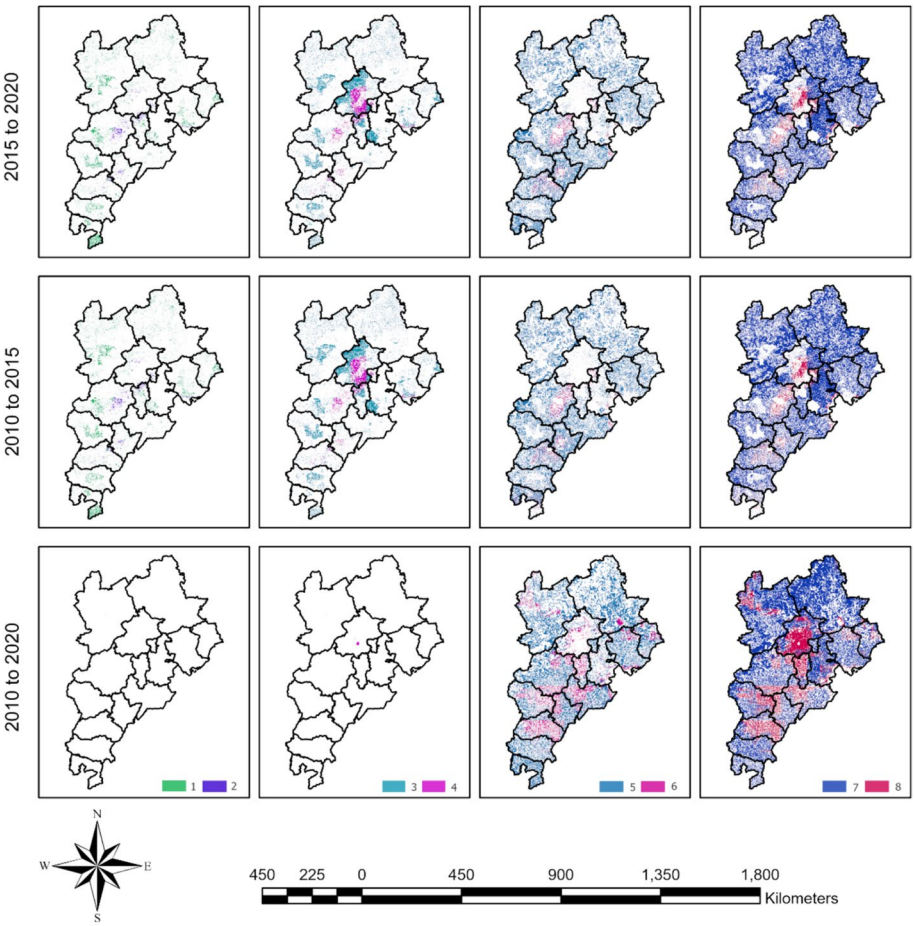

Fig. 3 The spatial distribution of nine types of changing patterns

suggests that in many regions, industrial pollution not only fails to stimulate economic growth but has triggered an economic downturn. Meanwhile, people continue to move to certain areas despite unfavorable economic and air quality conditions. Combined with spatial distribution visualization analysis of nine types of regions in Fig. 3 and the BTH Collaborative Development Strategy, the seventh category is mainly the northwestern ecological conservation areas of the region. This indicates that the seventh category areas are still ecologically better in the BTH region despite the rise in pollution, suggesting that reducing industrial pollution and maintaining a good ecology can attract population inflow. The diametrically opposite situation is the second-ranked category 5 area (economic decline, pollution increase, and population decline), which accounts for 26.964% and 27.175%, respectively from 2010 to 2015 and from 2015 to 2020. This means that the combination of increased industrial pollution and decreased population is prevalent in highly developed mega-city clusters. Combined with Fig. 3, the fifth category areas are almost excluded from Beijing, suggesting that the problem of population loss is more serious in the less developed cities in the BTH region as industrial pollution increases, as well as the economy, declines. At the same time, when analyzing the situation from 2010 to 2020 in a comprehensive manner, the proportions of category 7 and category 5 areas are stable in the first and second places, accounting for 48.096% and 29.569%, respectively, which indicates that the situation reflected by these two types of areas is stable and sustainable.

In the periods from 2010 to 2015 and from 2015 to 2020, the proportion of category 3 areas (economic increase, pollution increase, and population increase) ranks third, with 10.000% and 10.056%, respectively, indicating that a minority of regions can attract populations by offsetting the adverse effects of industrial pollution through economic growth. However, when analyzing the overall situation from 2010 to 2020, the proportion of category 3 areas is only 0.009%, suggesting that such a situation may be specific to the period, and in terms of long-term development, the BTH region still needs to address the issue of population exodus due to industrial pollution.

The first category (economic increase, pollution increase, and population decline) represents the next most prevalent regions, accounting for 5.148% and 5.220% in the periods 2010–2015 and 2015–2020, respectively. This underscores the challenge of economic development in countering population declines caused by intensified industrial pollution. Across the 2010–2020 period, Category 1 areas account for 0.005%, a lower percentage. According to the China Statistical Yearbook, the population of the BTH region was 105.97 million in 2020, 93.26 million in 2015, and 104.4 million in 2010. The population of the BTH region has experienced population reduction and growth between 2010 and 2020, which may have contributed to the decrease in the proportion of Category 1 areas. At the same time, type 1 is primarily distributed in the southeastern part of BTH, encompassing cities like Anyang, Hengshui, and Cangzhou according to Fig. 3. This, to a certain extent, reveals significant economic disparities between the southeastern and northwestern parts of BTH, where the overall economy in the southeast part is growing while the northwestern region is experiencing economic decline.

Regions where industrial pollution has decreased, such as category 8 (economic decline, pollution decline, and population increase) and category 4 (economic increase, pollution decline, and population increase), account for 3.184% and 2.919% in 2010–2015, and 2.168% and 2.168% in 2015–2020, respectively. These categories only contribute 13.254% and 0.051% over the combined 2010–2020 decade, indirectly suggesting that regions with reduced industrial pollution remain a minority. Notably, despite decreases in industrial pollution in some areas, they still experience population outflows. For example, the second and sixth categories constitute 0.672% and 1.820% from 2010 to 2015, and 0.672% and

1.566% from 2015 to 2020, totaling 0.004% and 7.963% for the decade from 2010 to 2020, respectively. However, these regions represent a relatively small share of the total, indicating that the overall trend suggests reducing industrial pollution could solve the issue of population exodus.

In summary, industrial pollution is increasing in most parts of the BTH region, and sometimes it is difficult for economic development to counteract the population loss caused by increased industrial pollution. In addition, the problem of population loss is more serious in the less developed cities in the BTH region. In the long run, the BTH region still needs to pay great attention to the phenomenon of population exodus due to pollution problems, and for this reason, each region should reduce industrial pollution and maintain a good ecological environment.

4.2 Effects of industrial pollutants and the economy on population

4.2.1 Noncollinearity results of variables

This study initially conducted ordinary least squares (OLS) analysis on the independent and dependent variables using the GIS Pro platform, the most commonly used and classic method. OLS analysis serves two primary purposes: it can detect whether the relationship between variables is linear or nonlinear and examines factors' correlation and potential collinearity. The results are presented in Table 5.

Collinearity, as measured by the variance inflation factor (VIF) greater than 10, generally suggests a high degree of covariance between the data, indicating potential issues in the regression model. Apart from PM_{10} and $PM_{2.5}$ factors for the period 2015–2020, the other factors do not exhibit severe collinearity at other time points (VIF > 10). However, the distribution differences between these two collinear factors are substantial in different regions, potentially significantly impacting population changes. Therefore, these factors are still included in further nonlinear ML model analysis.

4.2.2 Non-linear regression results of variables

This study split the data into a training set (70%) and a test set (30%) with a random state of 42. The hyperparameters of the LightGBM model are determined using a systematic

Table 5 The OLS analysis and VIF result of the change rate of variables

Variables	2015 to 2020			2010 to 2015			2010 to 2020		
	Coef	Sig	VIF	Coef	Sig	VIF	Coef	Sig	VIF
GDP	0.0007	0.00*	1.03	0.0019	0.00*	1.00	0.0028	0.00*	1.03
RD	0.0022	0.01*	1.00	0.0004	0.93	1.00	−0.0021	0.82	1.00
SO_2	−0.0014	0.60	1.84	0.0000	0.84	1.00	0.0000	0.45	1.00
O_3	−0.1413	0.00*	1.67	−0.1247	0.00*	1.00	−0.3978	0.00*	1.32
NO_2	0.0055	0.01*	1.08	0.0097	0.12	1.02	−0.0117	0.00*	1.02
PM_{10}	0.5610	0.00*	21.43	−0.8671	0.00*	1.34	−0.5117	0.00*	9.13
$PM_{2.5}$	−0.4370	0.00*	19.99	0.9237	0.00*	1.32	0.4609 s	0.00*	9.36
Old	0.3834	0.00*	1.01	0.4509	0.00*	1.01	0.4128	0.00*	1.02

grid search method available in the Scikit-learn library of Python. Based on the hyperparameter results, the optimal model parameters for the LightGBM model were determined to be learning rate = 0.15, max depth = 7, and n_estimators = 300.

Furthermore, we calculate the R^2, MAE, MSE, and RMSE (Root Mean Squared Error) to assess the predictive test set performance of the LightGBM model. The detailed results are shown in Table 6. Due to this study's relatively large sample size, considering various variables comprehensively, there are no apparent signs of overfitting or underfitting. While the overall R-square is not great enough, it is already better than the OLS model. As well as the other ML model evaluation metrics values are very good. As a result, the fitting performance of the LightGBM model can be considered satisfactory. The model is interpretable for the study's complex relationships between these variables.

To determine the impact of industrial pollution and urban economic variables on population density, this study combined the LightGBM model with the SHAP method to reveal the importance of each variable about population density. Specifically, we used LightGBM feature importance plots and SHAP summary plots to highlight the importance of each factor on population density in BTH, as shown in Fig. 4.

If we disregard the positive or negative effects on the outcome, the rankings of each variable are shown in Fig. 4-Left. Among these, the "old population density" significantly impacts "population density," as the age structure is closely related to population density. Between 2010 and 2015, the importance of other variables, from high to low, is as follows: $PM_{2.5}$, PM_{10}, GDP, O_3, SO_2, NO_2, and road density. Between 2015 and 2020, O_3, GDP, $PM_{2.5}$, SO_2, PM_{10}, NO_2, and road density, in descending order of importance. The combined analysis of the other variables for 2010–2020 is in order of importance: O_3, $PM_{2.5}$, GDP, SO_2, PM_{10}, NO_2, and road density. It can be preliminary analyzed that industrial pollution variables generally all have a greater impact on population density than economic variables at the three time points 2010, 2015, and 2020.

Figure 4-Right displays the specific observed positive and negative effects of each variable. It can be observed that three times in 2010, 2015, and 2020, a certain concentration of $PM_{2.5}$, PM_{10}, SO_2, and O_3 is more likely to harm population decline. RD exhibits dual effects, possibly because this variable has relatively balanced distributions across different regions, and their singular impact on the population could be more pronounced. Furthermore, increased road density contributes to the rise in population density.

4.2.3 Interactive effects between variables

The study further examines the interactive effects of depopulation between different variables, as shown in Fig. 5. For the figure on the left, there is a main effect on the diagonal, and interactions on either side of the diagonal, with the interaction increasing the further apart the dots

Table 6 The performance evaluation metrics of the LightGBM regression model

		R-square	MAE	MSE	RMSE
2015 to 2020	Train model	0.645	0.005	7.305	0.008
	Test model	0.616	0.005	7.145	0.008
2010 to 2015	Train model	0.647	0.005	6.892	0.008
	Test model	0.606	0.005	8.311	0.009
2010 to 2020	Train model	0.680	0.005	7.197	0.008
	Test model	0.620	0.005	8.474	0.009

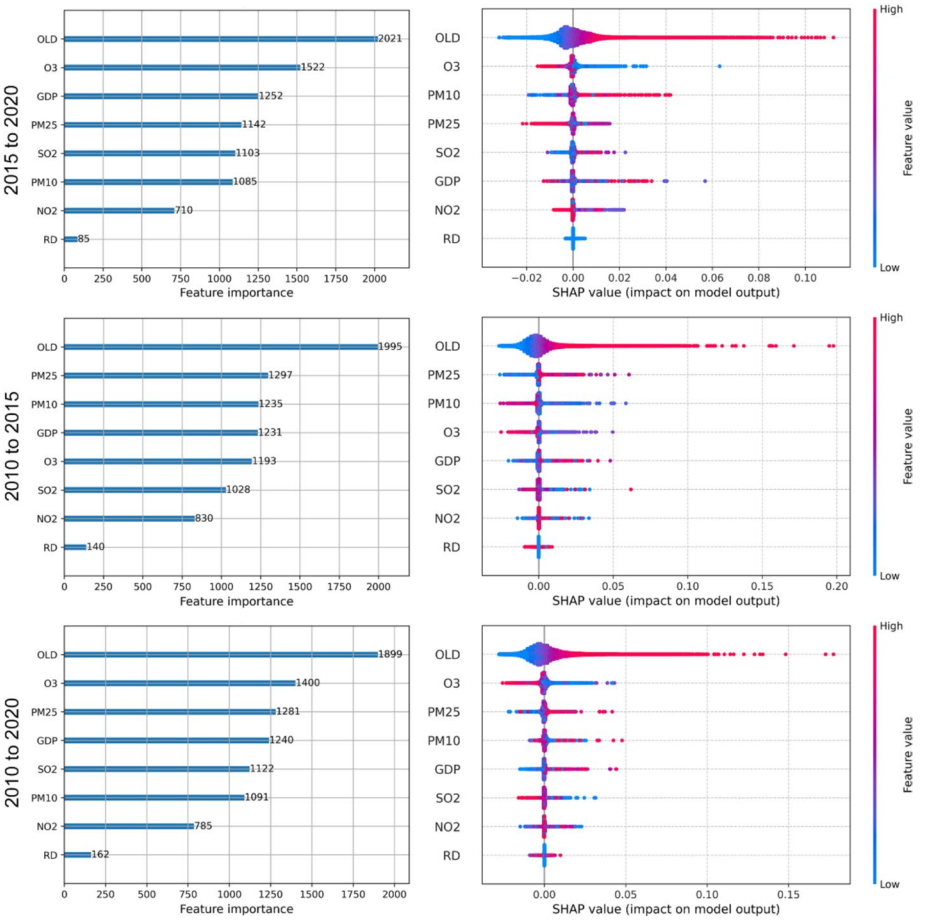

Fig. 4 Important ranking results of the variables. Left: Feature importance of variables based on the LightGBM model. Right: SHAP value of variables based on the SHAP explainer

are inside. The first thing that can be seen is that the interactive effects are relatively consistent across periods, which indicates that the main pollutant indicators for decline in population density are stable.

When considering the interaction between combined and single variables, it can be more easily observed that the negative impact of $PM_{2.5}$, PM_{10}, and O_3 on population density is significantly ahead, and their combinations also have a relatively negative effect on the population. Moreover, when industrial pollution factors are combined with economic development factors, such as GDP and PM_{10}, the positive impact of economic development is offset, resulting in a negative effect on population density.

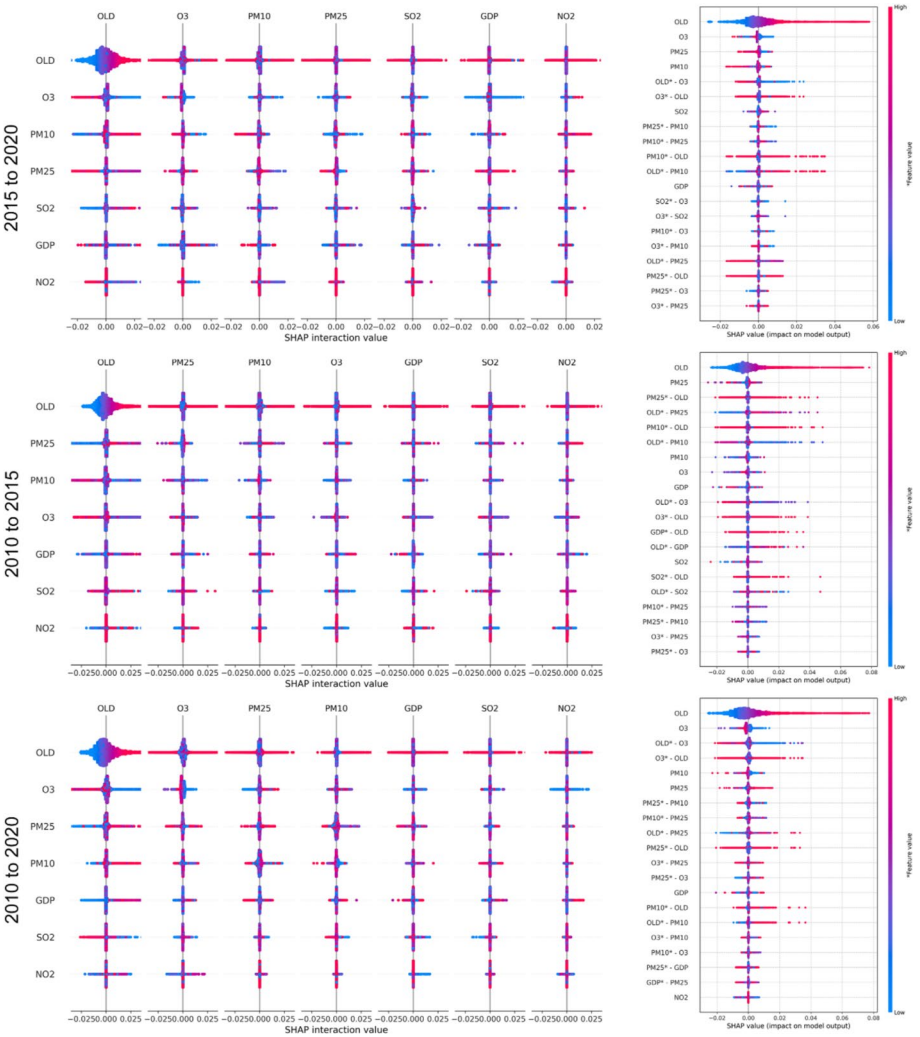

Fig. 5 Intersection effects of different variables. Left: SHAP interaction value of variables. Right: SHAP value of single and interactive variables

5 Discussion

According to the results findings, it can be observed that in the BTH region, only certain areas show a growth trend in population. At the same time, most other cities exhibit an economic decrease, pollution increase, and population outflow. This is consistent with previous research findings, where areas experiencing population outflow cover most counties outside the main urban areas of Beijing and Tianjin, forming an expansion zone in the suburbs of these mega-cities (Wu, 2019). Specifically, the upward trend in CPI in the BTH region also exhibits some coupling with the topography. The western and northern mountainous regions and the eastern areas show an increasing trend, while the central and

east plain areas show a decreasing trend. In contrast, the index shows a steady decline in the plain areas. The pollution patterns within the urban agglomeration exhibit significant differences.

This may be because China issued the "Air Pollution Prevention and Control Action Plan (Action Plan)" to improve air quality in 2013 and the supporting assessment methods and the implementation details for the Action Plan were issued in 2014 (Hu et al., 2020). Previous studies have shown that long-term exposure to these pollutants has significant adverse effects on human health, and is even considered one of the main causes of morbidity and mortality, especially in the BTH region, where deaths related to PM2.5 are particularly prominent (Liu et al., 2017). Since 2016, the Chinese government has invested heavily in promoting environmental technologies and strict emission standards. Most polluting factories and high-emission vehicles have been shut down and eliminated, and traditional solid fuels in household energy have gradually been replaced by cleaner modern energy sources. Over the years, the overall pollution levels in the BTH region have gradually decreased, mitigating population loss caused by air pollution and setting a good example for many other emerging market countries and developing nations.

However, despite the decline in pollution levels in core areas, regional disparities in pollution remain. Previous studies have shown that the implementation of air quality policies affects regional differences in air pollution improvement, with developed areas typically experiencing greater improvements than less developed areas. This is also reflected in the urban–rural divide, where urban areas generally see a higher reduction in PM2.5 levels compared to rural areas. Notably, the largest urban–rural disparity exists in the BTH region, where pollution control policies are the strictest. The implementation of these policies in urban areas is more stringent than in rural areas, leading to a greater reduction in PM2.5 levels in urban regions (Xiao et al., 2020).

Due to industrial relocation and outsourcing, PM2.5 emissions have increased in other parts of China (Fang et al., 2019). In the BTH urban agglomeration, heavily polluting heavy industries have been transferred from developed counties to less developed counties that are not urban centers. Cities such as Beijing and Tianjin, which are gradually undergoing industrial transformation, have been able to attract more population inflows. In contrast, less developed areas, particularly in the southern part of the BTH region, such as cities like Xingtai and Handan in Hebei, which heavily rely on low-end manufacturing industries, have a large number of coal-burning industrial enterprises in the surrounding counties and urban areas. Therefore, after 2015–2020, pollution levels in developed areas like Beijing and Tianjin significantly declined, while the pollution levels in peripheral, underdeveloped areas increased. Ultimately, the overall industrial transformation in the BTH region reduced pollution in the more developed areas, which, leveraging their inherent economic advantages, were better able to attract population inflows. In contrast, despite economic growth, the increase in pollutants in previously underdeveloped areas ultimately led to significant population outflow and subsequent economic decline.

So the increase of PM2.5, PM_{10}, and O_3 exhibits a strong negative effect on population density, consistent with previous research (Chen et al., 2022; Liu & Yu, 2020; Qin & Zhu, 2018). Furthermore, this negative effect cannot be counteracted by economic development. However, the negative effects of nitrogen and sulfur dioxide were insignificant. These results may be attributed to the policies implemented by the Chinese government during this period to control atmospheric pollutants. Measures such as switching from coal to cleaner energy sources like natural gas or electricity, banning small industrial boilers and coal-burning stoves for household heating, and promoting electric or hybrid vehicles have decreased NOx and SO2 concentrations and increased O_3 concentration. This has resulted

in a more severe ozone pollution problem, which is consistent with previous studies (Fang et al., 2020; Wang et al., 2014).

Notedly, the air pollutants in mega-cities like Beijing have decreased much faster than in other cities, indicating that larger cities have higher environmental regulatory requirements (Liang et al., 2019). The composite variation of CPI and population density exhibits an irregular radial outflow of population from core cities such as Beijing and Tianjin. This confirms the population decline in underdeveloped areas on the edge of the mega-region due to regional development imbalances (Wu, 2019).

In summary, the government should take further measures to control air pollution. Secondly, the government should continue to promote economic development and invest in infrastructure, including road construction. There should be a transition in the overall energy structure to reduce emissions. Additionally, when addressing air quality issues, the government should avoid using industrial relocation as the sole solution, as it may adversely affect underdeveloped cities. Comprehensive policies must be formulated to balance regions' economic and environmental needs, promoting sustainable development and improving population mobility.

6 Conclusions

This study focuses on the mechanism of economic and industrial pollution in urban populations. The impact of changes in the combined effect on population density in the BTH from 2010 to 2020 can be investigated at a microscale level. The main findings are as follows:

1. *Economic Development and Population Dynamics*: Economic development struggles to counter population decline exacerbated by increased industrial pollution. In most areas within the Beijing-Tianjin-Hebei (BTH) region, industrial pollution is on the rise. Additionally, less developed cities within the BTH region facing both economic downturns and increased pollution levels have experienced significant population outflows.
2. *Spatial Inequality in Combined Effects*: The combined effects in various regions within urban agglomeration are unequal, and the impact of industrial pollution factors under different type combinations is also unevenly distributed. Overall, air quality is more influenced by $PM_{2.5}$ and PM_{10}.
3. *Threshold Effects of Pollution on Population Density*: Once it reaches a certain concentration, industrial pollution hurts population density, and this negative effect remains unmitigated when interacting with economic factors. The adverse effects of $PM_{2.5,}$ PM_{10}, and O_3 are the most significant.

This study contributes to a deeper understanding of the complex impact of urban pollution on the population and further promotes sustainable and healthy urban development. However, there are some limitations. Due to data limitations, this study only compared data from 2010 and 2020. Since population mobility is a continuous long-term process, future research needs to study the combined effect on population density on a longer time scale. In addition, population density is influenced by multiple factors and is a complex process. It needs to be determined whether residents' perceptions of the impact of industrial pollution on population distribution align with satellite data. Additionally, since individuals in the floating population have different characteristics, In the future, it is possible to explore the differences in population sensitivity to air pollution, especially in combination with

the differences in the level of regional socio-economic characteristics, and analyze the heterogeneous impact of air pollution on residents' willingness to settle or migrate. Further empirical research to further investigate individuals' perceptions of industrial pollution could also be valuable inputs for policymaking, promoting healthy urban development and environmental regulation in China.

Funding Open access funding provided by The Hong Kong Polytechnic University.

Data availability The sources of the data used for the study have all been provided in the article.

References

Agboola, O. P., Alotaibi, B. S., Dodo, Y. A., Abuhussain, M. A., & Abuhussain, M. (2024). Built environment transformation in Nigeria: The effects of a regenerative framework. *Journal of Asian Architecture and Building Engineering, 23*(2), 789–812. https://doi.org/10.1080/13467581.2023.2238045

Alotaibi, B. S., Elnaklah, R., Agboola, O. P., Abuhussain, M. A., Tunay, M., Dodo, Y. A., Maghrabi, A., & Alyami, M. (2024). Enhancing Najran's sustainable smart city development in the face of urbanization challenges in Saudi- Arabia. *Journal of Asian Architecture and Building Engineering*. https://doi.org/10.1080/13467581.2024.2358203

Ananberg, S. C. (2023). Nitrogen dioxide surface-level annual average concentrations. NASA Goddard Earth Sciences Data and Information Services Center. https://doi.org/10.5067/J99FI2U38YRN

Balogun, A.-L., Tella, A., Baloo, L., & Adebisi, N. (2021). A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Climate, 40*, 100989. https://doi.org/10.1016/j.uclim.2021.100989

WorldPop, & Bondarenko, M. (2020). *Individual Countries 1 km UN Adjusted Population Density (2000-2020)* [Dataset]. University of Southampton. https://doi.org/10.5258/SOTON/WP00675

Brottrager, M., Crespo Cuaresma, J., Kniveton, D., & Ali, S. H. (2023). Natural resources modulate the nexus between environmental shocks and human mobility. *Nature Communications, 14*(1), 1. https://doi.org/10.1038/s41467-023-37074-y

Chen, S., Oliva, P., & Zhang, P. (2022). The effect of air pollution on migration: Evidence from China. *Journal of Development Economics, 156*, 102833. https://doi.org/10.1016/j.jdeveco.2022.102833

Cui, Y., Zhang, W., Bao, H., Wang, C., Cai, W., Yu, J., & Streets, D. G. (2019). Spatiotemporal dynamics of nitrogen dioxide pollution and urban development: Satellite observations over China, 2005–2016. *Resources, Conservation and Recycling, 142*, 59–68. https://doi.org/10.1016/j.resconrec.2018.11.015

Dong, F., Wang, Y., Zheng, L., Li, J., & Xie, S. (2020). Can industrial agglomeration promote pollution agglomeration? Evidence from China. *Journal of Cleaner Production, 246*, 118960. https://doi.org/10.1016/j.jclepro.2019.118960

Dvir, R., Vedlitz, A., & Ye, X. (2024). Worried (and) sick: How environmental hazards affect Americans' health-related risk attitudes. *Urban Informatics, 3*(1), 26. https://doi.org/10.1007/s44212-024-00057-5

Fan, H., Wang, Y., Wang, Y., & Coyte, P. C. (2022). The impact of environmental pollution on the physical health of middle-aged and older adults in China. *Environmental Science and Pollution Research, 29*(3), 4219–4231. https://doi.org/10.1007/s11356-021-15832-z

Fang, D., Chen, B., Hubacek, K., Ni, R., Chen, L., Feng, K., & Lin, J. (2019). Clean air for some: Unintended spillover effects of regional air pollution policies. *Science Advances, 5*(4), eaav4707. https://doi.org/10.1126/sciadv.aav4707

Fang, X., Xiao, H., Sun, H., Liu, C., Zhang, Z., Xie, Y., Liang, Y., & Wang, F. (2020). Characteristics of ground-level ozone from 2015 to 2018 in BTH area, China. *Atmosphere, 11*(2), 2. https://doi.org/10.3390/atmos11020130

Germani, A. R., Scaramozzino, P., Castaldo, A., & Talamo, G. (2021). Does air pollution influence internal migration? An empirical investigation on Italian provinces. *Environmental Science & Policy, 120*, 11–20. https://doi.org/10.1016/j.envsci.2021.02.005

Grace, K., Hertrich, V., Singare, D., & Husak, G. (2018). Examining rural Sahelian out-migration in the context of climate change: An analysis of the linkages between rainfall and out-migration in two Malian villages from 1981 to 2009. *World Development, 109*, 187–196. https://doi.org/10.1016/j.worlddev.2018.04.009

Gu, X., Lai, Z., Zhu, L., & Liu, X. (2025). Urban greenery distribution and its heat mitigation effect on outdoor jogging activities. *Urban Forestry & Urban Greening, 104*, 128655. https://doi.org/10.1016/j.ufug.2024.128655

Gu, X., Tang, X., Chen, T., & Liu, X. (2024a). Predicting the network shift of large urban agglomerations in China using the deep-learning gravity model: A perspective of population migration. *Cities, 145*, 104680. https://doi.org/10.1016/j.cities.2023.104680

Gu, X., Wu, Z., Liu, X., Qiao, R., & Jiang, Q. (2024b). Exploring the nonlinear interplay between urban morphology and nighttime thermal environment. *Sustainable Cities and Society, 101*, 105176. https://doi.org/10.1016/j.scs.2024.105176

Hao, Y., Zhang, Q., Zhong, M., & Li, B. (2015). Is there convergence in per capita SO2 emissions in China? An empirical study using city-level panel data. *Journal of Cleaner Production, 108*, 944–954. https://doi.org/10.1016/j.jclepro.2015.06.054

Hu, C., Hu, B., Shi, X., & Wu, Y. (2020). The roles of Beijing-Tianjin-Hebei coordinated development strategy in industrial energy and related pollutant emission intensities. *Sustainability, 12*(19), 19. https://doi.org/10.3390/su12197973

Huang, G., Zhang, J., Yu, J., & Shi, X. (2020). Impact of transportation infrastructure on industrial pollution in Chinese cities: A spatial econometric analysis. *Energy Economics, 92*, 104973. https://doi.org/10.1016/j.eneco.2020.104973

Jiang, F., & Lv, Y. (2022). *Changes in the spatial distribution of population in urban agglomerations: A spatial-temporal perspective*. https://doi.org/10.21203/rs.3.rs-1244609/v1

Jiang, L., He, S., Cui, Y., Zhou, H., & Kong, H. (2020). Effects of the socio-economic influencing factors on SO2 pollution in Chinese cities: A spatial econometric analysis based on satellite observed data. *Journal of Environmental Management, 268*, 110667. https://doi.org/10.1016/j.jenvman.2020.110667

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Kovats, R. S. (2012). Global health and global environmental governance—Research for policy. *Global Environmental Change, 22*(1), 1–2. https://doi.org/10.1016/j.gloenvcha.2011.11.011

Kuhn, H. W., & Tucker, A. W. (2016). *Contributions to the theory of games (AM-28)* (Vol. II). Princeton University Press.

Laouzai, A., & Ouafi, R. (2022). A prediction model for atmospheric pollution reduction from urban traffic. *Environment and Planning b: Urban Analytics and City Science, 49*(2), 566–584. https://doi.org/10.1177/23998083211005776

Li, B., & Wu, S. (2017). Effects of local and civil environmental regulation on green total factor productivity in China: A spatial Durbin econometric analysis. *Journal of Cleaner Production, 153*, 342–353. https://doi.org/10.1016/j.jclepro.2016.10.042

Liang, L., Wang, Z., & Li, J. (2019). The effect of urbanization on environmental pollution in rapidly developing urban agglomerations. *Journal of Cleaner Production, 237*, 117649. https://doi.org/10.1016/j.jclepro.2019.117649

Liu, B., Shi, Y., Li, D.-J., Wang, Y.-D., Fernandez, G., & Tsou, M.-H. (2020). An economic development evaluation based on the OpenStreetMap road network density: The case study of 85 cities in China. *ISPRS International Journal of Geo-Information, 9*(9), 9. https://doi.org/10.3390/ijgi9090517

Liu, M., Huang, Y., Ma, Z., Jin, Z., Liu, X., Wang, H., Liu, Y., Wang, J., Jantunen, M., Bi, J., & Kinney, P. L. (2017). Spatial and temporal trends in the mortality burden of air pollution in China: 2004–2012. *Environment International, 98*, 75–81. https://doi.org/10.1016/j.envint.2016.10.003

Liu, X., Lu, D., Zhang, A., Liu, Q., & Jiang, G. (2022). Data-driven machine learning in environmental pollution: Gains and problems. *Environmental Science & Technology, 56*(4), 2124–2133. https://doi.org/10.1021/acs.est.1c06157

Liu, Z., & Yu, L. (2020). Stay or leave? The role of air pollution in urban migration choices. *Ecological Economics, 177*, 106780. https://doi.org/10.1016/j.ecolecon.2020.106780

Lu, X., Zhang, Y., Li, J., & Duan, K. (2022). Measuring the urban land use efficiency of three urban agglomerations in China under carbon emissions. *Environmental Science and Pollution Research, 29*(24), 36443–36474. https://doi.org/10.1007/s11356-021-18124-8

Lyu, Y., & Jiang, F. (2022). Spatial and temporal distribution of population in urban agglomerations changes in China. *Scientific Reports, 12*(1), 1. https://doi.org/10.1038/s41598-022-12274-6

Ma, J., & Cheng, J. C. P. (2016). Data-driven study on the achievement of LEED credits using percentage of average score and association rule analysis. *Building and Environment, 98*, 121–132. https://doi.org/10.1016/j.buildenv.2016.01.005

Ma, J., Ding, Y., Cheng, J. C. P., Jiang, F., Tan, Y., Gan, V. J. L., & Wan, Z. (2020). Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. *Journal of Cleaner Production, 244*, 118955. https://doi.org/10.1016/j.jclepro.2019.118955

Meng, X., & Long, Y. (2022). Shrinking cities in China: Evidence from the latest two population censuses 2010–2020. *Environment and Planning a: Economy and Space, 54*(3), 449–453. https://doi.org/10.1177/0308518X221076499

Miao, Z., & Gu, X. (2024). Measuring the impact of industrial pollution on population in Beijing, China: An empirical analysis based on the GWR model. In C. Yuan (Ed.), *Proceedings of 5th international conference on resources and environmental research—ICRER 2023* (pp. 31–44). Springer Nature Switzerland.

Qin, Y., & Zhu, H. (2018). Run away? Air pollution and emigration interests in China. *Journal of Population Economics, 31*(1), 235–266. https://doi.org/10.1007/s00148-017-0653-0

Roth, A. E. (Ed.). (1988). *The shapley value: Essays in Honor of Lloyd S Shapley*. Cambridge University Press.

Shen, J., Wei, Y. D., & Yang, Z. (2017). The impact of environmental regulations on the location of pollution-intensive industries in China. *Journal of Cleaner Production, 148*, 785–794. https://doi.org/10.1016/j.jclepro.2017.02.050

Shuhua, Xu., Sun, C., Wei, H., & Hou, X. (2023). Road construction and air pollution: Analysis of road area ratio in China. *Applied Energy, 351*, 121794. https://doi.org/10.1016/j.apenergy.2023.121794

Son, Y., Osornio-Vargas, Á. R., O'Neill, M. S., Hystad, P., Texcalac-Sangrador, J. L., Ohman-Strickland, P., Meng, Q., & Schwander, S. (2018). Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *Science of the Total Environment, 639*, 40–48. https://doi.org/10.1016/j.scitotenv.2018.05.144

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Tella, A., Balogun, A.-L., & Faye, I. (2021). Spatio-temporal modelling of the influence of climatic variables and seasonal variation on PM_{10} in Malaysia using multivariate regression (MVR) and GIS. *Geomatics, Natural Hazards and Risk, 12*(1), 443–468. https://doi.org/10.1080/19475705.2021.1879942

Wan, L., Ye, X., Lee, J., Lu, X., Zheng, L., & Wu, K. (2015). Effects of urbanization on ecosystem service values in a mineral resource-based city. *Habitat International, 46*, 54–63. https://doi.org/10.1016/j.habitatint.2014.10.020

Wang, H., & Ge, Q. (2023). Spatial association network of PM2.5 and its influencing factors in the Beijing–Tianjin–Hebei urban agglomeration. *Environmental Science and Pollution Research, 30*(27), 70541–70557. https://doi.org/10.1007/s11356-023-27434-y

Wang, Y., Ying, Q., Hu, J., & Zhang, H. (2014). Spatial and temporal variations of six criteria air pollutants in 31 provincial capital cities in China during 2013–2014. *Environment International, 73*, 413–422. https://doi.org/10.1016/j.envint.2014.08.016

Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., & Cribb, M. (2021a). Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications. *Remote Sensing of Environment, 252*, 112136. https://doi.org/10.1016/j.rse.2020.112136

Wei, J., Li, Z., Xue, W., Sun, L., Fan, T., Liu, L., Su, T., & Cribb, M. (2021b). The ChinaHighPM10 dataset: Generation, validation, and spatiotemporal variations from 2015 to 2019 across China. *Environment International, 146*, 106290. https://doi.org/10.1016/j.envint.2020.106290

Wu, K. (2019). Urban shrinkage in the Beijing-Tianjin-Hebei region and yangtze River delta: Pattern, trajectory and factors. In Y. Long & S. Gao (Eds.), *Shrinking cities in China: The other facet of urbanization* (pp. 43–61). Springer. https://doi.org/10.1007/978-981-13-2646-2_3

Xiao, Q., Geng, G., Liang, F., Wang, X., Lv, Z., Lei, Y., Huang, X., Zhang, Q., Liu, Y., & He, K. (2020). Changes in spatial patterns of PM25 pollution in China 2000–2018: Impact of clean air policies. *Environment International, 141*, 105776. https://doi.org/10.1016/j.envint.2020.105776

Yan, C., Li, H., & Li, Z. (2022). Environmental pollution and economic growth: Evidence of SO2 emissions and GDP in China. *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2022.930780

Yu, W., Ye, T., Zhang, Y., Xu, R., Lei, Y., Chen, Z., Yang, Z., Zhang, Y., Song, J., Yue, X., Li, S., & Guo, Y. (2023). Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: A machine learning modelling study. *The Lancet Planetary Health, 7*(3), e209–e218. https://doi.org/10.1016/S2542-5196(23)00008-6

Zhao, Z., Lao, X., Gu, H., Yu, H., & Lei, P. (2021). How does air pollution affect urban settlement of the floating population in China? New evidence from a push-pull migration analysis. *BMC Public Health, 21*(1), 1696. https://doi.org/10.1186/s12889-021-11711-x

Authors and Affiliations

Xinyue Gu[1] · Zijian Miao[2] · Yan Mao[3] · Xintao Liu[1]

✉ Xintao Liu
xintao.liu@polyu.edu.hk

Xinyue Gu
xinyue.gu@connect.polyu.hk

Zijian Miao
m3032065997@163.com

Yan Mao
myy0716@outlook.com

[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

[2] School of Architecture, South China University of Technology, Guangzhou, China

[3] Institute for Public Policy, Zhejiang University, Hangzhou, China