GeoHealth

RESEARCH ARTICLE
10.1029/2023GH000794

Special Section:
Exploring the Links Between Air Quality and Lung Cancer

**Key Points:**
- The geographical scale of incidence rate and potential risk factors of lung cancer were unified at one geographic scale
- A Bayesian spatio-temporal interaction model was used to evaluate the relative risk of disease in different regions
- Lung cancer risk distribution was adjusted for multiple environmental factors

**Correspondence to:**
Y. Xu and H. Wei,
yanqing.xu@whu.edu.cn;
zzu_wei@163.com

**Citation:**
Liu, Y., Xu, Y., Li, Y., & Wei, H. (2023). Identifying the environmental determinants of lung cancer: A case study of Henan, China. *GeoHealth*, *7*, e2023GH000794. https://doi.org/10.1029/2023GH000794

Received 7 FEB 2023
Accepted 26 APR 2023



# Identifying the Environmental Determinants of Lung Cancer: A Case Study of Henan, China

**Yan Liu[1], Yanqing Xu[1], Yuchen Li[2], and Haitao Wei[3,4]**

[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, [2]MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK, [3]The School of the Geo-Science & Technology, Zhengzhou University, Zhengzhou, China, [4]Joint Laboratory of Eco-Meteorology, Zhengzhou University, Zhengzhou, China

**Abstract** Lung cancer has become one of the most prevalent cancers in the last several decades. Studies have documented that most cases of lung cancer are caused by inhaling environmental carcinogens while how external environmental factors lead to individual lung cancer is still an open issue as the pathogenesis may come from the combined action of multiple environmental factors, and such pathogenic mechanism may vary from region to region. Based on the data of lung cancer cases from hospitals at the county level in Henan from 2016 to 2020, we analyzed the response relationship between lung cancer incidence and physical ambient factors (air quality, meteorological conditions, soil vegetation) and socioeconomic factors (occupational environment, medical level, heating mode, smoking behavior). We used a Bayesian spatio-temporal interaction model to evaluate the relative risk of disease in different regions. The results showed that smoking was still the primary determinant of lung cancer, but the influence of air quality was increasing year by year, with meteorological conditions and occupational environment playing a synergistic role in this process. The high-risk areas were concentrated in the plains of East and Central Henan and the basin of South Henan, while the low-risk areas were concentrated in the hilly areas of North and West Henan, which were related to the topography of Henan. Our study provides a better understanding of the environmental determinants of lung cancer which will help refine existing prevention strategies and recognize the areas where actions are required to prevent environment and occupation related lung cancer.

**Plain Language Summary** The study investigated the relationship between environmental exposures (natural environment and human environment) and lung cancer incidence among residents in Henan Province, China. It also assessed the future development trend of lung cancer risk in the province through the screening of eight dominant environmental factors. Overall, there are obvious geographical differences in the incidence risk distribution, which are directly caused by smoking status and air pollution, and may be indirectly affected by factors such as climate, occupation and medical level. Our work is of positive significance for revealing the environmental incentives of lung cancer at the spatio-temporal level, and exploring a new model of regional cancer prevention and control centered on environmental optimization.

## 1. Introduction

Lung cancer is one of the major malignant tumors that endanger human life and health, accounting for about one-fifth of cancer deaths worldwide, and is the most deadly cancer in men and the second deadliest cancer in women (Bray et al., 2020; Torre et al., 2015). According to the global statistics on cancer released by the International Agency for Research on Cancer, there will be about 2.2 million new cases of lung cancer and about 1.8 million deaths worldwide in 2020, accounting for 11.4% and 18.0% of all cancer incidence and deaths, ranking second in cancer incidence and first in mortality (Sung et al., 2021). Of the 185 countries and 36 types of cancer included in the report, lung cancer was the most common cancer in men in 36 countries and the leading cause of cancer death in 93. With medical advances, the overall 5-year survival rate of lung cancer has improved slightly over time, but it is still poor, reaching only 10%–20% of lung cancer patients diagnosed between 2010 and 2014 in most countries (Allemani et al., 2018).

Compared with other parts of the world, the burden of lung cancer in China is the heaviest, accounting for 36.98% of the global cases and 39.21% of the deaths (He et al., 2020). Among the top 10 malignant tumors in China, the incidence and mortality of lung cancer account for 20.03% and 26.99%, respectively, making it the malignant tumor with the highest incidence and mortality in China (Q. Wang et al., 2020). The serious problem of lung cancer has brought a heavy social and economic burden to China. From 1996 to 2011, the direct medical expenses of lung cancer patients ranged from 1,382 to 3,661 dollars, with an average annual growth rate of 2.2%, the situation continued to deteriorate in the following years (Q. Wang et al., 2020). It is worth noting that due to the vast territory and diverse natural and social environment, the morbidity and mortality of lung cancer in different regions of China may vary considerably. The morbidity and mortality rates in the eastern and central parts of China are higher than those in western China, and the urban population has a higher risk of disease than the rural population (He et al., 2020; S. Liu et al., 2018). Henan is a large province in central China located at the junction of the central and western regions and the open coastal areas. As a populous province dominated by agriculture, Henan has always been a high-incidence area of malignant tumors (S. Liu et al., 2016). Among them, the mortality rate of lung cancer is the first among malignant tumors throughout the year, and is the highest among males (Cheng et al., 2012). With the development of the economy and the aging population, lung cancer has become a serious disease threatening the health of residents in this province.

Environmental factors refer to the space of human existence and various natural or social factors that can directly or indirectly affect human life and development. Environmental imbalance can easily lead to the destruction of physiological functions, further affecting human health and inducing related diseases (Singer, 2013). The air we breathe every day contains a large number of particles and gases that can damage the lungs through specific and non-specific mechanisms (Seow et al., 2014). These include carcinogens that we are exposed to in our living and working environments. Inhaling carcinogenic substances in the environment increases the risk for malignant and non-malignant respiratory disease and is in fact responsible for the majority of lung cancer cases. Smoking is generally considered to be a key risk factor inducing lung cancer as it exposes the lungs to a rich mix of particular carcinogenic agents and other harmful substances, while due to effective tobacco control and a series of health education and publicity measures, the contribution of smoking behavior to lung cancer has been decreasing (Cao & Chen, 2019). Moreover, environmental factors that can trigger and aggravate respiratory and cardiopulmonary diseases have attracted increased attention. For example, exposure to air pollutants (M. Liu et al., 2017), arsenic intake from groundwater (Poinen-Rughooputh et al., 2016), and occupational exposure to silica, asbestos, residential radon, and other occupational respiratory carcinogens (Alberg & Samet, 2003; Darby et al., 2005; Mendez et al., 2017) have all been confirmed to promote the occurrence of lung cancer to a certain extent. Since most aspects of the environment are modifiable and people's behavior can be changed, this offers great potential for lung cancer prevention. Hence, identifying the key environmental risks affecting human lung cancer and corresponding preventive interventions are feasible keys to reducing the threat of lung cancer to human health and improving the quality of life.

However, the existing environmental determinants of lung cancer studies have only examined a single or limited number of environmental perspectives, while the causes of interactions/synergies across different environmental factors are poorly understood and thus hamper prevention strategies. Hence, it is necessary to conduct a comprehensive analysis to evaluate the association between a variety of potential environmental exposures and lung cancer risk, which may refine the understanding of causes across multiple levels and help inform strategies to prevent lung cancer. Moreover, lung cancer prevention strategies may also need to be tailored specifically to different populations and communities, as the mechanisms by which environmental causes of lung cancer vary across regions and populations. Geographic information science (GIS) and technology and spatial data analytics provide a new possibility to study the environmental determinants of diseases on a large geographical scale. We estimate the Bayesian Spatio-temporal interaction model with lung cancer cases from hospitals at the county level in Henan from 2016 to 2020 along with a list of environmental variables derived from GIS, we also evaluate the relative risk of disease in different regions.

## 2. Method and Data

### 2.1. Study Area

Henan province is located along the middle and lower reaches of the Yellow River in central and eastern China, with a permanent population of about 99.37 million, and is the middle zone of China's economic development from east to west. It is high in the west and low in the east, with rich and diverse landforms, consisting of plains, mountains, hills, basins, and water surfaces. Henan has a continental monsoon climate from the north subtropical zone to the warm temperate zone. Summer is hot and rainy, while winter is cold and dry. The average annual temperature is 12.1–15.7°C, and the average annual precipitation is 533–1381 mm, mostly from June to August.

10.1029/2023GH000794

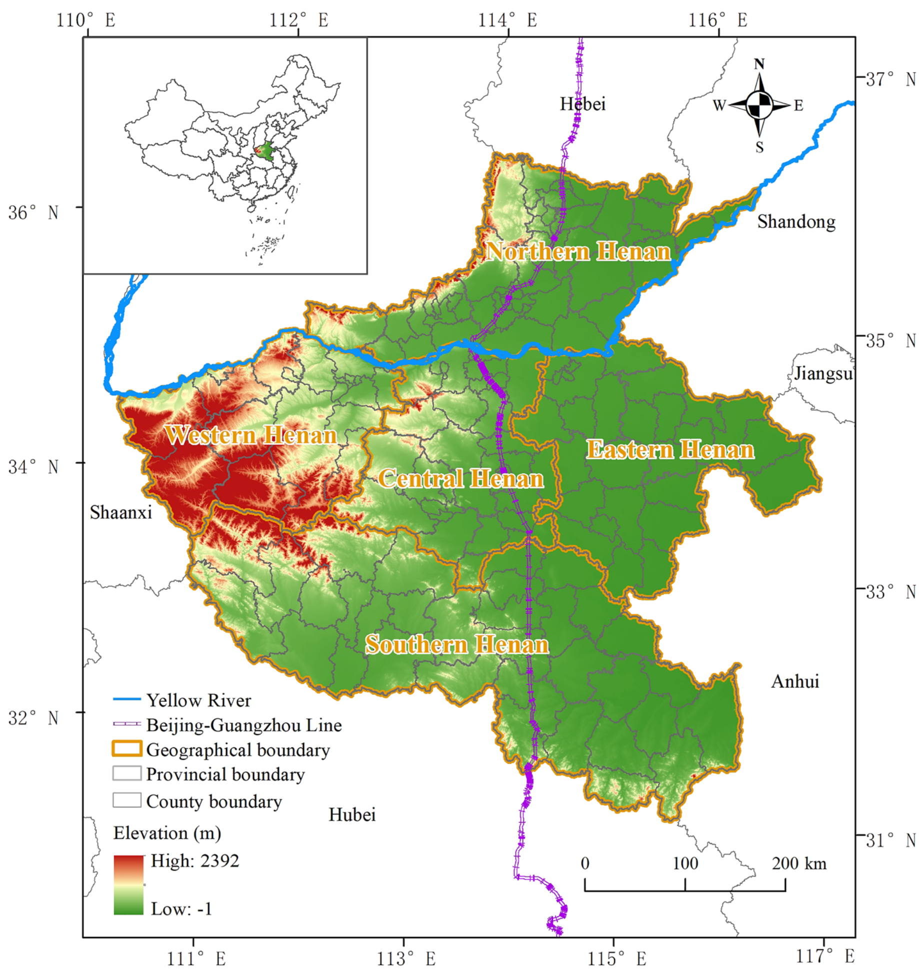


**Figure 1.** Overview of the study area.

The province has jurisdiction over 17 prefecture-level cities, 1 county-level city, and 157 county-level administrative regions, and can be divided into five regions: Central Henan, Eastern Henan, Western Henan, Southern Henan, and Northern Henan according to the geographical location, topography, rivers, roads, and other factors of the cities in Henan (Figure 1).

### 2.2. Data Sources

Lung cancer data from 2016 to 2020 were collected from confirmed lung cancer medical records of 329 hospitals above the county level in Henan Province, including 35 Class A hospitals, 9 Class B hospitals, and 285 Class C hospitals. The relevant data were desensitized and only included the patient's gender, age, address, administrative division code, and time of diagnosis. We used a multi-source online geocoding system (Peng et al., 2020; H. Zhang et al., 2020) based on Chinese syllogisms to visualize lung cancer case data in space. The spatial coordinates of 149,417 case addresses were obtained through a step-by-step coding method of "Tencent Map-Baidu Map-Gaode Map," and the total address matching rate was 96.14% (Figure 2). Finally, the lung cancer incidence map with a grid size of 1 km × 1 km was drawn by the kernel density estimation (KDE) method. And lung cancer incidence is expressed as ratio-based statistics and normalized to a population size of 100,000.

Environmental factors have always been the key factors affecting human survival and development. Therefore, we reviewed previous studies on the relationship between environment and lung cancer, initially screening 31 environmental variables that may have an impact on lung cancer from two aspects (Figure 3): natural environment (air quality, meteorological conditions, soil vegetation) and human environment (occupational environment, medical level, heating mode and smoking).

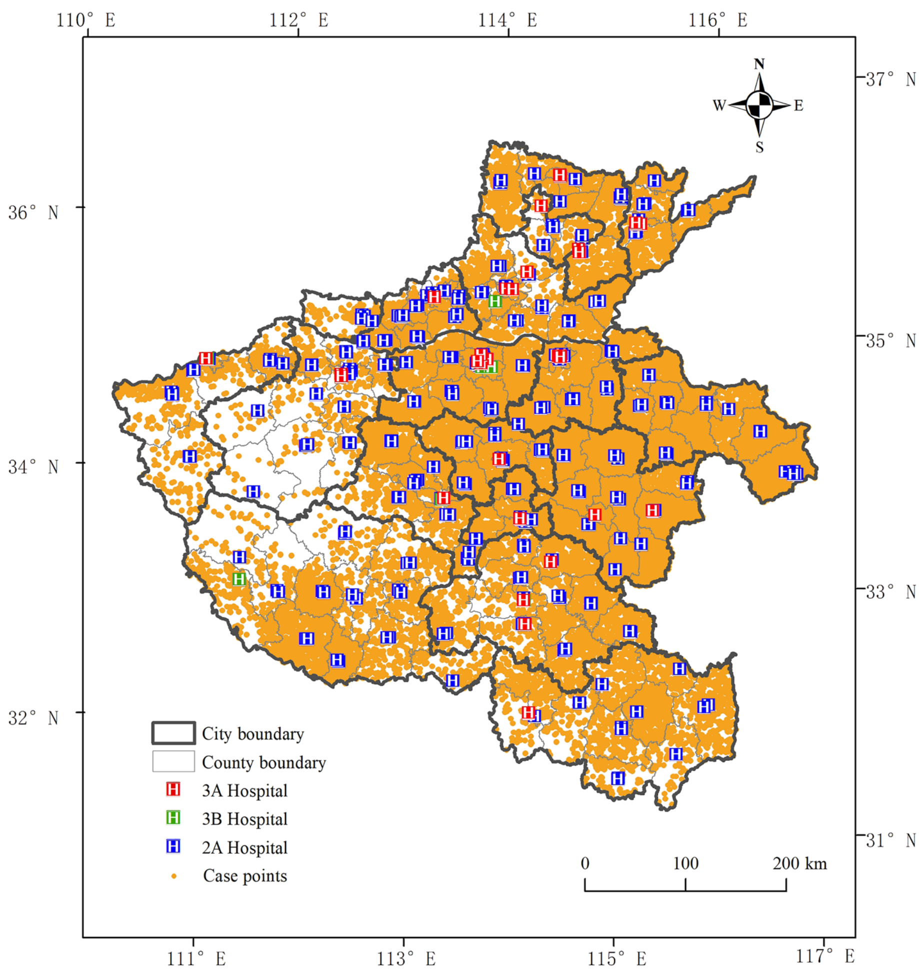


**Figure 2.** Spatial distribution of case points and registered hospitals.

Data on the annual mean concentration of six air pollutants with a spatial resolution of 10 km from 2009 to 2020 were obtained from the Chinahighairunsustainable series data set (Wei et al., 2021). Simultaneous meteorological data with a spatial resolution of 1 km were obtained from the national earth system science data center (http://www.geodata.cn). Soil attribute data were obtained from the Basic Attribute data set of China High-resolution National Soil Information Grid (2010–2018). NDVI from 2009 to 2020 were obtained from the Resource and Environmental Sciences and Data Center of Chinese Academy of Sciences (https://www.resdc.cn), with a spatial resolution of 1 km. Medical accessibility was calculated in AccessMod 5 (Munoz & Kallestal, 2012), and 2017 heating data were obtained from the second pollution survey project of Henan Province. In addition, other social data from 2009 to 2020 came from the Statistical Yearbook of Henan Province of the corresponding years, and some missing data were supplemented by combining local statistical yearbooks and census data of various cities in Henan. Descriptive statistics and spatial autocorrelation test (Moran's I) (B. Guo et al., 2021) for all data are shown in Table 1.

### 2.3. Geodetector

Geodetector is a new set of statistical methods to detect the spatial heterogeneity of geographical phenomena and reveal the driving factors behind them (J.-F. Wang et al., 2016). It includes four modules: the factor detector, the interaction detector, the risk detector, and the ecological detector. Among them, the factor detector is used to quantitatively detect whether a certain geographical factor is the cause of the difference in the spatial distribution of a certain index value, and the weight, and its expression is:

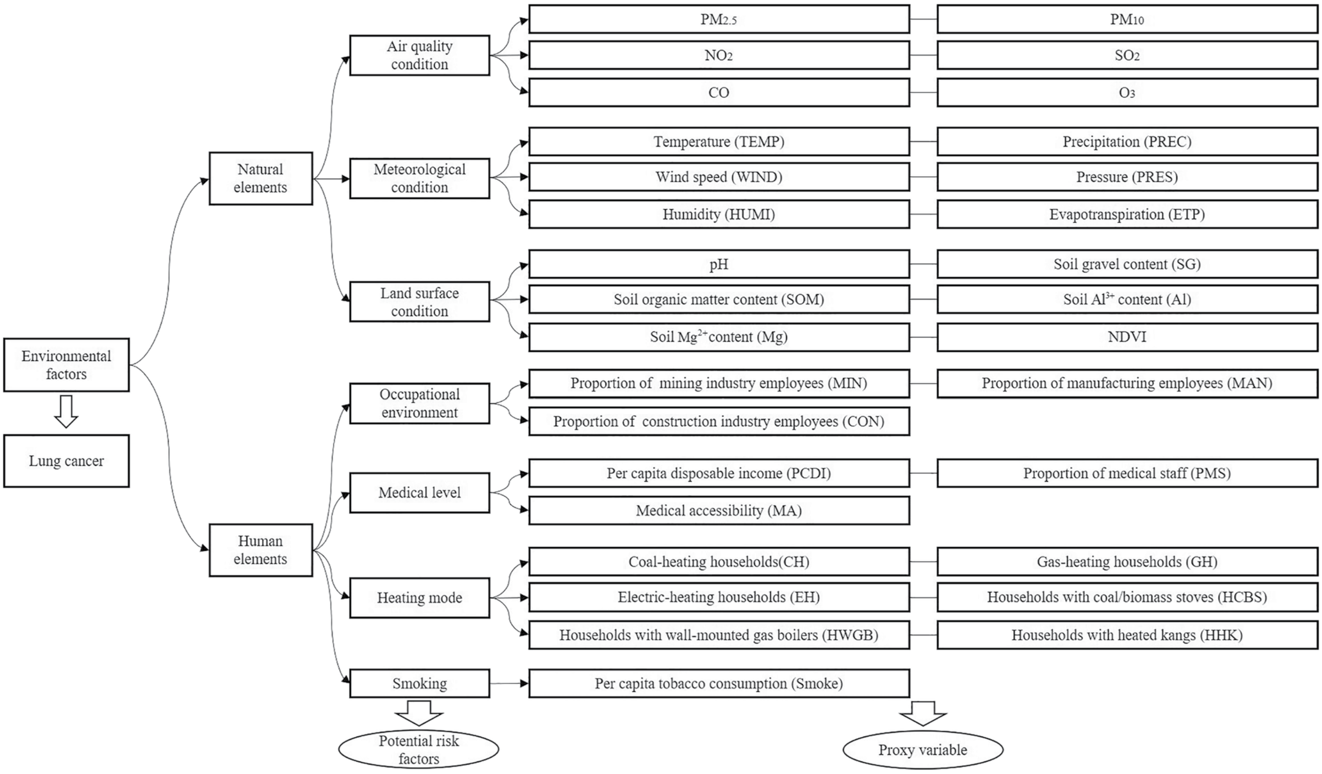


**Figure 3.** Potential risk factors and their proxy variables for lung cancer.

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{\text{SSW}}{\text{SST}} \tag{1}$$

where $h = 1, \dots, L$ is the stratification of target variable $Y$ or impact factor $X$, $N_h$, and $N$ are the number of units in the layer and the whole area respectively; $\sigma_h^2$ and $\sigma^2$ are the variances of the values of layer $h$ and region $Y$, respectively. SSW and SST are the sum of intra-layer variance and the total variance of the whole area. The value $q$ is used to measure the explanatory power of the independent variable to the dependent variable, $q \in [0, 1]$, the larger the $q$ value, the stronger the explanatory power of the independent variable to the dependent variable.

### 2.4. Bayesian Spatiotemporal Interaction Model

Bayesian spatiotemporal interaction model (BSTIM) is a model based on Bayesian statistical thinking to analyze spatiotemporal data (J. J. Li et al., 2020). This model uses a variety of random effect terms to analyze the observed temporal and spatial changes, solving the problem of variance inhomogeneity that may occur in time and space (G. G. Li et al., 2014). In addition, because this method can make full use of sample information and prior information to estimate the posterior distribution of spatiotemporal parameters, the confidence is higher and the result more robust (X. Zhang et al., 2018).

Considering the scattered, random, and non-negative integers of lung cancer incidence data, we assume that the number of lung cancer cases in $i$ ($i = 1, 2, \dots, 158$) districts and counties in year $j$ ($j = 1, 2, 3$) follows the Poisson distribution:

$$y_{ij} \sim \text{Poisson}(\mu_{ij}) \tag{2}$$

The parameter $\mu_{ij}$ of the Poisson distribution can be expressed as:

$$\mu_{\text{ij}} = e_{ij}\theta_{ij} \tag{3}$$

**Table 1**
*Descriptive Statistics of Potential Risk Factors and Spatial Autocorrelation Test*

| Environment variable | Min | Max | Mean | Std | Moran's I |
|---|---|---|---|---|---|
| $PM_{2.5}$ (μg/m$^3$) | 42.88 | 89.24 | 67.28 | 8.49 | 0.95* |
| $PM_{10}$ (μg/m$^3$) | 74.40 | 148.12 | 110.54 | 13.82 | 0.97* |
| $NO_2$ (μg/m$^3$) | 21.88 | 54.47 | 35.02 | 5.80 | 0.98* |
| $SO_2$ (μg/m$^3$) | 12.85 | 44.45 | 24.81 | 5.87 | 0.98* |
| CO (μg/m$^3$) | 0.74 | 1.96 | 1.22 | 0.20 | 0.98* |
| $O_3$ (μg/m$^3$) | 88.53 | 107.58 | 98.92 | 3.89 | 0.96* |
| TEMP (°C) | 5.76 | 16.89 | 15.20 | 1.21 | 0.99* |
| PREC (mm) | 535.44 | 1410.10 | 727.11 | 144.52 | 0.99* |
| WIND (m/s) | 0.36 | 2.05 | 1.10 | 0.17 | 0.98* |
| PRES (kPa) | 82.78 | 101.63 | 99.03 | 3.56 | 0.87* |
| HUMI (%) | 50.35 | 80.58 | 63.50 | 4.17 | 0.99* |
| ETP (mm) | 602.05 | 1058.76 | 987.03 | 50.12 | 0.98* |
| PH | 4.93 | 9.34 | 7.51 | 0.79 | 0.87* |
| SG (%) | 6.08 | 85.76 | 36.09 | 19.51 | 0.75* |
| SOM (g/kg) | 0.13 | 10.79 | 1.75 | 1.13 | 0.74* |
| Al (cmol/kg) | 0.01 | 2.63 | 0.37 | 0.42 | 0.71* |
| Mg (cmol/kg) | 0.35 | 4.21 | 1.71 | 0.90 | 0.79* |
| NDVI | 0.18 | 0.91 | 0.78 | 0.09 | 0.81* |
| MIN (%) | 0 | 36.48 | 1.99 | 4.88 | 0.14* |
| MAN (%) | 1.11 | 40.68 | 10.20 | 7.96 | 0.45* |
| CON (%) | 0.72 | 18.20 | 5.18 | 2.74 | 0.23* |
| PCDI ($10^5$¥) | 1.17 | 3.94 | 2.04 | 0.58 | 0.43* |
| PMS ($1/10^4$) | 21.32 | 151.12 | 61.11 | 33.24 | 0.21* |
| MA (minute) | 0 | 718 | 37.41 | 53.83 | 0.96* |
| CH (%) | 0 | 76.93 | 16.73 | 20.32 | 0.53* |
| GH (%) | 0 | 93.39 | 15.51 | 14.34 | 0.12* |
| EH (%) | 0.58 | 99.81 | 38.06 | 20.11 | 0.24* |
| HCBS (%) | 0 | 58.11 | 7.87 | 10.20 | 0.27* |
| HWGB (%) | 0 | 35.57 | 1.98 | 3.93 | 0.23* |
| HHK (%) | 0 | 15.40 | 0.69 | 1.90 | 0.42* |
| Smoke (¥) | 113.98 | 325.84 | 187.86 | 63.66 | 0.52* |

*Note.* * means 1% significance level.

In the formula, $e_{ij}$ represents the number of expected cases of disease, standardized according to the sex ratio, and represents the expected number of cases in each region calculated based on the incidence rates of men and women in the total number of cases in all regions in year $j$. $\theta_{ij}$ represents the ratio of the actual number of cases in year $j$ to the expected number of cases in area $i$, that is, the relative risk of lung cancer (RR), which is also a parameter that this research focuses on. The mathematical form of BSTIM is as follows:

$$\log(\theta_{\mathrm{ij}}) = \alpha_0 + \alpha_1 \times \mathrm{time}_j + u_i + \upsilon_i + g_i + \mathrm{psi}_{ij} + \beta_k \times x_{ijk} \quad (4)$$

In the formula, $\alpha_0$ is the intercept. $V_i$ represents uncorrelated spatial heterogeneity effects, that is, random effects caused by non-spatial factors. $u_i$ represents the relative spatial heterogeneity effect, that is, the random effect caused by spatial factors. $g_i$ represents the autoregressive effect. $\mathrm{time}_j$ represents the time effect in the $j$-th year, and $\alpha_1$ represents the time effect coefficient. $\mathrm{psi}_{ij}$ represents the space-time interaction effect. $x_{ijk}$ is the value of the impact factor $k$ in the $j$-th year of the $i$-th district and county, and $\beta_k$ is the regression coefficient corresponding to the impact factor $k$.

## 3. Results

### 3.1. Spatial Distribution Characteristics of Lung Cancer and Environmental Factors at a Unified Geographic Scale

Traditional disease maps are usually drawn based on administrative boundaries or regular grid cells, whose internal details are easily ignored and cannot solve the problem of abrupt changes at grid boundaries, therefore, the disease map is more suitable to be expressed in a spatially continuous way (Goovaerts, 2010). Among them, the methods based on Bayesian statistics and geostatistics occupy the main body of the research, whose basic principle is to process the case information based on the idea of "interpolation" or "smoothing," forming a disease map that is easy to interpret, spatially continuous, and smooth (Kang et al., 2016).

Geocoding provides high-precision case point data for this study. Based on the location information of these point data, KDE can be used to better avoid the mutation problem at the boundary and the homogenization problem inside the grid. Compared with point density estimation, KDE believes that the influence of a certain point on its surroundings is not uniform but gradually decays with the increase of distance, which is more consistent with the laws of geography (Carlos et al., 2010). In addition, since the point element is processed into a smooth surface with integral 1, the cumulative value of each location is the number of cases at that location rather than the total number of cases within the window, the denominator of the incidence calculation should also be the population at the same location. Combined with the 1 km spatial resolution population distribution grid map, the calculation formula of lung cancer incidence is as follows:

$$\mathrm{Incidence}(x, y) = \frac{1}{\mathrm{POP}_{(x,y)}} \sum_{i=1}^{n} K\left(\frac{d_i}{h_i}\right) \quad (5)$$

Where, $\mathrm{Incidence}(x, y)$ is the incidence at the location $(x, y)$; $\mathrm{POP}_{(x,y)}$ is the population at location $(x, y)$. The 1 km spatial resolution incidence map obtained based on this method is shown in Figure 4.

Geographic scale transformation is the process of transforming geographic data or information from one scale to another, which can be an up-scale transformation (scale expansion) or down-scale transformation (scale

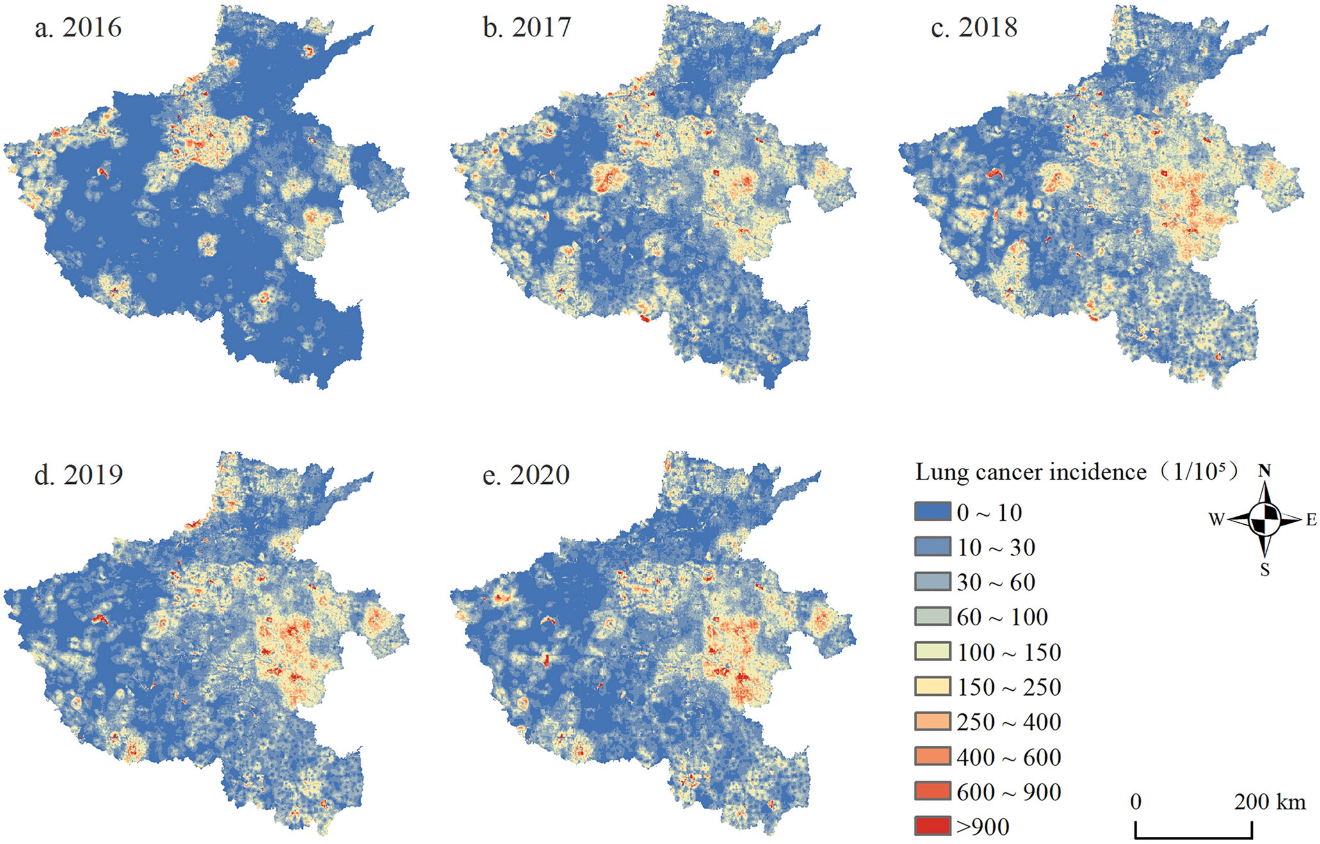


**Figure 4.** Lung cancer incidence map with 1 km spatial resolution in Henan province, 2016–2020.

contraction) (D. Li et al., 2005). For example, among different sources of environmental factor data used in this study, the physical geographic data are raster image data while the socio-economic data are mostly collected by counting units (county-level administrative scale), leading to the different spatial scale indicators from 30 arcseconds, 1, 10 km, to the size of counting units. Different observation scales may affect the study of causality between variables, resulting in deviations or even errors in the acquired rules or knowledge. Therefore, it is necessary to convert all environmental factors into a unified geospatial scale.

In order to clarify the impact of environmental factors on the incidence of lung cancer from a more microscopic perspective, and considering that the incidence data have been made into a raster map with a resolution of 1 km, we unified the spatial resolution of environmental data into 1 km based on the ideas of scale expansion and scale contraction. The scale expansion is realized by resampling, and the soil data is processed in Arcgis10.4. The methods of scale shrinkage mainly include centroid assignment, area weighting, regression, and interpolation (Hallisey et al., 2017). According to the characteristics of the collected data, two methods of Kriging interpolation and multi-factor weight distribution are used to scale the corresponding data. The scale conversion results are shown in Figure 5.

### 3.2. The Influence of Environmental Factors on the Incidence of Lung Cancer

On the basis of the unified data spatial scale, the differentiation and factor detection module of Geodetector was used to study the effects of seven categories of 31 proxy variables on the incidence of lung cancer from two aspects of physical geographic factors and social economic factors. Considering that the occurrence of cancer is often affected by long-term exposure to 10–20a carcinogenic factors, studies have shown that the impact of pollutants on the incidence of lung cancer has an obvious lag, and the 8-year lag period has the strongest correlation (H. Guo et al., 2020; Han et al., 2017). Therefore, the exposure period of environmental variables was

10.1029/2023GH000794

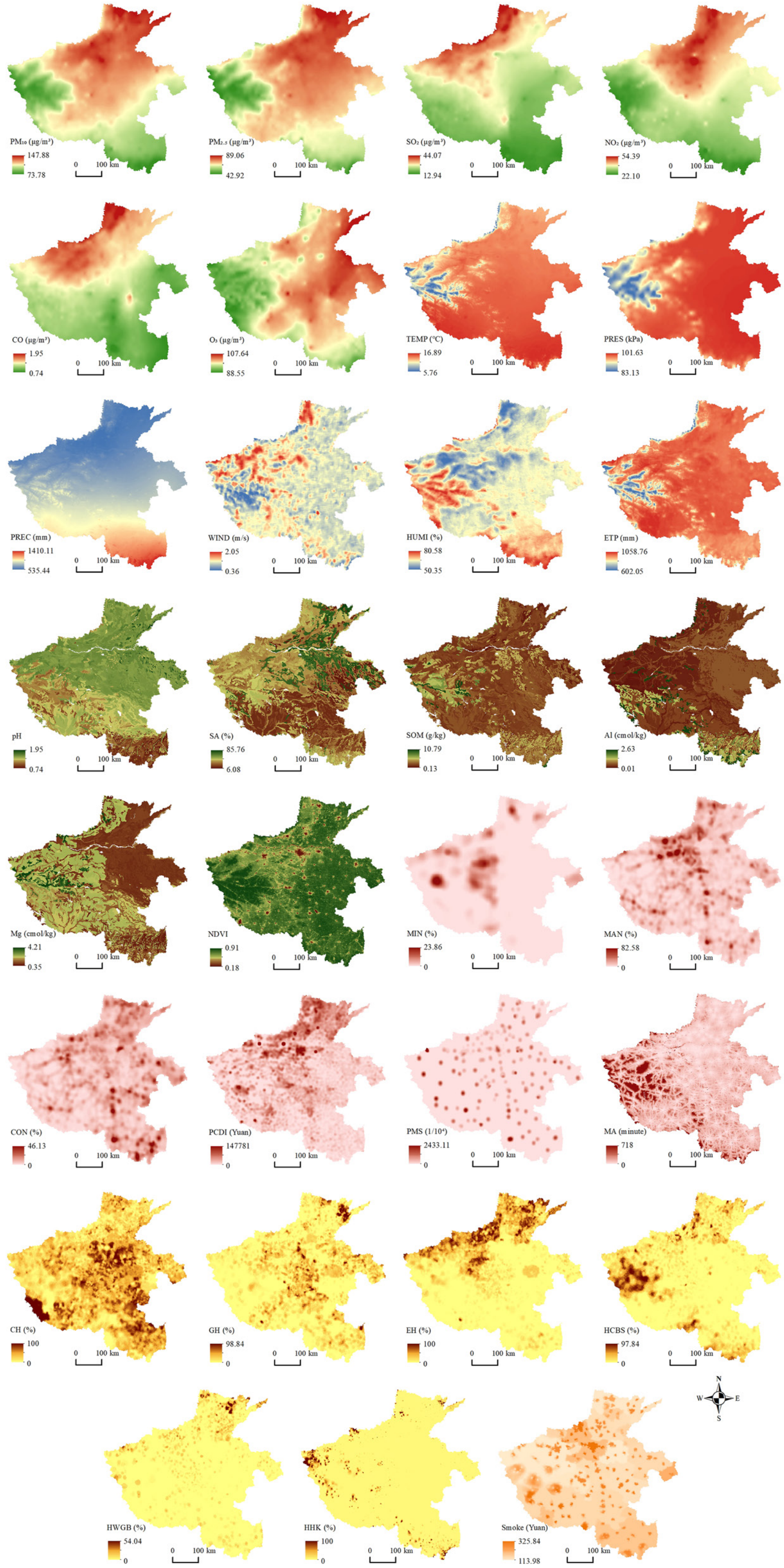


**Figure 5.** Spatial distribution of environmental factors at unified geographical scale from 2009 to 2020.

**Table 2**
*Factor Detection Results (All Genders)*

| Variable | *q* | *p* | Variable | *q* | *p* | Variable | *q* | *p* |
|---|---|---|---|---|---|---|---|---|
| $PM_{10}$ | 0.028 | 0.000 | TEMP | 0.022 | 0.000 | PH | 0.018 | 0.000 |
| $PM_{2.5}$ | 0.037 | 0.000 | PREC | 0.030 | 0.000 | SG | 0.019 | 0.000 |
| $NO_2$ | 0.032 | 0.000 | WIND | 0.028 | 0.000 | SOM | 0.013 | 0.000 |
| $SO_2$ | 0.033 | 0.000 | PRES | 0.013 | 0.000 | Al | 0.031 | 0.000 |
| CO | 0.022 | 0.000 | HUMI | 0.021 | 0.000 | Mg | 0.010 | 0.004 |
| $O_3$ | 0.033 | 0.000 | ETP | 0.019 | 0.000 | NDVI | 0.007 | 0.069 |
| MIN | 0.012 | 0.000 | CH | 0.026 | 0.000 | Smoke | 0.054 | 0.000 |
| MAN | 0.004 | 0.335 | GH | 0.006 | 0.112 | | | |
| CON | 0.007 | 0.063 | EH | 0.006 | 0.118 | | | |
| PCDI | 0.013 | 0.000 | HCBS | 0.018 | 0.000 | | | |
| PMS | 0.005 | 0.262 | HWGB | 0.006 | 0.105 | | | |
| MA | 0.006 | 0.098 | HHK | 0.003 | 0.761 | | | |

set as 8 years. Taking the etiology detection of lung cancer incidence in 2016 as an example, the environmental data used in the analysis was the mean value of the corresponding environmental data from 2009 to 2016.

Factor detection is used to measure the degree of univariate explanation of lung cancer incidence. The larger the value of $q$ is, the stronger the consistency between the dependent variable and the independent variable. Table 2 lists the results of univariate analysis. In general, the order of influence degree of various factors was as follows: smoking behavior > air quality > meteorological conditions > soil vegetation > heating mode > medical level > occupational environment. With significance test $p < 0.05$ as the condition for screening, univariate explanatory ability of lung cancer spatial differentiation was divided into primary variables ($q > 0.05$ and $p < 0.05$), secondary variables ($0.05 > q > 0.03$ and $p < 0.05$), tertiary variables ($0.03 > q > 0.02$ and $p < 0.05$), quaternary variables ($0.02 > q > 0.01$ and $p < 0.05$), and irrelevant variables ($q < 0.01$ or $p > 0.05$) based on $q$ value, and non-irrelevant variables were ranked from high to low (Figure 6).

The result shown the primary variable is per capita tobacco consumption, indicating that smoking behavior was still a decisive factor affecting the incidence of lung cancer. Secondary variables include $PM_{2.5}$, $O_3$, $SO_2$, $NO_2$,

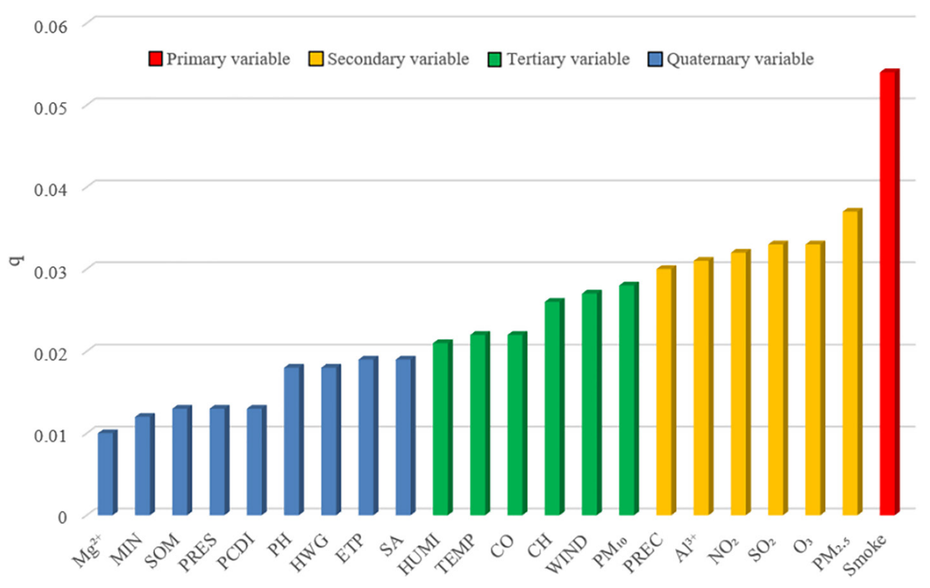


**Figure 6.** Significantly correlated univariate explanatory power ranking of environmental factors.

**Table 3**
*Factor Detection Results (Male)*

| Variable | *q* | *p* | Variable | *q* | *p* | Variable | *q* | *p* |
|---|---|---|---|---|---|---|---|---|
| $PM_{2.5}$ | 0.077 | 0.000 | TEMP | 0.063 | 0.000 | PH | 0.039 | 0.000 |
| $PM_{10}$ | 0.054 | 0.000 | PREC | 0.058 | 0.000 | SG | 0.045 | 0.000 |
| $NO_2$ | 0.057 | 0.000 | WIND | 0.041 | 0.000 | SOM | 0.024 | 0.000 |
| $SO_2$ | 0.055 | 0.000 | PRES | 0.022 | 0.000 | Al | 0.054 | 0.000 |
| CO | 0.043 | 0.000 | HUMI | 0.038 | 0.000 | Mg | 0.022 | 0.000 |
| $O_3$ | 0.075 | 0.000 | ETP | 0.050 | 0.000 | NDVI | 0.016 | 0.000 |
| MIN | 0.033 | 0.000 | CH | 0.032 | 0.000 | Smoke | 0.094 | 0.000 |
| MAN | 0.014 | 0.000 | GH | 0.009 | 0.032 | | | |
| CON | 0.004 | 0.464 | EH | 0.012 | 0.003 | | | |
| PCDI | 0.011 | 0.006 | HCBS | 0.061 | 0.000 | | | |
| PMS | 0.021 | 0.000 | HWGB | 0.012 | 0.002 | | | |
| MA | 0.023 | 0.000 | HHK | 0.009 | 0.047 | | | |

soil $Al^{3+}$ content, and annual precipitation, all of which are physical geographical factors. This means that under the current research scale, the influence of natural environmental factors on the incidence of lung cancer is higher than that of social environmental factors. Except for soil $Al^{3+}$ content, all the important variables were directly or indirectly related to the concentration of air pollutants. By comparing Figures 4 and 5, it can be found that compared with cities with better air quality such as western Henan and southern Henan, residents in areas with high air pollution exposure in northern Henan and Middle Henan have a higher incidence of lung cancer; more attention should be paid to the relationship between air quality and the incidence of lung cancer.

Tertiary variables are also given priority to natural factors, such as $PM_{10}$, CO, wind speed, temperature, and humidity, the exception was coal-fired heating method, as the only significant social factor, because of its relative maneuverability (natural factors are difficult to be improved in a short period of time by active means, while some social factors can be rectified by government compulsory measures, which is more operable), it is bound to become the focus of disease prevention and control in the next step. Quaternary variables are not strongly correlated, including soil gravel content, households with wall-mounted gas boilers, evapotranspiration, soil pH, per capita disposable income, pressure, soil organic matter content, the proportion of mining industry employees, the content of magnesium $Mg^{2+}$, in both natural factors and social factors. Unrelated variables belong to a category of variables that are independent of the spatial differentiation of lung cancer incidence. For example, NDVI is difficult to establish a direct relationship with the incidence of lung cancer due to the regional differentiation of high vegetation cover area and high density population gathering area.

Historically, lung cancer incidence can affect size differences and latencies between sex (Siegfried, 2022). Considering that when constructing BSTIM, standardization by sex ratio is required to estimate the expected number of cases, so sex-specific analysis is crucial to exclude differences caused by sex. In view of this, we used factor detectors to calculate the explanatory power of potential risk factors for different genders on the incidence of lung cancer, and the results are shown in Tables 3 and 4. The results showed that the explanatory power of each environmental factor in sex-disaggregated etiology detection was slightly improved, but it was similar to the order of explanatory power of all sexes. Among them, it is worth noting that smoking is still the first influencing variable for men, followed by air pollutants; but for women, although smoking is still one of the important influencing factors, its importance has dropped to after air pollutants. This work will help to enhance the credibility of the results of subsequent analyses.

### 3.3. Lung Cancer Risk Distribution Adjusted for Multiple Environmental Factors

It is necessary to diagnose the collinearity of variables before using BSTIM to avoid serious multicollinearity. Collinearity is generally tested by correlation or linearity between statistical variables. Before constructing the model, we conducted stepwise regression analysis on the initial variables and used variance inflation factor (VIF) to test the collinearity among variables. When VIF > 10, this variable was considered to have a strong collinearity

**Table 4**
*Factor Detection Results (Female)*

| Variable | *q* | *p* | Variable | *q* | *p* | Variable | *q* | *p* |
|---|---|---|---|---|---|---|---|---|
| $PM_{2.5}$ | 0.081 | 0.000 | TEMP | 0.064 | 0.000 | PH | 0.038 | 0.000 |
| $PM_{10}$ | 0.062 | 0.000 | PREC | 0.061 | 0.000 | SG | 0.054 | 0.000 |
| $NO_2$ | 0.055 | 0.000 | WIND | 0.043 | 0.000 | SOM | 0.021 | 0.000 |
| $SO_2$ | 0.055 | 0.000 | PRES | 0.021 | 0.000 | Al | 0.066 | 0.000 |
| CO | 0.030 | 0.000 | HUMI | 0.037 | 0.000 | Mg | 0.032 | 0.000 |
| $O_3$ | 0.087 | 0.000 | ETP | 0.047 | 0.000 | NDVI | 0.014 | 0.000 |
| MIN | 0.035 | 0.000 | CH | 0.030 | 0.000 | Smoke | 0.069 | 0.000 |
| MAN | 0.013 | 0.000 | GH | 0.010 | 0.020 | | | |
| CON | 0.003 | 0.636 | EH | 0.009 | 0.026 | | | |
| PCDI | 0.009 | 0.023 | HCBS | 0.056 | 0.000 | | | |
| PMS | 0.023 | 0.000 | HWGB | 0.010 | 0.010 | | | |
| MA | 0.024 | 0.000 | HHK | 0.011 | 0.012 | | | |

relationship with other variables and was eliminated. The obtained variables with VIF greater than 10 have been eliminated, and the remaining variables are $PM_{2.5}$, $NO_2$, TEMP, SG, PCDI, CH, HCBS, and Smoke. These eight variables will be used as input parameters for the final model.

The analysis results of the eight environmental indicators included in the model are shown in Table 5. Among them, the influence of temperature and per capita disposable income is negatively correlated, and the rest of the indicators are positively correlated. In addition, the posterior mean represents the corresponding change in lung cancer risk for each 1-unit change in each covariate. For example, for every one yuan increase in per capita tobacco consumption, the risk of lung cancer increases by about 0.1046 standard units. It can be found that the posterior means of $PM_{2.5}$ and $NO_2$ under the same unit are similar, $NO_2$ is slightly larger than $PM_{2.5}$. This may be related to the difference in the variation range of the two concentrations, and the variation range of the $NO_2$ concentration with less content in the air is smaller. The Monte Carlo errors of all covariates are less than the standard deviation, proving the accuracy and rationality of the model.

Exploring the spatial pattern of disease risk is of great significance to better understand the occurrence, prevention, and potential risk factors of disease, and to help the government determine the regional focus of prevention and control and for health care planning. The fine-scale incidence map and relative risk chart showed a higher level of detail. In order to better reflect the spatio-temporal differentiation of lung cancer risk in Henan Province, the RR value was divided into grade 1–6 from small to large (Figure 7). The color of the layer indicates the incidence or relative risk of the disease. The closer to blue, the lower the value, and the closer to red, the higher the value. From the visualization results of the spatial distribution of lung cancer risk in Henan Province from 2016 to 2020, the distribution of lung cancer in Henan showed great imbalance.

**Table 5**
*The Effect of Covariates on the Incidence of Lung Cancer*

| Variable (unit) | Mean | Standard deviation | MC error | 2.5% quantile | Median | 97.5% quantile |
|---|---|---|---|---|---|---|
| $PM_{2.5}$(μg/m$^3$) | 0.0609 | 0.0048 | 0.0003 | 0.0531 | 0.0604 | 0.0692 |
| $NO_2$(μg/m$^3$) | 0.0765 | 0.0155 | 0.0009 | 0.0484 | 0.0706 | 0.1074 |
| TEMP (°C) | −0.0644 | 0.0458 | 0.0026 | −0.0697 | −0.0622 | −0.0527 |
| SG(%) | 0.0810 | 0.0087 | 0.0005 | 0.0663 | 0.0796 | 0.0989 |
| PCDI($10^4$ ¥) | −1.2000 | 0.3428 | 0.0193 | −1.6900 | −1.2290 | −0.6003 |
| CH(%) | 0.0554 | 0.0106 | 0.0006 | 0.0347 | 0.0550 | 0.0736 |
| HCBS(%) | 0.0883 | 0.0318 | 0.0018 | 0.0355 | 0.0819 | 0.1529 |
| Smoke(¥) | 0.1046 | 0.0022 | 0.0001 | 0.1010 | 0.1044 | 0.1082 |

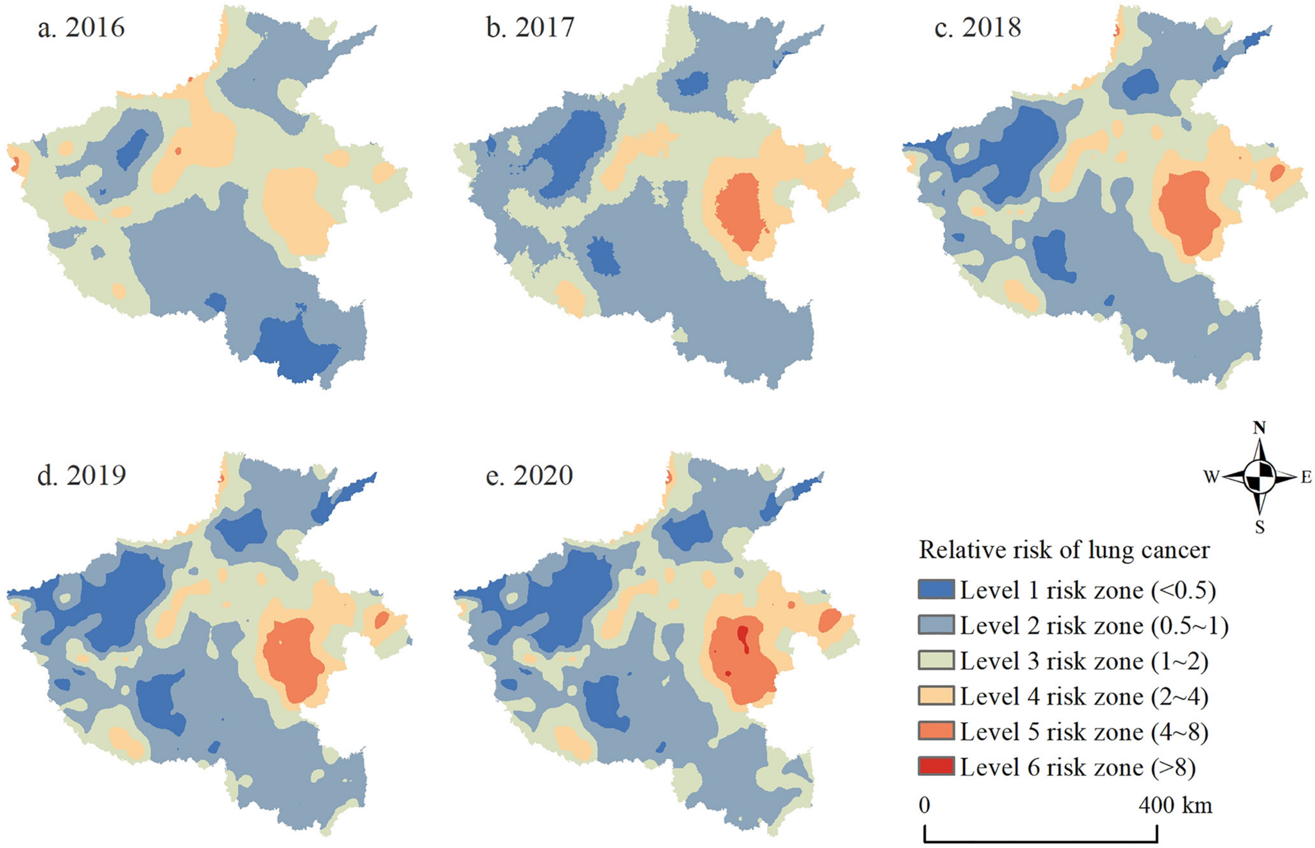


**Figure 7.** Spatial distribution of relative risk of lung cancer in Henan, 2016–2020.

The high-risk areas were mainly distributed in the Huang-Huai-Hai plain area of central and Eastern Henan and the Nanyang Basin area of Southern Henan. The incidence risk of most areas was greater than 1, that is, the actual incidence risk was higher than the expected risk. On one hand, the heating energy in these areas is mainly coal and natural gas, and the combustion of dirty energy deteriorates the indoor air quality (Figure 5). On the other hand, as the level of industrial greening is at a lower level, unbalanced and high-pollution development makes the problems of low resource utilization efficiency and aggravated environmental pollution become increasingly prominent (B. B. Li & Zhang, 2019). High concentrations of $PM_{2.5}$, $NO_2$, and other pollutants have accumulated in the region due to the blocking of mountains in the west. The low-incidence areas were concentrated in the mountainous and hilly areas of southern, northern, and western Henan, and the relative risk of the disease remained below 1 for many years. These areas have high altitudes and wide views, and pollutants are easy to diffuse and dilute (Jans et al., 2018). Moreover, the heating mode is mainly heat sink (Figure 5), this type of heating is generally unified by collective heating, burning energy using a water heating boiler, and separate heating equipment so that residents in the indoor environment avoid exposure to indoor pollutants caused by heating. In addition, the proportion of the second industry in southern Henan and western Henan is low, and the industrial pollution is not serious. However, Hebi city in the northwest of Henan has benefited from the government's relatively sufficient investment in public health services (Figure 5), which has improved the local chronic disease prevention and control as well as residents' awareness and attention. The above factors combined led to the spatial distribution pattern of lung cancer risk in the province, which was higher in the north and lower in the south, and higher in the west and lower in the east.

In the past 5 years, the center of risk of the disease has been moving eastward, forming a new gathering area centered in Zhoukou City, eastern Henan Province. This may be related to the shift of the center of gravity of pollution caused by the shift of the economic center of gravity. In the early development stage of Henan Province, the northern and western Henan regions rose rapidly with its rich mineral resources. The rapid development of heavy industry driven by coal combustion not only promoted the prosperity of the local economy, but also resulted in the serious discharge of industrial waste gas, wastewater, and waste residue, causing a heavy burden of disease for local residents (Y. Liu et al., 2022). However, after the 2008 economic crisis, the economic development of old industrial cities slowed down, and the economy of some counties and cities in eastern Henan began to improve, the economic center of the province shifted significantly to the southeast (Qiao et al., 2017). Meanwhile, Zhengzhou, the capital of Henan province and an important transportation hub in China, is enjoying an industrial boom and has been suffering from air pollution for years, ranking among the top 10 cities with the worst air pollution in China. According to analysis by Geng et al. (2013) on the chemical composition and source of $PM_{2.5}$ in Zhengzhou in 2010, the contribution rate of three main sources (soil dust, coal burning, and secondary aerosols) to $PM_{2.5}$ was 26%, 23%, and 24%, respectively, and the resulting high air pollution exposure was the main cause of the high risk of lung cancer in Zhengzhou over the years (Geng et al., 2013).

## 4. Conclusion

The current study has quantitatively analyzed of the ways and degrees of different environmental factors affecting the incidence of lung cancer. Our results added scientific evidence in exploring the internal mechanism of environmental influence on the incidence of lung cancer, and further contribute to providing references for effective lung cancer prevention.

1. Smoking is still the leading cause of lung cancer. Studies have shown that 75%–90% of lung cancer is related to smoking, tobacco smoke is the single biggest risk factor for lung cancer, and the risk of lung cancer in smokers is 30 times that of non-smokers (Geng et al., 2013). In addition, smoking not only harms smokers themselves, second-hand smoking and environmental smoking are the main sources of cancer and premature death in non-smokers, especially in children (Teresa Perez-Warnisher et al., 2018).
2. Air pollution has gradually become an important factor affecting the incidence of lung cancer. The six selected air pollutants, represented by $PM_{2.5}$ and $NO_2$, were positively and significantly associated with lung cancer incidence. The spatial differentiation of the two pollutants is consistent with the incidence of lung cancer, with low values concentrated in southern cities and high values concentrated in northern and central cities, which is consistent with previous studies (Chuai et al., 2020). High in the North, low in the south distribution pattern may be caused by two reasons: first, the economy of north Henan province depends on mineral resources and industry more than the southern region. In resource-based cities, such as Anyang, Hebi city, and north Henan Jiaozuo, energy consumption per unit Gross domestic product (GDP) is higher than the province and the national average, leading to higher emissions (Y. Zhang et al., 2020). Second, it may be related to the difference in meteorological conditions between the North and the South. Due to the vertical and horizontal turbulence effect, the air quality of cities with lower temperatures in the north is worse (Xu et al., 2017).
3. Meteorological conditions and occupational environment are synergistic factors affecting the incidence of lung cancer. Meteorological elements play an important role in the process of transport, diffusion, accumulation, and sedimentation of air pollutants, while occupational environment reflects the possibility of exposure to carcinogens at work. They do not directly affect each other but cooperate with other environmental factors to promote or inhibit the occurrence and development of diseases. Among the proxy variables reflecting meteorological conditions, the univariate influence of precipitation is the strongest, with a value of 0.03, followed by wind speed (0.028), temperature (0.022), humidity (0.021), evapotranspiration (0.019), and air pressure (0.013).

Precipitation is the scavenger of air pollutants, especially for particulate pollutants such as $PM_{2.5}$ and $PM_{10}$. The longer the precipitation time, the greater the percentage of positive removal (Naresh, 2003). Wind speed is the main power source to promote the dilution and diffusion of air pollutants. Under static wind conditions or when the wind speed is small, the dilution and diffusion ability of the atmosphere is weak and air pollutants easily accumulate, aggravating pollution in local areas (Yang et al., 2020). The influence of temperature on air pollutants may be bidirectional. Low temperature can not only aggravate surface air pollution by inhibiting the dilution of air pollutants and increasing the sources of air pollutants (heating), but also reduce the exposure risk of air pollutants of people by restricting outdoor activities (H. Guo et al., 2021). Both humidity and evapotranspiration are calculated based on the correlation with other climatic factors, their modes of action on air pollutants and diseases are difficult to determine, and their effects are not as obvious as those of other meteorological conditions.

Among workers of different occupations, the workers of the mining industry are more susceptible to the threat of lung cancer, with the univariate explanatory power of 0.012. Miners are the workers most likely to be exposed to carcinogens in the process of operation. Carcinogens, such as the dust of various metals and their compounds, carcinogenic hydrocarbons, radioactive substances such as uranium and their derivatives, arsenic, coal tar, asbestos, and radon gas, can all induce lung cancer (Poinen-Rughooputh et al., 2016). Miners also have a high smoking rate because of the tough conditions and psychological stress in mines. Therefore, influenced by long-term occupational exposure to carcinogens and bad living habits, the incidence and mortality of pneumoconiosis, cor pulmonales, and lung cancer in mining employees are far higher than the national average level (X. Wang et al., 2014).

4. Soil and vegetation are uncertain factors affecting the incidence of lung cancer. Previous studies have shown that soil contamination such as polycyclic aromatic hydrocarbons (Lue et al., 2009) and selenium deficiency (Jaworska et al., 2013) are associated with the development of lung cancer, but there is no evidence of a direct link between the physical or chemical properties of the soil itself and lung cancer. From the results of univariate analysis, $Al^{3+}$ content ($q = 0.031$), sand and gravel content (0.019), PH value (0.018), organic matter content (0.013), and the spatial differentiation of lung cancer incidence were similar to some extent. Among them, $Al^{3+}$ content, PH value, and organic matter content were excluded by stepwise regression analysis because they failed to pass the significance test instead of the collinearity test. Cation exchange and organic matter adsorption can promote or inhibit the degradation process of soil pollutants, which are easily affected by pH value. Therefore, it is speculated that the physical and chemical properties of soil may indirectly affect the incidence of lung cancer through this relationship with soil pollutants, but this effect lacks exact theoretical support and relationship model, still needing further in-depth exploration. In comparison, the influence of sand and gravel content in soil on the incidence of lung cancer is clearer. Soil with high sand content is more likely to produce dust, which is the largest source of $PM_{2.5}$ in Zhengzhou (S. Wang et al., 2017). NDVI is related to the vegetation cover and greening level of a region, and previous studies have shown that it may affect the occurrence of lung cancer (Lin et al., 2014). However, the results of this study showed that the correlation between NDVI and the incidence of lung cancer was not strong. One of the reasons may be the regional differentiation between the high vegetation cover area and the residential area at the spatial analysis scale of 1 km.
5. The level of health care is the key factor driving the incidence of lung cancer. Generally speaking, the medical and health level of a region can be divided into two levels: the individual level refers to the medical consumption capacity that individuals can afford, and the overall level refers to the medical resources and public health service level that the local government can provide. At the individual level, poor people are more affected by unequal exposure to pollutants than rich people because they invest less in self-protective products (Ouyang et al., 2018). In the past, GDP per capita was often used as an indicator to evaluate different levels of household health care (S.-C. Wang et al., 2017) or life satisfaction of family members (You et al., 2018), but this indicator is not as good as disposable income per capita to represent the consumption spending ability of residents. Therefore, per capita disposable income was used to represent the medical consumption ability at the individual level, and the proportion of medical staff and medical accessibility were used to represent the level of medical and health services at the overall level.

The results show that the univariate influence of per capita disposable income is significant, and the way of influence is different in different regions. For energy-driven heavy industrial cities, such as Anyang City, the higher the resident income the more serious the local industrial pollution, and there is a significant positive correlation between the per capita disposable income and the incidence of lung cancer. For agricultural cities with primary industries, such as Zhoukou city, where residents' low income means they spend less on self-protection products, there is a significant negative correlation between per capita disposable income and lung cancer incidence. However, on the whole, the relationship between the two is mainly negative, and high-income people can avoid the occurrence of cancer to a certain extent. For example, Hebi city, also one of the heavy industrial cities in the north, is a cold spot with a much lower incidence of lung cancer than the neighboring cities due to its high proportion of medical staff and strong health awareness among residents. In conclusion, the high level of medical and health conditions can curb the occurrence and development of lung cancer, thereby forcing the incidence of lung cancer in this area.

6. Heating mode is a controllable factor to further mitigate lung cancer. The results showed that among the three types of heating energy (coal, natural gas, and electricity), coal, as the only non-clean energy, has the risk of inducing lung cancer. At present, rural residents in developing countries apparently prefer cheap, low-quality bulk coal with high ash and sulfur content to clean and washed coal, given the daily costs of heating and

cooking (Ahmad & Puppim de Oliveira, 2015). Low quality bulk coal combustion easily produces $SO_2$, $NO_x$, and a large amount of dust. Moreover, the burning of bulk coal is basically a low-level direct emission, which makes it easier to inhale pollutants and has a more direct impact on human health, often leading to bronchitis, asthma, emphysema, and even lung cancer.

As society develops, in addition to looking for cleaner energy to replace traditional solid fuel, people will increasingly choose to turn to more convenient and modern energy use. Heat energy utilization efficiency can represent different levels of the energy conversion process, which is usually affected by economic development, urbanization level, and living standard (Cai & Jiang, 2008). The results of Geodetector showed that the order of single-factor influence from high to low was: HCBS (0.018) > HHK (0.008) > HWGB (0.006). In addition, the interaction between CH and HCBS has a significant role in increasing the incidence of lung cancer. This suggests that heating methods based on coal/biomass stoves increase the risk of lung cancer, and the combination of stoves and coal burning has the potential to further enhance this effect. In addition to cheap and poor-quality bulk coal, the use of inefficient traditional stoves will lead to incomplete combustion accompanied by heavy smoke, which easily spreads and seriously pollutes the surrounding environment. Due to the lack of desulfurization and denitrification equipment, the concentration of sulfur dioxide ($SO_2$) and nitrogen oxide ($NO_x$) in flue gas is high (J. Wang et al., 2016). As a result, even if households adopt natural gas and other clean energy sources, they still face an increasing risk of lung disease because their neighbors still use traditional stoves and solid fuels.

Of course, there are still some limitations in the research and guide us to make improvements in future work. First, the pathogenesis of lung cancer is complex and diverse, and it is a multifactorial disease caused by the interaction of genetics, behavior and environment (Buonanno et al., 2017). Since it is difficult to collect family medical history, genetics, lifestyle and other information of each person in the province, the current work mainly analyzes the impact of environmental factors that are easy to observe on a large scale and have a more intuitive impact on the incidence of lung cancer. Second, although previous studies have found a significant 8-year lag effect of pollutants on lung cancer (H. Guo et al., 2020; Han et al., 2017), the 8-year lag used in this study may still be too short to account for the potentially long latency of lung cancer. We hope to address this limitation in the future as work on lung cancer surveillance continues.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

Data on the annual mean concentration of six air pollutants with a spatial resolution of 10 km from 2009 to 2020 were obtained from the Chinahighairunsustainable series data set (https://zenodo.org/record/6398971#.ZCY9UnZBy38). Synchronous meteorological data with a spatial resolution of 1 km (such as the average temperature data set: http://www.geodata.cn/data/datadetails.html?dataguid=164304785536614&docid=1076) have been obtained from the National Earth System Science Data Center, and other meteorological data sets can also be obtained from the atmospheric sphere classification of the data center. Soil attribute data were obtained from the Basic Attribute data set of China High-resolution National Soil Information Grid (http://www.geodata.cn/data/datadetails.html?dataguid=122556887941975&docid=4045). NDVI from 2009 to 2020 were obtained from the Resource and Environmental Sciences and Data Center of Chinese Academy of Sciences (https://www.resdc.cn/data.aspx?DATAID=257), with a spatial resolution of 1 km. Medical accessibility was calculated in AccessMod 5 (https://www.accessmod.org/). Heating data were obtained from the second pollution survey project of Henan Province, but this data is not publicly available. In addition, other social data from 2009 to 2020 came from the Statistical Yearbook of Henan Province of the corresponding years, and all statistical yearbooks can be obtained from the Henan Provincial Bureau of Statistics (https://tjj.henan.gov.cn/tjfw/tjcbw/tjnj/). Due to the data confidentiality policy signed with the cooperating hospital, the admission data of lung cancer patients are not open to the public.

**Acknowledgments**

This study was supported by the Interdisciplinary Innovative Talents Foundation from Renmin Hospital of Wuhan University (JCRCYG-2022-013).

## References

Ahmad, S., & Puppim de Oliveira, J. A. (2015). Fuel switching in slum and non-slum households in urban India. *Journal of Cleaner Production*, *94*, 130–136. https://doi.org/10.1016/j.jclepro.2015.01.072

Alberg, A. J., & Samet, J. M. (2003). Epidemiology of lung cancer. *Chest*, *123*(1), 21S–49S. https://doi.org/10.1378/chest.123.1_suppl.21s

Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Niksic, M., et al. (2018). Global surveillance of trends in cancer survival 2000-14 (concord-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*, *391*(10125), 1023–1075. https://doi.org/10.1016/s0140-6736(17)33326-3

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2020). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*, 394.

Buonanno, G., Stabile, L., Morawska, L., Giovinco, G., & Querol, X. (2017). Do air quality targets really represent safe limits for lung cancer risk? *Science of the Total Environment*, *580*, 74–82. https://doi.org/10.1016/j.scitotenv.2016.11.216

Cai, J., & Jiang, Z. (2008). Changing of energy consumption patterns from rural households to urban households in China: An example from Shaanxi Province, China. *Renewable & Sustainable Energy Reviews*, *12*(6), 1667–1680. https://doi.org/10.1016/j.rser.2007.03.002

Cao, M., & Chen, W. (2019). Epidemiology of lung cancer in China. *Thoracic Cancer*, *10*(1), 3–7. https://doi.org/10.1111/1759-7714.12916

Carlos, H. A., Shi, X., Sargent, J., Tanski, S., & Berke, E. M. (2010). Density estimation and adaptive bandwidths: A primer for public health practitioners. *International Journal of Health Geographics*, *9*(1), 39. https://doi.org/10.1186/1476-072x-9-39

Cheng, Z., Wang, W., Dai, L.-L., & Kang, Y. (2012). MTHFR c667t polymorphism association with lung cancer risk in Henan province: A case-control study. *Asian Pacific Journal of Cancer Prevention*, *13*(6), 2491–2494. https://doi.org/10.7314/apjcp.2012.13.6.2491

Chuai, X., Fan, C., Wang, M., Wang, J., & Han, Y. (2020). A study of the socioeconomic forces driving air pollution based on a DPSIR model in Henan Province, China. *Sustainability*, *12*(1), 252. https://doi.org/10.3390/su12010252

Darby, S., Hill, D., Auvinen, A., Barros-Dios, J. M., Baysson, H., Bochicchio, F., et al. (2005). Radon in homes and risk of lung cancer: Collaborative analysis of individual data from 13 European case-control studies. *BMJ*, *330*(7485), 223. https://doi.org/10.1136/bmj.38308.477650.63

Geng, N., Wang, J., Xu, Y., Zhang, W., Chen, C., & Zhang, R. (2013). PM2.5 in an industrial district of Zhengzhou, China: Chemical composition and source apportionment. *Particuology*, *11*(1), 99–109. https://doi.org/10.1016/j.partic.2012.08.004

Goovaerts, P. (2010). Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. *Mathematical Geosciences*, *42*(5), 535–554. https://doi.org/10.1007/s11004-010-9286-5

Guo, B., Wang, Y., Pei, L., Yu, Y., Liu, F., Zhang, D. H., et al. (2021). Determining the effects of socioeconomic and environmental determinants on chronic obstructive pulmonary disease (COPD) mortality using geographically and temporally weighted regression model across Xi'an during 2014-2016. *Science of the Total Environment*, *756*, 15. https://doi.org/10.1016/j.scitotenv.2020.143869

Guo, H., Li, W., & Wu, J. (2020). Ambient PM2.5 and annual lung cancer incidence: A nationwide study in 295 Chinese counties. *International Journal of Environmental Research and Public Health*, *17*(5), 1481. https://doi.org/10.3390/ijerph17051481

Guo, H., Li, X., Li, W., Wu, J., Wang, S., & Wei, J. (2021). Climatic modification effects on the association between PM1 and lung cancer incidence in China. *BMC Public Health*, *21*(1), 880. https://doi.org/10.1186/s12889-021-10912-8

Hallisey, E., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., et al. (2017). Transforming geographic scale: A comparison of combined population and areal weighting to other interpolation methods. *International Journal of Health Geographics*, *16*(1), 29. https://doi.org/10.1186/s12942-017-0102-z

Han, X., Liu, Y., Gao, H., Ma, J., Mao, X., Wang, Y., & Ma, X. (2017). Forecasting PM2.5 induced male lung cancer morbidity in China using satellite retrieved PM2.5 and spatial analysis. *Science of the Total Environment*, *607*, 1009–1017. https://doi.org/10.1016/j.scitotenv.2017.07.061

He, S., Li, H., Cao, M., Sun, D., Lei, L., Li, N., et al. (2020). Trends and risk factors of lung cancer in China. *Chinese Journal of Cancer Research*, *32*(6), 683–694. https://doi.org/10.21147/j.issn.1000-9604.2020.06.02

Jans, J., Johansson, P., & Nilsson, J. P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of Health Economics*, *61*, 220–232. https://doi.org/10.1016/j.jhealeco.2018.08.002

Jaworska, K., Gupta, S., Durda, K., Muszynska, M., Sukiennicki, G., Jaworowska, E., et al. (2013). A low selenium level is associated with lung and laryngeal cancers. *PLoS One*, *8*(3), e59051. https://doi.org/10.1371/journal.pone.0059051

Kang, S. Y., Cramb, S. M., White, N. M., Ball, S. J., & Mengersen, K. L. (2016). Making the most of spatial information in health: A tutorial in Bayesian disease mapping for areal data. *Geospatial Health*, *11*(2), 190–198. https://doi.org/10.4081/gh.2016.428

Li, B., & Zhang, J. (2019). Evaluation of greenization and analysis of spatio-temporal characteristics in Henan province. *Areal Research and Development*, *38*, 27–32.

Li, D. L., Zhang, H. G., & Huang, W. G., & IEEE. (2005). Method of multi-scale transformation: An experiments of coastline. In *Proceedings of the 25th IEEE international geoscience and remote sensing symposium (IGARSS 2005)* (pp. 5357–5360).

Li, G., Haining, R., Richardson, S., & Best, N. (2014). Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach. *Spatial Statistics*, *9*, 180–191. https://doi.org/10.1016/j.spasta.2014.03.006

Li, J., Chen, X., Han, X., & Zhang, G. (2020). Spatiotemporal matching between medical resources and population ageing in China from 2008 to 2017. *BMC Public Health*, *20*(1), 845. https://doi.org/10.1186/s12889-020-08976-z

Lin, X.-L., Chen, Y., Gong, W.-W., Wu, Z.-F., Zou, B.-B., Zhao, J.-S., et al. (2014). Geographic distribution and epidemiology of lung cancer during 2011 in Zhejiang province of China. *Asian Pacific Journal of Cancer Prevention*, *15*(13), 5299–5303. https://doi.org/10.7314/apjcp.2014.15.13.5299

Liu, M., Huang, Y., Ma, Z., Jin, Z., Liu, X., Wang, H., et al. (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, S., Chen, Q., Guo, L., Cao, X., Sun, X., Chen, W., & He, J. (2018). Incidence and mortality of lung cancer in China, 2008-2012. *Chinese Journal of Cancer Research*, *30*(6), 580–587. https://doi.org/10.21147/j.issn.1000-9604.2018.06.02

Liu, S., Chen, Q., Quan, P., Zhang, M., Zhang, S., Guo, L., et al. (2016). Cancer incidence and mortality in Henan province, 2012. *Chinese Journal of Cancer Research*, *28*(3), 275–285. https://doi.org/10.21147/j.issn.1000-9604.2016.03.02

Liu, Y., Tian, Z., He, X., Wang, X., & Wei, H. (2022). Short-term effects of indoor and outdoor air pollution on the lung cancer morbidity in Henan Province, Central China. *Environmental Geochemistry and Health*, *44*(8), 2711–2731. https://doi.org/10.1007/s10653-021-01072-0

Lue, J., Xu, R., Zhang, Q., Liu, J., Liao, C., & Wei, F. (2009). Primary investigation of the pollution status of polycyclic aromatic hydrocarbons (PAHs) in water and soil of Xuanwei and Fuyuan, Yunnan Province, China. *Chinese Science Bulletin*, *54*(19), 3528–3535. https://doi.org/10.1007/s11434-009-0499-2

Mendez, W. M., Jr., Eftim, S., Cohen, J., Warren, I., Cowden, J., Lee, J. S., & Sams, R. (2017). Relationships between arsenic concentrations in drinking water and lung and bladder cancer incidence in us counties. *Journal of Exposure Science and Environmental Epidemiology*, *27*(3), 235–243. https://doi.org/10.1038/jes.2016.58

Munoz, U. H., & Kallestal, C. (2012). Geographical accessibility and spatial coverage modeling of the primary health care network in the western province of Rwanda. *International Journal of Health Geographics*, *11*(1), 40. https://doi.org/10.1186/1476-072x-11-40

Naresh, R. (2003). Qualitative analysis of a nonlinear model for removal of air pollutants. *International Journal of Nonlinear Sciences and Numerical Stimulation*, *4*, 379–385. https://doi.org/10.1515/ijnsns.2003.4.4.379

Ouyang, W., Gao, B., Cheng, H., Hao, Z., & Wu, N. (2018). Exposure inequality assessment for PM2.5 and the potential association with environmental health in Beijing. *Science of the Total Environment*, *635*, 769–778. https://doi.org/10.1016/j.scitotenv.2018.04.190

Peng, M., Li, Z., Liu, H., Meng, C., & Li, Y. (2020). *Weighted geocoding method based on Chinese word segmentation and its application to spatial positioning of COVID- 19 epidemic prevention and control* (Vol. 45, pp. 808–815). Geomatics and Information Science of Wuhan University.

Poinen-Rughooputh, S., Rughooputh, M. S., Guo, Y., Rong, Y., & Chen, W. (2016). Occupational exposure to silica dust and risk of lung cancer: An updated meta-analysis of epidemiological studies. *BMC Public Health*, *16*(1), 1137. https://doi.org/10.1186/s12889-016-3791-5

Qiao, G., Pan, S., & Qiao, J. (2017). Dynamic trajectory and contrast analysis about the center of environmental pollution and social economic in Henan province. *Areal Research and Development*, *36*, 23–28.

Seow, W. J., Hu, W., Vermeulen, R., Hosgood, H. D., III., Downward, G. S., Chapman, R. S., et al. (2014). Household air pollution and lung cancer in China: A review of studies in Xuanwei. *Chinese Journal of Cancer*, *33*, 471–475. https://doi.org/10.5732/cjc.014.10132

Siegfried, J. M. (2022). Sex and gender differences in lung cancer and chronic obstructive lung disease. *Endocrinology*, *163*(2). https://doi.org/10.1210/endocr/bqab254

Singer, M. (2013). Respiratory health and ecosyndemics in a time of global warming. *Health Sociology Review*, *22*, 98–111. https://doi.org/10.5172/hesr.2013.1000

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Teresa Perez-Warnisher, M., Carballosa de Miguel, M. P., & Seijo, L. M. (2018). Tobacco use worldwide: Legislative efforts to curb consumption. *Annals of Global Health*, *84*(4), 571–579. https://doi.org/10.29024/aogh.2362

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *Ca: A Cancer Journal for Clinicians*, *65*(2), 87–108. https://doi.org/10.3322/caac.21262

Wang, J., Lou, H. H., Yang, F., & Cheng, F. (2016). Development and performance evaluation of a clean-burning stove. *Journal of Cleaner Production*, *134*, 447–455. https://doi.org/10.1016/j.jclepro.2016.01.068

Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators*, *67*, 250–256. https://doi.org/10.1016/j.ecolind.2016.02.052

Wang, Q., Shen, Q., Zhang, Z., Cai, C., Lu, H., Zhou, X., et al. (2020). Prediction of gene mutation in lung cancer based on deep learning and histomorphology analysis. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, *37*, 10–18.

Wang, S., Yu, S., Li, P., Wang, L., Mehmood, K., Liu, W., et al. (2017). A study of characteristics and origins of haze pollution in Zhengzhou, China, based on observations and hybrid receptor models. *Aerosol and Air Quality Research*, *17*(2), 513–528. https://doi.org/10.4209/aaqr.2016.06.0238

Wang, S.-C., Sung, W.-W., Kao, Y.-L., Hsieh, T.-Y., Chen, W.-J., Chen, S.-L., & Chang, H. R. (2017). The gender difference and mortality-to-incidence ratio relate to health care disparities in bladder cancer: National estimates from 33 countries. *Scientific Reports*, *7*(1), 4360. https://doi.org/10.1038/s41598-017-04083-z

Wang, X., Lin, S., Yano, E., Yu, I. T. S., Courtice, M., Lan, Y., & Christiani, D. C. (2014). Exposure-specific lung cancer risks in Chinese chrysotile textile workers and mining workers. *Lung Cancer*, *85*(2), 119–124. https://doi.org/10.1016/j.lungcan.2014.04.011

Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., et al. (2021). 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

Xu, L. J., Zhou, J. X., Guo, Y., Wu, T. M., Chen, T. T., Zhong, Q. J., et al. (2017). Spatiotemporal pattern of air quality index and its associated factors in 31 Chinese provincial capital cities. *Air Quality Atmosphere and Health*, *10*(5), 601–609. https://doi.org/10.1007/s11869-016-0454-8

Yang, J., Shi, B., Shi, Y., Marvin, S., Zheng, Y., & Xia, G. (2020). Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. *Sustainable Cities and Society*, *54*, 101941. https://doi.org/10.1016/j.scs.2019.101941

You, W., Ruhli, F. J., Henneberg, R. J., & Henneberg, M. (2018). Greater family size is associated with less cancer risk: An ecological analysis of 178 countries. *BMC Cancer*, *18*(1), 924. https://doi.org/10.1186/s12885-018-4837-0

Zhang, H., Xiao, L., Zhou, Y., & Zhao, D. (2020). Method for aggregating geocoding and toponymic poi search services from multiple map APIs. *Geography and Geo-Information Science*, *36*, 1–7.

Zhang, X., Xu, C., & Xiao, G. (2018). Space-time heterogeneity of hand, foot and mouth disease in children and its potential driving factors in Henan, China. *BMC Infectious Diseases*, *18*(1), 638. https://doi.org/10.1186/s12879-018-3546-2

Zhang, Y., Li, X., Wang, S., Guo, J., & Lv, G. (2020). A multi-objective zoning framework for mineral resources development and management: A case study in Henan province, China. *Natural Resources Research*, *29*(5), 3103–3119. https://doi.org/10.1007/s11053-020-09652-0