applied sciences

MDPI

Article

Air Quality Research Based on B-Spline Functional Linear Model: A Case Study of Fujian Province, China

Yihan Xu [1,2], Tiange You [1,2,*], Yuanyao Wen [1,2], Jing Ning [1,2], Yanglan Xiao [1,2] and Huirou Shen [1,2]

1 College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 1211153027@fafu.edu.cn (Y.X.); 1201153021@fafu.edu.cn (Y.W.); ningjing@fafu.edu.cn (J.N.); 1221153030@fafu.edu.cn (Y.X.); 1221153025@fafu.edu.cn (H.S.)
2 Fujian Statistical Information Research Center, Fuzhou 350000, China
* Correspondence: 000q101026@fafu.edu.cn

Abstract: It is generally accepted that air quality is closely related to human health. In this study, to investigate the dynamic characteristics of air quality and explore the driving factors of air pollution, the Air Quality Index (AQI) and concentration data of six air pollutants (CO, NO_2, O_3, $PM_{2.5}$, PM_{10}, and SO_2) were fitted to functional curves using the B-spline basis function. Compared with discrete data, functional data can better express the dynamic characteristics of data and reduce information loss. Additionally, functional clustering based on the principal component coefficient was established to analyze the spatiotemporal dynamic characteristics of air quality, and a functional linear model was established to analyze the relationship between pollutants and anthropogenic factors. The results showed that air pollutants in Fujian Province were found to have certain temporal and spatial heterogeneity, among which the seasonal characteristics of NO_2 and O_3 (high in summer, low in winter) were opposite to those of the other pollutants considered. The spatial distribution of air pollution was low (high) pollution in inland (coastal) areas, and the primary air pollutants in Fujian Province were PM_{10} and $PM_{2.5}$. The functional linear model indicated that anthropogenic factors (e.g., vehicle numbers and emissions of industrial NO_X emissions) were found to have a notable impact on air pollutants. The findings of this study could act as a reference in support of air pollution control.

Keywords: air quality; functional data; functional clustering; functional linear model; B-spline basis function

check for updates

Citation: Xu, Y.; You, T.; Wen, Y.; Ning, J.; Xiao, Y.; Shen, H. Air Quality Research Based on B-Spline Functional Linear Model: A Case Study of Fujian Province, China. *Appl. Sci.* **2023**, *13*, 11206. https://doi.org/10.3390/app132011206

Academic Editors: Hyo Choi, Milton S. Speer and Selahattin Incecik

Received: 15 September 2023
Revised: 6 October 2023
Accepted: 10 October 2023
Published: 12 October 2023

1. Introduction

In recent decades, air pollution has become one of the most serious challenges to public health globally [1]. According to the 2022 World Air Quality Report [2], only six countries met the World Health Organization's annual $PM_{2.5}$ guideline value of $\leq 5\ \mu g/m^3$: Australia, Estonia, Finland, Grenada, Iceland, and New Zealand. Additionally, the harm caused by air pollution has been growing rapidly, causing a range of related diseases and even deaths. According to the World Health Organization [3], 3.2 million people die prematurely each year due to illnesses caused by air pollution. Among these, 32% are attributed to ischemic heart disease, 23% to stroke, 21% to lower respiratory infections, 19% to chronic obstructive pulmonary disease, and 6% to lung cancer. Air pollution is mainly caused by natural sources and human activities, and air pollution caused by human activities usually lasts for a long time and is widespread. Industrialization and urbanization pose the major threats to the quality of the air we breathe [4]. Fortunately, air pollution is a preventable risk, and reducing pollution at its source can yield rapid and substantial improvements. Therefore, an analysis of both air pollution and its influence factors is crucial for controlling air pollution and promoting the sustainable development of the ecological environment.

The Air Quality Index (AQI) and its six pollutant indicators have long been the focus of research into air pollution. Analyzing the spatiotemporal distribution characteristics of air pollution is a prerequisite for the formulation of reasonable plans for pollution control action [5,6]. Clarifying the correlations between different air pollutants can also effectively identify the main air pollutants in specific regions [7–9]. Previous evaluations of air quality based on discrete data were mainly focused on the establishment of multiple regression models [10,11], the use of the Moran's Index [12], and other methods suitable for statistical analysis and comprehensive evaluation. Machine learning tools can also be used to systematically monitor multiple air pollutants [13–17]. It can extract relevant information by artificial intelligence in order to predict air pollution concentration and assess air pollution. Moreover, the rebound effects of economic development, technological advancement, or other factors are evident causes of air pollution [18], and several studies have examined the impact of anthropogenic factors on air pollution and the distribution of specific pollutants using spatial analysis methods [19,20], the random forest method [21], econometric tools [22,23], and the Environmental Kuznets Curve [24,25]. However, the concentration of air pollutants was a continuous state on a time scale, and the above analysis was performed using discrete air quality data, so it cannot accurately express the continuous variation of data. In contrast, functional data can capture the continuous curve or functions of such data and rely on fewer assumptions [26,27]. This flexibility enables the modeling and analysis of complex relationships and patterns in functional data, which may not be easily captured by traditional methods. Studying the spatial and temporal heterogeneity of air pollutants using functional data to achieve low information loss can reflect the dynamic continuous features of the data in greater depth [28]. By treating data as functions, functional data can capture the full shape and dynamics of data, providing more accurate representations of underlying patterns and trends. The emergence of functional data has greatly enriched statistical analysis methods and provided more powerful tools for a better analysis and solution of practical problems. A functional data analysis, which was proposed by Ramsey [29] and then further refined by other statisticians, can represent data in a functional form. On this basis, a functional cluster analysis has gradually developed from the basic cluster analysis approach. A cluster analysis is an important data analysis technique used in data mining, the purpose of which is to categorize data according to their intrinsic attributes [30]. The functional cluster analysis involves clustering methods based on direct distance and functional clustering methods based on the dimensionality reduction [31]. Given the unique inherent structure of the data, the selection of the appropriate analysis method can produce explanatory clustering results. Moreover, a functional linear model was developed to analyze the relationship between the independent variables and the dependent variables. In comparison with the traditional linear model, the functional linear model can determine dynamic characteristics and analyze the direction of influence and intensity of factors [32,33].

Fujian Province, an important province in eastern China, has had the fastest growing economy in China over the past decade. In the process of its economic and social development, the concentrations of air pollutants in Fujian Province, especially O_3, have also shown a slow upward trend [34]. Given the limitations of previous research and the urgency of air pollution control in Fujian Province, the primary objectives of this study were as follows: (1) to clarify the air quality characteristics in Fujian Province using functional data, (2) to explore the driving factors of air pollution, and (3) to present practical suggestions for air pollution prevention. To fulfill those objectives, appropriate B-spline basis functions were chosen to fit the AQI and six pollutants data. The roughness penalty method was used to prevent over-fitting and to ensure the smoothness of the function. A functional clustering method based on principal component coefficients was also applied to analyze the spatial differences of pollutants, and the driving factors of air pollution were determined using the functional linear model.

2. Materials and Methods

2.1. Study Area

Fujian Province lies on the southeast coast of China (23°31′–28°18′ N, 115°50′–120°43′ E) and covers an area of 1.24×10^5 km^2 (Figure 1). Fujian Province has nine prefecturelevel cities: Fuzhou, Longyan, Ningde, Putian, Quanzhou, Sanming, Xiamen, Zhangzhou, and Nanping. The selection of nine cities in Fujian Province for this study can provide a comprehensive perspective on an air quality analysis. The selected cities cover a wide range of geographical locations within Fujian Province, including coastal areas, mountainous regions, and urban centers. The nine cities have varying population densities, and diverse economic activities, such as Quanzhou's manufacturing industry or Xiamen's tourism. Areas of mountains and hills account for more than 80% of the total area of Fujian Province. The region has a subtropical maritime monsoon climate. Because the region is at the intersection of land and sea, its atmospheric pollution mechanism is complicated. The convergence of land and sea introduces various pollution sources into the atmosphere, and the interaction of pollutants emitted from land and sea can lead to chemical reactions in the atmosphere. Additionally, the interaction between land and sea creates complex atmospheric circulation patterns. This circulation can carry pollutants from inland sources to the coast, and vice versa, leading to the accumulation of pollutants and increased complexity in their dispersion. In 2022, the region's GDP was 5.310985 trillion billion yuan (RMB), ranking first in eastern China, with an average annual growth rate of 6.4% [35].

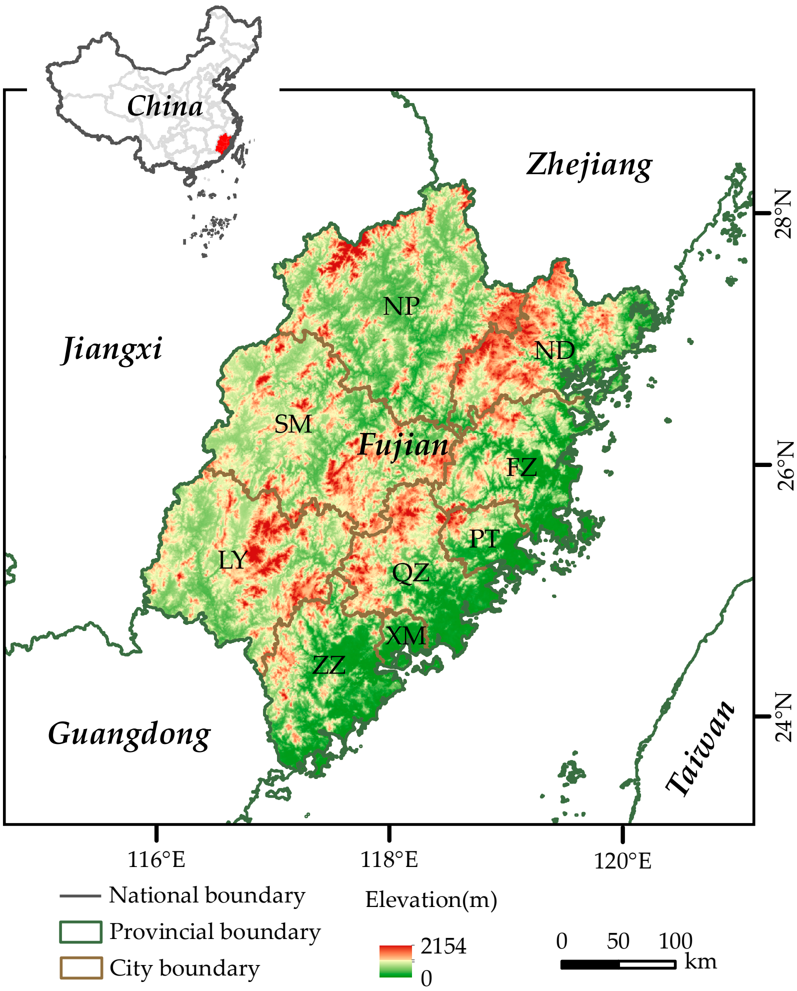

Figure 1. Location of the study area (FZ: Fuzhou city, LY: Longyan city, ND: Ningde city, PT: Putian city, QZ: Quanzhou city, SM: Sanming city, XM: Xiamen city, ZZ: Zhangzhou city, NP: Nanping city).

2.2. Data Sources and Preprocessing

The data used in this study included daily AQI values and concentration values of six pollutant indicators from the nine cities in Fujian Province obtained from 1 January 2015 to 31 December 2021. The air pollutants included CO, NO_2, O_3, $PM_{2.5}$, PM_{10}, and SO_2. Additionally, data of vehicle ownership and industrial NO_X emissions were selected to analyze the drivers of air pollution. The data were collected from the Weather Aftermath website of China (http://www.tianqihoubao.com, accessed on 1 January 2022) and the Statistical Yearbooks of China (http://www.stats.gov.cn/, accessed on 1 May 2022).

The AQI quantitatively describes a dimensionless index of air quality conditions, which is regulated by national regulations and applies only in China. The AQI describes the combined urban air quality conditions using categories (Table 1) based on the maximum Individual Air Quality Index (IAQI) as follows [36]:

$$IAQI_i = \frac{IAQI_U - IAQI_L}{P_U - P_L}(N_i - P_L) + IAQI_L \quad (1)$$

where $IAQI_i$ is the $IAQI$ of pollutant i, N_i is the concentration value of pollutant i, P_U (P_L) is the high-value (or low-value) limit for the pollutant whose individual concentration value is comparable with that of pollutant i, $IAQI_U$ is the $IAQI$ corresponding to P_U, $IAQI_L$ is the $IAQI$ corresponding to P_L, and AQI is the maximum of the $IAQI$ corresponding to the six air pollutants.

Table 1. Air quality classification standards.

Air Quality Index	**0–50**	**51–100**	**101–150**	**151–200**	**201–300**	**>300**
Categories	Excellent I	Good II	Light pollution III	Moderate pollution IV	Heavy pollution V	Severe pollution VI

2.3. Framework of Methodology

This study was conducted in three steps. Firstly, the basis function of the data was determined by observing the errors and periodicity of the data, and the smoothing parameters were further established by the generalized cross-validation (GCV) method to realize the functionalization of the air quality data. Secondly, functional clustering based on principal component coefficients was used to analyze the spatiotemporal dynamic characteristics of air pollutants. Through the principal component weight function, the principal component score was calculated and clustered on this basis. Finally, a functional linear model was established for analyzing the response relationship between air pollutants and anthropogenic activities. The method framework of this study is shown in Figure 2.

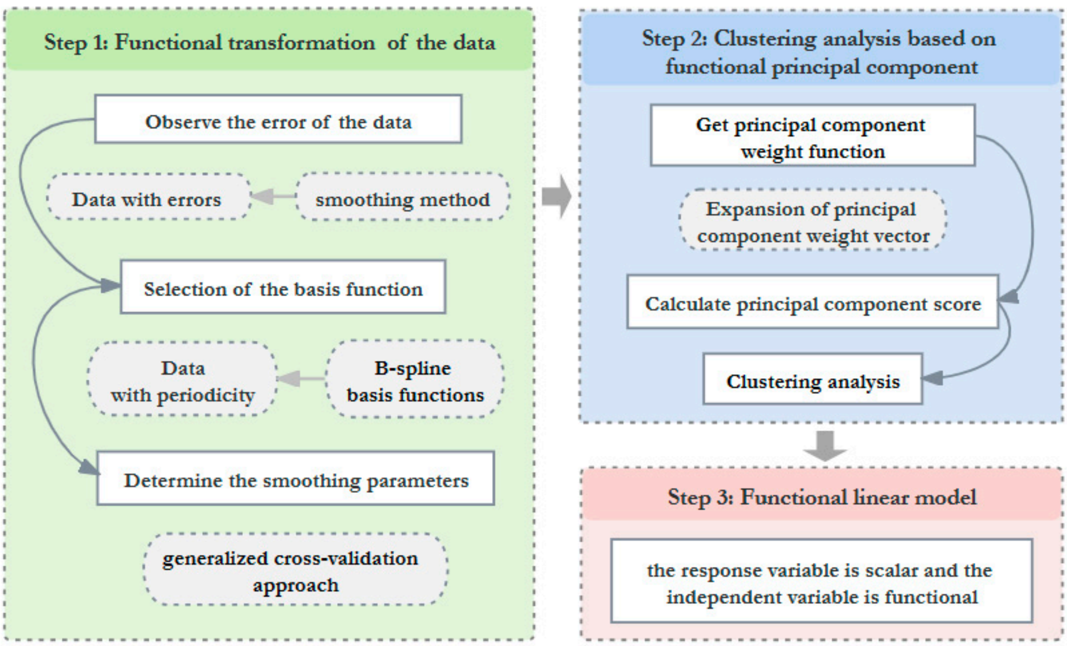

Figure 2. The flowchart of the method in the study.

2.4. Functionalized Processing of Data

With the continuous upgrading of data observation equipment, an increasing number of data types are being collected, and the data format is becoming more complex. At this time, the conclusions drawn by traditional statistical analysis methods often deviate greatly from the actual situation, and functional data are introduced to solve this problem. Functional data are used to convert the observed discrete data into a continuous curve through the basis function and apply it to the statistical analysis model to obtain results that are closer to the actual conclusion. Compared to traditional methods, it is able to take advantage of additional information such as dynamic changes, smoothness, etc. To functionalize the data, the process is as follows.

Step 1: Observe the error of the data.

Generally, observations with errors were smoothed and observations without errors were interpolated [37]. The air quality monitoring stations recorded the data with errors; therefore, the smoothing method was more appropriate to fit the data.

Step 2: Selection of the basis function.

The Fourier basis function is suitable for fitting data with periodicity, and the B-spline basis function for fitting non-periodic data [38]. Given that the periodicity of the air quality data obtained in Fujian Province was not obvious (Figure S1), the B-spline basis function was used to functionalize the data in this study. If a set of basis functions is defined as $\phi_{m,k}(t)$, the estimate can be calculated as follows [39]:

$$\hat{x}(t) = \sum_{k=1}^{K} c_k \phi_{m,k}(t) \tag{2}$$

$$\phi_{m,k}(t) = \begin{cases} 1, t_m \leq t \leq t_{m+1} \\ 0, otherwise \end{cases} \tag{3}$$

$$\phi_{m,k}(t) = \frac{t - t_m}{t_{m+k-1} - t_m} \phi_{m,k-1}(t) + \frac{t_{m+k} - t}{t_{m+k} - t_{m+1}} \phi_{m+1,k-1}(t) \tag{4}$$

where c_k is the coefficient matrix, t_m is the time point of the observed data, k is the order of the B-spline basis function, and K is the number of basis functions. In this study, fourth-order B-spline basis functions were set to fit the air quality data, so the value of k was four.

Step 3: Determine the smoothing parameters.

To balance the relationship between goodness-of-fit and smoothness, the roughness penalty method was used. If $x(t)$ is assumed to be derivable, then the mean square error with a penalty term can be defined as follows [39]:

$$PENSSE_{\lambda}(x|y) = \sum_{i=1}^{N} \{y_i - x(t_i)\}^2 + \lambda \cdot PEN_2(x) \tag{5}$$

where N is the number of raw data, λ is the smoothing parameter, y_i is the observed data, and the penalty term function $PEN_2(x)$ is $\int_{\mathrm{T}} \{D^2 x(t)\}^2 dt$. This is calculated in terms of the curvature of the function, where $D^2(\cdot)$ is the second-order derivative of the function. The higher the $PEN_2(x)$, the greater the volatility of the function.

Considering the m-order derivative function $D^m x(t)$ of function $x(t)$, a more general formula for characterizing the degree of roughness (fluctuation) of the function was calculated as follows [39]:

$$PEN_m(x) = \int_T \{D^m x(s)\}^2 ds = c' R c \tag{6}$$

$$\hat{c} = (\Phi' W \Phi)^{-1} \Phi' W y \tag{7}$$

$$\hat{y} = \Phi \hat{c} = \Phi (\Phi' W \Phi)^{-1} \Phi' W y \tag{8}$$

Thus, the sum of the squares of the more general residuals with penalties corresponding to Equation (5) was calculated as follows:

$$PENSSE_m(y|c) = (y - \Phi c)' W (y - \Phi c) + \lambda c' R c \tag{9}$$

$$\hat{c} = (\Phi' W \Phi + \lambda R)^{-1} \Phi' W y \tag{10}$$

$$\hat{y} = \Phi(\Phi' W \Phi + \lambda R)^{-1} \Phi' W y \tag{11}$$

where $R = \int_T D^m \phi(s) D^m \phi'(s) ds$, $\phi(s)$ is a column vector formed by $\phi_k(t)(k = 1, \cdots, K)$; c is the optimal solution where the weighted least squares method $(y - \Phi c)' W (y - \Phi c)$ reaches the minimum value, i.e., $\hat{c}$; y is the vector of the data columns that can be tuned; column vector $\hat{y}$ is formed from the fitted value of y; and W is the weight matrix. When $\lambda = 0$, Equation (10) is equal to Equation (11). By changing the value of λ, the degree of data smoothing was adjusted.

To determine the optimal value of the smoothing parameter λ, the GCV approach was used. GCV, with the effect of the degrees of freedom on the cross-validation values considered, was calculated as follows [40]:

$$GCV(\lambda) = \frac{1}{N} \sum_{i=1}^{N} GCV_i(\lambda) = \left(\frac{N}{N - df(\lambda)}\right)\left(\frac{SSE}{N - df(\lambda)}\right) \tag{12}$$

where $df(\lambda)$ is the fitted degree of the freedom of λ, $df(\lambda) = trace[H(\lambda)]$; $H(\lambda)$ is a function of λ and $\hat{x} = H(\lambda)y$, $H = \phi(\phi'\phi + \lambda R)^{-1}\phi'$, and $R = \int \phi(t)\phi'(t)dt$; and SSE are the sum of the squared errors. When the value of the GCV taken is the smallest, the optimal smoothing parameter is obtained.

2.5. Clustering Analysis Based on Functional Principal Component

To observe the spatial similarity in the air pollutants, a functional cluster analysis was performed with the air quality data. In this study, a cluster analysis based on the principal component coefficient was used. By converting the high-dimensional space problems into low-dimensional space problems, clustering algorithms were able to overcome the curse of dimensionality and improve the efficiency and effectiveness of the clustering process. This allows for better understanding and interpretation of the data, making it easier to identify similarities and differences among data points. Functional principal components are similar to the idea of solving principal components in a general multivariate statistical analysis. The solution of the principal component weight vector was transformed into the solution of the principal component weight function. The principal component score can be calculated using Equation (13), and the k-th principal component score of the i-th fitted curve can be denoted as follows [39]:

$$f_{ik} = \int \xi_k(t) x_i(t) dt \tag{13}$$

where $\xi(t)$ is the combined weight, $i = 1, 2, 3, \cdots, N$. In the functional principal component analysis, the combined weights ξ_k in a multivariate statistical analysis become the function values $\xi_k(t)$, and $x_i(t)$ represents the functional data. The constraint satisfied by the weight function $\xi_j(t)$ of the principal component of the j-th function of the i-th fitted curve was calculated as follows:

$$\begin{cases} \max \frac{1}{8} \sum\limits_{i=1}^{9} f_{ij}^2 = \max \frac{1}{8} \sum\limits_{i=1}^{9} \left(\int \xi_j(t) x_i(t) dt\right)^2 \\ \int_T (\xi_j(t))^2 dt = 1 \\ \int_T \xi_j(t)\xi_1(t)dt = \cdots = \int_T \xi_j(t)\xi_{j-1}(t)dt = 0 \end{cases} \tag{14}$$

The weight functions can be expanded by basis functions, and the covariance functions of variables $x(t)$ and $x(s)$ are $v(s,t)$. The characteristic equation satisfied by the weight function $\xi(t)$ of the principal component of each function was calculated as follows:

$$\int_T \xi(t)v(s,t)dt = \lambda\xi(s) \tag{15}$$

where λ is the eigenvalue and $\xi(s)$ is the characteristic function corresponding to λ. The integral transformation of the weight function $\xi(s)$ into V can be defined as follows:

$$V\xi(s) = \int v(s,t)\beta(t)dt \tag{16}$$

where V is the covariance operator. Therefore, the problem of finding the largest characteristic function can be transformed into solving the characteristic equation as follows:

$$V\xi = \lambda\xi \tag{17}$$

where ξ is no longer an eigenvector, but a characteristic function.

In this study, a cluster analysis was conducted using the Euclidean distance. The Euclidean distance between the fitted function curves $x_i(t)$ and $x_j(t)$ in the functional data analysis was calculated as follows [31]:

$$D^2(x_i(t), x_j(t)) = \int_a^b (x_i(t) - x_j(t))^2 dt = (b-a)[(c_{i0} - c_{j0})^2 + (c_{i1} - c_{j1})^2 + \cdots + (c_{iK} - c_{jK})^2] \tag{18}$$

where c_{iK} and c_{jK} are vectors of the expansion coefficients for $x_i(t)$ and $x_j(t)$, respectively. The distance between the function curves depends on the coefficients of the function expansion.

2.6. Functional Linear Model

This study used the functional linear model to analyze the driving factors of air pollution. The functional linear model is an extension of the multiple regression model. It transforms the regression coefficients in the traditional model into a function of time t, which is the slope function. This approach was adopted in this study because the functional linear model has a stronger forecasting capability and can analyze the direction of influence and intensity of the factors.

The types of the functional linear model were mainly as follows: (1) the response variable was scalar and the independent variable was functional; (2) the response variable was functional and the independent variable was scalar; and (3) both the response variable and the independent variable were functional variables. The first type of functional linear model used in this study was calculated as follows [41]:

$$Y_i = \beta_0(t) + \int \beta(t)X_i(t)dt + \varepsilon_i \tag{19}$$

$$\beta(t) = \sum_{k=1}^{P} \beta_k\phi_k(t) \tag{20}$$

Using the basis function coefficient expansion, the roughness penalty was used to estimate $\beta(t)$, where Y_i is the scalar response variable, $\beta_0(t)$ is the intercept term, $\beta(t)$ is the slope function to be estimated, $X_i(t)$ is the functional independent variable, ε_i is the random perturbation term, $i = 1, 2, ..., n$, β_k is the coefficient matrix, and $\phi_k(t)$ represents a set of basis functions.

3. Results

3.1. Proportion of Days with the Primary Pollutants

Figure 3 shows the proportion of days with primary pollutants in the nine cities of Fujian Province from 2015 to 2021, in which primary pollutants were not counted when the

AQI level reached the excellent category (that is, less than or equal to 50). Other categories included days when there was more than one primary pollutant (e.g., PM_{10} and $PM_{2.5}$ were the primary pollutants on a day), and days when the primary pollutant was SO_2 or CO. It is evident that except for Putian and Nanping, other cities had the largest proportion of days with the PM_{10} pollutant as the primary pollutant, and the second largest proportion of days with the $PM_{2.5}$ pollutant as the primary pollutant. Conversely, Putian and Nanping had the largest proportion of days with the $PM_{2.5}$ pollutant as the primary pollutant, and the second largest proportion of days with the PM_{10} pollutant as the primary pollutant. Therefore, they indicated that the primary air pollutants in Fujian Province were mainly PM_{10} and $PM_{2.5}$. It is worth noting that the proportion of days in Putian with the O_3 pollutant as the primary pollutant exceeded 20%, and the pollution situation was more serious than in other cities.

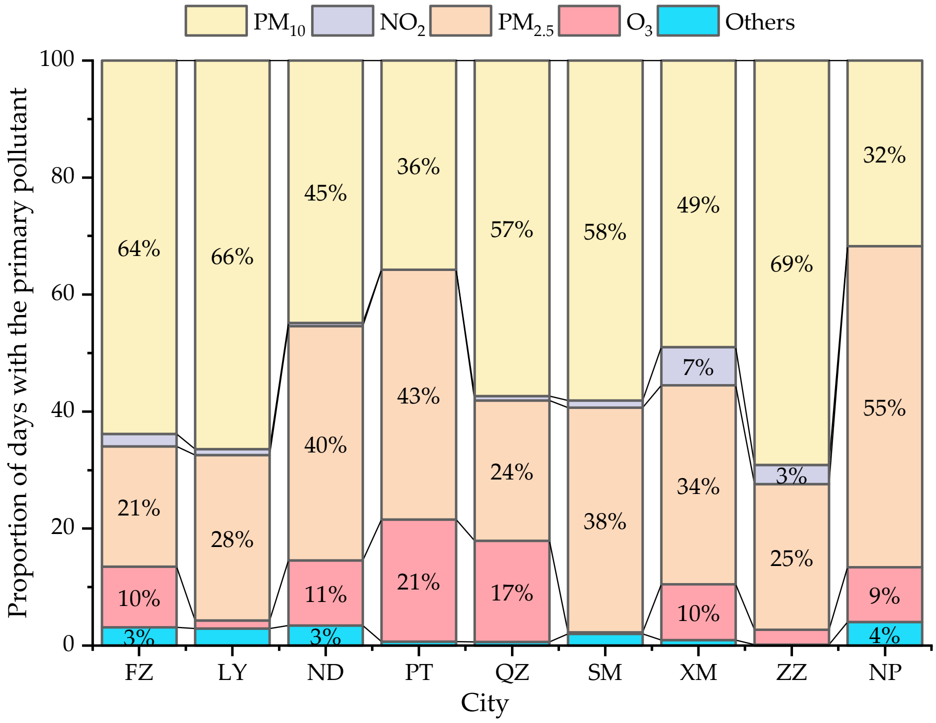

Figure 3. Proportion of days with the primary pollutants in nine cities of Fujian Province from 1 January 2015 to 31 December 2021.

3.2. Spatiotemporal Characteristics of Air Pollution

By the GCV approach, the optimal smoothing parameter of AQI was obtained (Figure 4). At this time, $\log_{10}(\lambda)$ was taken as -0.5, that is, the optimal smoothing parameter of AQI was $10^{-0.5}$. Similarly, the GCV method was used to determine the optimal smoothing parameter for each of the six pollutants (Table 2). Through comparison with Figure S1, it is evident that the functionalized curves retained the information of the original data and had a certain smoothness (Figure 5). Thus, they were considered to accurately capture the features of the changes in air quality in Fujian Province.

Table 2. Optimal parameters of the six pollutants were determined using the generalized cross-validation approach.

Pollutant	Optimal Smoothing Parameters
CO	$10^{0.1}$
NO_2	$10^{0.2}$
O_3	$10^{-0.8}$
$PM_{2.5}$	$10^{-0.2}$
PM_{10}	$10^{-0.5}$
SO_2	$10^{-1.7}$

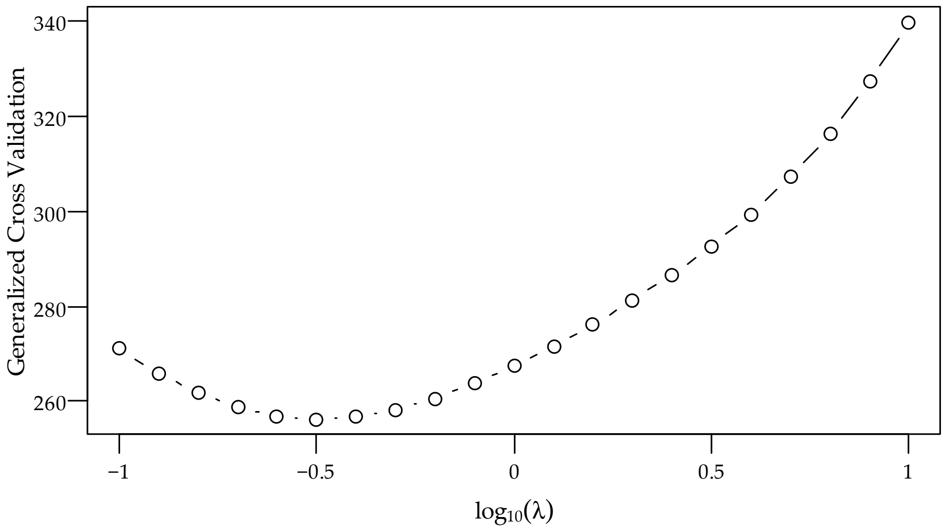

Figure 4. Optimal parameters of the Air Quality Index were determined using the generalized cross-validation approach.

Table 3 illustrates the principal component variance contribution results of the AQI and the six pollutants. The cumulative contribution rate exceeded 85%, indicating that the principal component analysis results were excellent and most of the information contained within the original data was extracted. The dynamic change of the first three principal components of the extracted AQI and six pollutants is displayed in Figure 6. It was indicated that the CO, O_3, $PM_{2.5}$, PM_{10}, and SO_2 pollutants were mainly affected by the first principal component, and the AQI and NO_2 pollutants were mainly affected by the third principal component. The weight functions of the first principal component of the O_3, $PM_{2.5}$, and PM_{10} fluctuated greatly but remained above zero. It was indicated that the concentration of these pollutants was greater than the mean, and the greater the positive fluctuation, the higher the pollutant concentration and the worse the air quality. For the first principal component, the weight function of both $PM_{2.5}$ and PM_{10} showed a trend of slow increase and then slow decline. The weight function of O_3 displayed a slow and continuous upward trend, and the weight functions of the remaining pollutants showed a trend of year-by-year reduction. It is evident that both $PM_{2.5}$ and PM_{10} pollution were generally more serious in the spring and winter. Conversely, O_3 pollution showed a clear upward trend in summer, reaching its highest concentration in autumn. The weight functions of CO and SO_2 fluctuated little, and they have remained at a relatively stable concentration since 2017, with less obvious pollution. For the third principal component, the weight function of NO_2 showed a trend opposite to the weight function of its first principal component, with higher concentrations in the summer. In detail, from 2015 to 2019, most of the principal component scores of NO_2 were negative, and the principal component scores of NO_2 after 2020 were greater than zero. The negative value here indicated that the concentration of the NO_2 pollutant was less than the mean, and the greater the negative fluctuation, the lower the pollutant concentration, and the better the air quality.

Table 3. Proportion of the principal component variance contribution of the Air Quality Index and the six pollutants (FPC 1: first functional principal component; FPC 2: second functional principal component; FPC 3: third functional principal component).

Pollutant	FPC 1	FPC 2	FPC 3	Accumulation
Air Quality Index	34.7%	13.7%	40.6%	89%
CO	52.1%	27.3%	16.9%	96.3%
NO_2	30.6%	28.9%	31.9%	91.4%
O_3	45.5%	12.4%	33.6%	91.5%
$PM_{2.5}$	54.9%	22.9%	10.3%	88.1%
PM_{10}	61.3%	19.3%	10%	90.6%
SO_2	42.4%	26.6%	18.2%	87.2%

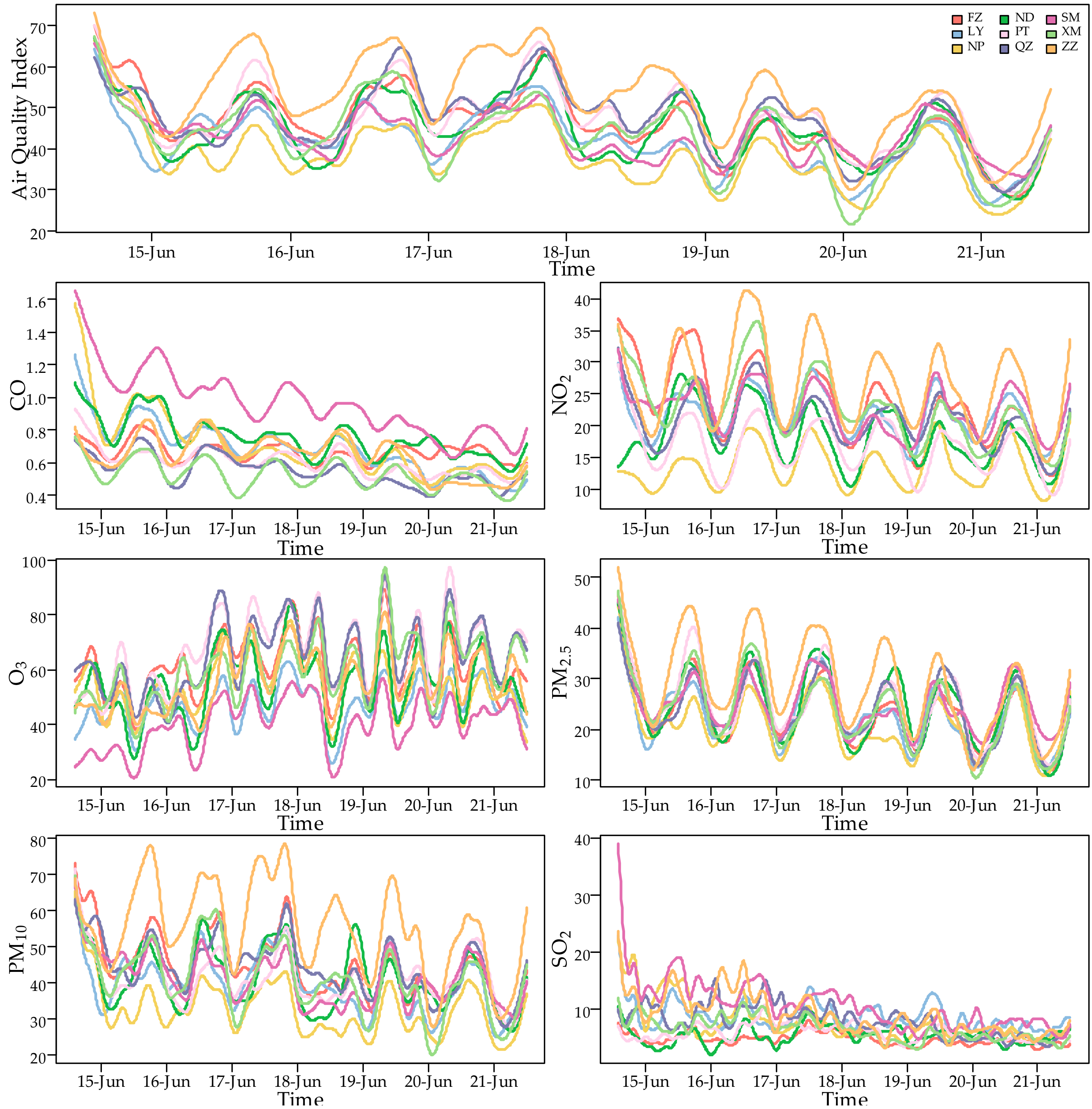

Figure 5. Curves of the smoothed Air Quality Index and six air pollutants in Fujian Province (unit: mg/m^3, except for CO: ug/m^3).

Figure 7 displays the functional clustering results of the AQI and the six pollutants. In this paper, the air pollution levels of the contaminated area were divided into four categories according to the cluster spacing of the principal component scores, including low, slight, medium, and serious air pollution. Air pollution had notable spatial differences, with the air quality in coastal areas being worse than that in inland areas. Overall, the pollution levels of $PM_{2.5}$, PM_{10}, and O_3 pollutants were lower in inland areas than in coastal areas. The pollution levels of CO and SO_2 pollutants had opposite results. Furthermore, the pollution levels of NO_2 pollutants showed a stepped distribution from the north toward the south. According to the division of air pollution levels in various cities, it can be concluded that the pollution levels of the NO_2, $PM_{2.5}$, and PM_{10} pollutants in Nanping (northern Fujian Province) were all in low air pollution, only CO and SO_2 pollutants caused the medium air pollution level, and its AQI was at a low air pollution level. In contrast, Zhangzhou (southern Fujian Province) had three pollutants (NO_2, $PM_{2.5}$, and PM_{10}) that caused severe air pollution, and its AQI was at a serious air pollution level. Therefore, the level of air pollution was most serious in Zhangzhou and low in Nanping.

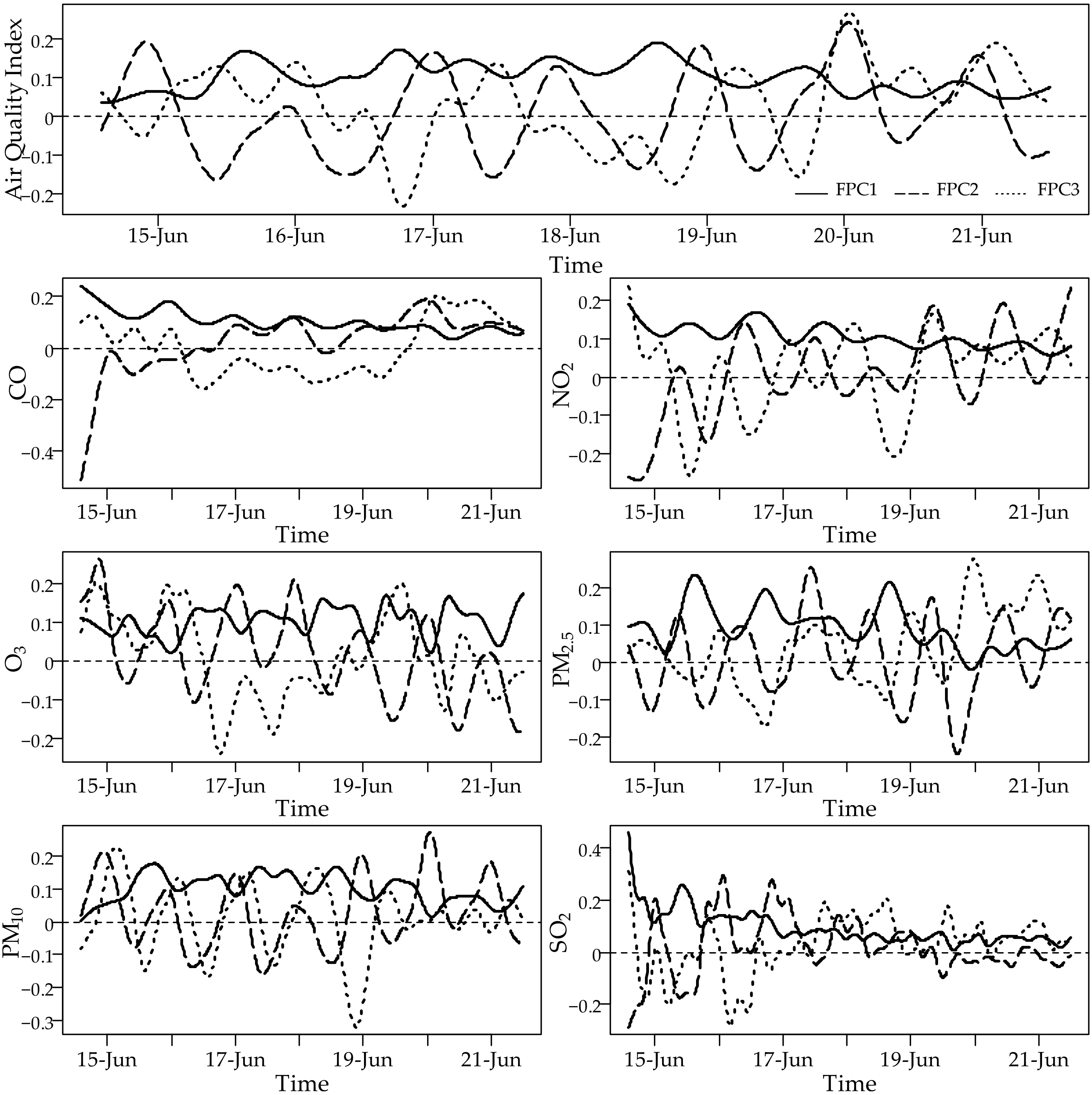

Figure 6. Dynamic change of principal component weight function of the Air Quality Index and the six pollutants in Fujian Province (unit: mg/m^3, except for CO: ug/m^3).

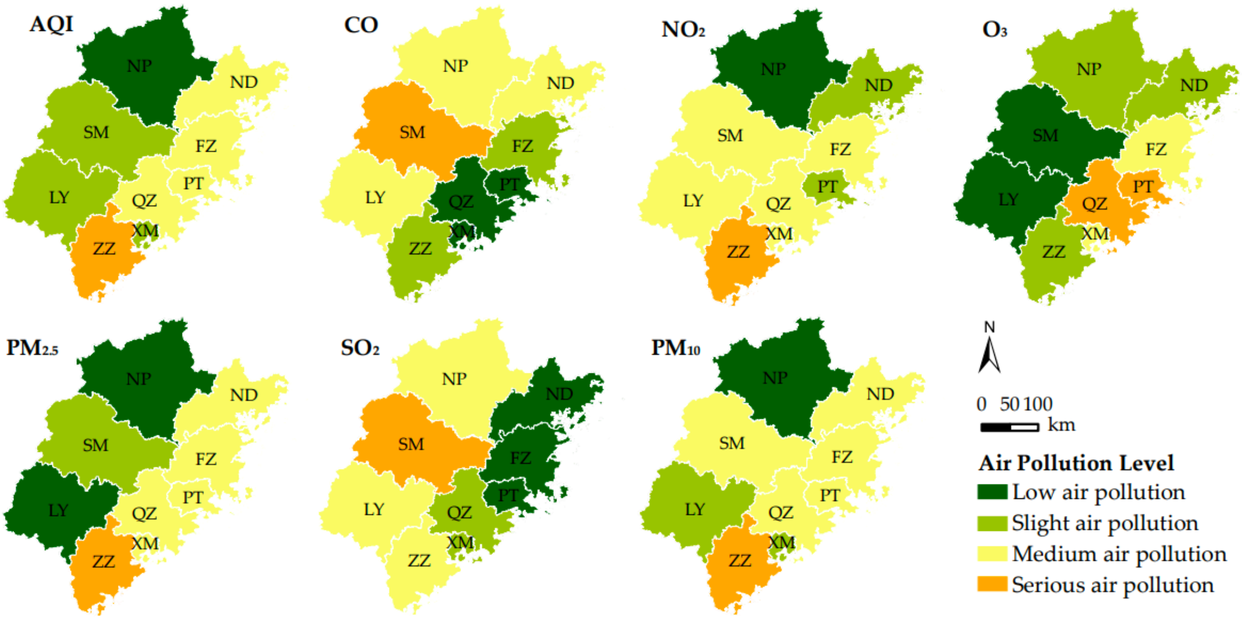

Figure 7. Clustering results of air pollution in Fujian Province.

3.3. Response Relationship between Anthropogenic Activities and Air Quality

Based on the functional linear model, the driving factors of air pollution in Fujian Province were analyzed. Man-made air pollution sources are pollution sources formed by human production and living activities, of which vehicle exhaust and industrial emissions are the main direct driving factors of air pollution [42], and vehicle ownership and industrial NOx emissions were selected as the scalar response variables. According to earlier research [43,44], the impact of PM_{10} on air quality is associated with industry, and the impact of NO_2 on air quality is associated with vehicle exhaust emissions. Therefore, PM_{10} and NO_2 in heavily polluted areas were selected as the covariates. The functional linear model of NO_2 and vehicle ownership was established, and the model fit was good (goodness-of-fit: 0.93). Similarly, the functional linear model of PM_{10} and industrial NOx emissions was established, and the model fit was also good (goodness-of-fit: 0.99). The coefficient estimation curves of the driving factors causing levels of air pollution are shown in Figures 8 and 9.

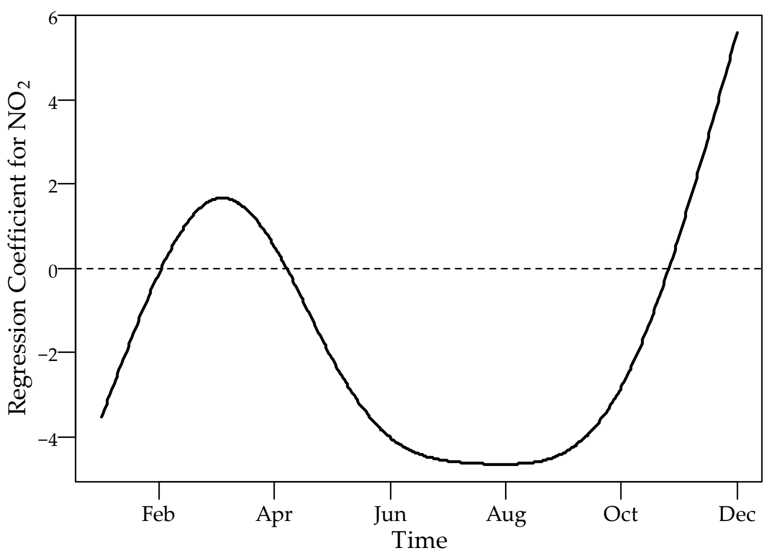

Figure 8. Dynamic changes of the influence degree of vehicle ownership on NO_2 based on the functional linear model.

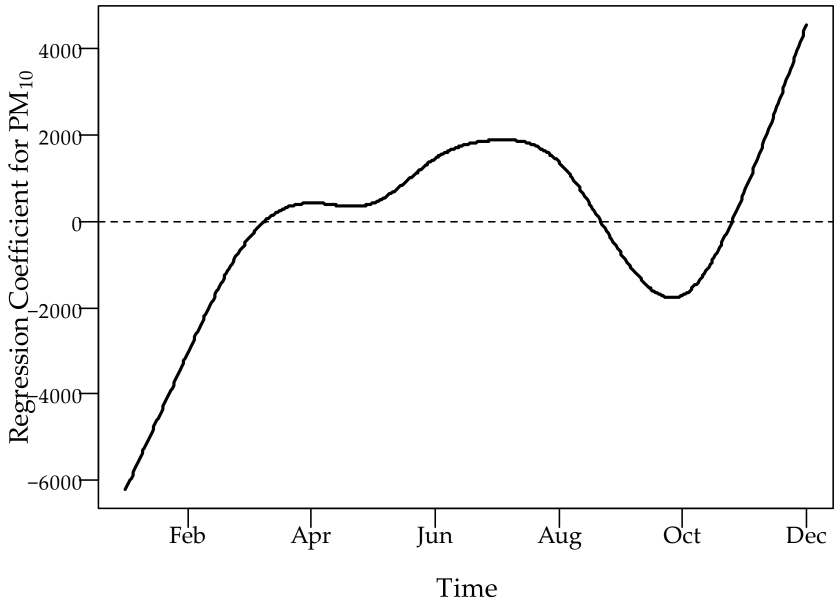

Figure 9. Dynamic changes of the influence degree of industrial NO_X emissions on PM_{10} based on the functional linear model.

Concretely, vehicle emissions had the most serious impact on NO_2 pollution in January, July, August, and December, and the impact of vehicle exhaust emissions on NO_2 pollution had certain seasonal characteristics. The impact magnitude increased rapidly in February–March and October–November, showing that the impact degree was more obvious in autumn and winter. Specifically, the impact of vehicle ownership on NO_2 pollution increased rapidly from February to March, which is possibly due to the factor of commuting patterns. The impact of vehicle ownership on NO_2 pollution increased rapidly from October to November, which is possibly related to the fact that vehicles tend to consume more fuel due to an increased use of heating systems and longer warm-up times in colder months. The impact of vehicle emissions on NO_2 pollutants decreased significantly from April to June, indicating that vehicle emissions had less impact on NO_2 pollutant concentrations in spring and summer. In the same way, industrial NO_X emissions had the most serious impact on PM_{10} pollution in January and December, and the impact of industrial NO_X emissions on PM_{10} pollution had certain seasonal characteristics. The impact magnitude increased rapidly in February–April and October–November, showing that the degree of influence was more obvious in winter. Specifically, the impact of industrial emissions on PM_{10} pollution increased rapidly from February to April, which is possibly related to factories restarting production following the Spring Festival.

4. Discussion

4.1. Dynamic Changes in Air Quality Associated with Natural Factors and Anthropogenic Activities

It is essential to investigate the effects of natural factors and anthropogenic activities on the dynamic variations in air quality. During the study period, the spatial and temporal dynamics of AQI and six pollutants indicated that air quality was related to seasonal changes. Concretely, air quality in the summer was substantially better than that in the winter, mainly because the climatic conditions in winter were harsh, particulate matter was not easily diffused, and energy consumption increased. In general, the climate in winter is relatively dry, the sunshine time short, and the atmospheric convection not active, which is not conducive to the diffusion of pollutants in the air [45]. In addition, winter jet stream can cause dust storms by picking up dry soil particles and carrying them over long distances and creating a temperature inversion [46]. A temperature inversion can lead to poor air quality as pollutants become trapped close to the surface, unable to disperse.

In contrast, in the summer, with the more active atmospheric convective activity, and the arrival of the summer monsoon bringing widespread rainfall, the pollution is more obviously removed [47]. Heavy rain can carry particles from smog to the ground, or it can dissolve some harmful gases in the rain to reduce air pollution. However, O_3 is derived from the secondary conversion of NO_X and volatile organic compounds at high temperatures and dry air [48]. Therefore, the hot weather in the summer might promote such a transformation. In addition, volatile organic compounds can undergo a series of reactions with NO_X under the condition of ultraviolet light irradiation, improve the oxidation of the atmosphere, and cause an increase in the concentration of O_3.

Industrial and vehicle emissions were found to be correlated with air pollution. Over the past decade, the economy of Fujian Province has been expanding, and its air pollution has also shown a fluctuating upward trend. With the development of the economy, the number of vehicles has increased, industrial zones have also sprung up in coastal areas, and a series of industrial clusters have formed. Industrial pollution sources and automobile exhaust pollution sources are the most serious sources of air pollution generated by human activities. The recent economic development of Fujian Province has been concentrated in the coastal areas. There are many large petrochemical parks, steel plants, nearly 20 large thermal power plants, over 30 waste incineration power plants, and other facilities distributed along the coastal areas. These emit a significant amount of pollutants. In addition, emissions from coastal ports, ships, and the logistics industry also contribute to pollution [49]. Against the backdrop of rapid urbanization and industrialization, the continuous expansion of cities has led to shorter distances between them, and urban expansion is positively correlated with air pollution [50]. As a result, pollutants are easily transported between coastal areas in Fujian Province, and their concentration declines slowly, leading to more severe pollution levels among cities in these coastal regions. However, from the end of 2019 to 2021, due to the impact of COVID-19 quarantine measures, factories shut down, transportation significantly reduced, and shipping halted, resulting in a noticeable improvement in overall air quality, which was consistent with the studies of Albayati et al. [51].

4.2. Advantages and Limitations of the Study

Air quality data have the characteristics of a function, and the use of discrete data for an analysis cannot well describe the dynamic characteristics and individual differences of air quality data. The functional data approach used in this study treated the observed data functions as an entity, not just as the order of individual observations. Through the unique method of functional data, more potential change laws can be elicited from a continuous perspective. The advantages of functional data are that it requires fewer assumptions, reduces the requirement for data acquisition frequency, and it can elicit more information. Moreover, in this study, the rough penalty method was also used to prevent overfitting and to ensure the smoothness of the function.

To analyze the differences in air pollution in different regions, this study used functional clustering based on the principal component coefficient to classify areas of air pollution. The functional cluster analysis approach broadens the application scope of a traditional cluster analysis by resolving the problem that traditional clustering techniques are only used to address static issues. The difficulty of functional data is that the data belong to an infinite dimensional space, and a principal component analysis is an effective dimensionality reduction method, which can use a few synthetic variables to summarize the original variables. Combining a principal component analysis with a cluster analysis overcomes the problem of collinearity among indicators. Additionally, it can also address issues with greater data dimensionality and larger data volumes that traditional methods cannot handle.

The functional linear model could be used to clarify the driving factors of air pollution. The functional linear model converted the regression coefficients in the traditional model into coefficient functions, thereby enriching the results of the model. The characteristics of the coefficient function could be used not only to explain the impact of the driving factors on air pollution, but also to obtain the trend of the degree of influence over time. However, the formation of air pollution is too complex to be addressed completely, and our limited data availability also prevented us from taking more environmental factors and anthropogenic activities into consideration. Therefore, such topics remain worthwhile items for further research.

5. Recommendations

Although the prevention and control of air pollution have been strengthened in recent years, regional and structural pollution problems are still prominent, and the emission of major pollutants is still at a high level, particularly PM_{10} and $PM_{2.5}$ (Figure 3). Some air pollutants have shown a fluctuating upward trend in recent years, particularly NO_2 and O_3 (Figure 6). Moreover, the quality of the atmospheric environment is greatly affected by changes in the natural environment and anthropogenic activities, and air pollution dynamically varies from region to region. Fujian Province wants to continuously improve its air quality, and especially to strengthen the coordinated control of $PM_{2.5}$, PM_{10}, NO_2, and $O_{3.}$ It is necessary to strengthen the coordinated control of multiple pollutants and regional collaborative management. Therefore, the recommendations for guiding air pollution control and sustainable ecological environment development are as follows. First, regions should propose feasible and precise measures to control air pollution. Due to the complexity of air pollution, different meteorological conditions, pollution sources, and pollutant concentrations in different regions, scientific and precise treatment measures can be proposed according to the actual situation to effectively promote the improvement of air pollution. Second, the industrial layout should be adjusted as soon as possible and treatments for the purification of pollutant discharges should be strengthened. A reasonable industrial structure should be based on the principle of sustainable development, and it is necessary to prevent the excessive exploitation of resources and excessive damage to the environment. Finally, the population should be encouraged to use new energy vehicles or public transport to reduce exhaust emissions. Automobile exhaust emission sources have become the main source of NO_X emissions in many cities. Therefore, it is necessary not only to increase the ratio of pure electric vehicles for manned vehicles, but also to accelerate the ratio of pure electric vehicles from exhaust emission sources such as trucks, and to prohibit the use of motor vehicles that exceed the standard.

6. Conclusions

Understanding the dynamic changes of air pollutants and identifying the drivers of air pollution are essential for improving air quality and promoting the sustainable development of the ecological environment. In this study, air quality data were functionalized using the B-spline basis function. The smoothing parameters were determined using the roughness penalty method, which can ensure the smoothness of the function curve while also ensuring accuracy. A functional cluster analysis based on the principal component coefficient and functional linear model were used to analyze both the spatiotemporal dynamic characteristics and the driving factors of air pollution in Fujian Province. The results showed that AQI and the air pollutants presented notable spatiotemporal heterogeneity within the study area. Moreover, the functional linear model fit was good, and the impact of anthropogenic factors on air pollutants had certain seasonal characteristics. The functional linear model could elucidate the degree of influence of temporal change through the coefficient function. These conclusions could be taken as a guide for effectively controlling air pollution and promoting the sustainable development of the ecological environment.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app132011206/s1, Figure S1: Change in the Air Quality Index, CO, NO_2, O_3, $PM_{2.5}$, PM_{10}, and SO_2 concentrations in Fujian Province from 1 January 2015 to 31 December 2021 (unit: mg/m^3, except for CO: ug/m^3).

Author Contributions: Conceptualization, Y.X. (Yihan Xu) and T.Y.; methodology, Y.X. (Yihan Xu) and Y.W.; software, Y.X. (Yihan Xu); validation, H.S., J.N. and Y.X. (Yanglan Xiao); formal analysis, Y.W.; resources, J.N.; data curation, Y.X. (Yihan Xu); writing—original draft preparation, Y.X. (Yihan Xu) and Y.W.; writing—review and editing, Y.X. (Yihan Xu); supervision, T.Y.; project administration, T.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Social Science Foundation of Fujian Province, China (Grant No. FJ2018B063).

Data Availability Statement: Not applicable.

Acknowledgments: We offer our thanks to the reviewers and editor for their insightful suggestions regarding this work. We would like to express our gratitude to Xuemin You for her help of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ouyang, H.; Tang, X.; Kumar, R.; Zhang, R.; Brasseur, G.; Churchill, B.; Alam, M.; Kan, H.; Liao, H.; Zhu, T.; et al. Toward Better and Healthier Air Quality: Implementation of WHO 2021 Global Air Quality Guidelines in Asia. *Bull. Am. Meteorol. Soc.* **2022**, *103*, E1696–E1703. [CrossRef]
2. IQAir. *IQAir World Air Quality Report 2022*; IQAir: Goldach, Switzerland, 2022; p. 11.
3. World Health Organization. Household Air Pollution. Available online: https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health (accessed on 28 November 2022).
4. Kilinc, B.K. The Effects of Industry Increase and Urbanization on Air Pollutants in Turkey: A Nonlinear Air Quality Model. *Appl. Ecol. Environ. Res.* **2019**, *17*, 9889–9903. [CrossRef]
5. Liu, Q.; Li, X.; Liu, T.; Zhao, X. Spatio-Temporal Correlation Analysis of Air Quality in China: Evidence from Provincial Capitals Data. *Sustainability* **2020**, *12*, 2486. [CrossRef]
6. Harrison, R.M.; Laxen, D.; Moorcroft, S.; Laxen, K. Processes affecting concentrations of fine particulate matter ($PM_{2.5}$) in the UK atmosphere. *Atmos. Environ.* **2012**, *46*, 115–124. [CrossRef]
7. Zhao, Y.; Zhang, X.; Chen, M.; Gao, S.; Li, R. Regional Variation of urban air quality in China and its dominant factors. *Acta Geogr. Sin.* **2021**, *76*, 2814–2829. [CrossRef]
8. Chu, B.; Ma, Q.; Liu, J.; Ma, J.; Zhang, P.; Chen, T.; Feng, Q.; Wang, C.; Yang, N.; Ma, H.; et al. Air Pollutant Correlations in China: Secondary Air Pollutant Responses to NOx and SO_2 Control. *Environ. Sci. Technol. Lett.* **2020**, *7*, 695–700. [CrossRef]
9. Li, T.; Lu, W.; Zhao, J.; Zeng, S. Analysis of Chengdu Air Quality Pollution Based on Logistic Sequential Multi-classification of Distance between Classes. *Sci. Technol. Eng.* **2022**, *22*, 409–415.
10. Jiao, D.; Sun, Z. Regression Analysis of Air Quality Index. *Period. Ocean Univ. China* **2018**, *48*, 228–234.
11. Tan, S.T.; Mohamed, N.; Ng, L.C.; Aik, J. Air quality in underground metro station commuter platforms in Singapore: A cross-sectional analysis of factors influencing commuter exposure levels. *Atmos. Environ.* **2022**, *273*, 118962. [CrossRef]
12. Jin, Z.; Gao, X.; Li, B.; Zhai, D.; Xu, J.; Li, F. Spatio-temporal distribution pattern and influencing factors of air quality in Sichuan-Chongqing region. *Acta Ecol. Sin.* **2022**, *42*, 4379–4388.
13. Lei, T.M.T.; Siu, S.W.I.; Monjardino, J.; Mendes, L.; Ferreira, F. Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao. *Atmosphere* **2022**, *13*, 1412. [CrossRef]
14. Xie, S.; Zhou, Z.; Li, G. Relationship between PM(2.5) Concentration and Meteorological Factors in Nanjing. *Sci. Technol. Eng.* **2020**, *20*, 460–466.
15. Navares, R.; Aznarte, J.L. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecol. Inform.* **2020**, *55*, 101019. [CrossRef]
16. Phruksahiran, N. Improvement of air quality index prediction using geographically weighted predictor methodology. *Urban Clim.* **2021**, *38*, 100890. [CrossRef]
17. Sun, X.; Xu, W.; Jiang, H.; Wang, Q. A deep multitask learning approach for air quality prediction. *Ann. Oper. Res.* **2021**, *303*, 51–79. [CrossRef]
18. Fu, J.; Zhou, F. Chinese provincial air quality measurement and its influencing factors. *Urban Probl.* **2020**, 20–27. [CrossRef]
19. Zhao, W.; Liu, H.; Pu, H. The Spatiotemporal Evolution and Socio-economic Driving Forces of Air Quality in China during the 13th Five-Year Plan Period. *J. Southwest Univ. Nat. Sci. Ed.* **2022**, *44*, 99–109.
20. Li, M.; Zhang, M.; Du, C.; Chen, Y. Study on the spatial spillover effects of cement production on air pollution in China. *Sci. Total Environ.* **2020**, *748*, 141421. [CrossRef] [PubMed]
21. Xia, X.; Chen, J.; Wang, J.; Cheng, X. PM (2.5) Concentration Influencing Factors in China Based on the Random Forest Model. *Environ. Sci.* **2020**, *41*, 2057–2065. [CrossRef]
22. Song, B.; Shi, H.; Wang, M.; Gu, R. The research on the effect of digital economy development on urban air quality. *Front. Environ. Sci.* **2022**, *10*, 993353. [CrossRef]
23. Rahman, M.M.; Alam, K. The roles of globalization, renewable energy and technological innovation in improving air quality: Evidence from the world's 60 most open countries. *Energy Rep.* **2022**, *8*, 9889–9898. [CrossRef]

24. Xiao, J.; Liu, X.; Yang, W. Status quo of air quality and its relationship with economic development in seven cities of China. *J. Environ. Occup. Med.* **2019**, *36*, 533–539.
25. Zhao, Y.; Wang, S.; Zhou, C. Understanding the relation between urbanization and the eco-environment in China's Yangtze River Delta using an improved EKC model and coupling analysis. *Sci. Total Environ.* **2016**, *571*, 862–875. [CrossRef] [PubMed]
26. Yan, M. The Statistical Analysis of Functional Data: Thoughts, Methods and the Applications. *Stat. Res.* **2007**, *24*, 87–94.
27. Huang, H.; Qi, W. Massive Semi-Structured Data: Collection, Storage and Analysis—Based on the Practice of Real-time Air Quality Data Processing. *Stat. Res.* **2014**, *31*, 10–16.
28. King, M.C.; Staicu, A.M.; Davis, J.M.; Reich, B.J.; Eder, B. A Functional Data Analysis of Spatiotemporal Trends and Variation in Fine Particulate Matter. *Atmos. Environ. 1994* **2018**, *184*, 233–243. [CrossRef]
29. Ramsay, J.O. When the data are functions. *Psychometrika* **1982**, *47*, 379–396. [CrossRef]
30. Aggarwal, C.C.; Reddy, C.K. *Data Clustering: Algorithms and Applications*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2013; p. 652.
31. Wang, D.; Zhu, J.; Liu, X.; He, L. Review and Prospect of Functional Data Clustering Analysis. *J. Appl. Stat. Manag.* **2018**, *37*, 51–63.
32. Liang, Y.; Liu, L.; Lu, Y. Characterization of air pollution in Beijing, Tianjin and Hebei based on functional data clustering. *World Surv. Res.* **2017**, 43–48. [CrossRef]
33. Wu, Q.; Zhou, L.; Sun, J.; Wang, N.; Yu, Q. Research on characteristic of air quality change in Zhejiang Province—Based on functional data analysis. *J. Shandong Univ. Nat. Sci.* **2021**, *56*, 53–64.
34. Fujian Provincial Department of Ecology and Environment. Fujian Province's 14th Five-Year Plan for Air Quality Improvement. Available online: https://sthjt.fujian.gov.cn/zwgk/ghjh1/zcqfzgh/202303/t20230315_6131424.htm (accessed on 27 February 2022).
35. Fujian Provincial Bureau of Statistics. Statistical Communiqué on National Economic and Social Development of Fujian Province in 2022. Available online: https://tjj.fujian.gov.cn/xxgk/tjgb/202303/t20230313_6130081.htm (accessed on 14 March 2023).
36. *HJ 633-2012*; Technical Regulation on Ambient Air Quality Index(on Trial). Ministry of Environmental Protection of the People's Republic of China: Beijing, China, 2012.
37. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*, 2nd ed.; Springer: New York, NY, USA, 2004.
38. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R, Uncorrected ed.*; Springer: New York, NY, USA, 2013; p. xiv, 426 pages: Illustrations (some color).
39. Yan, M. *Methods of Functional Data Analysis and Its Economic Applications*; Financial and Economic Publishing House: Beijing, China, 2014; p. 191.
40. Craven, P.; Wahba, G.; University of Wisconsin—Madison. Department of, S. *Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation*; Department of Statistics, University of Wisconsin: Madison, WI, USA, 1977.
41. Ramsay, J.; Hooker, G.; Graves, S.; SpringerLink. *Functional Data Analysis with R and MATLAB*, 1st ed.; Springer: New York, NY, USA, 2009.
42. Ali, M.; Athar, M. Impact of transport and industrial emissions on the ambient air quality of Lahore City, Pakistan. *Env. Monit. Assess.* **2010**, *171*, 353–363. [CrossRef] [PubMed]
43. Freire, S.M.; Relvas, H.; Lopes, M. Impact of traffic emissions on air quality in Cabo Verde. *Env. Monit Assess* **2020**, *192*, 726. [CrossRef] [PubMed]
44. Mutlu, A. Air quality impact of particulate matter (PM(10)) releases from an industrial source. *Env. Monit Assess* **2020**, *192*, 547. [CrossRef]
45. Handhayani, T. An integrated analysis of air pollution and meteorological conditions in Jakarta. *Sci. Rep.* **2023**, *13*, 5798. [CrossRef]
46. Guo, J.; Chen, Y. Patterns of seasonal change in air pollution and suggestions on pollution control in Beijing. *J. Ecol. Environ.* **2009**, *18*, 952–956.
47. Li, X.; Zhang, M.; Wang, S.; Zhao, A.; Ma, Q. Variation Characteristics and Influencing Factors of Air Pollution Index in China. *Environ. Sci.* **2012**, *33*, 1936–1943.
48. Yan, C.; Zhou, X.; Zhang, H.; Huang, X.; Sheng, Y.; Nie, M.; Ding, M. Spatiotemporal Distribution of Air Pollution and Its Impacted Factors in Nanchang City. *Resour. Environ. Yangtze River Basin* **2019**, *28*, 1446–1459.
49. Fameli, K.M.; Kotrikla, A.M.; Psanis, C.; Biskos, G.; Polydoropoulou, A. Estimation of the emissions by transport in two port cities of the northeastern Mediterranean, Greece. *Environ. Pollut.* **2020**, *257*, 113598. [CrossRef] [PubMed]
50. Lu, J.; Li, B.; Li, H.; Al-Barakani, A. Expansion of city scale, traffic modes, traffic congestion, and air pollution. *Cities* **2021**, *108*, 102974. [CrossRef]
51. Albayati, N.; Waisi, B.; Al-Furaiji, M.; Kadhom, M.; Alalwan, H. Effect of COVID-19 on air quality and pollution in different countries. *J. Transp. Health* **2021**, *21*, 101061. [CrossRef] [PubMed]