International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 1009-1019

https://doi.org/10.70517/ijhsa46377

Study on the effects of atmospheric pollutants on allergic skin diseases in Nanchong City based on the combination of principal component analysis and regression modeling

Kangheng Tang[1] and Xiaojie Ding[1,*]
[1] Department of Clinical Medicine, The Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, 637000, China
Corresponding authors: (e-mail: 15183569733@163.com).

Abstract The study was conducted to explore the effects of atmospheric pollutants on allergic skin diseases, combining principal component analysis and multiple linear regression modeling to statistically analyze the data of atmospheric pollutants (PM2.5, PM10, etc.) and outpatient cases of allergic skin diseases in hospitals in Nanchong City from 2020 to 2024. The main components of air pollutants were extracted by dimensionality reduction using principal component analysis to solve the problem of multicollinearity among variables. Regression models were also constructed to analyze the association between principal components and the incidence of allergic skin diseases. The results showed that the principal component analysis extracted three principal components in the atmospheric pollutants with a cumulative variance contribution of 91.97%.The annual average concentrations of the atmospheric pollutants PM2.5, PM10, NO_2, SO_2, CO, and O3 were 31.267 μg/m³, 51.109 μg/m³, and 24.252 μg/m³, respectively, from 2020 to 2024, 7.765 μg/m³, 0.951 mg/m³, and 131.452 μg/m³. The regression model showed that a 1-unit increase in atmospheric CO concentration was associated with a significant increase in the risk of allergic skin visits by 6.22. The regression model provided a good fit with errors between the predicted and true values of the number of visits for urticaria, eczema, and contact dermatitis in 2 general hospitals ranging between 0 and 1. The study confirmed that atmospheric oxidizing pollutants are the main environmental risk factors for allergic skin diseases in Nanchong City.

Index Terms Principal component analysis, Regression model, Atmospheric pollution, Allergic skin disease

I. Introduction

With the rapid development of China's economy and the acceleration of urbanization, the ensuing problem of environmental pollution, especially air pollution, has become increasingly prominent, and it has become a major environmental threat affecting the health of the population [1], [2]. As the skin can resist microorganisms, ultraviolet rays, and chemical stimuli externally, and prevent the loss of water, electrolytes, and nutrients internally, it ensures that the human body maintains the relative stability of the internal environment in the complex and changing external environment [3]-[5]. Therefore, its as the body's largest organ, and the first barrier to the body's defense function, is the main pathway of exposure to atmospheric pollutants, the correlation between atmospheric pollutants and skin diseases research is worthy of attention [6]-[8]. Atmospheric pollutants, mainly sulfur oxides (SOx), nitrogen oxides (NOx), atmospheric particulate matter (PM), and ozone, can directly or indirectly affect the health of the body, mainly in the respiratory system, cardiovascular system, and reproductive system [9]-[11]. At the same time, there is growing evidence that air pollution seriously affects skin aging, accelerates skin aging, and disrupts the skin barrier function, which leads to an increased risk of skin diseases such as dermatitis, eczema, urticaria, acne, and psoriasis [12]-[15].

Allergic skin diseases (ASD), mainly mediated by IgE, are a group of inflammatory skin diseases with abnormalities of innate and adaptive immunity [16]. The main clinical manifestation of the disease is intense itching, which severely interferes with the patient's daily life and work [17]. ASD is prevalent in both the pediatric and adult populations, and its high healthcare costs impose a huge economic burden on individuals, families, and society [18], [19]. ASD was thought to have a high prevalence in developed countries only in the past, but in the last three decades the prevalence of ASD in developing countries has increased in tandem with the However, the prevalence of ASD in developing countries has risen along with the process of international industrial transfer in the last 30 years [20]. A single genetic factor cannot fully explain the rapid increase in global ASD prevalence in recent decades, and environmental factors may play a crucial role in the development of ASD [21]-[23]. Therefore, exploring the mechanism of atmospheric pollutants on ASD may provide more preventive ideas for patients at high risk of allergic skin diseases [24], [25].

In this paper, the diagnosed cases of allergic skin diseases in several general hospitals in Nanchong City from 2020 to 2024 and the data of atmospheric pollutants including PM2.5, PM10, NO_2, SO_2, CO, and O_3 during the period period were collected. The covariance matrix of the atmospheric pollutant data was calculated using principal component analysis, and eigenvalue decomposition was performed to screen out the principal components therein according to the size of the eigenvalues and solve the multicollinearity problem. On this basis, a regression model was constructed between the principal components of the air pollutants and the incidence of allergic skin diseases to derive the relationship between the two, and the coefficient of determination R^2 was combined to explain the degree of fit of the model.

II. Research information

Allergic skin diseases (ASD) [26] are inflammatory diseases of the skin caused by allergic reactions. Urticaria, eczema, and contact dermatitis are the most common allergic skin diseases, and there is a close association between their onset and ambient air quality. The aim of this study was to explore the effects of atmospheric pollutants on allergic skin diseases in Nanchong City, and to provide a reference basis for the prevention and management of allergic diseases.

II. A.Study area

Located in the northeast of Sichuan Basin and the middle reaches of Jialing River, Nanchong is the second most populous city in Sichuan Province and China's excellent tourist city. It is also a national regional center city, the northern center city of Chengdu-Chongqing Economic Zone and the regional center city of northeast Sichuan. The geographic location is 105°27′~106°58′ east longitude and 30°35′~31°51′ north latitude. The city has 3 districts (Shunqing District, Gaoping District, Jialing District) and 1 city and 5 counties (Langzhong City, Nanbu County, Yingshan County, Peng'an County, Yilong County, Xichong County), with a population of 7,032,100 people and a land area of 12,500 square kilometers. Nanchong City is a subtropical monsoon humid climate zone, with abundant rainfall, abundant sunshine, four distinct seasons, hot summer climate, precipitation concentration, and slightly cooler and wetter in winter and other characteristics.

II. B.Study population

The data on allergic dermatology outpatient visits in the study were obtained from the information systems of 2 general hospitals in Nanchong. These general hospitals serve the majority of residents in the main city area of Nanchong, with an annual outpatient volume of approximately 3 million visits. Records of urticaria, eczema, and contact dermatitis visits from January 1, 2020, to December 31, 2024, were retrieved from the hospital information systems. Patient data included basic information (age, sex, and place of residence) and diagnosis (name of diagnosis and date of diagnosis.) Dermatologists at the 2 hospitals diagnosed definitively and according to the International Classification Of Diseases (10th revision) as urticaria, eczema, or contact dermatitis.

We excluded any patients with missing information and patients whose residence was not in the main city of Nanchong, and only included all patients with complete information and residence in the main city of Nanchong.

The Medical Ethics Committee of our hospital approved our study. As this study did not involve tissue samples or commercial interests and was a retrospective observational study, informed consent was waived. We are committed to keeping patients' personal information strictly confidential, and all research was conducted in accordance with the Declaration of Helsinki.

II. C.Environmental data

Daily atmospheric pollutant data for the time period from January 1, 2020 to December 31, 2024 are collected through the National Environmental Monitoring General Station of China. Daily concentrations of atmospheric pollutants (PM2.5, PM10, SO_2, NO_2, CO, and O_3) are measured in real time at a number of fixed-site stationary centers covering the main urban area of Nanchong City. Because these stations are designed to be located away from sources of traffic or industrial air pollution, the measurements are representative of levels throughout the city.

III. Research methodology

III. A. Atmospheric pollutant principal component extraction model

III. A. 1) Principal component analysis methods

Principal Component Analysis (PCA) [27] is an unsupervised dimensionality reduction method which is widely used in multiple linear regression analysis. The purpose of principal component analysis is to find linear combinations of a small number of predictors that can be used to reflect information from the original data without losing too much information. This statistical method solves the problem of multicollinearity by converting a broad set of correlated variables into a smaller number of uncorrelated variables, which are called principal components.

Let X be a $n \times p$ matrix, where n is a number of observations and p is a set of predictor variables $X_1, X_2, \cdots, X_p$, where X denotes random observations of $X_1, X_2, \cdots, X_p$. The goal of principal component analysis is to select a subset of predictor variables that reflects most of the information in the original data set, and an important step involves solving the covariance matrix of the data to obtain eigenvalues and eigenvectors. Let σ_{ij} denote the covariance of the variables X_i and X_j in the data matrix, and the matrix (σ_{ij}) can be written as Σ, which is denoted as the variance-covariance array. The values of the diagonal elements of the matrix Σ are the variances of X_i and the covariance matrix Σ is a symmetric square matrix.

The principal components (PCs) of principal component analysis (PCA) are linear combinations of the original variables $x_{i1}, x_{i2}, \cdots, x_{ip}$, denoted as $\sum a_i x_i$, where a_i is the scalar. If $\sum |a_i| = 1$, a principal component is a normalized linear combination of the original predictor variables in the data matrix. The principal component analysis method solves the problem of variable correlation by choosing a series of orthogonal principal components, that is, all principal components are uncorrelated with each other; the first principal component, PC1, is uncorrelated with PC2, PC2 is uncorrelated with PC3, and so on. There are two commonly used methods for solving principal components:

(1) Eigen-decomposition based on covariance matrix to solve principal components

Assuming that the eigenvalues of the covariance matrix Σ of X are $\lambda_1, \lambda_2, \cdots, \lambda_p$, where $\lambda_1 > \lambda_2 > \cdots > \lambda_p$, and the orthogonal eigenvectors corresponding to each eigenvalue are $b_1, b_2, \cdots, b_p$, then the i th principal component $t_i = x_{i1}b_{i1} + x_{i2}b_{i2} + \cdots + x_{ip}b_{ip}$, with $i = 1, 2, \cdots, n$, and p is the number of predictor variables. The selection of the number of principal components k is based on the cumulative variance contribution rate $\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$, which ultimately results in a cumulative contribution rate of the first k principal components of $\geq 85\%$.

(2) Solving principal components by singular value decomposition based on covariance matrix

Singular value decomposition (SVD) is a commonly used matrix decomposition method. The specific steps of using singular value decomposition for matrix A are: firstly, calculate the eigenvalues and corresponding eigenvectors of AA^T, unitize these eigenvectors, and combine them into matrix U. Then calculate the eigenvalues and corresponding eigenvectors of $A^T A$, unitize these eigenvectors and combine them into matrix V. Finally, the square root of the eigenvalues of AA^T or $A^T A$ is solved for, and a matrix D is formed with the above values as diagonal elements. Then for matrix $A_{m \times n}$, the singular value decomposition is expressed as:

$$A = UDV^T \tag{1}$$

where U is the left singular matrix, containing the orthogonal vectors. D is the diagonal matrix, where the elements on the diagonal are called singular values and the other elements are 0. V^T is the right singular matrix, which also contains orthogonal vectors. The principal components are expressed as $Y = XU$.

III. A. 2) Principal component extraction process

Principal Component Analysis converts high dimensional data into low dimensional data through linear transformation while retaining the main features of the data. In this process, redundant data are eliminated and the overall data accuracy is improved. The specific steps are as follows:

(1) Normalize the atmospheric pollutant data [28] to ensure that the data are consistent in the scale of each feature, so as to avoid the influence of different scales and ensure the comparability of the data.

(2) Calculate the covariance matrix of the normalized pollutant data set, which can describe the degree of correlation between each feature.

(3) Decompose the eigenvalues of the covariance matrix and calculate the eigenvalues and eigenvectors of the components of atmospheric pollutants, which can describe the direction and magnitude of changes in the data.

(4) Select and rank the eigenvalues to select the principal components. According to the magnitude of the eigenvalues, the number of K-dimensional features to be retained is selected and the K-dimensional eigenvectors are sorted in order to select the most important features.

(5) Project the high dimensional raw data into a lower dimensional space and compute to get the final reduced dimensional data.

In the application of extracting principal components of atmospheric pollutants, the number of K-dimensional features selected to be retained is usually smaller than the dimensionality of the original data, so that an effective dimensionality reduction effect can be achieved while retaining the main information of the data. The process of principal component analysis implementation is shown in Figure 1.

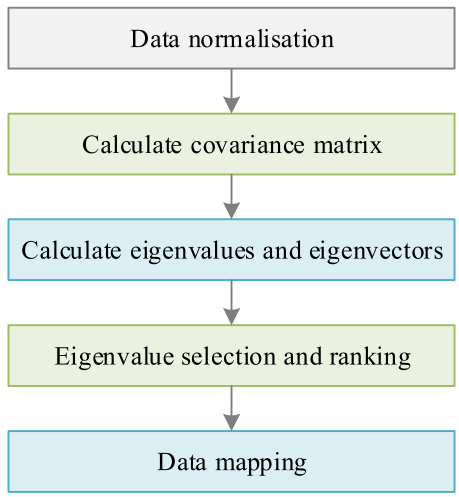

Figure 1: Principal Component Analysis Implementation Process

III. B. Multiple linear regression models

Regression analysis [29] is a statistical method that utilizes the relationship between two or more variables so that one dependent variable can be predicted using another variable or variables.

When one independent variable is used to explain the change in the dependent variable, what is used is one-way linear regression. When more than one independent variable is used to explain the dependent variable, the regression used is multiple regression. When there is a linear relationship between multiple independent variables and the dependent variable, it is multiple linear regression [30].

III. B. 1) Model setup

If there are p independent variables used to explain the dependent variable Y: $X_1, X_2, \cdots, X_p$, then the equation between them is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{2}$$

where ε is the error term and $\beta_0, \beta_1, \cdots, \beta_p$ are unknown parameters. The n sets of observations $\left(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i\right)$ of the independent variables $X_1, X_2, \cdots, X_p$ with $i = 1, 2, \cdots, n$, are satisfied:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon \tag{3}$$

There are certain settings for the variables in Eq.

First of all, the independent variables $x_1, x_2, \cdots, x_p$ are not random variables in the multivariate linear equation. Also the error term ε_i, $i = 1, 2, \cdots, n$ is a random variable, which is generated by a combination of factors. Since the error random variable ε is random, the dependent variable Y is also random. We first assume that the error term ε_i, $i = 1, 2, \cdots, n$ satisfies the following assumptions, viz:

$$\begin{aligned} &E\left(\varepsilon_i\right) = 0 \\ &\operatorname{var}\left(\varepsilon_i\right) = \sigma^2 \\ &\operatorname{cov}\left(\varepsilon_i \varepsilon_j\right) = 0, i \neq j \end{aligned} \tag{4}$$

Set the dependent variable about the independent variable in the form of equation (2).

The regression model is now written in matrix form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{5}$$

Noting the column vectors or matrices in (5) as y, X, β, and ε in that order, the matrix form of (5) can again be written as:

$$y = X\beta + \varepsilon \tag{6}$$

where y is a $n \times 1$-dimensional observed variable. X is a known matrix in $n \times (p+1)$ dimension. β is the $(p+1) \times 1$-dimensional vector of unknown parameters. ε is the vector of random errors. The Markov assumption is that:

$$E(\varepsilon) = 0, \operatorname{cov}(\varepsilon) = \sigma^2 I_n \tag{7}$$

In summary, the linear regression model is:

$$y = X\beta + \varepsilon, E(\varepsilon) = 0, \operatorname{cov}(\varepsilon) = \sigma^2 I_n \tag{8}$$

III. B. 2) Least squares estimation of model parameters

For regression analysis, a question that must be investigated is how the parameters should be estimated. For the problem of how to estimate the parameter vector β using the observations of the independent and dependent variables, the method usually used is Gaussian least squares. The idea is that according to the regression model above, it is known that $X\beta$ is the predicted value of the dependent variable obtained from the model's prediction of multiple independent variables in the training sample, but the value of the dependent variable is actually known to be y. So the distance between the actual y and the predicted $X\beta$ is this:

$$Q(\beta) = \|y - X\beta\|^2 = (y - X\beta)^T (y - X\beta) \tag{9}$$

Then find β such that $Q(\beta)$ is minimized. Now expand (9) as:

$$Q(\beta) = y^T y - 2y^T X\beta + \beta^T X^T X\beta \tag{10}$$

Take the partial derivative with respect to β so that its value is zero to obtain the system of equations:

$$X^T X\beta = X^T y \tag{11}$$

Solution for:

$$\hat{\beta} = \left(X^T X\right)^{-} X^T y \tag{12}$$

Here $\left(X^T X\right)^{-}$ is any generalized inverse of $X^T X$. According to the extremal theory of calculus, $\hat{\beta}$ is just a stationary point of the function $Q(\beta)$. It remains to be shown that $\hat{\beta}$ indeed minimizes $Q(\beta)$. For any β there is:

$$\begin{aligned} Q(\beta) &= \|y - X\beta\|^2 = \left\|y - X\hat{\beta} + X\left(\hat{\beta} - \beta\right)\right\|^2 \\ &= \left\|y - X\hat{\beta}\right\|^2 + \left(\hat{\beta} - \beta\right)^T X^T X\left(\hat{\beta} - \beta\right) \\ &+ 2\left(\hat{\beta} - \beta\right)^T X^T \left(y - X\hat{\beta}\right) \end{aligned} \tag{13}$$

Since $\hat{\beta}$ satisfies (11), $X^T\left(y-X\hat{\beta}\right)=0$, which follows:

$$\|y-X\beta\|^2=\left\|y-X\hat{\beta}\right\|^2+\left(\hat{\beta}-\beta\right)^T X^T X\left(\hat{\beta}-\beta\right) \tag{14}$$

Again, since $X^T X$ is a positive definite array, and so:

$$Q(\beta)=\|y-X\beta\|^2 \geq \left\|y-X\hat{\beta}\right\|^2=Q\left(\hat{\beta}\right) \tag{15}$$

This shows that $\hat{\beta}$ does minimize $Q(\beta)$:

$$\hat{\beta}=\left(X^T X\right)^{-1} X^T y \tag{16}$$

At this point, $\hat{\beta}$ is called the least squares estimate of β.

III. B. 3) Decision factors

An extremely important metric in regression analysis is the coefficient of determination R^2, which measures how well the independent variables $x_1,x_2,\cdots,x_p$ fit the dependent variable y. The R^2 is the proportion of the sum of squared deviations explained by the model to the total sum of squared deviations, i.e:

$$R^2=\frac{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}^2\right)} \tag{17}$$

Obviously $0 \leq R^2 \leq 1$, the larger its value, the better the model fit.

III. B. 4) Residual analysis

Residuals are commonly used to detect and assess the degree of discrepancy between a hypothesized model (including the assumptions made) and the observed data. Residuals are utilized to evaluate the model as well as to determine the presence or absence of outliers in the data.

For regression equations:

$$y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_p x_{ip}+\varepsilon \tag{18}$$

The residuals are defined as $e_i=y_i-\hat{y}_i$. There are as many residual values as there are data. The residual is the difference between the actual observation y and the regression value, and the residual e_i can be thought of as an estimate of the error term ε_i. The residual $e_i=y_i-\hat{y}_i=y_i-\hat{\beta}_0-\hat{\beta}_1 x_i$, and the error term $\varepsilon_i=y_i-\beta_0-\beta_1 x_i$.

The expectation of the residuals e_i is equal to zero and the variance:

$$\operatorname{var}(e_i)=\sigma^2(1-h_{ii}) \tag{19}$$

where $h_{ii}=\frac{1}{n}+\frac{(x_i-\bar{x})^2}{L_{xx}}$, called the leverage value. The residuals also satisfy the constraints $\sum_{i=1}^{n} e_i=0$ and $\sum_{i=1}^{n} x_i e_i=0$. This indicates that the residuals $e_1,e_2,\cdots,e_n$ are correlated, not independent.

According to the above properties, residual analysis contains the following main elements:

(1) Whether the residuals obey a normal distribution with zero mean.

(2) Whether the residuals obey a normal distribution with equal variances.

(3) Whether the sequence of residuals is independent.

(4) The use of residuals to detect outliers present in the sample.

IV. Effects of atmospheric pollutants on allergic skin diseases

IV. A. Atmospheric pollutant principal component extraction results

Principal component analysis is a method of eliminating covariance between variables, a statistical method of transforming initial variables with certain correlation into a set of uncorrelated indicators through the process of dimensionality reduction. That is, using fewer indicators to replace and reflect the original more information, the processing done is a mathematical linear transformation, and these integrated indicators are the main components of the original multiple indicators.

Using principal component analysis to extract the principal components of the original variables PM2.5 (X_1), PM10 (X_2), NO_2 (X_3), SO_2 (X_4), CO (X_5), O_3 (X_6) and other six atmospheric pollutant indicators. Table 1 shows the data information of the principal components of the atmospheric pollutants, including the order of the eigenroots from the largest to the smallest, the contribution rate of each principal component and the cumulative contribution rate. Figure 2 shows the gravel plot of principal components and eigenvalues. From the table, it can be seen that the eigenroot of the 1st principal component is 3.362 and the total variation is 56.03%. The eigenroot of the 2nd principal component is 1.475 and it explains 24.58% of the total variation. The eigenroot of the 3rd principal component is 0.681 and the first 3 eigenroots have a cumulative contribution of 91.97%. Based on the cumulative contribution of more than 85%, i.e., these principal components collectively retain more than 85% of the original amount of information, the first 3 principal components are selected here. Meanwhile, combined with the gravel plot, the trend of the inflection point of the characteristic curve can also indicate that the first 3 principal components should be selected to replace the original variables for principal component analysis.

Table 1: Data on the main components of atmospheric pollutants

variable	1	2	3	4	5	6
Characteristic root	3.362	1.475	0.681	0.303	0.119	0.06
Contribution rate	56.03%	24.58%	11.35%	5.05%	1.98%	1.00%
Cumulative contribution rate	56.03%	80.62%	91.97%	97.02%	99.00%	100%

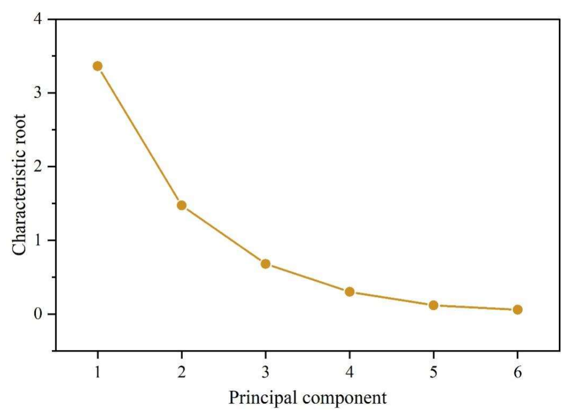

Figure 2: The rubble of the main composition and eigenvalues

The matrix of factor score coefficients in Table 2 is the result of principal component analysis, from which the original variables of each geographic factor of the principal components can be expressed as linear equations of three principal components with the following equations:

$$\begin{aligned}
Z_1 &= 0.148stdX_1 + 0.485stdX_2 + 0.147stdX_3 + 0.049stdX_4 \\
&\quad + 0.188stdX_5 + 0.259stdX_6 \\
Z_2 &= 0.022stdX_1 + 0.372stdX_2 + 0.305stdX_3 + 0.278stdX_4 \\
&\quad + 0.044stdX_5 + 0.141stdX_6 \\
Z_3 &= 0.435stdX_1 + 0.152stdX_2 + 0.261stdX_3 + 0.417stdX_4 \\
&\quad + 0.511stdX_5 + 0.224stdX_6
\end{aligned} \tag{20}$$

where $stdX_i\left(i=1,2,3,4,5,6,7,8,9\right)$ denotes the standard indicator variable.

Table 2: Factor score coefficient matrix

	Principal component		
	1	2	3
X_1	0.148	0.022	0.022
X_2	0.485	0.372	0.372
X_3	0.147	0.305	0.305
X_4	0.049	0.278	0.278
X_5	0.188	0.044	0.044
X_6	0.259	0.141	0.141

Table 3 shows the descriptive statistics of each atmospheric pollutant indicator from 2020 to 2024. The annual average concentrations of atmospheric oxidizing pollutants PM2.5, PM10, NO_2, SO_2, CO, and O_3 were 31.267 μg/m³, 51.109 μg/m³, 24.252 μg/m³, 7.765 μg/m³, 0.951 mg/m³, and 131.452 μg/m³, respectively. The standardized indicator variables can be derived from the means and standard deviations in the table:

$$\begin{aligned} stdX_1 &= (X_1 - 31.267)/10.752 \\ stdX_2 &= (X_2 - 57.109)/9.842 \\ stdX_3 &= (X_3 - 24.252)/7.236 \\ stdX_4 &= (X_4 - 7.765)/1.544 \\ stdX_5 &= (X_5 - 0.951)/0.348 \\ stdX_6 &= (X_6 - 131.452)/26.916 \end{aligned} \tag{21}$$

Table 3: Descriptive statistics of atmospheric indicators

Indicators	Mean value	Standard deviation
X_1	31.267	10.752
X_2	57.109	9.842
X_3	24.252	7.236
X_4	7.765	1.544
X_5	0.951	0.348
X_6	131.452	26.916

The regression equations of principal components Z_1, Z_2, and Z_3 with the dependent variable contact dermatitis and other allergic skin diseases incidence reference value Y were obtained using principal component analysis as follows:

$$Y = 23.19 + 8.91Z_1 + 4.17Z_2 + 1.01Z_3 \pm 2.16 \tag{22}$$

Substituting Eq. (20) into the above equation yields a linear regression equation for the dependent variable Y and the standard independent variable $stdX1 \sim stdX6$:

$$\begin{aligned} Y = 23.19 &+ 1.53stdX_1 + 12.58stdX_2 + 4.25tdX_3 \\ &+ 2.10stdX_4 + 2.17stdX_5 + 4.25stdX_6 \pm 2.16 \end{aligned} \tag{23}$$

Reducing the criterion variables to independent variables and substituting Eq. (21) into Eq. (23) yields a linear regression equation for allergic skin diseases and atmospheric oxidized pollutants as:

$$\begin{aligned} Y = &-105.73 + 0.14X_1 + 1.28X_2 \\ &+ 0.59X_3 + 1.36X_4 + 6.22X_5 \\ &+ 0.16X_6 \pm 2.16\,(F = 16.781, P = 0.000) \end{aligned} \tag{24}$$

where Y is the reference value of the risk of allergic skin disease visit, X_1 is the average annual PM2.5 concentration (μg/m³), X_2 is the average annual PM10 concentration (μg/m³), X_3 is the average annual NO_2

concentration (μg/m³), X_4 is the average annual SO_2 concentration (μg/m³), X_5 is the average annual CO concentration (mg/ m³), X_6 is the O_3 concentration (μg/m³), $F = 16.781$, $P = 0.00 < 0.01$ indicating that the regression equation is highly significant.

IV. B. Allergic Skin Disease Risk Regression Prediction

In this section, six indicators of air pollutants, namely PM2.5 (X_1), PM10 (X_2), NO_2 (X_3), SO_2 (X_4), CO (X_5), and O_3 (X_6), were used as independent variables, and the number of ASD visits was used as the dependent variable. And ASD was subdivided into urticaria, eczema and contact dermatitis, and a multiple linear regression model was constructed to analyze the strength of the association between air pollutants and the risk of ASD visits.

Table 4 shows the regression prediction results of the effect of atmospheric pollutants on the risk of ASD visits in the total population. In the table, the F-value for model 1 was 16.781, indicating that each unit increase in the concentration of the six atmospheric pollutants had a statistically significant effect on the relative risk of ASD visits. Among them, the atmospheric CO concentration has the strongest positive association with the risk of ASD outpatient visits, with a correlation coefficient of 6.22, and passes the significance test at the 0.001 level. That is, for every 1-unit increase in atmospheric CO concentration, the relative risk of an ASD visit increased significantly by 6.22. The discriminant coefficient R^2 of the model reached 0.317, explaining 31.7% of the phenomenon.

According to disease typing, this study found that the contamination models for urticaria and eczema were overall similar to those for contact dermatitis and ASD. From models 2 to 4, the risk of urticaria, eczema with contact dermatitis visits by each atmospheric pollutant passed the test of significance at the 0.05 level or even lower. Atmospheric CO concentration had the greatest effect on the risk of urticaria, eczema and contact dermatitis, with regression coefficients ranging from 6.03 to 6.38. The explanatory strength of R^2 reached 0.275 to 0.334. In conclusion, the increase in the concentration of air pollutants significantly enhanced the risk of allergic skin diseases in Nanchong City.

Table 4: The regression prediction of the risk effect of ASD visits

Independent variable	Model 1	Model 2	Model 3	Model 4
	ASD	Urticaria	Eczema	Contact dermatitis
X_1	0.14***	0.19**	0.11*	0.18**
X_2	1.28***	1.42***	1.23***	1.14***
X_3	0.59***	0.55***	0.68*	0.34**
X_4	1.36***	1.26**	1.37**	1.13***
X_5	6.22***	6.03***	6.38***	6.27***
X_6	0.16***	0.12**	0.06**	0.24**
R^2	0.317	0.275	0.334	0.296
F	16.781	11.497	13.275	17.425

Note: *, **, *** represent passing the test of significance at the 0.5, 0.01, and 0.001 levels, respectively.

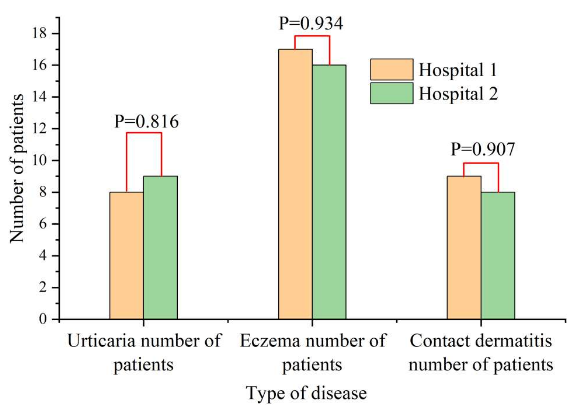

Figure 3: The comparison of the predicted value and the real value of the patient

Figure 3 shows the results of the comparison between the predicted and true values of the number of visits for urticaria, eczema and contact dermatitis in 2 general hospitals in Nanchong City, China, derived from the regression model. From the figure, it can be seen that the predicted number of urticaria, eczema and contact dermatitis visits in each hospital has a relatively high degree of fit to the true value, with a maximum error of 1 person. The paired-sample two-sided t-test of the predicted and true values of the number of visits for urticaria, eczema and contact dermatitis in each hospital showed that the P-values between the predicted and true values ranged from 0.816 to 0.934, which were all greater than 0.05, indicating that there was no significant difference between the two.

V. Conclusion

The article mines the data of atmospheric pollutants in Nanchong City in recent years as well as the diagnosed cases of allergic dermatosis in several hospitals, focusing on the exploration of the influence of atmospheric pollutants on allergic dermatosis in Nanchong City. The data of air pollutants are normalized, and the principal component analysis is used to calculate the characteristic roots of the multidimensional variables and extract the main components of them. The correlation between the principal components of the air pollutants and the incidence of allergic skin diseases was further analyzed by combining the regression analysis model, and the degree of influence of the principal components on the incidence of allergic skin diseases was also predicted.

The study utilized principal component analysis to obtain three principal components in the air pollution indicators, with variance contribution rates of 56.03%, 24.58%, and 11.35%, respectively. According to the descriptive statistics of air pollutants in Nanchong City from 2020 to 2024, the linear regression equation of allergic skin diseases and air oxidized pollutants was obtained as follows: $Y = -105.73 + 0.14X_1 + 1.28X_2 + 0.59X_3 + 1.36X_4 + 6.22X_5 + 0.16X_6 \pm 2.16$. The regression results showed that elevated concentrations of air pollutants in the atmosphere exhibited a significant positive correlation with the risk of allergic skin visits. The predicted values for the number of hospital visits for urticaria, eczema and contact dermatitis were not significantly different from the true values, with p-values greater than 0.05 between them.

References

[1] Song, C., Wu, L., Xie, Y., He, J., Chen, X., Wang, T., ... & Mao, H. (2017). Air pollution in China: Status and spatiotemporal variations. Environmental pollution, 227, 334-347.
[2] Liu, W., Xu, Z., & Yang, T. (2018). Health effects of air pollution in China. International journal of environmental research and public health, 15(7), 1471.
[3] Dąbrowska, A. K., Spano, F., Derler, S., Adlhart, C., Spencer, N. D., & Rossi, R. M. (2018). The relationship between skin function, barrier properties, and body-dependent factors. Skin Research and Technology, 24(2), 165-174.
[4] Hofmann, E., Schwarz, A., Fink, J., Kamolz, L. P., & Kotzbeck, P. (2023). Modelling the complexity of human skin in vitro. Biomedicines, 11(3), 794.
[5] Gallo, R. L. (2017). Human skin is the largest epithelial surface for interaction with microbes. Journal of Investigative Dermatology, 137(6), 1213-1214.
[6] Bocheva, G., Slominski, R. M., & Slominski, A. T. (2023). Environmental air pollutants affecting skin functions with systemic implications. International journal of molecular sciences, 24(13), 10502.
[7] Araviiskaia, E., Berardesca, E., Bieber, T., Gontijo, G., Sanchez Viera, M., Marrot, L., ... & Dreno, B. (2019). The impact of airborne pollution on skin. Journal of the European Academy of Dermatology and Venereology, 33(8), 1496-1505.
[8] Abolhasani, R., Araghi, F., Tabary, M., Aryannejad, A., Mashinchi, B., & Robati, R. M. (2021). The impact of air pollution on skin and related disorders: a comprehensive review. Dermatologic therapy, 34(2), e14840.
[9] Tran, H. M., Tsai, F. J., Lee, Y. L., Chang, J. H., Chang, L. T., Chang, T. Y., ... & Chuang, H. C. (2023). The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. Science of the Total Environment, 898, 166340.
[10] Franklin, B. A., Brook, R., & Pope III, C. A. (2015). Air pollution and cardiovascular disease. Current problems in cardiology, 40(5), 207-238.
[11] Li, Y., Du, J., Lin, S., He, H., Jia, R., & Liu, W. (2022). Air pollution increased risk of reproductive system diseases: a 5-year outcome analysis of different pollutants in different seasons, ages, and genders. Environmental Science and Pollution Research, 1-10.
[12] Kim, K. E., Cho, D., & Park, H. J. (2016). Air pollution and skin diseases: Adverse effects of airborne particulate matter on various skin diseases. Life sciences, 152, 126-134.
[13] Roberts, W. (2021). Air pollution and skin disorders. International journal of women's dermatology, 7(1), 91-97.
[14] Puri, P., Nandar, S. K., Kathuria, S., & Ramesh, V. (2017). Effects of air pollution on the skin: A review. Indian journal of dermatology, venereology and leprology, 83, 415.
[15] Gu, X., Li, Z., & Su, J. (2024). Air pollution and skin diseases: A comprehensive evaluation of the associated mechanism. Ecotoxicology and Environmental Safety, 278, 116429.
[16] Wang, J., Suárez-Fariñas, M., Estrada, Y., Parker, M. L., Greenlees, L., Stephens, G., ... & Howell, M. D. (2017). Identification of unique proteomic signatures in allergic and non-allergic skin disease. Clinical & Experimental Allergy, 47(11), 1456-1467.
[17] Huang, J., Zheng, W. H., Huang, H. C., Ran, Y. G., Liu, Y., & Huang, P. (2023). Particulate matter, nitrogen dioxide, and sulfur dioxide and their associations with allergic skin diseases: A systematic review and meta-analysis. Atmospheric Pollution Research, 14(7), 101804.
[18] Wang, I. J., Tung, T. H., Tang, C. S., & Zhao, Z. H. (2016). Allergens, air pollutants, and childhood allergic diseases. International journal of hygiene and environmental health, 219(1), 66-71.
[19] Faraz, K., Seely, M., & Marano, A. L. (2024). The role of the environment in allergic skin disease. Current Allergy and Asthma Reports, 24(6), 323-330.

[20] Yong, S. B., Gau, S. Y., Guo, Y. C., & Wei, J. C. C. (2022). Allergy from perspective of environmental pollution effects: from an aspect of atopic dermatitis, immune system, and atmospheric hazards—a narrative review of current evidences. Environmental Science and Pollution Research, 29(38), 57091-57101.

[21] Takano, H., & Inoue, K. I. (2017). Environmental pollution and allergies. Journal of Toxicologic Pathology, 30(3), 193-199.

[22] Kim, B. E., Hui-Beckman, J. W., Nevid, M. Z., Goleva, E., & Leung, D. Y. (2024). Air pollutants contribute to epithelial barrier dysfunction and allergic diseases. Annals of Allergy, Asthma & Immunology, 132(4), 433-439.

[23] Pan, Z., Dai, Y., Akar-Ghibril, N., Simpson, J., Ren, H., Zhang, L., ... & Sun, J. L. (2023). Impact of air pollution on atopic dermatitis: a comprehensive review. Clinical Reviews in Allergy & Immunology, 65(2), 121-135.

[24] Park, T. H., Park, S., Cho, M. K., & Kim, S. (2022). Associations of particulate matter with atopic dermatitis and chronic inflammatory skin diseases in South Korea. Clinical and experimental dermatology, 47(2), 325-334.

[25] Park, S. K., Kim, J. S., & Seo, H. M. (2022). Exposure to air pollution and incidence of atopic dermatitis in the general population: a national population-based retrospective cohort study. Journal of the American Academy of Dermatology, 87(6), 1321-1327.

[26] Natthanan Ruangkijpaisal,Adithep Sawatchai,Potjanee Kiewngam,Wanlapa Jotikasthira,Ratchaneewan Sinitkul,Watcharoot Kanchongkittiphon & Wiparat Manuyakorn. (2025). Effect of Fine Particulate Matter on Allergic Skin Diseases in Children. The Journal of Allergy and Clinical Immunology,155(2S),AB310-AB310.

[27] Z. M. Shibzukhov. (2025). About One Robust Variant of Principal Component Analysis. Pattern Recognition and Image Analysis,34(4),1150-1159.

[28] Yindong Song,Linfeng Xiang,Xiuwei Cheng,Dongshu Yang & Kambiz Vafai. (2025). Normalization of wettability and heat transfer coefficient curve prediction model on biphilic surfaces. Experimental Thermal and Fluid Science,164,111427-111427.

[29] G. R. Ramya & P. Bagavathi Sivakumar. (2025). Discovering Social Media Influencers Using a Deep Regression Analysis. SN Computer Science,6(4),377-377.

[30] Bożena Gajdzik,Radosław Wolniak,Anna Sączewska Piotrowska & Wiesław Wes Grebski. (2025). Polish Steel Production Under Conditions of Decarbonization—Steel Volume Forecasts Using Time Series and Multiple Linear Regression. Energies,18(7),1627-1627.