



Effect of geographical factors on reference values of creatine kinase isoenzyme

Xinrui Pang¹ · Miao Ge¹

Received: 8 May 2022 / Revised: 16 October 2022 / Accepted: 10 January 2023 / Published online: 21 March 2023
© The Author(s) under exclusive licence to International Society of Biometeorology 2023

Abstract

The aim of this study was to investigate the geographical spatial distribution of creatine kinase isoenzyme (CK-MB) in order to provide a scientific basis for clinical examination. The reference values of CK-MB of 8697 healthy adults in 137 cities in China were collected by reading a large number of literates. Moran index was used to determine the spatial relationship, and 24 factors were selected, which belonged to terrain, climate, and soil indexes. Correlation analysis was conducted between CK-MB and geographical factors to determine significance, and 9 significance factors were extracted. Based on R language to evaluate the degree of multicollinearity of the model, CK-MB Ridge model, Lasso model, and PCA model were established, through calculating the relative error to choose the best model PCA, testing the normality of the predicted values, and choosing the disjunctive kriging interpolation to make the geographical distribution. The results show that CK-MB reference values of healthy adults were generally correlated with latitude, annual sunshine duration, annual mean relative humidity, annual precipitation amount, and annual range of air temperature and significantly correlated with annual mean air temperature, topsoil gravel content, topsoil cation exchange capacity in clay, and topsoil cation exchange capacity in silt. The geospatial distribution map shows that on the whole, it is higher in the north and lower in the south, and gradually increases from the southeast coastal area to the northwest inland area. If the geographical factors are obtained in a location, the CK-MB model can be used to predict the CK-MB of healthy adults in the region, which provides a reference for us to consider regional differences in clinical diagnosis.

Keywords Creatine kinase isoenzyme (CK-MB) · Spatial autocorrelation · Ridge regression model · Lasso regression model · Principal component analysis · Geostatistical analysis

Introduction

Creatine kinase isoenzyme (CK-MB) is one of the main serum indexes for clinical detection of myocardial ischemia injury and myocardial infarction (Xu et al. 2014). It is the most characteristic index among the five indexes of myocardial enzyme spectrum (Wei et al. 2021). CK-MB can also be used as an effective biomarker to predict the occurrence and development of acute pancreatitis (Perry Elliott et al. 2021). Acute coronary syndromes can be diagnosed by its specificity and sensitivity (Baroni 2020). The current

coronavirus disease (COVID-19), caused by severe acute Respiratory syndrome coronavirus type 2 infection, has rapidly spread worldwide. Korean scholars compared patients in a hospital in Daegu and found that there were differences in arrhythmia and myocardial enzyme spectrum indexes. When CK-MB > 6.3, it could be used as an independent predictor of severe disease in COVID-19 patients. Appropriate evaluation of prognostic factors and close monitoring, and provision of necessary interventions to high-risk patients at appropriate times, can reduce the fatality rate from COVID-19 (Jang JG et al. 2020).

CK-MB medical health reference value as the most intuitive distinction between myocardial infarction patients and healthy people standard, clinical medicine in the establishment of reference value in the study of innovative practice. At present, researchers in China, by comparing different CK-MB detection methods, the results of reference values are more referential. In order to diagnose early viral

✉ Miao Ge
1551509996@qq.com
Xinrui Pang
2768739671@qq.com

¹ College of Tourist and Environment Science, Shaanxi Normal University, Xi'an, Shaanxi 710119, China

myocarditis and ischemic children, Chinese domestic scholars (Chen and Deng 2019), foreign scholars (Soldin et al. 1999). The reference ranges of CK-MB for children of different ages in Hunan, China, and Washington, USA, were established respectively. At present, Habte ML proposed that the change of CK-MB reference value is related to the influence of non-environmental factors, such as gender, age, living habits, human obesity degree, and blood lipid concentration (Habte et al. 2020). Domestic and foreign studies on CK-MB reference value and regional environment are still in a blank state. There are only a few studies on CK-MB pathology, and there is a lack of unified and accurate reference value standards to detect and prevent the incidence of cardiovascular diseases in various countries. China is located in $73^{\circ}33'E \sim 135^{\circ}05'E$, $3^{\circ}51'N \sim 53^{\circ}33'N$, spanning many temperature zones and wet and dry zones, with obvious differences between different regions. According to these characteristics, it is very necessary to develop a standard system for the reference range of CK-MB reference value which is accurate and applicable to different regions in China.

In this paper, a set of geographical distribution criteria for predicting the reference values of CK-MB among healthy adults in 2322 cities (counties) in China was established by considering the comprehensive effects of multiple geographical factors, so as to provide scientific basis for clinical diagnosis. Based on hydrology, climate, topography, and soil, spatial autocorrelation test was carried out on the measured values of CK-MB, and the geographical correlation between regions and medical indicators was obtained, which ensured the reasonable feasibility of the prediction model. Through the geographical distribution law of CK-MB reference value of healthy people and the quantitative calculation of the prediction model, a unified model standard was developed nationwide to prevent the incidence of cardiovascular diseases in different regions.

Materials

Reference source of creatine kinase isoenzyme (CK-MB)

Through searching the Chinese periodical net, China Excellent Master and Doctor Dissertation Full-text Database, China Conference Full-text Database and hospital measurement data. Based on the retrieval of “Reference value of creatine kinase isoenzyme” and “CK-MB,” we collected the CK-MB data of healthy adults for physical examination from 137 cities in China. Most of data are distributed in the second and third steps of China, which coincide with the population distribution density in China with a total of 8697 cases. Distribution of sample points is shown in (Fig. 1). The

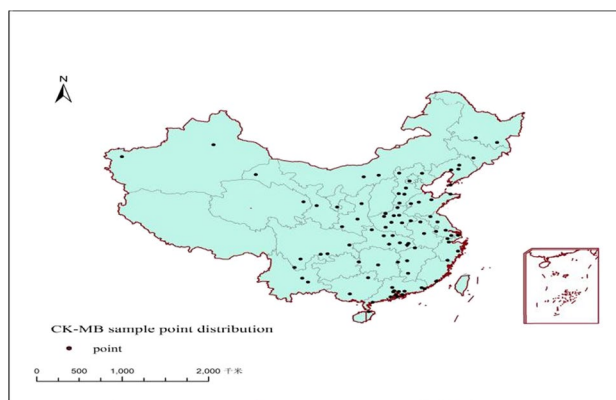


Fig. 1 Spatial distribution of CK-MB reference values in healthy adults

eastern data were more than the western data, and the data of remote areas such as Lhasa and Xinjiang were relatively scarce, and the data of Hong Kong, Macao, and Taiwan were lacking.

The geographical factors

China has a varied climate and complex terrain; the geographical differences are wide and obvious; therefore, this article selects three broad types of geographical environment factors named topographic, climate, and soil, including 24 sub-indexes (Table 1).

Method and analysis

Spatial autocorrelation

In order to further judge whether medical data is geospatial-dependent, global or local indicators are generally used to judge. Among them, global spatial autocorrelation (DeWitt et al. 2021; Wang 2020b) focuses more on the spatial distribution of attribute values of an object in the whole region. We conducted global spatial autocorrelation analysis on CK-MB reference sample data of healthy adults in arc GIS, and obtained Z value and P value by Moran's I index analysis. Moran's I value is within the range of $[-1, 1]$. The closer the value is to 1, the stronger the positive correlation is; otherwise, the stronger the negative correlation is. Z value represents the spatial distribution state of the whole region, and the significance of P value can prove whether the data is randomly distributed.

Through the calculated data (Table 2 and Fig. 2), we can know that at the confidence level of 0.01, $|Z|=9.6465 \geq 2.54$, the spatial autocorrelation of the reference value of CK-MB is significant. It will be seen from the results that it is

Table 1 The description of geographical factors

Geographical factors		Unit	Source
Topographic index	Longitude (X_1)	Degree($^{\circ}$)	The measured data of State Bureau of Surveying and Mapping
	Latitude (X_2)	Degree($^{\circ}$)	
	Altitude (X_3)	Meter(m)	
Climate index	Annual sunshine duration (X_4)	Hour (h)	China Meteorological Data Science Data Sharing Network
	Annual mean air temperature (X_5)	Centigrade ($^{\circ}\text{C}$)	
	Annual mean relative humidity (X_6)	Percentage (%)	
	Annual precipitation amount (X_7)	Millimeter (mm)	
	Annual range of air temperature (X_8)	Centigrade ($^{\circ}\text{C}$)	
	Annual mean wind speed (X_9)	Meter per-second(m/s)	
Soil index	Topsoil sand fraction (X_{10})	Weight percentage (% wt)	The World Data Soil Bank of the United Nations Food and Agriculture Organization (FAO) and the Viana International System for Applications (IIASA)
	Topsoil silt fraction (X_{11})	Weight percentage(%wt)	
	Topsoil clay fraction (X_{12})	Weight percentage(%wt)	
	Topsoil reference capacity (X_{13})	Kilogram per cubic meter(kg/dm ³)	
	Topsoil bulk density (X_{14})	Kilogram per cubic meter(kg/dm ³)	
	Topsoil gravel content (X_{15})	Volume percentage(%vol)	
	Topsoil organic carbon (X_{16})	Weight percentage(% wt)	
	Topsoil pH (X_{17})		
	Topsoil (clay) cation exchange capacity (X_{18})	Centimol per kilogram (cmol/kg)	
	Topsoil (silt) cation exchange capacity (X_{19})	Centimol per kilogram	
	Topsoil base saturation (X_{20})	Percentage (%)	
	Topsoil total exchange capacity (X_{21})	Centimol per kilogram (cmol/kg)	
	Topsoil calcium carbonate (X_{22})	Centimol per kilogram	
	topsoil calcium sulphate(X_{23})	centimol per kilogram (cmol/kg)	
	Surface soil alkalinity (X_{24})	d-Siemens per (cmol/kg)	

Table 2 Global Moran index report

Moran's <i>I</i>	Z value	P value
0.299	9.6465	0.00718

affected by the space geographical environment; different environments make the different geographical distributions of CK-MB.

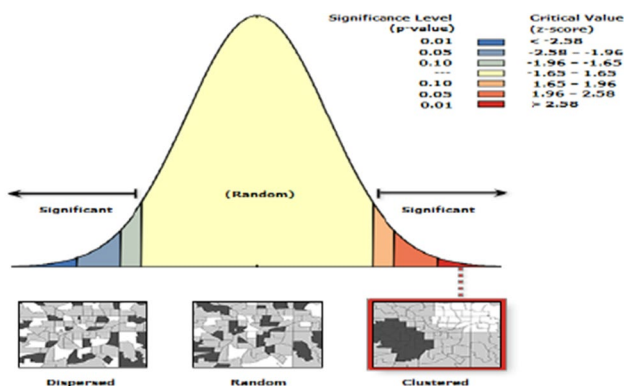


Fig. 2 Spatial autocorrelation analysis of CK-MB in healthy adults

Correlation analysis

Correlation analysis can further find the geographical factors with significant correlation, which lays a foundation for the establishment of subsequent prediction models. Among them, the common correlation coefficients are Pearson, Spearman, Kendall, etc. The data presented non-normal distribution through K-S test under the non-parametric analysis method. To solve the non-linearity of sample data, Spearman correlation coefficient was selected to judge the monotone form among geographical factors.

Based on SPSS25.0, with 24 geographical indicators as independent variables and CK-MB as dependent variable, the correlation coefficient (*r*) and significance coefficient (*P*) between CK-MB of healthy adults in various regions of China and 24 geographical indicators can be obtained (Table 3). Among the 24 geographical factors selected, there are 9 geographical factors that are correlated with the CK-MB reference value of healthy adults, among which the following 5 geographical factors are generally related: latitude (X_2), annual sunshine duration (X_4), annual mean relative humidity (X_6), annual precipitation

Table 3 The *r* and *P* values of geographical factors and CK-MB

Geographical factors		<i>r</i>	<i>P</i>
Topographic index	Longitude (X_1)	0.222*	0.015
	Latitude (X_2)	0.279**	0.002
	Altitude (X_3)	-0.106	0.249
Climate index	Annual sunshine duration (X_4)	0.215*	0.018
	Annual mean air temperature (X_5)	-0.221*	0.015
	Annual mean relative humidity (X_6)	-0.124	0.177
	Annual precipitation amount (X_7)	-0.191*	0.036
	Annual range of air temperature (X_8)	0.247**	0.007
	Annual mean wind speed (X_9)	0.258**	0.004
	Soil index	Topsoil sand fraction (X_{10})	0.173
Topsoil silt fraction (X_{11})		0.215*	0.018
Topsoil clay fraction (X_{12})		-0.223*	0.015
Topsoil reference capacity (X_{13})		0.215*	0.018
Topsoil bulk density (X_{14})		0.137	0.135
Topsoil gravel content (X_{15})		-0.143	0.120
Topsoil organic carbon (X_{16})		0.007	0.935
Topsoil pH (X_{17})		0.137	0.134
Topsoil (clay) cation exchange (X_{18}) (X ₁₇) capacity (X_{18})		0.011	0.908
Topsoil (silt) cation exchange capacity (X_{19})		0.225*	0.014
Topsoil base saturation (X_{20})		-0.039	0.670
Topsoil total exchange capacity (X_{21})		-0.003	0.975
Topsoil calcium carbonate (T_CaCO3) (X_{22})		0.109	0.234
Topsoil calcium sulphate(T_CaSO4) (X_{23})		0.051	0.578
Surface soil alkalinity (X_{24})		0.026	0.777

** Indicates a significant correlation, within 99% confidence interval; * indicates a significant correlation, within a 95% confidence interval

amount (X_7), and annual range of air temperature (X_8). The four geographical factors with significant correlation are as follows: annual mean air temperature (X_5), topsoil gravel content (X_{15}), topsoil (clay) cation exchange(X_{18}), and topsoil (silt) cation exchange capacity (X_{19}).

In R language (Muhammad et al. 2021), the correlation coefficients of 9 related geographical factors were taken as independent variables and dependent variables to judge whether there was multicollinearity influence among geographical factors (Fig. 3).

Latitude has a strong collinearity effect with annual mean temperature, annual precipitation, and annual temperature difference. Therefore, the least square regression analysis is excluded, which leads to the inconsistency between the results and reality. Therefore, appropriate model and method are further selected for optimal prediction.

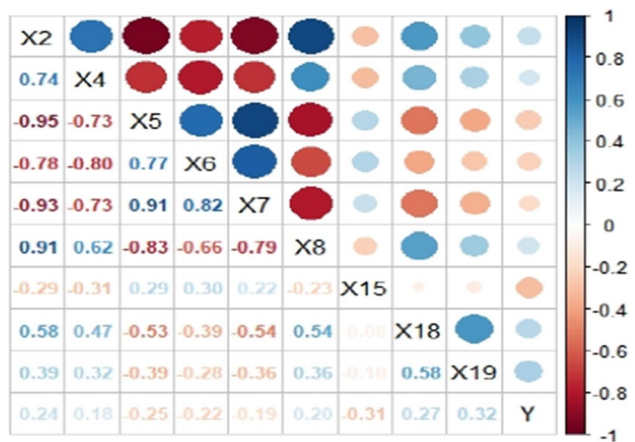


Fig. 3 Heat map of correlation coefficient of geographical factors with correlation

Ridge regression analysis

For the problem of multicollinearity among factors, A.E. Hoerl first proposed an improved least square estimation method (Hoerl and Kennard 2000), called ridge estimation, in 1962. In

1970, Hall, in collaboration with Kennard, developed a further systematic development of ridge estimation. Ridge estimation is a kind of biased estimation modified to the least square estimation method. There exists a ridge regression estimation with β as $k \geq 0, b(k) = (X'X + kI) - 1X'Y$. When $k=0$, the ridge regression equation is the least square estimation. When there are significant multiple correlations among the independent variables, the regression coefficient can be controlled to reduce the error (Choi et al. 2019).

Based on R language and R Studio programming environment, 9 geographical factors with correlation were loaded and substituted. As shown in Fig. 4(a), the error value became flat when $\text{Log}(\lambda) = 2$, and the error of the prediction regression equation was the smallest. The regression equation obtained from Fig. 4(b) and programming code was:

$$\hat{Y} = 0.0332X_2 + 0.0225X_4 - 0.0271X_5 - 0.0204X_6 + 0.0107X_7 + 0.0447X_8 - 0.0667X_{15} + 0.0886X_{18} + 0.0762X_{19}$$

Lasso regression model

Lasso regression model aims to solve many variables in the prediction process, and the influence of “over fitting” and multicollinearity brought about by OLS linear fitting, so as to improve the explanatory power of the model. Tibshirani proposed Lasso method for the first time. Lasso makes the dimension of independent variable compressed by adjusting lambda. The number of samples was selected based on the punishment method, and the original data

was compressed to compress the relative coefficient of non-significant variables to 0. Such methods avoid over-interpretation of the current sample and explore rules that apply to the population (Tibshirani et al. 2005).

Based on R language and R Studio programming environment, when the 9 relevant geographical factor indicators that are useful to the reference value of CK-MB obtained from Fig. 5(a) are reduced to 4, the prediction regression equation has the smallest error, and the corresponding $\text{Log}(\lambda) = 0.1414$. The regression equation obtained from the programming code in Fig. 5(b) is as follows:

$$\hat{Y} = 0.00224X_8 - 0.0744X_{15} + 0.188X_{18} + 0.0806X_{19}$$

Principal component analysis model

Feasibility of PCA

Prior to PCA, correlation exploration among factors was conducted (Fig. 2). We obtained that the correlation coefficient between latitude and annual mean temperature was 0.95 (> 0.90), the correlation coefficient between latitude and annual precipitation was 0.93 (> 0.90), and the correlation coefficient between latitude and annual temperature was 0.91 (> 0.90). The existence of serious collinearity is suitable for principal component analysis (Beattie and Esmonde-White 2021).

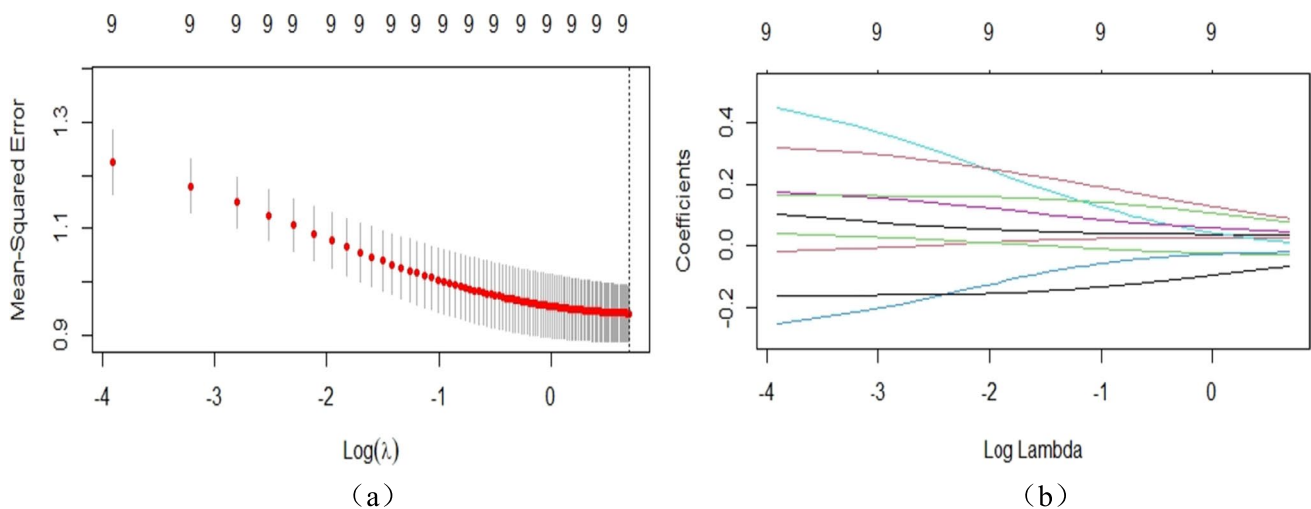


Fig. 4 CK-MB reference ridge map for healthy adults Fig. 4 (a) CK-MB reference Ridge regression prediction error Fig. 4 (b) CK-MB reference Ridge regression factor contribution

Principal component extraction and regression equation establishment

First, KMO statistics and Bartlett test of spherical were used for judgment (Table 4). The results showed that the closer KMO was to 1, the stronger the correlation between variables was, and the original variables were suitable for factor analysis. Bartlett’s spherical test value $P < 0.001$ negates the null hypothesis, indicating that the correlation matrix among variables is not the identity matrix. In other words, it is considered that there is a correlation between variables, and it is necessary to carry out factor analysis. In Table 4, KMO

value is 0.779 and Sig. value is 0.000, which is in line with the condition of factor analysis.

SPSS25.0 was applied to the geographical index—latitude (X_2), annual sunshine duration (X_4), annual mean air temperature (X_5), annual mean relative humidity (X_6), annual precipitation amount (X_7), and annual range of air temperature (X_8), topsoil gravel content (X_{15}), topsoil (clay) cation exchange (X_{18}), and topsoil (silt) cation exchange capacity (X_{19}) were extracted as principal components. The correlation coefficient matrix of geographical indicators (Table 5) and variance decomposition accumulation table (Table 6) were obtained.

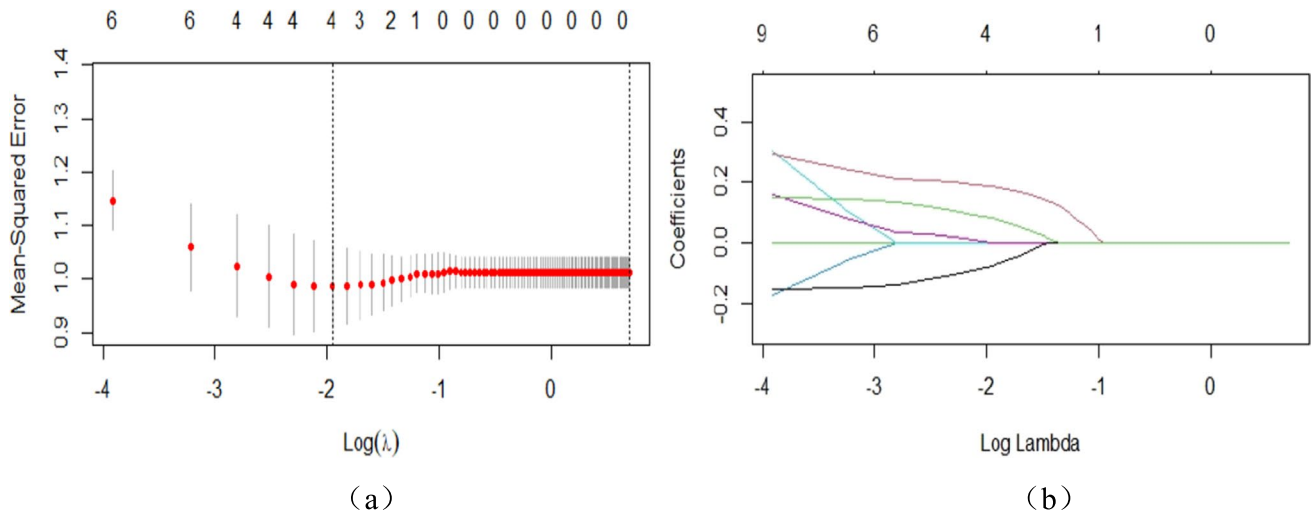


Fig. 5 CK-MB reference ridge map for healthy adults Fig. 5 (a) CK-MB reference Lasso regression prediction error Fig. 5 (b) CK-MB reference Lasso regression factor contribution

Table 4 KMO statistics and Bartlett’s test for sphericity

KMO statistics and Bartlett’s test			
Sampling the Kaiser–Meyer–Olkin measure of adequacy			0.779
Bartlett’s test for sphericity		The approximate chi-square	917.181
		df	36
		Sig.	0.000

Table 5 Correlation coefficient matrix of geographical indicators

	X_2	X_4	X_5	X_6	X_7	X_8	X_{15}	X_{18}	X_{19}
X_2	1.000	0.650	−0.928	−0.737	−0.885	0.919	0.117	−0.071	0.053
X_4	0.650	1.000	−0.657	−0.783	−0.619	0.500	−0.012	−0.197	0.026
X_5	−0.928	−0.657	1.000	0.732	0.829	−0.812	−0.039	0.027	−0.009
X_6	−0.737	−0.783	0.732	1.000	0.779	−0.554	−0.048	0.189	0.006
X_7	−0.885	−0.619	0.829	0.779	1.000	−0.750	−0.106	0.026	−0.028
X_8	0.919	0.500	−0.812	−0.554	−0.750	1.000	0.117	−0.082	0.086
X_{15}	0.117	−0.012	−0.039	−0.048	−0.106	0.117	1.000	0.034	0.038
X_{18}	−0.071	−0.197	0.027	0.189	0.026	−0.082	0.034	1.000	0.564
X_{19}	0.053	0.026	−0.009	0.006	−0.028	0.086	0.038	0.564	1.000

Table 6 Variance decomposition cumulative table

Total variance of the interpretation							
Ingredients	Initial eigenvalue			Extract the sum of squares and load it			Rotate the sum of squares and load the total
	Total	% Variance	%Cumulative	Total	% Variance	% Cumulative	
1	4.755	52.833	52.833	4.755	52.833	52.833	2.181
2	1.592	17.692	70.524	1.592	17.692	70.524	1.262
3	1.008	11.201	81.725	1.008	11.201	81.725	1.004
4	0.667	7.414	89.139				
5	0.450	5.002	94.141				
6	0.255	2.831	96.972				
7	0.136	1.514	98.485				
8	0.111	1.232	99.717				
9	0.025	0.283	100.000				

According to the results in Table 6, the cumulative contribution rate of the three factors reached 81.725%, indicating that 81.725% of all geographical variables had been included and the effect was good. When the fourth characteristic root is taken, the characteristic value of the curve has dropped below 1, so three common factors are selected to replace the original nine geographical variables to study the influence on the reference value of adult CK-MB. By extracting the first three principal components, the combined relationship between the first three principal components and geographical factors can be obtained according to the component score coefficient matrix obtained (Table 7):

$$\begin{aligned}
 Z_1 = & -0.966stdX_2 - 0.782stdX_4 + 0.933stdX_5 + 0.859stdX_6 \\
 & + 0.916stdX_7 - 0.858stdX_8 - 0.1stdX_{15} + 0.129stdX_{18} \\
 & - 0.028stdX_{19}, Z_2 = 0.069stdX_2 - 0.116stdX_4 - 0.052stdX_5 \\
 & + 0.106stdX_6 - 0.064stdX_7 + 0.096stdX_8 + 0.15stdX_{15} \\
 & + 0.879stdX_{18} + 0.867stdX_{19}, Z_3 = 0.049stdX_2 - 0.177stdX_4 \\
 & + 0.043stdX_5 + 0.087stdX_6 - 0.026stdX_7 + 0.092stdX_8 \\
 & + 0.966stdX_{15} - 0.079stdX_{18} - 0.131stdX_{19}
 \end{aligned}$$

In the formula, Z_i ($i = 1, 2, 3$) represents the principal component; $StdX_i$ ($i = 1, 2, 3, \dots, 19$) represents the standard

Table 7 Principal component score coefficient matrix

	Z_1	Z_2	Z_3
X_2	-0.966	0.069	0.049
X_4	-0.782	-0.116	-0.177
X_5	0.933	-0.052	0.043
X_6	0.859	0.106	0.087
X_7	0.916	-0.064	-0.026
X_8	-0.858	0.096	0.092
X_{15}	-0.100	0.147	0.966
X_{18}	0.129	0.879	-0.079
X_{19}	-0.028	0.867	-0.131

indicator variable. The standard indicator variables are standardized from the original variables, which can be obtained according to Table 8: $stdX_2 = (X_2 - 31.609)/6.310$; $stdX_4 = (X_4 - 2147.831)/436.556$; $stdX_5 = (X_5 - 14.896)/4.833$; $stdX_6 = (X_6 - 69.850)/9.252$; $stdX_7 = (X_7 - 989.517)/491.148$; $stdX_8 = (X_8 - 24.368)/7.159$; $stdX_{15} = (X_{15} - 1.360)/0.055$; $stdX_{18} = (X_{18} - 6.956)/1.104$; $stdX_{19} = (X_{19} - 45.8)/17.953$.

Taking $Z_1, Z_2,$ and Z_3 as independent variables and the reference value of CK-MB for healthy middle-aged people as dependent variables, the regression equation was obtained:

$\hat{Y} = 9.04 - 1.047Z_1 - 0.292Z_2 - 1.448Z_3 \pm 0.089$. Finally, the linear regression equation between CK-MB index and geographical index is obtained by putting the original variable into the multiple regression equation with the original variable:

$$\begin{aligned}
 \hat{Y} = & 19.49 + 0.123X_2 + 0.00214X_4 \\
 & - 0.197X_5 - 0.081X_6 - 0.00179X_7 + 0.0940X_8 \\
 & - 6.929X_{15} - 0.194X_{18} - 0.00190X_{19}
 \end{aligned}$$

Model quality assessment

In order to select the appropriate model, the optimal ridge regression method, Lasso method, and principal component analysis method of root mean square error, mean absolute error, and Theil inequality coefficient are compared to eliminate the influence of multicollinearity, and finally get the model with the best prediction accuracy for the optimal prediction model of CK-MB reference value.

Root mean square error (RMSE) is the standard error. Is the square root of the ratio of the square of the deviation between the predicted value Y_i and the measured value Y_i to the predicted number N . RMSE can better reflect the prediction accuracy of the model, and the smaller RMSE value is, the better the prediction result will be.

Table 8 Descriptive statistics of geographic indicators

Geographical indicators	X_2	X_4	X_5	X_6	X_7	X_8	X_{15}	X_{18}	X_{19}
Mean	31.609	2147.831	14.896	69.850	989.517	24.368	1.360	6.956	45.8
Standard deviation	6.310	436.556	4.833	9.252	491.148	7.159	0.055	1.104	17.953

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2}$$

Mean absolute error (MAE) excludes the occurrence of maximum or minimum error values and excludes the particularity of predicted values, so as to better reflect the actual situation of predicted values. The smaller the MAE value is, the better the model prediction result is.

$$MAE = \frac{1}{N} \sum_{i=1}^N |(Y_i - y_i)|$$

Theil inequality coefficient (TIC), its evaluation standard is more accurate than RMSE, where $TIC \in [0,1]$. The smaller the TIC value is, the better the model prediction result is.

$$TIC = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - y_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N Y_i^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}}$$

Furthermore, to establish the forecast model of the TIC data reference value of Chinese healthy adult, using test data, the effect of three kinds of models to forecast assessment (Table 9), according to the results of principal component analysis of RMSE and MAE, TIC value is the smallest, so we choose principal component analysis as Chinese healthy adult CK-MB reference optimal prediction model.

Geostatistical analysis

The predicted CK-MB values of 2322 counties and cities in China were imported into the vectorized map of China, and the normal distribution of the data was explored through histogram analysis and normal Q-Q map, and the normal distribution was replaced by the mean value.

The trend distribution chart of predicted CK-MB values in China was further drawn (Fig. 6). The predicted indexes

Table 9 Measure of model prediction accuracy

Evaluation of measure-ment	Ridge	Lasso	PCA
RMSE	61.36	7.003	6.506
MAE	60.29	5.712	5.192
TIC	0.761	0.448	0.324

of longitude, latitude, and 2322 observation points were set as X, Y, and Z axes, with the X axis being east and the Y axis being north. The trend distribution of predicted values of CK-MB presents a second-order polynomial on the whole. In the east–west direction of (X-axis), it presents a concave curve, which rapidly drops from west to east and then tends to be flat. In the (Y axis), it presents an upward convex curve from the south to the north, rising first and then falling from the south to the north. The predicted values of healthy adults are basically consistent with the trend analysis results of sample data. In order to further observe the change of reference values in different geographical regions, the geospatial distribution map of the predicted values of CK-MB is also needed.

Judging from the histogram results, the predicted values of CK-MB in China do not obey normal distribution, and the disjunction Kriging should be used in the interpolation. As can be seen from Fig. 7, the overall geographical distribution characteristics of the predicted values of CK-MB for healthy adults are higher in the north and lower in the south, gradually increasing from the south-east coastal area to the northwest inland area. The general trend is similar to the average temperature in winter and the distribution law of precipitation in north and south of China. The lowest values are located in the tropical and subtropical regions of Guangzhou Guangxi and the south of Yunnan-Guizhou Plateau. The lower values are widely distributed in the south of Qinhuai River and the south of the Tropic of Cancer. The highest values are located in the northwest and northeast plain north of 40°N latitude of China. The predicted distribution law of CK-MB for healthy adults is the most obvious in latitude, which is

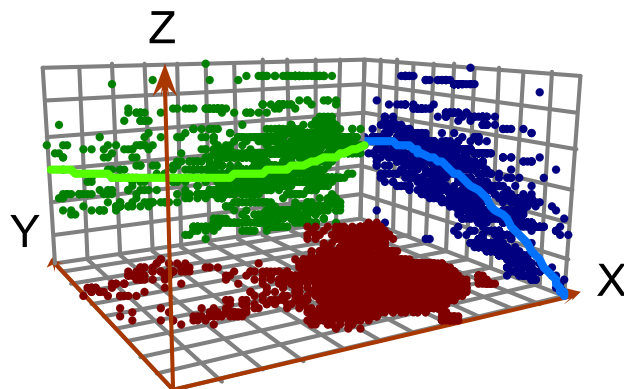
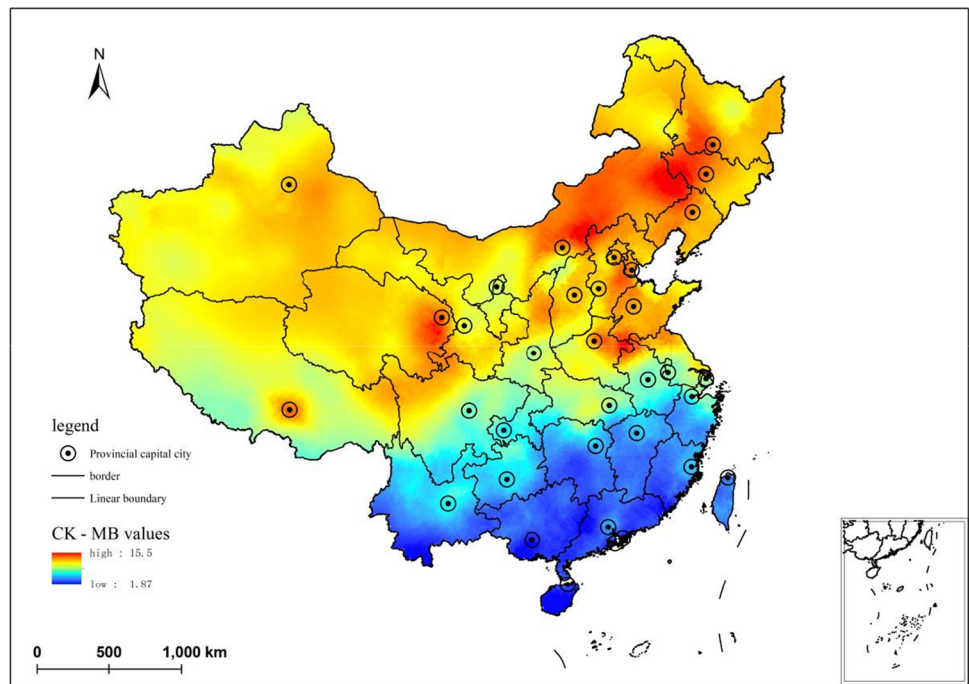


Fig. 6 Trend surface analysis of CK-MB predictive values in Chinese healthy adults

Fig. 7 Geographic distribution of CK-MB predicted values among healthy adults in China



consistent with the distribution change of annual precipitation and temperature in China.

Discussion

According to the “Theory of Three Causes” proposed in *Neijing* and *Discourse on Time and Diseases*, that is, due to time, person, and place, in different natural geographical environment; warm, hot, and cold climate; and people’s physiological indicators, metabolic function is different, and geographical environment and health are closely related. Based on the comprehensive discussion of spatial autocorrelation, correlation analysis, and geostatistical analysis, this paper concluded that the reference values of CK-MB for healthy adults in China were correlated with 9 geographical factors. After the establishment of principal component analysis model prediction and interpolation mapping, the overall pattern is higher in the north and lower in the south, gradually increasing from the southeast coast to the northwest inland. From the perspective of single-phase correlation coefficient combined with geographical factors, the discussion can be carried out from the following three aspects: terrain index, climate index, and soil index.

In terms of terrain index, latitude is generally positively correlated with the CK-MB reference value of healthy adults in China. China has a large territory and abundant resources, with a wide span of longitude and latitude, and latitude directly affects the distribution of temperature zones and climate types in the north and south (Nguyen et al. 2016).

China’s climate types are complex and diverse, with significant characteristics of continental monsoon climate, temperature decreases from south to north, and there is a large temperature difference between north and south. At the same longitude, the reference value of CK-MB in high latitude area is high and decreases to low latitude area in turn (Inclair 1989). Xue Wang proposed that in the northeast of China, winter is cold and long, and summer is warm and short (Wang et al. 2020a). Moreover, in cold areas, human body is often stimulated by low temperature, which is prone to erythrocyte aggregation, increased blood viscosity, capillary constriction, increased blood pressure, increased myocardial oxygen consumption, aggravated myocardial ischemia, and the corresponding CK-MB value is also high. Therefore, in Northwest China with large annual temperature difference, the value of residents in Northeast China is higher than that in Southern China with high annual average temperature, which is consistent with the spatial pattern of high temperature in the north and low temperature in the south of the geographical spatial distribution map.

In terms of climate indicators, the annual sunshine hours, annual average temperature, annual average relative humidity, annual precipitation, and annual temperature are non-negligible factors affecting the CK-MB reference value of healthy people in China. Annual sunshine duration is closely related to latitude and altitude, and geographical factors are interrelated and inseparable. Studies show that the longer the sunshine duration is, the higher the incidence of cardiovascular diseases, and the higher CK-MB value reflects the positive correlation effect of sunshine duration (Scarborough

et al. 2012). Relative humidity and annual precipitation are negatively correlated with CK-MB reference value. Studies have shown that low precipitation and relatively low air humidity can increase the incidence of cardiovascular diseases (Afanas'eva et al. 2010). Studies have shown^[25–26] that when Humidity is greater than 60%, the human body will feel uncomfortable, blood filling will increase, blood viscosity will decrease, and it will be easier to eliminate the metabolites in the blood vessels (Yin and Wang 2018). Therefore, the CK-MB level in the body will decrease, which reflects the negative correlation effect of humidity. Humidity is lower in northwest area of our country, because this CK-MB reference value is higher than the coastal area value with higher humidity (Goldie et al. 2018).

In terms of soil indexes, most of the CK-MB reference value is indirectly related to soil factors such as topsoil gravel content, topsoil (clay) cation exchange capacity, and topsoil (silt) cation exchange capacity, rather than factors such as longitude, latitude, temperature, and humidity that can directly affect the indexes. The cation exchange capacity of the topsoil (clay) and the cation exchange capacity of the topsoil (silt) are important indexes for evaluating soil fertility. The soil types in China have provincial and regional rules, presenting alternation changes under different land forms of plains, basins, and plateaus in China. Different types of crops are planted in different soil types in different regions (Duan et al. 2016). The soil in southern and northern regions is further differentiated from wheat and rice cultivation, which affects the eating habits of different people in China. For example, according to “Su Wen · On Fever Diseases,” “In the case of Yin deficiency, Yang is bound to co-exist, so when there is less air, it is hot and sweats.” In hot and dry areas, the body reduces blood flow by constricting the arteries, inhibiting the sweating function of the skin, and maintaining water balance in the body. The contraction of arterial vessels is bound to lead to the decrease of cardiovascular systolic function, leading to the decrease of CK-MB index. The reference value of CK-MB is negatively correlated with the topsoil gravel content, because the higher gravel content leads to the decline of soil organic matter and the weakening of soil fertility. Compared with people in fertile soil and water areas, the residents in arid desert areas of China have a single diet and higher lipid and protein, so the higher value of CK-MB will also increase the risk of cardiovascular diseases (Zhang and Yang 2016).

Conclusion

Spatial analysis shows that there is spatial autocorrelation dependence in CK-MB reference values of healthy Chinese adults. At the same time, combined with the correlation analysis, the latitude, sunshine duration, annual average relative

humidity, annual precipitation, annual poor temperature, annual average temperature, topsoil gravel content, topsoil (clay) cation exchange capacity, and topsoil (silt) cation exchange capacity were further determined. Reasonable reference values of CK-MB medical indicators in a certain region can be obtained by integrating the optimal model and local statistical analysis if the above indexes are known. The distribution pattern of CK-MB reference values in China is higher in the north and lower in the south, gradually increasing from southeast coastal areas to northwest inland areas. The reference value of CK-MB of healthy people in the region can be determined through the collected geographical indicators of any region, combined with the principal component analysis prediction model or the obtained geographical distribution map.

As medical index data are derived from online literature, there are problems such as uneven distribution of data points and data deviation. The data of the eastern region is slightly more than that of the western region, which is the result of the differences in population density, economic development level, national policy support, and medical resource level in the eastern and western regions of China. Furthermore, explore models suitable for medical geography, and improve the prediction accuracy of data by combining machine learning, epidemiological learning, public health, and other disciplines. Therefore, with the progress of research, more data should be combined to focus on the selection of human geographical factors, and factors affecting medical indicators at different levels should be considered, so as to improve the practical value and scientific research value brought by research.

Funding Project 41761100 is supported by the National Natural Science Foundation of China. Project 2019JM-408 is supported by the Natural Science Basic Research Program of Shaanxi Province.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Afanas'eva GN, Panova TN, Dedova AV et al (2010) The influence of certain meteorological factors on mortality from complications of arterial hypertension. *Sud Med Ekspert* 53(5):10–12
- Baroni S (2020) Lack of utility of CK-MB in combination with troponin I for the diagnosis of acute coronary syndrome. *Intern Emerg Med* 15(7):1345–1346
- Beattie JR, Esmonde-White FWL (2021) Exploration of principal component analysis: deriving principal component analysis visually using spectra. *Appl Spectrosc* 75(4):361–375. <https://doi.org/10.1177/0003702820987847>
- Chen J, Deng Y (2019) Diagnostic performance of serum CK-MB, TNF- α and hs-CRP in children with viral myocarditis. *Open Life Sciences* 14(1):38–42. <https://doi.org/10.1515/biol-2019-0005>
- Choi SH, Jung H et al (2019) Ridge fuzzy regression model. *Int J Fuzzy Syst* 21(7):2077–2090. <https://doi.org/10.1016/J.WASMAN.2021.04.054>

- DeWitt TJ, Fuentes JI, Ioerger TR et al (2021) Rectifying I: three point and continuous fit of the spatial autocorrelation metric, Moran's I, to ideal form. *Landscape Ecol*. <https://doi.org/10.1007/s10980-021-01256-0>
- Duan Q, Lee J, Liu Y et al (2016) Distribution of heavy metal pollution in surface soil samples in china: a graphical review. *Bull Environ Contam Toxicol* 97(3):303–309. <https://doi.org/10.1007/s00128-016-1857-9>
- Elliott P, Cowie MR, Franke J et al (2021) Development, validation, and implementation of biomarker testing in cardiovascular medicine state-of-the-art: proceedings of the European Society of Cardiology—Cardiovascular Round Table. *Cardiovasc Res* 117(05):1248–1256. <https://doi.org/10.1093/cvr/cvaa272>
- Goldie J, Alexander L, Lewis SC et al (2018) Changes in relative fit of human heat stress indices to cardiovascular, respiratory, and renal hospitalizations across five Australian urban populations. *Int J Biometeorol* 62:423–432. <https://doi.org/10.1007/s00484-017-1451-9>
- Habte ML, Melka DS, Degef M et al (2020) Comparison of lipid profile, liver enzymes, creatine kinase and lactate dehydrogenase among type II diabetes mellitus patients on statin therapy. *Diabetes Metab Syndr Obes* 18(13):763–773. <https://doi.org/10.2147/DMSO.S234382>
- Hoerl A, Kennard R (2000) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 42(1):80–86. <https://doi.org/10.2307/1271436d>
- Inclair H (1989) Latitude and ischaemic heart disease. *Lancet* 1(8643):895. [https://doi.org/10.1016/s0140-6736\(89\)92880-8](https://doi.org/10.1016/s0140-6736(89)92880-8)
- Jang JG, Hur J, Choi EY et al (2020). Prognostic factors for severe coronavirus disease 2019 in Daegu, Korea. *J Korean Med Sci* 35(23):e209. <https://doi.org/10.3346/jkms>
- Muhammad Q, Månsson, et al (2021) On some beta ridge regression estimators: method, simulation and application. *J Stat Comput Simul* 91(9):1699–1712. <https://doi.org/10.1080/00949655.2020.1867549>
- Nguyen JL, Yang W, Ito K et al (2016) Seasonal influenza infections and cardiovascular disease mortality. *JAMA Cardiol* 1(3):274–281. <https://doi.org/10.1001/jamacardio.2016.0433S>
- Scarborough P, Allender S, Rayner M et al (2012) Contribution of climate and air pollution to variation in coronary heart disease mortality rates in England. *PLoS One* 7(3):e32787. <https://doi.org/10.1371/journal.pone.0032787>
- Soldin SJ et al (1999) Pediatric reference ranges for creatine kinase, CKMB, troponin I, iron, and cortisol. *Clin Biochem* 32(1):77–80. [https://doi.org/10.1016/S0009-9120\(98\)00084-8](https://doi.org/10.1016/S0009-9120(98)00084-8)
- Tibshirani R, Saunders M, Rosset S et al (2005) Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society* 67(1):91–108. <https://doi.org/10.1111/j.1467-9868.2005.00490.x>
- Wang X, Jiang Y, Bai Y et al (2020a) Association between air temperature and the incidence of acute coronary heart disease in Northeast China. *Clin Interv Aging* 15:47–52. <https://doi.org/10.2147/CIA.S235941>
- Wang YS, Wang JM, Wang WB (2020b) Temporal-spatial distribution of tuberculosis in China, 2004–2016. *Zhonghua Liu Xing Bing Xue Za Zhi* 41(4):526–531. <https://doi.org/10.3760/cma.j.cn112338-20190614-00441>
- Wei W, Zhang L, Zhang Y et al (2021) Predictive value of creatine kinase MB for contrast-induced acute kidney injury among myocardial infarction patients. *BMC Cardiovasc Disord* 21:337. <https://doi.org/10.1186/s12872-021-02155-7>
- Xu C, Zhang T, Zhu B et al (2014) Diagnostic role of postmortem CK-MB in cardiac death: a systematic review and meta-analysis. *Forensic Sci Med Pathol* 16(2):287–294. <https://doi.org/10.1007/s12024-020-00232-5>
- Yin Q, Wang J (2018) A better indicator to measure the effects of meteorological factors on cardiovascular mortality: heat index. *Environ Sci Pollut Res Int* 25(23):22842–22849. <https://doi.org/10.1007/s11356-018-2396-1>
- Zhang M, Yang XJ (2016) Effects of a high fat diet on intestinal microbiota and gastrointestinal diseases. *World J Gastroenterol* 22(40):8905–8909. <https://doi.org/10.3748/wjg.v22.i40.8905>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.