International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7009-7022

https://doi.org/10.70517/ijhsa463601

Research on the Application of Multi-source Data Analysis and Intelligent Modeling Technology in Training in College Sports Athletics

Yan Shen[1,*] and Qinghua Zhu[2]
[1] Department of Physical Education and Research, Hunan Institute of Technology, Hengyang, Hunan, 421002, China
[2] College of Physical Education, University of South China, Hengyang, Hunan, 421001, China
Corresponding authors: (e-mail: qq7897273@163.com).

Abstract With the development of big data technology, the application of multi-source data analysis and intelligent modeling technology in the field of track and field training can realize the accurate assessment of training quality, intelligent monitoring and prediction of functional changes, which can help to formulate personalized training programs, improve the scientificity and effect of training, and promote the transformation of data in the field of track and field training. Based on factor analysis and time series theory, this paper constructs a framework for the application of multi-source data analysis and intelligent modeling technology in college sports track and field. The study collects S collegiate sports track and field training data through questionnaires, uses factor analysis for multi-source data processing and evaluation, and uses time series theory to construct a training function monitoring and prediction model. The results of the study showed that: the standardized Cronbach's α coefficient of the questionnaire was 0.986, which was highly consistent and reliable; the KMO value was 0.861, which was suitable for factor analysis; three principal components were extracted from the principal component analysis, and the cumulative variance explained rate reached 84.42%; the time series smoothness test of the function indicator hemoglobin (HGB) showed that its ADF test statistic was - 3.90368, which is less than the critical value of 5% significant level and suitable for modeling; the predicted value of HGB based on the ARMA(1,1) model is highly consistent with the true value, with an average predicted value of 150 g/L, and the variation of residual variations is controlled within the range of [-1,1]. The study proved that the application of multi-source data analysis and intelligent modeling technology to college sports track and field training can achieve scientific evaluation of training quality and intelligent monitoring of functional changes, provide data support for the development of personalized training programs, and promote the development of track and field training in the direction of more scientific and intelligent.

Index Terms Multi-source data analysis, Intelligent modeling, Factor analysis, Time series, Athletic training, Function monitoring

I. Introduction

In the process of rapid social development, sports training in colleges and universities has also progressed [1]. Accompanied by the gradual deepening of learning, students have a more profound understanding of the content related to athletics [2], [3]. In general colleges and universities, track and field sports training for students can have a positive effect on the improvement of students' physical and psychological quality [4], [5]. However, at present, many colleges and universities still have more problems in track and field training [6]. Teachers' lack of attention to track and field training, insufficient student participation and other related problems have led to the long-term existence of this problem, which affects the smooth implementation of educational reform work in colleges and universities [7]-[9]. Therefore, relevant educators need to pay enough attention to track and field training. According to the characteristics of students and the talent cultivation program developed by the school, students are provided with targeted guidance [10]-[12]. With the development of big data and artificial intelligence, the application of multi-source data analysis and intelligent modeling technology in college sports athletics provides technical support for the development of college athletics [13], [14].

Multi-source data analysis of college sports athletics, that is, through the collection of students' physical quality, athletic performance and other data, and its processing and analysis, to find out its deficiencies in order to take appropriate measures [15]-[17]. The intelligent modeling technology of college sports athletics, on the other hand, is an intelligent transformation and technical enhancement of the traditional training models, methods and other aspects, which can realize the personalization of training, accurate assessment and effective guidance, etc [18]-

[20]. The application of multi-source data analysis and intelligent modeling technology in training is the innovation of the era of college track and field reform, and it is also a necessary way to improve the training effect [21], [22].

As a basic sport, physical education track and field occupies an important position in physical education in colleges and universities. In recent years, the modernization process of physical education has been accelerating, but the field of college sports track and field training still faces many challenges. Traditional training methods rely too much on coaching experience and lack of scientific basis; training data collection is fragmented and underutilized for analysis; training evaluation system is single and difficult to comprehensively reflect the status of athletes; personalized training programs are lacking and difficult to meet the needs of different athletes. These problems seriously constrain the improvement of college sports track and field training level. Under the background of the data era, the introduction of multi-source data analysis and intelligent modeling technology into the field of sports track and field training, and the construction of a scientific and efficient training system has become an important direction of current sports training research. Related studies at home and abroad show that data-driven training methods can effectively improve the training effect. Foreign scholars have realized the quantitative assessment and prediction of training effect by mining and analyzing multi-source data such as athletes' physiological indexes and training loads. Domestic scholars have also begun to explore the application of big data technology in sports training, but there are still deficiencies in the integration of multi-source data and intelligent modeling. Athletic training in college sports involves many factors such as athletes' physical quality, technical level, psychological state and other factors, which need to be considered comprehensively. Therefore, how to construct an effective multi-source data analysis framework, how to use intelligent modeling technology to achieve accurate monitoring and prediction of the training process, and how to optimize personalized training programs based on the results of data analysis are key issues that need to be solved in the current study.

This study takes track and field training in S colleges and universities as the research object, collects multi-source data of track and field training through questionnaires, uses factor analysis to explore the key factors affecting the training quality; constructs the intelligent monitoring and prediction model of the changes in track and field training functions based on the time series theory; and realizes the scientific evaluation of the quality of track and field training and the intelligent monitoring of the changes in the functions, so as to provide data support for the formulation of the personalized training program. The research has both theoretical innovation, combining factor analysis and time series theory, expanding the method of sports training data analysis; and practical value, providing a feasible data-based transformation program for college sports track and field training, which can provide reference for promoting the development of sports training field in a more scientific and intelligent direction.

II. Factor Analysis

In this paper, factor analysis will be used as a method for analyzing multi-source data in college sports athletics [23].

II. A. Factor analysis model

$$\begin{cases} x_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1k}F_k + \varepsilon_1 \\ x_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2k}F_k + \varepsilon_2 \\ \vdots \\ x_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pk}F_k + \varepsilon_p \end{cases} \tag{1}$$

in matrix shorthand:

$$X = AF + \varepsilon \tag{2}$$

where $X = (x_1, x_2, \cdots, x_p)^T$ is an observable p -dimensional stochastic vector; and $F = (F_1, F_2, \cdots, F_k)^T (k < p)$ is an unobservable vector. Call $F = (F_j)^T (j = 1, 2, \cdots, k)$ a common factor of X ; $A = (a_{ij})$ refers to the factor loading matrix, where a_i is the loading of the $i(i = 1, 2, \cdots, p)$ vector on the $j(j = 1, 2, \cdots, k)$ common factor; and $\varepsilon = (\varepsilon_i)^T (i = 1, 2, \cdots, p)$ is a special factor, a factor not explained by the common factors in $X = (x_i)^T (i = 1, 2, \cdots, p)$, which acts only on X.

Let the k -dimensional unobservable vector $F \sim N(0, R)$, $R > 0$ be the correlation coefficient matrix of F; the p -dimensional special factor $\varepsilon \sim N(0, \Psi^2)$, $\Psi^2 = diag(\psi_1^2, \psi_2^2, \cdots, \psi_p^2)$ is the covariance matrix of ε; F is independent of ε, then model (1) is said to be a diagonal cross-factor model.

Assumptions:

$$\begin{cases} E(F)=0 & Cov(F)=E(FF^T)=I \\ E(\varepsilon)=0 & Cov(\varepsilon)=E(\varepsilon\varepsilon^T)=\Psi^2=diag(\psi_1^2,\psi_2^2,\cdots,\psi_p^2) \\ Cov(\varepsilon,F)=0 \end{cases} \tag{3}$$

Without loss of generality, the centered orthogonal factor model, assuming that the original data matrix was centered, is:

$$X-\mu=AF+\varepsilon \tag{4}$$

$\mu=(\mu_1,\mu_2,\cdots,\mu_p)^T$ is the mean vector of $X=(x_1,x_2,\cdots,x_p)^T$.

Obviously, the orthogonal factor analysis model is a special case of the oblique factor analysis model when the correlation coefficient matrix $R=I$ of the common factor F. Since the oblique factor model poses additional estimation difficulties and is highly arbitrary in the study, the following studies in this paper are for the orthogonal factor model. In order to avoid the influence of the quantiles on the factor analysis, the raw data are generally standardized.

II. B.Meaning of parameters in the factor analysis model

Assuming that the covariance array of the original data matrix X in the centered orthogonal factor model is Σ, the decomposition of Σ is as follows:

$$\begin{aligned} \Sigma &= Cov(X)=E[(X-\mu)(X-\mu)^T] \\ &= E[(AF+\varepsilon)(AF+\varepsilon)^T] \\ &= E(AFF^TA^T+AF\varepsilon^T+\varepsilon F^TA^T+\varepsilon\varepsilon^T) \\ &= ACov(F)A^T+Cov(\varepsilon) \\ &= AA^T+\Psi^2 \end{aligned} \tag{5}$$

Weighing in:

$$h_i^2=\sum_{j=1}^{k}a_{ij}^2, i=1,2,\cdots,p \tag{6}$$

is a common degree of x_i, then:

$$x_i=\sum_{j=1}^{k}a_{ij}F_j+\varepsilon_i, i=1,2,\cdots,p \tag{7}$$

Find the variance for each side of equation (7), i.e:

$$\begin{aligned} Var(x_i) &= \sum_{j=1}^{k}\operatorname{var}(a_{ij}F_j)+\operatorname{var}(\varepsilon_i) \\ &= \sum_{j=1}^{k}a_{ij}^2\operatorname{var}(F_j)+\psi_i^2 \\ &= \sum_{j=1}^{k}a_{ij}^2+\psi_i^2 \\ &= h_i^2+\psi_i^2 \end{aligned} \tag{8}$$

As can be seen from equation (8), the variance of x_i consists of two parts: firstly, the variable commonality h_i^2, which represents the contribution of all common factors to the overall variance of the variable x_i, and the closer h_i^2 is to 1, it indicates that all the original information of x_i is almost entirely explained by the common factor; followed by the special factor variance ψ_i^2, which is only related to the change of the variable x_i, the smaller the value of the special factor variance, the stronger the explanatory ability of the factor analysis model, the more accurate the estimation, and the better the mapping of the space of original variables to the space of public factors.

From equations (6) and (8), it can be seen that factor analysis can be used to study and approximate the covariance of the sample data based on the estimation of the factor loading matrix A and the special factor covariance Ψ^2. Since the covariance matrix Σ after standardizing the variable X is equal in value to the original correlation coefficient matrix R, i.e., equation (6) can be written as:

$$R = AA^T + \Psi^2 \tag{9}$$

Order:

$$R^* = AA^T \tag{10}$$

Then it is claimed:

$$R^* = AA^T = R - \Psi^2 \tag{11}$$

is the approximate correlation coefficient matrix, which is identical to the original correlation coefficient matrix R except that the elements on the diagonal are different.

According to $R^* = AA^T$, the correlation coefficients x_i and x_j of r_{ij} can be expressed as:

$$r_{ij} = \sum_{m=1}^{k} a_{im} a_{jm}, \quad i, j = 1, 2, \cdots, p \tag{12}$$

From equation (12), r_{ij} is the intra-row product corresponding to the factor loading matrix, indicating that the factor loading array A contains a large amount of information about the variables.

Weighing:

$$V_j^2 = \sum_{i=1}^{p} a_{ij}^2, \quad j = 1, 2, \cdots, k \tag{13}$$

is the variance contribution of the common factor F_j to x_i, which refers to the sum of the variance contributions given by the same common factor F_j to each of the variables, and is a very useful metric for weighing the common factors.

II. C.Estimation of factor loading matrices

The estimation method of the factor loading matrix used in this section is the Alpha factor method [24]. As with the iterative principal factor method, the common degree $h_i^2 (i = 1, 2, \cdots, p)$ is not known at the beginning, so the iterative computation procedure of the Alpha factor method can be shown as follows:

Step1, give an initial value of the common degree matrix $H^2(i,i) = h_i^2$:

$$H^2(i,i) = R(i,i) - \frac{1}{R^{-1}(i,i)} = 1 - \frac{1}{R^{-1}(i,i)} \tag{14}$$

where $R(i,i)$ denotes the $i(i = 1, 2, \cdots, p)$ th diagonal element of the correlation coefficient matrix R, $R^{-1}(i,i)$ denotes the $i(i = 1, 2, \cdots, p)$ th diagonal element of the inverse matrix R^{-1} of the correlation coefficient matrix, and the other methods of giving initial values of the commonality degree can be referred to the method in the iterative principal factorization method.

Step2, calculate the matrix:

$$U^2 = I - H^2 \tag{15}$$

where I is the unit matrix;

Step3, Let:

$$B = H^{-1}(R - U^2)H^{-1} \tag{16}$$

Calculate the eigenvalues of matrix B:

$$\Lambda = (\lambda_1, \lambda_2, \cdots, \lambda_p) \tag{17}$$

and the corresponding eigenvector V, the first k eigenvalues are selected according to the number k of common factors:

$$\Lambda_k = (\lambda_1, \lambda_2, \cdots, \lambda_k) \tag{18}$$

and the corresponding k eigenvectors V_k.

Step4, calculate the factor loading matrix:

$$\tilde{A} = HV_k \Lambda_k^{1/2} \tag{19}$$

Step5, compute the common degree of the factor loading matrix $\tilde{A}$, resulting in $\tilde{H}^2$.

Step6, replace $\tilde{H}^2$ with H^2 in step2 and continue iterating until the change in common degree is extremely small to stop the iteration.

The number of factors k in the alpha factorization method is chosen based on the number of positive values of α, such that $\sum_{i=1}^{p}\sum_{j=1}^{p} C_i$ represents the variance of the factor scores as a whole, f_k refers to the k th factor, and A_k is the weight vector, then:

$$f_k = XA_k \tag{20}$$

The Cronbach's alpha coefficient for Factor f_k is:

$$\alpha_k = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p} C_{ii}}{\sum_{i=1}^{p}\sum_{j=1}^{p} C_{ij}}\right) \tag{21}$$

The molecule in Eq:

$$\sum_{i=1}^{p} C_{ij} = \sum_{i=1}^{p} A_{ki}^2 \tag{22}$$

The denominator $\sum_{i=1}^{p}\sum_{j=1}^{p} C_i$ is the overall variance of the factors, and the standardized factor overall variance is 1, so equation (21) can be reduced to:

$$\alpha_k = \frac{p}{p-1}\left(1 - \sum_{i=1}^{p} A_{ki}^2\right) \tag{23}$$

In the principal component method, based on the eigenvalues Λ and the corresponding eigenvectors V, there are:

$$A = V\Lambda^{-1/2} \tag{24}$$

So:

$$\sum_{i=1}^{p} A_{ki}^2 = \lambda_k^{-1} \tag{25}$$

Then Eq. (23) can be reduced to:

$$\alpha_k = \frac{p}{p-1}\left(1 - \frac{1}{\lambda_k}\right) \tag{26}$$

The advantages of the Alpha Factor method are that the estimated factor loading matrices work best when both the sample size and the variables are small, and that it is also possible to generalize from one sample of a variable to the full set of the variable.

III. Multi-source data analysis of college sports track and field

Through the questionnaire survey to collect the track and field training data of S colleges and universities, we analyze the multivariate data of track and field training by using the factor analysis method proposed above, to provide theoretical support and practical guidance for the transformation of data in the field of track and field training, to broaden the idea of training, to update the training methods, to promote the field of track and field training towards a more scientific and smarter direction, and to provide a more accurate and personalized support and guidance for the growth and development of athletes. It also provides more precise and personalized support and guidance for athletes' growth and development.

III. A. Reliability test

In order to facilitate the statistical analysis of the questionnaire data, the questionnaire was initially divided into five groups, and the standardized Cronbach's α coefficient of the questionnaire was calculated to ensure the reliability of the questionnaire data. The results are shown in Table 1, the standardized Cronbach's α coefficient of the questionnaire is 0.986, which indicates that the questions of the questionnaire have a high degree of consistency and reliability in evaluating the same concepts or themes, and the questionnaire has a good stability and credibility, and can effectively assess the respondents' views and attitudes towards the method of evaluating and improving the quality of track and field training based on the data analysis.

Table 1: Results of reliability test

Cronbach 's α coefficient	Standardized Cronbach 's α coefficient	Number of items	Number of samples
0.986	0.986	5	255

III. B. Analysis of descriptive statistics results

Descriptive statistics can give a general idea of the respondents' evaluation of the quality of track and field training, as shown in Table 2. The minimum value of the five groups of data is 15, the maximum value is 25, and the median is 20, so the range of data distribution is relatively consistent. The mean values of the AA, BB, CC, DD and EE groups were 23.355, 23.372, 23.484, 23.845, 23.824 respectively, and the data of each group were close to the median, indicating that the respondents' evaluation of the quality of track and field training was higher but with some variability. The standard deviation is an index to measure the degree of dispersion of data distribution, from the table we can see that there are some differences in the degree of dispersion of data distribution, which lays the foundation for further in-depth analysis.

Table 2: Results of descriptive statistics

Group	Maximum value	Minimum value	Mean value	standard deviation	Median	Variance
AA	25	15	23.355	2.158	20	4.491
BB	25	15	23.372	2.183	20	4.666
CC	25	15	23.484	2.02	20	4.119
DD	25	15	23.845	1.884	20	3.779
EE	25	15	23.824	2.047	20	3.961

III. C. Principal component analysis results

Factor analysis can be conducted to better understand the underlying structure and relationship behind the data, so as to analyze the key factors affecting the quality of track and field training in a more in-depth manner, and to provide a basis for the development of effective improvement strategies. Before factor analysis, KMO test and Bartlett's test of sphericity were performed, as shown in Table 3. KMO test was used to assess the correlation between variables, the closer the value was to 1, the stronger the correlation between variables, which was suitable for factor analysis, and Bartlett's test of sphericity was used to test whether there was correlation between variables, $P<0.05$ indicated that there was significant correlation between variables, which was suitable for factor analysis. Bartlett's test of sphericity is also used to test whether there is correlation between the variables, $P<0.05$ indicates that there is significant correlation between the variables, which is suitable for factor analysis.

From the table, we can see that the KMO value is 0.861, which is close to 1, indicating that there is a strong correlation between the variables, and it is suitable for factor analysis.Bartlett's test of sphericity has a p-value of 0.001, which shows significance at the 1% level, and it rejects the original hypothesis, indicating that there is indeed a correlation between the variables, and the degree of factor analysis is effective and suitable.

Table 3: Results of validity analysis

KMO test and Bartlett sphericity test		
KMO value		0.861
Bartlett sphericity test	Approximate chi-square	4432.055
	df	180.000
	P	0.001***

The total variance explained results are specifically shown in Table 4. As can be seen from the table, three principal components were obtained after principal component analysis, i.e., training program design, technical guidance, and nutritional security, and their eigenroots were 7.48, 6.701, and 2.703, respectively, with the variance explained of 37.4%, 33.505%, and 13.515%, respectively, and the cumulative variance explained of 84.42%, which means that these three principal components could better summarize the The variance in the evaluation of the quality of training in each sport and track and field involved in the questionnaire and in the raw data better reflects the concerns of the respondents.

Table 4: Total variance interpretation

Ingredients	Latent root			Variance interpretation rate after rotation		
	Characteristic root	Percentage of variance explanation rate(%)	Cumulative percentage(%)	Characteristic root	Percentage of variance explanation rate(%)	Cumulative percentage(%)
1	14.184	70.92	70.92	7.48	37.4	37.4
2	1.818	9.09	80.01	6.701	33.505	70.905
3	0.882	4.41	84.42	2.703	13.515	84.42
4	0.507	2.535	86.955	-	-	-
5	0.406	2.03	88.985	-	-	-
6	0.308	1.54	90.525	-	-	-
7	0.307	1.535	92.06	-	-	-
8	0.222	1.11	93.17	-	-	-
9	0.195	0.975	94.145	-	-	-
10	0.191	0.955	95.1	-	-	-
11	0.153	0.765	95.865	-	-	-
12	0.145	0.725	96.59	-	-	-
13	0.137	0.685	97.275	-	-	-
14	0.132	0.66	97.935	-	-	-
15	0.123	0.615	98.55	-	-	-
16	0.116	0.58	99.13	-	-	-
17	0.073	0.365	99.495	-	-	-
18	0.063	0.315	99.81	-	-	-
19	0.035	0.175	99.985	-	-	-
20	0.003	0.015	100	-	-	-

IV. Intelligent Modeling Techniques for Sports Athletics Functional Data

In this chapter, we will propose intelligent modeling techniques for sports track and field data based on time series theory to achieve smooth series modeling of sports track and field training in colleges and universities.

IV. A. AR model

A model with the following structure is called a p-order autoregressive model, abbreviated as AR(p):

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \\ \phi_p \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ Ex_s \varepsilon_t = 0, \forall s < t \end{cases} \tag{27}$$

The AR (p) model has three constraints:

Condition 1: This restriction ensures that the highest order of the model is (p);

Condition II: This restriction effectively requires that the sequence of random disturbances $\{\varepsilon_i\}$ be a zero-mean white noise sequence $\varepsilon_t \sim WN(0,\sigma_\varepsilon^2)$;

Condition 3: This restriction states that random disturbances in the current period are independent of past sequence values. It is common to default to the restriction of the default equation (27), shorthanding the AR (p) model as:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \tag{28}$$

When $\varphi_0 = 0$, the autoregressive model (27) is also known as a centered AR (p) model. Any non-centralized AR (p) model can be transformed into a centralized AR (p) model by transformation. Introducing the delay operator, the centralized AR (p) model can again be abbreviated as:

$$\Phi(B)x_t = \varepsilon_t \tag{29}$$

where $\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$, is known as the p-order autoregressive coefficient polynomial.

The recursive formula for the autocorrelation coefficient of the smooth AR model is:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \tag{30}$$

According to Eq. (30), it is easy to see that the expression for the autocorrelation coefficient of the AR (p) model is in fact a chi-squared difference equation of order p. Then the generalized solution of the autocorrelation coefficient with lag of any k order is:

$$\rho_k = \sum_{i=1}^{p} c_i \lambda_i^k \tag{31}$$

where $|\lambda_i| < 1 (i = 1, \cdots, p)$ is the characteristic root of this difference equation; $c_1, \cdots, c_p$ are arbitrary constants that are not all zero.

With this generalized form, it is easy to introduce that ρ_k always has non-zero values and will not be constant and equal to zero after k is greater than some constant, a property known as the trailing property.

For a smooth AR (p) model:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \tag{32}$$

IV. B. MA model

A model with the following structure is called a q-order moving average model, abbreviated as MA (q):

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \end{cases} \tag{33}$$

The MA (q) model has two constraints:

Condition I: This restriction ensures that the highest order of the model is (q);

Condition II: This restriction requires that the random disturbance sequence $\{\varepsilon_t\}$ be a zero-mean white noise sequence.

Denote the model shorthand as:

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{34}$$

When $\mu = 0$, model (33) is called a centered MA (q) model. Any non-centralized MA (q) model can be transformed into a centralized MA (q) model by a simple displacement. Introducing the delay operator, the centralized MA (q) model can again be abbreviated as:

$$x_t = \Theta(B)\varepsilon_t \tag{35}$$

where $\Theta(B)=1-\theta_l B-\cdots-\theta_q B^q$, is called the q th order moving average coefficient polynomial.

The autocorrelation coefficients are q -order truncated:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & k=0 \\ -\theta_k + \sum_{i=1}^{q-k}\theta_i\theta_{k+1}, & 1\le k\le q. \\ 1+\theta_1^2+\cdots+\theta_q^2 \end{cases} \tag{36}$$

Lag k biased autocorrelation coefficient:

$$\phi_{kk} = (-\theta_t\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q})(-\theta_1\varepsilon_{t-k-1}-\cdots-\theta_q\varepsilon_{t-k-q+1}) \tag{37}$$

In the case of the MA (q) model, $x_{t-1},\cdots,x_{t-k+1}$ is given, which is equivalent to $\varepsilon_{t-1},\cdots,\varepsilon_{t-k-q+1}$ given, the model lags k white correlation coefficient ϕ_k is $\varepsilon_{t-1},\cdots,\varepsilon_{t-k-q+1}$, $\varepsilon_{t-1},\cdots,\varepsilon_{t-k-q+1}$ is not always equal to zero, $\theta_1,\cdots,\theta_q$ is again arbitrarily given, so ϕ_{ik} It is not possible to be constant zero after a finite order, i.e., the partial autocorrelation coefficient of the MA (q) model is tailed.

IV. C. ARMA model

A model with the following structure is called an autoregressive moving average model (ARMA model), abbreviated as ARMA (p,q) [25]:

$$\begin{cases} x_t = \phi_0+\phi_1 x_{t-1}+\cdots+\phi_p x_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q} \\ \phi_p\ne 0, \theta_q\ne 0 \\ E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s)=0, s\ne t \\ Ex_s\varepsilon_t=0, \forall s<t \end{cases} \tag{38}$$

When $\varphi_0=0$, the model is called a centered ARMA (p,q) model. By default default conditions, the centered ARMA (p,q) model can be abbreviated as:

$$x_t = \phi_1 x_{t-1}+\cdots+\phi_p x_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q} \tag{39}$$

Introducing the delay operator, the ARMA (p,q) model is abbreviated as:

$$\Phi(B)x_t = \Theta(B)\varepsilon_t \tag{40}$$

where $\Phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$, is a p -order autoregressive coefficient polynomial;

$\Theta(B)=1-\theta_1 B-\cdots-\theta_q B^q$, a q th-order moving average coefficient polynomial, and $\Phi(B),\Theta(B)$ mutually prime.

When the roots of $\Phi(B)=0$ and $\Theta(B)=0$ are outside the unit circle, the ARMA (p,q) model is called a smooth reversible model.

Autocorrelation coefficient:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2} \tag{41}$$

Among them:

$$\begin{cases} G_0 = 1 \\ G_k = \sum_{j=1}^{k}\phi_j' G_{k-j} - \theta_{j'}, k\ge 1 \end{cases}, \phi_j' = \begin{cases} \phi_j, 1\le j\le q \\ 0, j>q \end{cases}, \theta_j' = \begin{cases} \theta_j, 1\le j\le q \\ 0, j>q \end{cases} \tag{42}$$

Based on the expression of autocorrelation coefficient it is easy to judge that the autocorrelation coefficient of ARMA (p,q) model is not truncated, which is consistent with the property that ARMA (p,q) model can be transformed into an infinite-order moving average model. Similarly, based on the fact that the ARMA (p,q) model can be transformed into an infinite-order autoregressive model, it can be judged that its partial correlation coefficient is also not truncated.

IV. D. Smooth Sequence Modeling for Sports Athletic Training

If a certain sequence of observations can be determined to be a smooth non-white noise sequence by sequence preprocessing, the model can be used to model the sequence. In this paper, the basic steps for modeling the smooth sequence of sports track and field training are:

1) Find the values of the sample autocorrelation coefficient (ACF) and the sample partial autocorrelation coefficient (PACF) for the sequence of observations.

Sample autocorrelation coefficient:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{n}(x_i - \overline{x})^2}, \forall 0 < k < n \tag{43}$$

Sample biased autocorrelation coefficient:

$$\hat{\phi}_{kk} = \frac{\hat{D}_k}{\hat{D}}, \forall 0 < k < n \tag{44}$$

Among them:

$$\hat{D} = \begin{vmatrix} 1 & r_1 & \cdots & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-2} \\ \vdots & \vdots & & \vdots \\ r_{k-1} & r_{k-2} & \cdots & 1 \end{vmatrix}, \hat{D}_k = \begin{vmatrix} 1 & r_1 & \cdots & r_1 \\ r_1 & 1 & \cdots & r_2 \\ \vdots & \vdots & & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_k \end{vmatrix} \tag{45}$$

2) Based on the nature of the sample autocorrelation coefficients and partial autocorrelation coefficients, an ARMA (p,q) model of appropriate order is selected for fitting.

3) Test the validity of the model. If the fitted model fails the test, turn to step 2 and re-select the model to be fitted again.

A good fitting model should be able to extract almost all the sample-related information in the sequence of observations, in other words, the fitting residual term will no longer contain any relevant information, that is, the residual sequence should be white noise sequence, so the significance of the model test that is the residual sequence of white noise test. The original and alternative hypotheses are:

$$\begin{aligned} &H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0, \forall m \geq 1 \\ &H_1 : \text{At least there exists a certain } \rho_k \neq 0, \forall m \geq 1, k \leq m \end{aligned} \tag{46}$$

The test statistic is the LB test statistic:

$$LB = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \overset{appr.}{\sim} \chi^2(m-p-q-1), \forall m > 0 \tag{47}$$

If the original hypothesis is rejected, the fitted model is not significant; if the original hypothesis cannot be rejected, the fitted model is significantly valid.

4) Model optimization.

When a fitted model passes the test, it means that the model can effectively fit the fluctuation of the observation series under a certain confidence level, but this effective model is not unique. To solve this problem, the AIC information criterion is introduced for model optimization.

The AIC criterion is defined as:

$$\text{AIC=-2ln(the maximum likelihood function value of the model)} \\ \text{+2(the number of unknown parameters in the model)} \tag{48}$$

The AIC function for the centered ARMA (p,q) model is:

$$AIC = n\ln(\hat{\sigma}_{\varepsilon}^{2}) + 2(p+q+1) \tag{49}$$

The AIC function for the decentralized ARMA (p,q) model is:

$$AIC = n\ln(\hat{\sigma}_{\varepsilon}^{2}) + 2(p+q+2) \tag{50}$$

To improve the AIC criterion, the BIC criterion is proposed.Bayes theory also yields the same discriminant criterion called the SBC criterion.The SBC criterion is defined as:

$$\text{SBC=-2ln(the maximum likelihood function value of the model)} \\ \text{+1n(n)(the number of unknown parameters in the model} \tag{51}$$

It has been shown theoretically that the SBC criterion is a conjugate estimate of the true order of the optimal model.

The SBC function for the centered ARMA (p,q) model is:

$$SBC = n\ln(\hat{\sigma}_{\varepsilon}^{2}) + \ln(n)(p+q+1) \tag{52}$$

The SBC function for the decentered ARMA (p,q) model is:

$$SBC = n\ln(\hat{\sigma}_{\varepsilon}^{2}) + \ln(n)(p+q+2) \tag{53}$$

The model that minimizes the AIC or SBC function among all the models that pass the test is the relatively optimal model.

5) Using the fitted model, predict the future movement of the series.

The l-step predicted values are:

$$\hat{x}_t(l) = \begin{cases} \phi_1\hat{x}_t(l-1)+\cdots+\phi_p\hat{x}_t(l-p)-\sum_{i=l}^{q}\theta_1\varepsilon_{t+l-i}, & l \le q \\ \phi_1\hat{x}_t(l-1)+\cdots+\phi_p\hat{x}_t(l-p), & l > q \end{cases} \tag{54}$$

Among them:

$$\hat{x}_t(k) = \begin{cases} \hat{x}_t(k), k \ge 1 \\ x_{t+k}, k \le 0 \end{cases} \tag{55}$$

The variance of the l-step prediction error is:

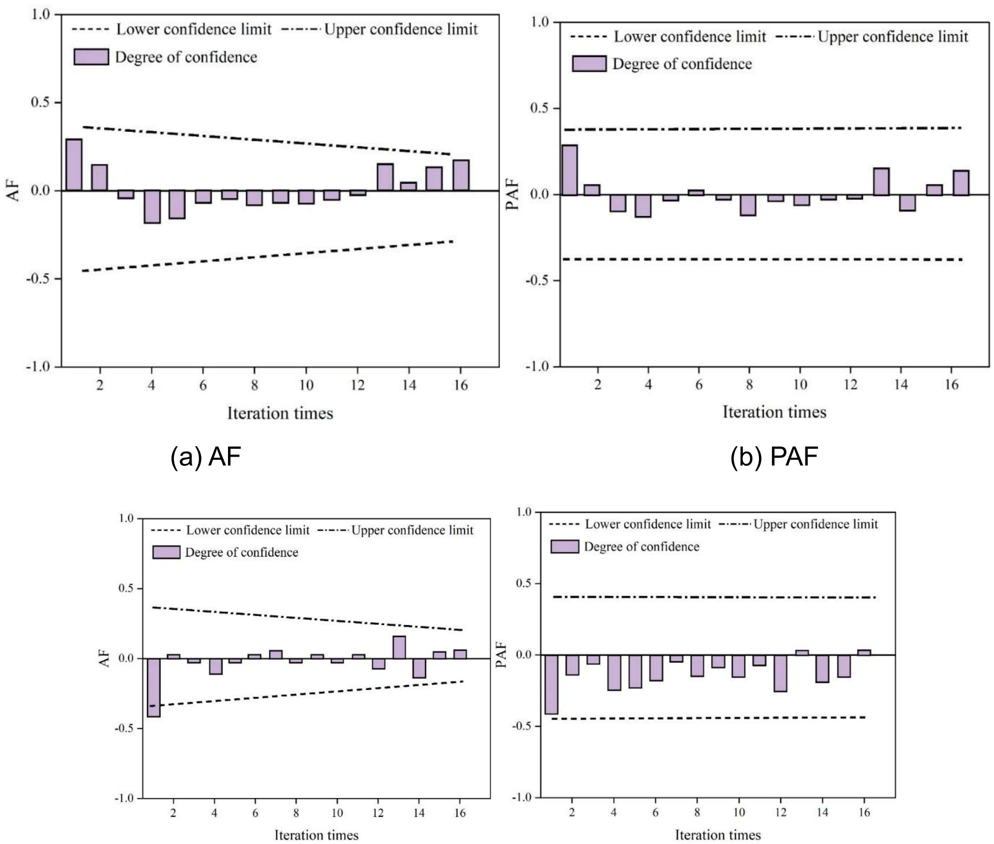
$$Var[e_t(l)] = (1+G_1^2+\cdots+G_{l-1}^2)\sigma_{\varepsilon}^2 \tag{56}$$

V. Application Analysis of Intelligent Modeling Technology for Sports Athletics Training

This chapter will apply the intelligent modeling technology of sports track and field data proposed above to the sports track and field training in colleges and universities to realize the intelligent monitoring and prediction of the functional changes in track and field training, and provide data support for the customization of personalized training programs for students' sports track and field training.

V. A. Smoothness judgment analysis

The study randomly selected a player Z from the track and field team of S university as the research subject, and the corresponding hemoglobin (HGB) changes of Z from May 2022 to October 2024 were selected in order to perform the smoothness judgment of the functional TS model. In order to verify whether the Z∗HGB level is a smooth time series, the study conducts a smoothness test of the TS at the HGB level by the judging principle of AF and PAF. The results before and after the 1st order differencing of Z∗HGB can be obtained, as shown in Fig. 1, Figs. (a) and (b) correspond to the AF and PAF, respectively, and Figs. (c) and (d) correspond to the AF and PAF after the 1st order differencing, respectively. According to the figure, it can be seen that the AF of Z∗HGB lags for 2 periods and then gradually converges to 0, and the AF and PAF are in the confidence intervals, therefore, it can be determined that the level of Z∗HGB is a kind of stochastic smooth time series. After the TS of Z∗HGB is differenced by order 1, the corresponding AF and PAF do not change significantly, which indicates that the time series of Z∗ HGB does not have a trend change.



(a) AF (b) PAF

(c) AF after 1-order difference (d) PAF after 1-order difference

Figure 1: The results before and after the first-order difference of Z∗HGB

In order to further make an accurate determination of the smoothness of the Z∗HGB level time series, the study uses the unit root test for additional validation, as shown in Table 5. As can be seen from the table, the value of the test t-statistic is -3.90368, which is smaller than the critical value at the 5% significance level, which indicates that the series is in a smooth state and can be used for modeling. Since the AM(a) model is a linear equation estimation, it is easier to obtain the parameter estimation results compared with the other two models. After comprehensive consideration, there are three models of AM (1), AMA(1,1) and AMA(1,2) for motor function TS to choose. Eviews software was used to estimate the parameters of the above models, and AMA(1,1) was selected as the model for TS at HGB level by considering the overall fitting effect of the model, the coefficient of determination, the SC criterion and the AC criterion.

Table 5: ADF test

-		T statistic	P
ADF test		-3.90368	0.0253
ADF critical value	1%	-3.61448	-
	5%	-2.86994	-
	10%	-2.55697	-

V. B. Short-term forecast analysis

After determining the AMA(1,1) model, short-term predictions of HGB levels can be made. The corresponding model fitting effect and residual variation results are specifically shown in Figure 2. Where the true value of HGB is derived from the actual measurement of the athletes by the monitoring system. It can be observed from the figure that the predicted values of the AMA model fit the real values very well, and the average real value of HGB was 148 g/L, and the average predicted value was 150 g/L, which indicated that the model could accurately reflect the changes of HGB levels. In addition, the residual variations of the model were small in [-1,1], which indicated that the AMA(1,1) model was fitted very well.

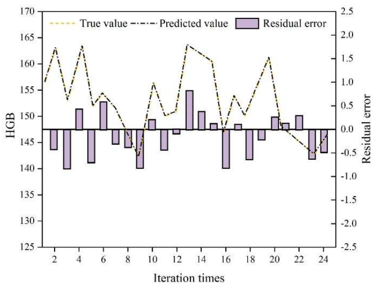


Figure 2: Fitting effect and residual change results of AMA (1,1) model

V. C. Analysis of application effects

In order to verify the effect of the proposed intelligent modeling technology of sports track and field data in the practical application of sports track and field training, the training data of team member Z in August and October 2024 for a period of 20 days in the plateau area were selected for the study. The basal heart rate (BHR) and blood oxygen saturation (BOS) were selected for the study, and the monitoring and changing results of BHR and BOS of team member Z during the training process can be obtained, which are shown in Fig. 3. As can be seen from the figure, in general, the BHR of Z reached the highest level on the 4th day after arriving at the plateau for training, with an average of 54.01 beats/minute. As the number of training days increased, the BHR would show a sharp decline, and after the 12th day, the BHR would show a relatively stable trend.The trend of BOS was opposite to that of BHR, and the BOS was the lowest on the 4th day of training, with an average of 92.55%, and then it would show an increasing trend, but in general, it was less than 96%. Based on the real-time changes in Z's physical function, the coach can make adjustments to his training and develop a personalized training plan for future training.

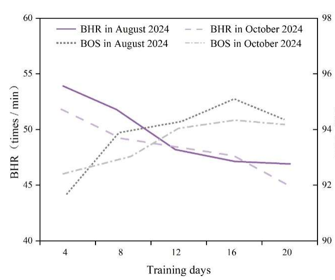


Figure 3: Monitoring changes in BHR and BOS results

VI. Conclusion

The application of multi-source data analysis and intelligent modeling technology in college sports track and field training is effective. The questionnaire reliability test showed that the standardized Cronbach's α coefficient was 0.986, indicating that the questionnaire had high reliability; the KMO value was 0.861, and the P-value of Bartlett's sphericity test was 0.001, which made it suitable for factor analysis. Three main factors were extracted by principal component analysis: training program design, technical guidance, and nutritional protection, and their characteristic roots were 7.48, 6.701, and 2.703, respectively, with a cumulative variance explained rate of 84.42%, which better reflected the respondents' concerns. The time series analysis showed that the hemoglobin (HGB) level of track and field team member Z was a random smooth time series, and the ADF test statistic was -3.90368, which was less than the critical value of 5% significant level; the average predicted value of HGB based on the ARMA(1,1) model was 150g/L, which was very close to the true value of 148g/L, and the prediction accuracy was high. During plateau training, the intelligent monitoring system tracked the changes of basal heart rate (BHR) and blood oxygen saturation (BOS) of team member Z in real time, and found that on the 4th day, BHR reached the highest value of 54.01 beats/minute and BOS reached the lowest value of 92.55%. The application of multi-source data analysis and intelligent modeling technology for sports and athletics training in colleges and universities provides scientific evaluation and precise monitoring means, realizes the data management of the training process, provides a reliable basis for the formulation of personalized training programs, and effectively promotes the development of sports and athletics training in colleges and universities in the direction of scientification and intelligence.

Funding

This work was supported by:

1. Research and practice of "three knowledge" teaching project and teaching method in track and field teaching reform in Hunan Province in 2023 (HNJG-20231318).

2. 2024 Scientific Research Project of Hunan Provincial Department of Education and high-quality Development of Physical Education in Universities (24C0547).

References

[1] Wang, J. (2022). Influence of physical training on the physical quality of university students. Revista Brasileira de Medicina do Esporte, 29(spe1), e2022_0184.
[2] Wang, H. (2019, January). On the Value Exploration of Track and Field Teaching in Colleges and Universities Based on the Concept of" Process Evaluation". In 3rd International Seminar on Education Innovation and Economic Management (SEIEM 2018) (pp. 521-523). Atlantis Press.
[3] Peng, S. (2021). Research on the Coordinated Development of Sports and Education in Universities. International Journal of Social Science and Education Research, 4(10), 534-537.
[4] Olesov, N. P., Borisov, E. S., & Okoneshnikova, M. A. (2015). State and prospects of development of track and field athletics in Northeastern Federal University. Theory and Practice of Physical Culture, (10), 11-11.
[5] Wang, Y. (2022). COOPERATIVE EDUCATION STRATEGY OF TRACK AND FIELD TEACHING IN COLLEGES AND UNIVERSITIES AND ITS IMPACT ON COLLEGE STUDENTS'PSYCHOLOGICAL ANXIETY. Psychiatria Danubina, 34(suppl 1), 555-557.
[6] Yu, A. (2017, October). Research on fuzzy comprehensive evaluation of track and field project level in China. In 2017 International Conference on Robots & Intelligent System (ICRIS) (pp. 253-256). IEEE.
[7] Wang, B. (2020). Analysis on the Innovation of Track and Field Teaching Practice in Universities. Frontiers in Sport Research, 2(8).
[8] Lin, W. (2016). Human resources management of track and field web course in college physical education. International Journal of Emerging Technologies in Learning (Online), 11(4), 95.
[9] Wang, W., & Yin, G. (2021, April). Analysis and research on the application of internet technology in sports track and field teaching. In Journal of Physics: Conference Series (Vol. 1881, No. 4, p. 042026). IOP Publishing.
[10] Wang, W. (2022). Analysis of Teaching Tactics Characteristics of Track and Field Sports Training in Colleges and Universities Based on Deep Neural Network. Computational intelligence and neuroscience, 2022(1), 1932596.
[11] Kelley, C. P., Soboroff, S. D., Katayama, A. D., Pfeiffer, M., & Lovaglia, M. J. (2018). Institutional reforms and the recoupling of academic and athletic performance in high-profile college sports. Sport Journal, 21(1), 1-25.
[12] Lan, S. (2023). The reform of physical education in colleges and universities under the background of physical education integration. International Journal of New Developments in Education, 5(26), 11.
[13] Zimbalist, A. (2017). Reforming college sports and a constrained, conditional antitrust exemption. Managerial and Decision Economics, 38(5), 634-643.
[14] Gu, Q. (2022). Research on the Reform and Innovation of Physical Education Management in Colleges and Universities under the New Situation. Open Journal of Social Sciences, 10(12), 459-465.
[15] Shi, J. (2021). Sport action recognition by fusing multi‐source sensor information. Internet Technology Letters, 4(3), e279.
[16] Dong, X. (2021). Physical training information system of college sports based on big data mobile terminal. Mobile Information Systems, 2021(1), 4109794.
[17] Zhou, S. A. (2024, December). Multi-source data fusion analysis model for intelligent operation of stadiums. In Proceedings of the 2024 International Conference on Sports Technology and Performance Analysis (pp. 346-352).
[18] Dan, J., Zheng, Y., & Hu, J. (2022). Research on sports training model based on intelligent data aggregation processing in internet of things. Cluster Computing, 25(1), 727-734.
[19] Zhang, Y., Duan, W., Villanueva, L. E., & Chen, S. (2023). Transforming sports training through the integration of internet technology and artificial intelligence. Soft Computing, 27(20), 15409-15423.
[20] Nie, D., & Liu, J. (2022). Application of multiagent technology in intelligent distributed sports training simulation system. Wireless Communications and Mobile Computing, 2022(1), 8286371.
[21] Rajšp, A., & Fister Jr, I. (2020). A systematic literature review of intelligent data analysis methods for smart sport training. Applied Sciences, 10(9), 3013.
[22] Wei, S., Wang, K., & Li, X. (2022). Design and implementation of college sports training system based on artificial intelligence. International Journal of System Assurance Engineering and Management, 13(Suppl 3), 971-977.
[23] Chengxu Chen,Xiaoze Du,Lizhong Yang & Alessandro Romagnoli. (2025). Multi-objective optimization of cascaded packed bed thermal energy storage unit based on response surface and factor analysis methods. Applied Energy,386,125598-125598.
[24] Yin Jin,Zhang Boyu,Geng Xin,Yin Zimin & Lin Hongyan. (2023). Research on the Intelligent manufacturing Path of Equipment Manufacturing Enterprises Based on Factor Load Matrix Model. Procedia Computer Science,225,2942-2950.
[25] Xiong Yin. (2024). Linkage Analysis of Business Administration and Organisational Effectiveness of Enterprises Applying the ARMA Model. Applied Mathematics and Nonlinear Sciences,9(1).