

Analyzing Tennis Match Momentum: Factor Analysis of Player Performance in the 2023 Wimbledon Men's Singles Tournament

Shaohang Chen *, Jiacheng Chen, Jinzhi Li, Lei Li, Bo He

School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang, China, 050043

* Corresponding Author Email: chenshaohang1217@163.com

Abstract. This paper discusses a model developed for capturing the flow of play in tennis matches, particularly focusing on player performance and match momentum using factor analysis and data from the 2023 Wimbledon Tennis Men's Singles Tournament. The model involves categorizing data labels into serve, return, and score sessions, extracting key technical indicators, conducting factor analysis, and quantifying momentum scores. The analysis includes correlation tests, variance contribution rates, and factor score coefficient matrices to derive principal factors and calculate player performance scores. The momentum score expression and its application in analyzing match flow and player performance are also discussed, validated using match results from the tournament.

Keywords: Tennis Match Analysis, Factor Analysis, Match Momentum.

1. Introduction

In modern tennis, an in-depth understanding of player performance and match trends is crucial for developing tactical strategies and predicting match outcomes. However, it is not easy to accurately capture the dynamics of the match and the key factors of player performance. Therefore, this study aims to develop a model for quantifying player performance and match momentum based on data from the 2023 Wimbledon men's singles tournament using a factor analysis approach, thus providing a new perspective for strategic analysis of the sport of tennis [1].

In tennis, a player's performance and match momentum are affected by numerous factors, such as the performance of serve, return, and point. Existing studies are limited to analyses of individual factors and lack a comprehensive assessment of overall match trends and player performance. Therefore, the problem of this study is how to use factor analysis methods to extract key factors from multiple technical indicators and establish a model to quantify players' match performance and match momentum [2-3].

Firstly, by classifying and pre-processing the data of 2023 Wimbledon men's singles matches, the match process was divided into serving session, return session and scoring session. Then, the factor analysis method was used to extract the main factors from them and assess the weight proportion of these factors in the player's performance and match momentum. Eventually, a quantitative model of players' match performance and match momentum is established to provide a reliable tool and method for the strategic analysis of tennis matches.

2. Model for Capturing the Flow of Play

In this section, based on the existing data, we employ factor analysis to determine the principal factors and their respective weight proportions during different sessions of the match. Using these principal factors, we aim to quantify "momentum", thereby establishing a model to capture the match progression and player performance [4].

2.1. Data Extraction and Preprocessing

Based on the data provided for the 2023 Wimbledon Tennis Men's Singles Tournament after the first two rounds, we categorize the original data labels into three classes according to the match progression: the serve session, the return session, and the score session, as illustrated in Figure 1.

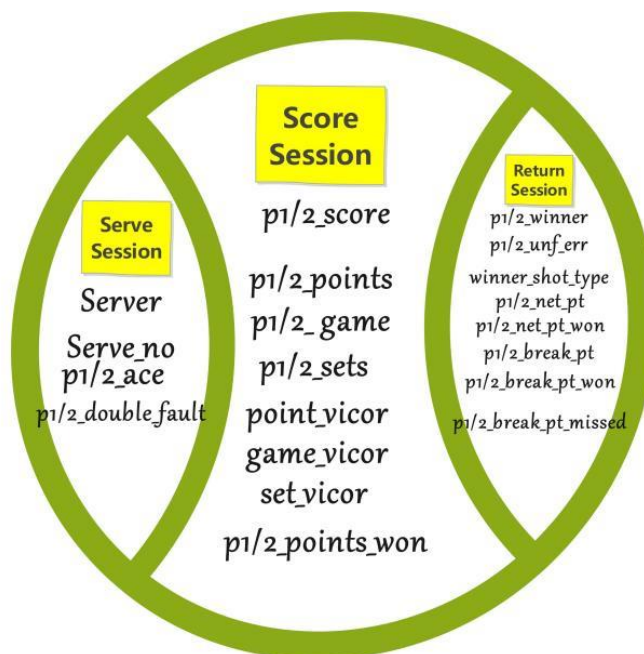


Figure 1: Labels of original data for the Serve (left), Return (middle) and Score (right) session.

2.2. Factor Analysis Model

2.2.1. Factors Extraction

Player performance is constrained by various factors [5], and through selection and extraction processing, the following technical indicators are as shown in Figure 2.

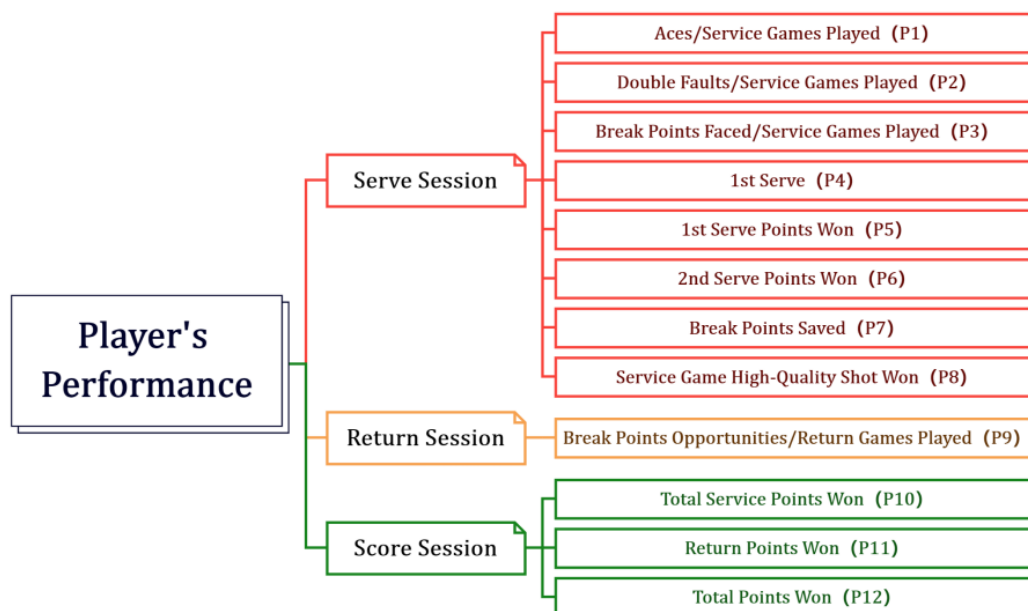


Figure 2: Player Performance Factor Breakdown.

The aforementioned 12 factor indicators are key elements in the evaluation of player performance, all of which are derived from the dataset calculations. The following will conduct a factor analysis on them.

2.2.2. Factor Analysis

We use SPSS to conduct factor analysis on the results calculated from the aforementioned technical indicators. Initially, we perform a KMO and Bartlett test to assess the correlation coefficients. This yields a heatmap of the correlation matrix for the factor indicators and the results of the KMO-Bartlett test as shown in Figure 3 and Table 1:

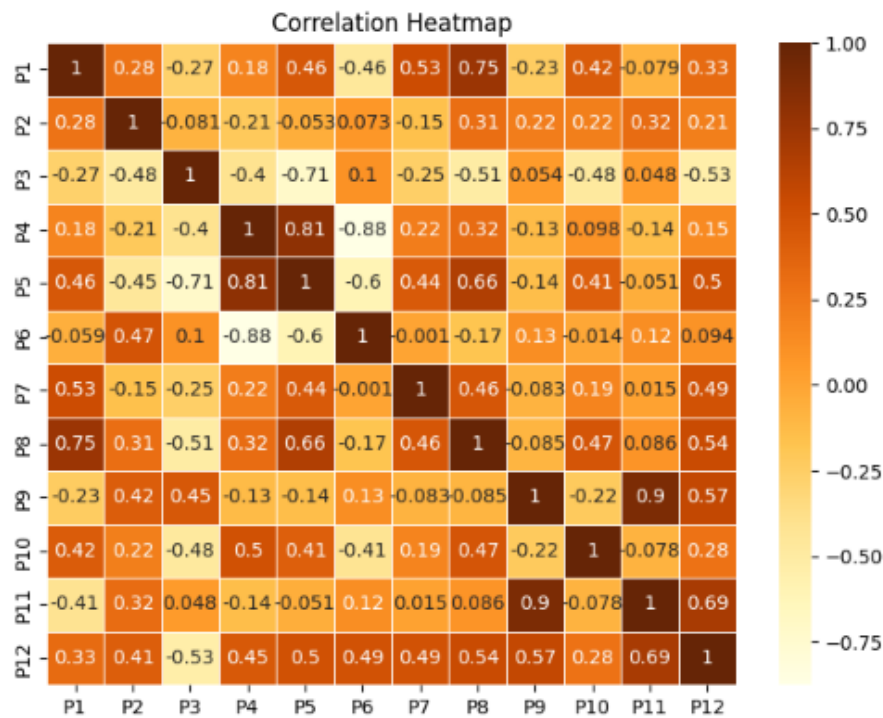


Figure 3: The Correlation Heatmap.

Table 1: KOM and Bartlett Sphericity Test Table.

KMO Test		0.737
Bartlett's Test of Sphericity	Approximate Chi-Square	341.578
	Degrees of Freedom	66
	Significance	0

Based on the correlation heatmap, it is evident that the majority of the factor indices have correlation coefficients above 0.4, indicating relevance. The KMO test yielded a value of 0.737, which is greater than 0.5[6-7], and the significance of the Bartlett's test of sphericity is 0, meaning that at a 95% confidence level, we reject the null hypothesis that the correlation coefficients are zero. In summary, there is a correlation between factor indices, and it is appropriate to proceed with factor analysis.

Table 2: Table of common factor variances.

Symbol	Starting Point	Extraction
P1	1.000	0.699
P2	1.000	0.825
P3	1.000	0.528
P4	1.000	0.964
P5	1.000	0.956
P6	1.000	0.892
P7	1.000	0.802
P8	1.000	0.800
P9	1.000	0.937
P10	1.000	0.571
P11	1.000	0.939
P12	1.000	0.959

As can be seen from table 2, the common factors account for a reaction degree of the variance for the majority of the factor indices reaching above 0.8, indicating the feasibility of representing these 12 indicators with common factors is high [8].

Table 3: Variance Devoting Rates.

Initial Eigenvalues				Sum of Squared Loadings for Rotation		
Component	Total	Percentage of Variance Contribution	Cumulative Percentage of Variance Contribution	Total	Percentage of Variance Contribution	Cumulative Percentage of Variance Contribution
1	4.214	35.119	35.119	3.506	29.215	29.215
2	2.702	22.518	57.637	2.558	21.317	50.532
3	1.804	15.030	72.667	2.525	21.041	71.573
4	1.151	9.592	82.259	1.282	10.686	82.259
5	0.934	7.781	90.039			
6	0.483	4.023	94.062			
7	0.321	2.676	96.738			
8	0.184	1.534	98.272			
9	0.101	0.839	99.111			
10	0.048	0.400	99.511			
11	0.042	0.349	99.860			
12	0.017	0.140	100.000			

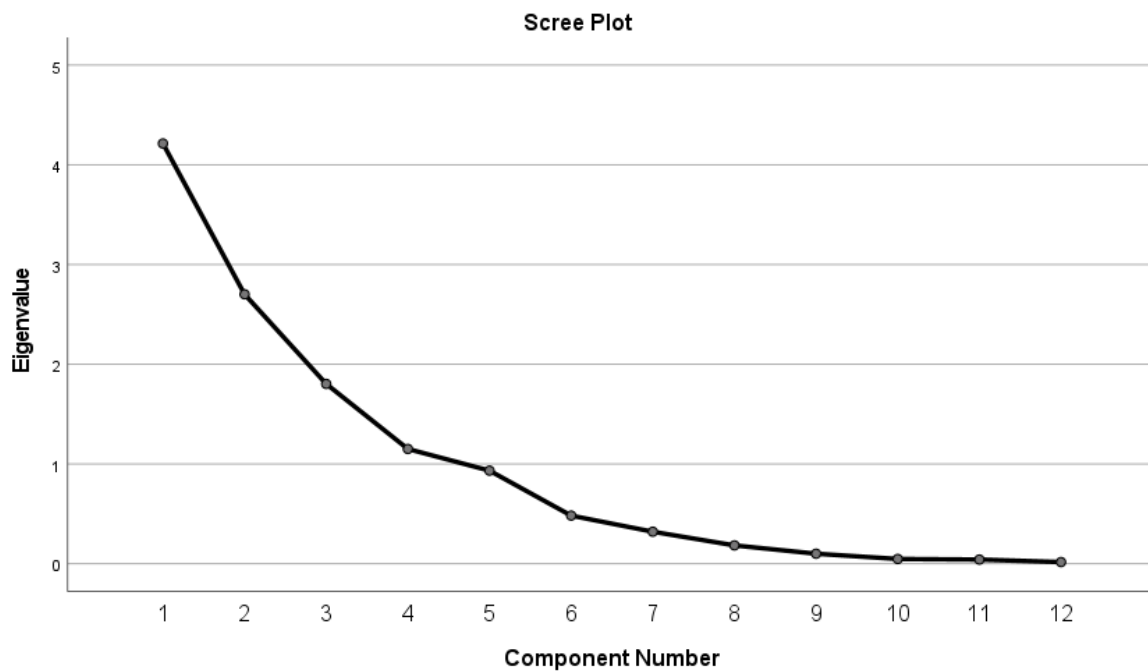


Figure 4: Scree Plot.

As shown in Table 3 and Figure 4, the 12 selected technical indicators show that the cumulative variance contribution rate of the first four factors reaches 82.259%, which is greater than 80%. Therefore, the first four factors are extracted as the principal factors, denoted as Z1, Z2, Z3, and Z4, respectively. The scree plot indicates that when the number of factors is greater than 4, the eigenvalues are less than 1, which also confirms that the number of principal components is 4. The cumulative variance of these four principal components accounts for 82.259% of the total variance, which significantly retains the information of the original indicators, thus reducing the original 12 technical indicators to 4 principal factor indicators [9-10].

In conclusion, we have derived the factor score coefficient matrix as shown in Table 4:

Table 4: Matrix of Component Score Coefficients.

Symbol	Factor			
	1	2	3	4
P1	0.276	-0.086	-0.115	0.008
P2	0.180	0.045	-0.031	-0.633
P3	-0.153	-0.024	-0.083	0.004
P4	-0.066	0.019	0.399	0.025
P5	0.096	0.028	0.247	0.063
P6	0.107	0.007	-0.439	0.184
P7	0.128	0.038	-0.138	0.559
P8	0.255	0.002	-0.006	-0.063
P9	-0.114	0.387	0.049	-0.021
P10	0.255	-0.102	-0.058	-0.217
P11	-0.049	0.380	0.016	-0.054
P12	0.110	0.295	-0.025	0.171

From the factor score coefficient matrix, the formula for calculating the player's match performance score can be deduced:

$$\begin{aligned}
 Z_1 &= 0.0275P_1 + 0.180P_2 + 0.153P_3 + \dots + 0.110P_{12} \\
 Z_2 &= -0.086P_1 + 0.045P_2 + (-0.024)P_3 + \dots + 0.295P_{12} \\
 Z_3 &= -0.115P_1 + (-0.031)P_2 + (-0.083)P_3 + \dots + (-0.025)P_{12} \\
 Z_4 &= 0.008P_1 + (-0.633)P_2 + 0.004P_3 + \dots + 0.171P_{12} \\
 Z_{\text{total}} &= \frac{(0.29215Z_1 + 0.21317Z_2 + 0.21041Z_3 + 0.10686Z_4)}{0.82259}
 \end{aligned} \tag{1}$$

According to the component matrix, the aforementioned principal factors are interpreted as follows: Z1 represents the serving game performance factor score, Z2 represents the receiving game performance factor score, Z3 represents the first serve error performance factor score, Z4 represents the player's serve stability factor score, and Z_ total represents the total score of influencing factors [1].

This model can effectively judge a player's scoring performance at different stages of a match.

2.3. Quantification of "Momentum"

Determination of the momentum score expression:

First, we use the principal factors determined through factor analysis to construct the expression of "momentum" for tennis athletes in different states:

$$M_t = f_t(\alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \alpha_4 Z_4) \tag{2}$$

Within the formula, Z1~ Z4 represent the evaluation factors for a player's in-game performance as derived from factor analysis, with $f_t()$ being the mapping function.

Then, in order to further ascertain the match state of a player at each time point during the match process, we convert the four principal factors into a scoring index for each moment through the mapping function. This index is differentiated into cumulative scores for the serving game, F1, and the receiving game, F2. Both of these variables can be calculated from the data indicators within the dataset.

Consequently, the expression for the momentum score Mt is defined as:

$$M = \begin{cases} \delta_1 \beta_1 F_1 \\ \delta_2 \beta_2 F_2 \end{cases} \tag{3}$$

In which δ_1 and δ_2 are the factor coefficients for the serving and receiving games respectively, and β_1 and β_2 represent the dynamic weighting proportions for the serving and receiving games. Given that the server's probability of winning is much higher, the serving game factor coefficient is

determined by the cumulative contribution ratio of the three main components Z1, Z3, and Z4, which are related to the serving game. The receiving game factor coefficient is determined by the contribution ratio of the main component Z2, which is related to the receiving game. The expressions are as follows, with the dynamic weighting proportions for the serving and receiving games set at $\beta_1 = 1.2$ and $\beta_2 = 0.8$, respectively. Since the probability of winning is much higher for the serving player, the dynamic weights at serve β_1 value is greater than the dynamic weight at serve receive β_2 value.

$$\delta_1 = \frac{\sum_{j=1}^4 \text{Percentage}_{z_i}}{\sum_{j=1}^4 \text{Percentage}_{z_j}} \quad (i = 1 \ 3 \ 4) \quad (4)$$

$$\delta_2 = \frac{\text{Percentage}_{Z_2}}{\sum_{j=1}^4 \text{Percentage}_{Z_j}} \quad (5)$$

Using Player1 as an example, the rules for determining the cumulative score F are shown in Figure 5:

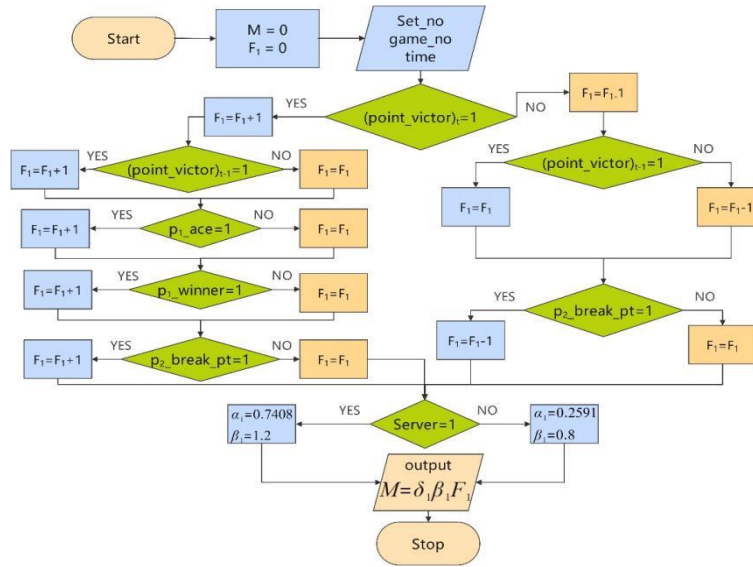


Figure 5: The calculation rules for "momentum".

2.4. Model Result Analysis

To validate the above model and better demonstrate the analysis of match flow and the correspondence with the actual match process, we selected the match results of the final between Carlos Alcaraz and Novak Djokovic as the basic dataset. We calculated the momentum score for each player at every moment and plotted the "momentum" time series graph for both players during that match, as shown in Figure 6:

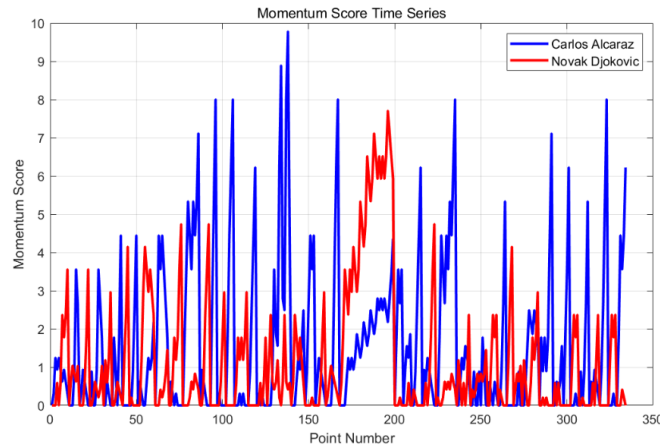


Figure 6: Momentum Score Time Series.

From the graph, it is evident that at the beginning of the match, there was not a significant difference in momentum between the two players. However, at certain moments, Novak Djokovic's momentum was noticeably higher than that of Carlos Alcaraz, leading to Djokovic winning the first set with a score of 6:1, where his momentum saw a significant increase. Subsequently, Alcaraz won with a score of 7:6, with his momentum score increasing and gradually surpassing that of Djokovic. Throughout the match, Carlos Alcaraz's momentum was generally higher than Novak Djokovic's, and the final result of Alcaraz winning the match also demonstrates that momentum has a strong correlation with winning to a certain extent.

3. Conclusions

In this study, based on the factor analysis method, this paper conducted an in-depth analysis of the data of the 2023 Wimbledon men's singles tournament and established a model to quantify players' match performance and match momentum. By factor analyzing the key technical indicators of serving, returning and scoring during the match, this paper successfully extracted the main factors and determined their weighting ratios in the players' performance and match momentum. By validating the model and analyzing the results of actual matches, we found that the model was able to effectively capture the dynamic changes in the course of the match and accurately predict the player's performance and match results.

The results of this study show that the use of factor analysis can effectively quantify player performance and match momentum in tennis matches, providing a new tactical analysis tool for coaches and players. Future research can further extend the data sources and factor analysis model to improve the predictive accuracy and applicability of the model, thus better guiding players' performance and tactical strategies in matches.

References

- [1] Ye Yi, Xu Xinhua, Chen Suna, et al. Research on academic performance analysis and prediction based on data mining [J]. Computer and Information Technology, 2024, 32 (02): 78-81.
- [2] Guanghai Zhang, Xuan Zu, Yang Wang. A three-classification prediction model for student achievement based on DBO-SVM [J]. Journal of Lanzhou College of Arts and Sciences (Natural Science Edition), 2024, 38 (02): 56-61.
- [3] Fan W, Yang Y, Dong Yanqiu, et al. The effect of reward motivation on self-deception in a competitive situation: eye movement evidence [J]. Psychological Science, 2024, 47 (02): 411-423.
- [4] Liu JW, Dong ZY. Predicting physical test scores based on different neural network models [J]. Information Technology, 2024, (01): 65-70+76.
- [5] Haohe Wang, Yue Zhang, Yuqi Jiang. Predictive analysis of college students' examination results based on data mining [J]. Journal of Wuyi College, 2024, 43 (01): 93-97.
- [6] Huan Xu, Qing Li, Daoming Xia, et al. Research on the prediction of students' performance in colleges and universities based on improved support vector regression model [J]. Journal of EJU, 2024, 31 (01): 102-105.
- [7] Ziling Chen. Research on student performance analysis and prediction based on data mining [D]. Zhejiang University of Science and Technology, 2024.
- [8] Du Jiusheng, Zhao Beibei, Hou Qiang. Research on learning behaviour and performance assessment model based on stepwise regression [J]. Surveying and Mapping Bulletin, 2023, (S2): 148-151.
- [9] Wang Xuan, Liu Y, Zhu Changxiang, et al. Random Forest predictive analysis between usual grades and success rate of graduate school for biotechnology students [J]. Science and Education Letters, 2023, (24): 99-102.
- [10] Li Yi, Zhao Yuanyuan. Design of Learning Behaviour Analysis and Achievement Prediction System for Higher Vocational Students [J]. Journal of Shijiazhuang Institute of Vocational Technology, 2023, 35 (06): 23-28.