Problem Chosen	2024	Team Control Number
C	MCM/ICM Summary Sheet	2401509

Gather Your Momentum: Model Capture of Momentum in Tennis

Summary

Momentum is the invisible indicator of the non-stop flow of competition in tennis matches. While it is rather hard to capture momentum variation during match for it is difficult to quantify, the research into momentum could possibly provide valuable guidance for tennis coaches and players who are eager to gain better results. In our research, we established the mathematic model to describe the flow of tennis matches based on momentum, and verified the significant role of it. Based on our momentum analysis, we further established model to predict swings in tennis matches, and determined the most related factors leading to swings.

First, we established the **Dynamic Match Flow Depiction Model** to capture the flow of play. Based on the features of the data set, we determined two main components of momentum: the mental state factor and the technical performance factor, and we modeled momentum of the player as the state vector $Z=[V,Q]$. We used **sliding window** and **Markov chain model** for determination of the mental state component V, and used the **Principal Component Analysis** for determination of the technical performance Q. Then we provided the **three-dimensional visualization** of the match flow of Wimbledon 2023 final match.

Next, to assess the claim of the coach, we separately conducted the **Swings in Play Analysis** and **Runs of Success Analysis**. We established the coach's random match flow model with Monte Carlo simulation, and took advantage of the Logistic function as evaluation model for the results of the coach's random model and our momentum model. Our model scored 7 and 8.9 for each player in Swings in Play Analysis, 5.6 and 5.7 for each player in Runs of Success Analysis, and the random model scored only correspondingly 1.4 and 0.6, 3.1 and 1.9. Additionally, we conducted **ARIMA regression** of both components, and found that the result was significantly different from Gaussian white noise, which proved that momentum itself is not simply random series, and has actual effect on swings in play and runs of success.

Then, based on our former analysis, we assumed that momentum is a **linear dynamic system.** We used Kalman Filter to predict swings in play, and further found 5 most related factors to the occurrence of swings, which are break points, catching a serve, error, physical strength and ace. Based on our findings, we provided three main suggestions for coaches and players.

Finally, we evaluated the model with **concurrency control model** and **confusion matrix.** The precision of our model prediction was 81.77% and the recall rate was 76.29%. The F1 score of the model was 0.78935. The model could perform well in most situations.

Additionally, we discussed extended typed of matches where our model could possibly be applicable. Our model may be applied to women's tennis match, matches in fields of different material, or other similar competitive sports like table tennis.

Keywords: Dynamic system, sliding window, Markov chain, Principal Component Analysis, Kalman Filter, Bayesian Changepoint Detection

Contents

1 Introduction

1.1 Problem Background

Like many other competitive sports, tennis is highly intensive, fast-paced and stochastic. The final outcome of a battle between tennis masters not only depends on strength, but is also a matter of "luck", the "feeling", or, the "momentum".

Some have been arguing about the existence of "pace" in many intense events where there is a non-stop flow, like tests and competitions. It seems like there are some critical moments that could get a person to catch up with the pace of a game, making him feel at ease or even leading him all the way to victory. In other words, the "momentum" is what makes swings occur when it gathers up, and you begin to reign superiority and feel that the Goddess of Victory is smiling at you. It sounds just like the VS bar in duel games. You get to win when you prevail in momentum.

There are indeed something indicating the validity of the notion of "momentum". For example, many of us may have such experience that we just "get the feel" after successfully catching that particular shot. The phenomenon could be attributed to complex possible factors including mental motivation and physical adaption, etc. However, the existence of such "momentum" is still rather ambiguous. The perception of momentum in games is much a subjective feeling. Meanwhile, it is always hard to identify when "momentum" is about to gather, and in what form it is coming in your favor. Is it a serve? Or the next shot? If players want to gain their momentum, it is vital to figure out what makes momentum, when it tends to rise, and pay special attention to the potential "critical moments".

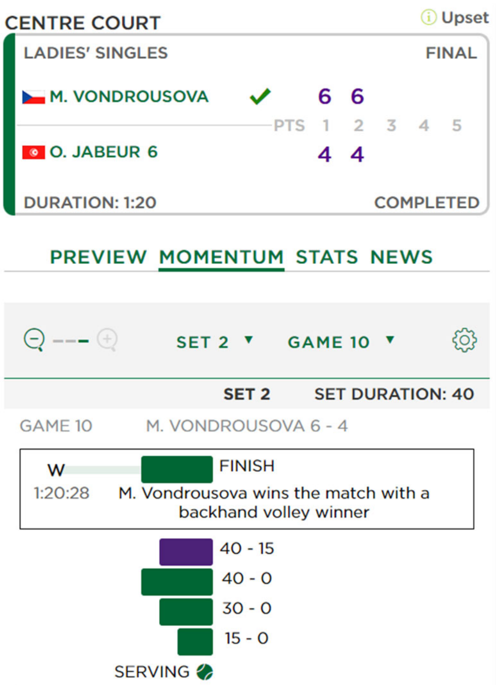


Figure 1. Momentum report from Wimbledon official website[1].

1.2 Restatement of the Problem

Our main goal is to verify the existence of momentum in tennis games through quantified analysis of Wimbledon 2023 men's matches. Then we want to design a feasible strategy to predict and capture momentum in games and put forward a paradigm for better momentum utilization in favor of tennis players.

Considering our primitive research goals and problem requirements, we summarized the problems as follows.

- Develop a model to capture the flow of a game as every point occurs. Apply the model to one or more of the Wimbledon 2023 men's matches. The model should be able to identify which player is playing better as well as how much better they are performing at a given time. Then present a visualization of the match flow based on the model.
- Use the model to assess the claim of a tennis coach that swings in play and runs of success by one player are random instead of a result influenced by momentum.
- Find out possible indicators that help determine swings. Specifically, complete the following two tasks:
  - ✧ Develop a model using data from at least one match that predicts the swings and find out if there are most related factors.
  - ✧ Provide advice for a player going to play a new match against a different rival.
- Test the model with data of one or more of the other matches, and evaluate the quality of swing prediction. If the model would sometimes perform poorly, find out some omitted factors. Evaluate how generalizable the model is to other matches.
- Provide a report of no more than 25 pages with a one- to two-page memo of the research result and advice for coaches as for the role of momentum and how to help players better utilize it.

1.3 Our Work

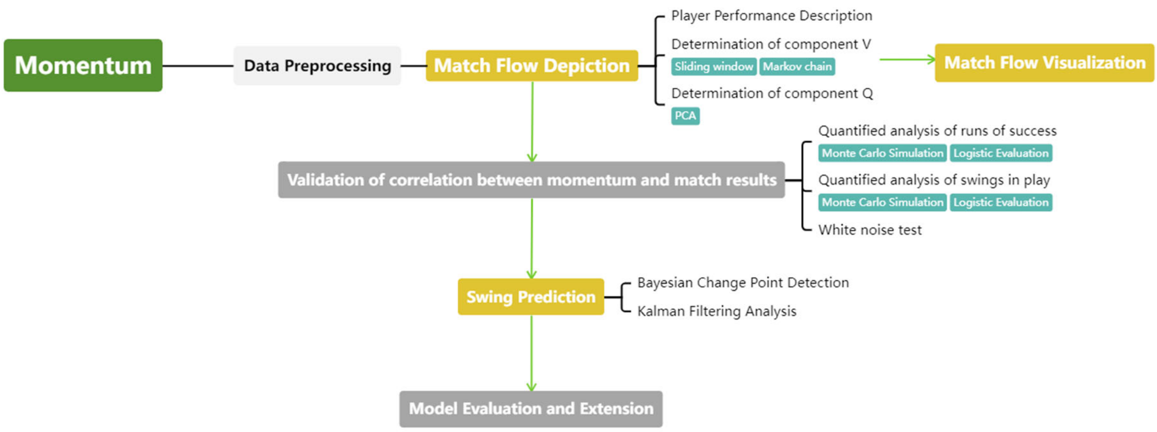


Figure 2. Work Overview.

2 Assumptions and Justifications

Our model is established on the basis of the following assumptions:

✓ **Data authenticity.**
The only data used in this paper is the provided data from Wimbledon 2023 men's matches. The validity of our research results is based on the assumption that all data is authentic with all matches played under Wimbledon's rules and regulations.

✓ **Possible abnormalities like injuries in games are not considered.**
We assume that there is no exceptional situation in all recorded matches that could have significant impact on the player's performance. Such situation includes a twist in the ankle, stimulant usage, and at the same time any mental trauma the player is going through. Considering that such events are especially rare in official contests, they were taken out of consideration so that we could focus on the effects of momentum.

✓ **The effects of players' mental state and technical skill on their performance satisfy the Markov property.**
We assume that the players' mental state and technical skill at each point have Markov property. In other words, the current performance of any player is only determined by his current mental state and the type of the current point, but not any past mental change and past points. We made this assumption because it is sensible that a player is influenced by his current state, and how he got this state matters little.

3 Notations and Glossary

3.1 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
Z	State vector of the player's performance.
V	Mental state factor in affecting player performance.
Q	Technical performance factor in affecting player performance.
V_0	Mental effect from memorized points of the game.
V_1	Mental effect from the score of prior game.
V_2	Mental effect from current estimated win rate.
mp	Number of memorized points in the sliding window.
Win_{N+1-i}	Whether the player won the i_{th} point before the current point.
$Q_{v_0}(mp,N+1-i)$	Weight of the i_{th} memorized point before the current point N.

3.2 Glossary

Technical Performance. The technical performance of the current point means the technical features of the shot, including whether the player is the server, whether the player hits an ace, the player's distance ran during the point, etc. For the complete list of data items involved in the technical performance variable, please refer to Appendix 1.

Mental state. The mental state of the player at current point is decided by the player's scoring history. It is determined by the current score of the game.

4 The Data

4.1 Data Description and Preprocessing

The only data used in this paper is the provided data of Wimbledon 2023 men's matches. The data covers every point after each match's first two rounds.

We found missing values in the items *serve_width*, *serve_depth* and *return_depth*. The missing values were simply dropped out because of lack of effective approach of filling. All values in the items *rally_count* and *serve_width* of the match indexed 1310 were invalid. The match was taken out of consideration when the model involved these two items.

5 Dynamic Match Flow Depiction Model

5.1 Description of Player Performance

Before looking into the effects of momentum on players' performance and the flow of play on the whole, we need to quantify our description of players' performance first.

We investigated the features of the data set for possible components that could most indicate a player's current performance at a time point, and we found that the data includes two main types of information: one is the player's scoring history, and the other is the technical performance of the current point. Considering that the player's current mental state and technique are most likely to affect the match results, and the player's mental state depends mainly on his current scoring, we chose the scoring history and technique factor as the two indicators of player's current performance.

A player's overall performance/momentum is denoted by the state vector Z, which is defined as follows:

$$Z = [\ V,\ Q\ ] \tag{1}$$

Here V and Q are seen as functions of the current point numbered N in this match. Separately, V is the player's mental state fluctuation estimated by his recent scoring and Q is the technical performance reflected by the current shot. The components have range [0,1]. When this player is superior to his rival in this factor, the value of this component is closer to 1.

5.1.1 Determination of component V

We applied the sliding window method and the Markov chain model to determine the component V. We used the weighted average of the player's recent scoring history to describe the overall situation of the player's scoring. V is represented by the equation below:

$$V = w_0 V_0 + w_1 V_1 + w_2 V_2 \tag{2}$$

where V_0 is the variable describing the mental effect from the performance of the player in recent memorized points of shots and is calculated as the weighted average of these memorized points. V_1 is the estimated mental effect from the score of the prior game. V_2 is the mental effect from the estimated possibility of winning based on the player's performance in former sets. Note that V_0 takes into consideration the time point when the point occurs, while V_1 and V_2 are only affected by the current score.

The determination of V_0 applied the sliding window technique. For each point we consider the effect of the past memorized points (denoted as mp) on the current point. The value of mp depends on the order of the point. Specifically, if the current point is the x_{th} point of the game, mp takes value:

$$mp = \begin{cases} 0, \ if\ the\ former\ 2\ points\ of\ the\ match \\ 3, \ if\ the\ former\ 2\ points\ of\ the\ game \\ \qquad but\ not\ the\ former\ 2\ points\ of\ the\ match \\ x, \ if\ not\ the\ former\ 2\ points\ of\ the\ game \end{cases} \tag{3}$$

Considering that points with more time proximity have more effect on the player's current mental state, V_0 is represented as the weighted average:

$$V_0 = \begin{cases} 0.5, \ mp = 0 \\ \sum_{i=1}^{mp} Win_{N+1-i} \cdot Q_{v_0}(mp, N+1-i), \ mp \neq 0 \end{cases} \tag{4}$$

Win_{N+1-i} is the virtual variable denoting whether the player wins the $i-1_{th}$ point before the current point in this match numbered N. The weight of the i_{th} memorized point Q_{v_0} before the current point N follows the definition presented below.

$$Q_{v_0}(mp, N+1-i) = \frac{mp+1-i}{mp(mp+1)/2} \tag{5}$$

Similarly, the determination of V_1 is also by weighted average. The weight is decided by the score of the prior game. Note that when the total score of the two players in the prior game is n, and the score of player 1 is m, The values of V_1 of the two players should add up to 1. Therefore, the weight is determined using combinatorial number.

$$V_1(n,m) = \frac{\sum_{i=0}^{m-1} C_{n-1}^{i}}{2^{n-1}} \tag{6}$$

so that V_1 satisfies the constraint:

$$V_1(n,m) + V_1(n, n-m) = 1 \tag{7}$$

The determination of V_2 used the Markov chain model. We assumed that the effects of players' mental state on their performance have Markov property. In other words, the estimated possibility, or the player's performance, should only depend on the player's current status. The match is modeled as Markov chain to predict the possibility of winning. The match status is described by the score. The value of V_2 equals the win rate under current status:

$$V_2(n,m) = win\ rate \tag{8}$$

Therefore, based on our modeling of the match status, the state transition process of Markov chain with each time step could be depicted by the graph below.

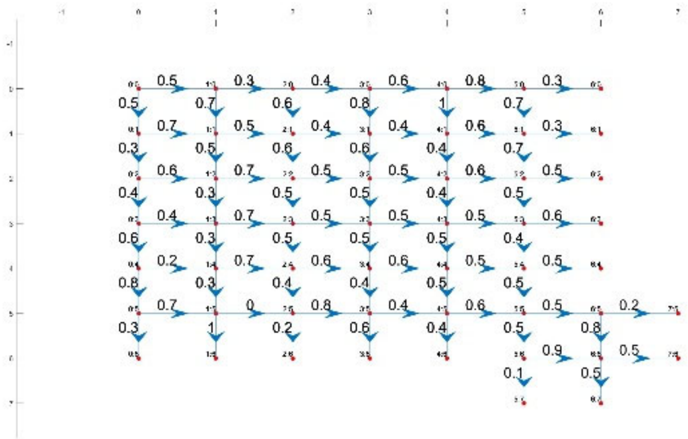


Figure 3. State transition diagram of Markov chain.

When the state space of Markov chain is finite, the Markov chain can be depicted with the state transition matrix. Arranging the single-step state transition probability of all possible states in the matrix, we got the state transition matrix:

$$P_{n,n+1} = (P_{i_n,i_{n+1}}) = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{9}$$

5.1.2 Determination of component *Q*

The component Q is determined using PCA (Principal component analysis) for data dimension reduction. Based on the 29 indicators reflecting technical performance obtained from processing of the data set, we generated 10 synthesized principal components which are the linear combination of the items (Appendix 1).

Assuming that there are p original factors and m principal components, the principal components are linear combination of original technical performance indicators:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{cases} \quad (m \leqslant p) \tag{10}$$

When all principal components are mutually independent and each is the linear combination that satisfies the condition and has largest variance, the dimension of indicator can be reduced while retaining information to the utmost.

5.2 Model Solution and Match Flow Visualization

Solution of component V

Based on MATLAB solution, set w_0, w_1, w_2 as 0.7, 0.2, 0.1.

For the solution of V_2, under each possible score, the estimated win rate is presented in the table:

Table 2: Estimated win rate under different score

	0	1	2	3	4	5	6
0	0.50	0.57	0.57	0.92	0.99	0.99	-
1	0.43	0.50	0.66	0.79	0.99	0.99	-
2	0.22	0.34	0.50	0.69	0.90	0.95	-
3	0.08	0.21	0.31	0.50	0.72	0.98	-
4	0.01	0.15	0.10	0.28	0.50	0.76	-
5	0.01	0.01	0.05	0.02	0.24	0.50	0.65
6	-	-	-	-	-	0.35	0.50

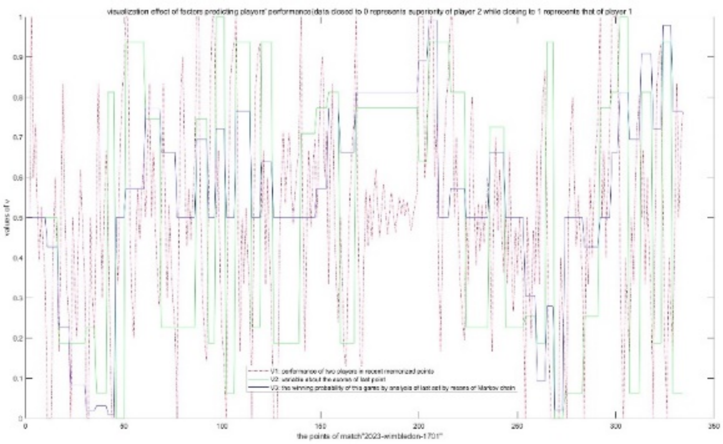


Figure 4: Visualization of value of component V.

Solution of component Q

In order to ensure model accuracy, we chose to generate 10 principal components. Use SPSS solution to determine the principal components. PCA result and total sum of contribution to variance is presented below. The variance explained by all principal components accounts for 80% of total variance.

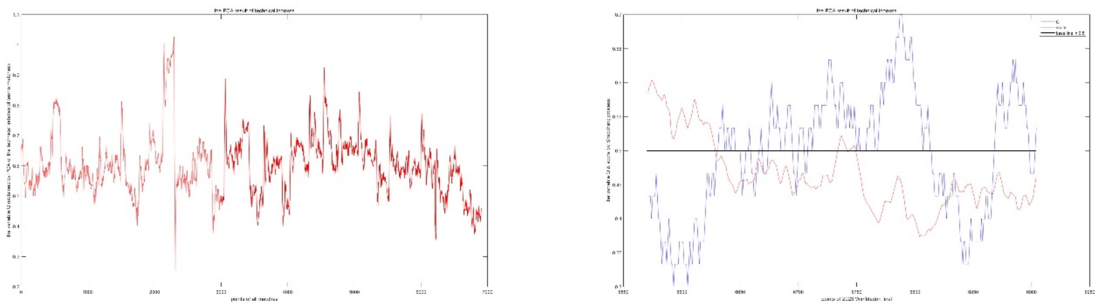


Figure 5: Visualization of value of Q in all matches and in match 2023-Wimbledon-1701.

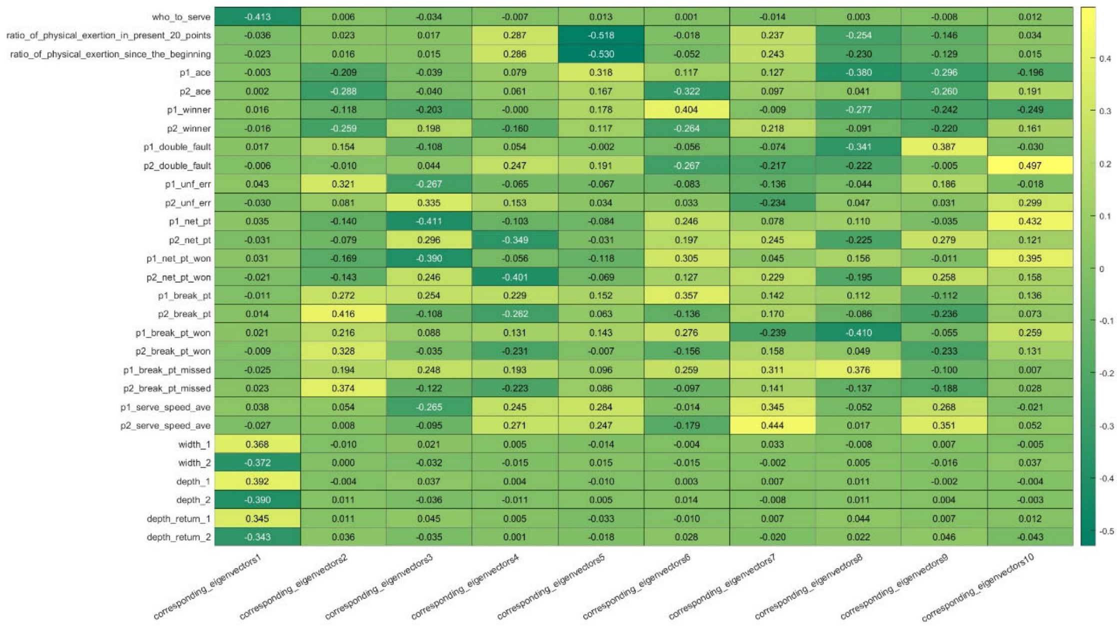


Figure 6: Heat map of PCA result of Q.

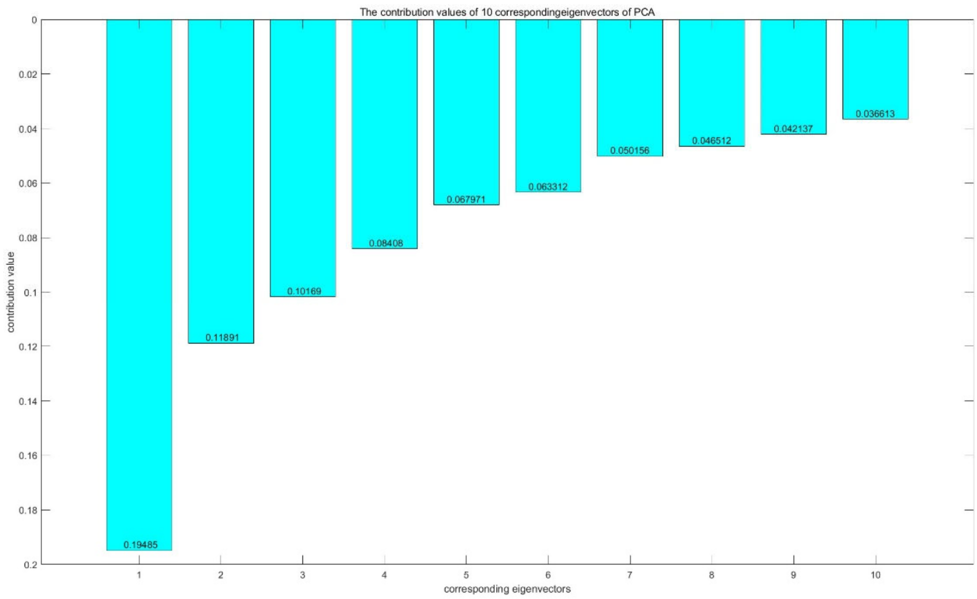


Figure 7. PCA result and principal components contribution to total variance.

We applied the model to the match 2023-Wimbledon-1701 between Carlos Alcaraz and Novak Djokovic. Components are calculated after smoothing process and normalization. The visualized match flow is presented below.

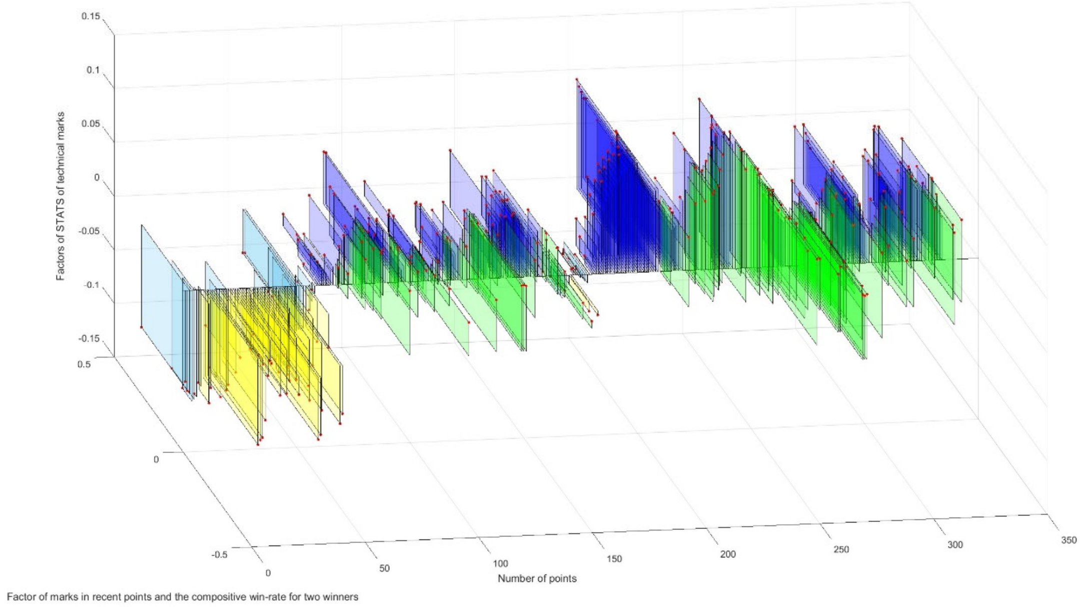


Figure 8. Visualization of the match flow of match 2023-Wimbledon-1701.

According to the visualization, in the first set, the yellow blocks significantly prevail over blue blocks, showing that Djokovic was dominating in this set. In the second set, there are mostly green blocks and the values are close to axis, indicating that the two were evenly matched in power in this set. Similarly, in the third set, there are significantly more blue blocks, it means that Alcaraz was performing remarkably better. In the fourth set, there are more green blocks and the flow turned in favor of Djokovic. In the fifth set, the blue blocks slightly surpass the green blocks, and the set ended up in Alcaraz's victory.

The result depicts the real situation of the match well. It demonstrates the feasibility of our model.

6 Correlation Analysis of Momentum and Match Result

6.1 Swings in Play Analysis

Swing Quantification

To verify the correlation between momentum and the swings and results of the match, we conducted separate research. For investigation into the effects of momentum on swings in matches, we first completed quantification of swings in play. When the difference of player scores reaches 5, we consider that a swing happens.

Monte Carlo Simulation

To establish the random match flow model of the coach without consideration of effects of momentum, we conducted Monte Carlo simulation method to simulate random swing occurrence in match. Number of simulations was set to 200.

We considered the strength of player, whether the player serves the ball and random factors in the simulation model. The player's strength is approximately evaluated by the score of the final game, which we considered as most representative of player's relative strength. Empirically, the server tends to have higher probability of winning. When the player is the server, we take this factor into consideration and multiply the player's result with an advantageous factor.

The result of Monte Carlo simulation (corresponding to the coach's random match flow model) is compared to our match flow depiction model. The mean and variance of match swing estimation of the two models are:

Table 3: Model estimation results of player 1.

	Mean	*Variance*
Random Match Flow Model	0.568	0.130
Momentum Match Flow Model	0.654	0.020

Table 4: Model estimation results of player 2.

	Mean	*Variance*
Random Match Flow Model	0.508	0.100
Momentum Match Flow Model	0.317	0.012

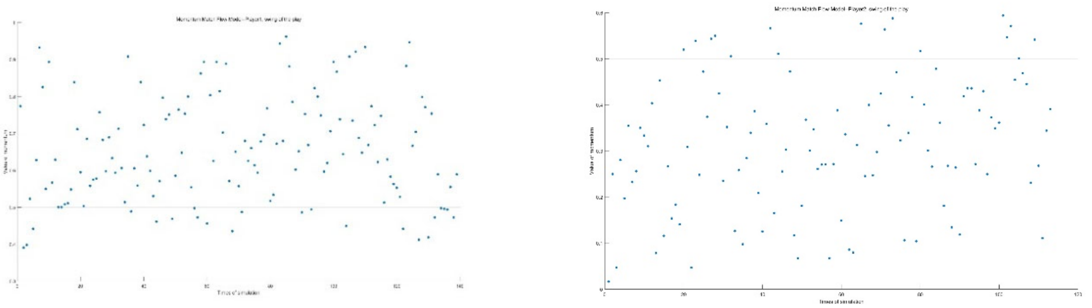


Figure 9. Scatter plot visualization results of Momentum match flow model.

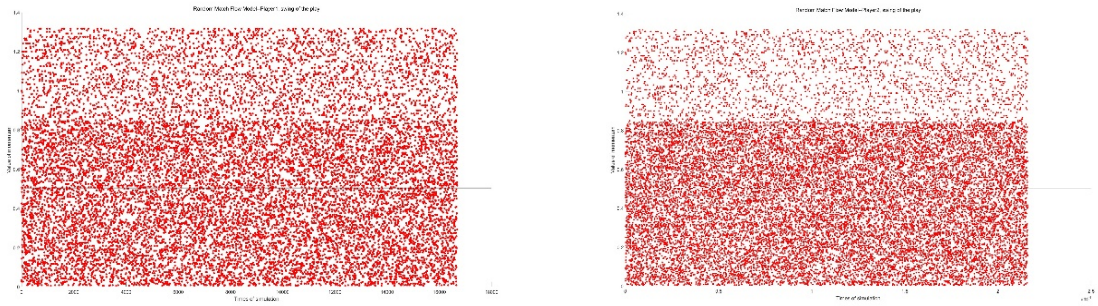


Figure 10. Scatter plot visualization results of Random match flow model.

The variance show that the estimation of our model based on momentum is more stable and reliable.

Logistic Evaluation Model

For comprehensive evaluation of the estimation of the two models, we applied the evaluation model based on the Logistic function. A simple Logistic function can be expressed by the following equation. It is continuous and monotonically increasing. When $t=0$, $P(t)=0.5$. The function tends to 1 when the argument tends to $+\infty$ and tends to 0 when the argument tends to $-\infty$.

$$P(t) = \frac{1}{1+e^{-t}} \tag{11}$$

Consider the inverse of Logistic function:

$$y = \ln x - \ln(1-x) \tag{12}$$

When estimation result tends to 1, the function's value tends to $+\infty$. Meanwhile, when the estimation tends to -1, the function tends to $-\infty$. To avoid overly significant impact of the estimation of a single point on the overall score of the model, a single scoring in the model is restricted in [-2,2]. When the scoring goes beyond the range, it is counted as the boundary value.

We evaluated all results of each model and calculated the average score as the model's overall score. The Logistic score of our model estimating result of player 1 and 2 were 7 and 8.9, while the random match flow model scored 1.4 and 0.6. The momentum match flow model performed significantly better in swing estimation.

6.2 Runs of Success Analysis

Runs of success is the consecutive scoring of player.

Similar to analysis of swings in play, we first conducted 200 times of Monte Carlo simulation to establish the coach's random match flow model. Then we compared the results of the two models. The variance shows that the result of our model is more stable.

Table 5: Model estimation results of player 1.

	Mean	*Variance*
Random Match Flow Model	0.612	0.130
Momentum Match Flow Model	0.624	0.020

Table 6: Model estimation results of player 2.

	Mean	*Variance*
Random Match Flow Model	0.468	0.080
Momentum Match Flow Model	0.387	0.020

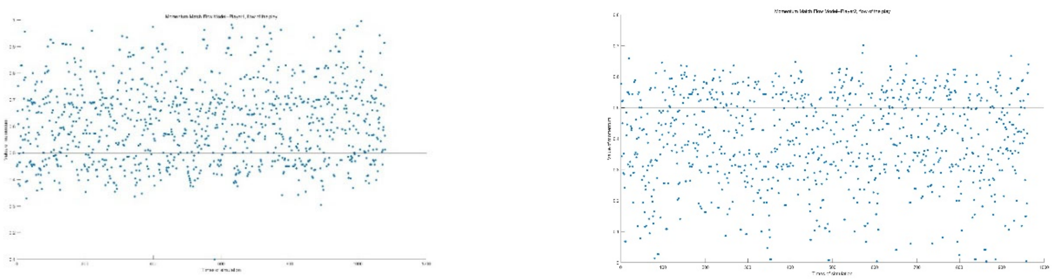


Figure 11. Scatter plot visualization results of Momentum match flow model.

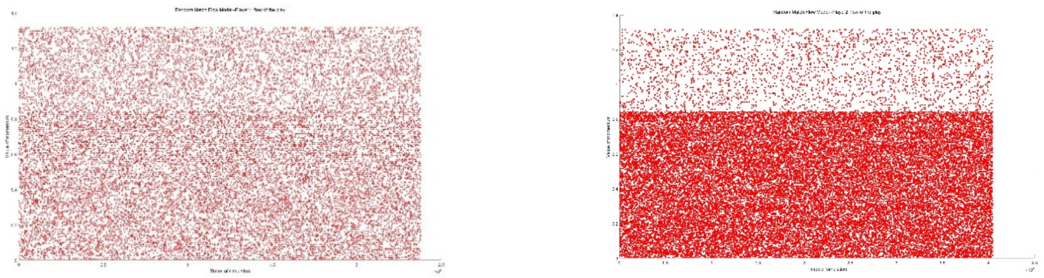


Figure 12. Scatter plot visualization results of Random match flow model.

The Logistic score of our model estimation result of player 1 and 2 were 5.6 and 5.7, while the random match flow model scored 3.1 and 1.9. It demonstrates that our model based on momentum is also significantly better in describing occurrence of runs of success. These results provide valid proof for the important role of momentum in tennis.

6.3 ARIMA regression

We assumed that momentum itself is not simply random series as well. To verify the hypothesis, we modeled momentum with ARIMA time series model. Make null hypothesis that the time series of momentum is significantly different from Gaussian white noise.

For the technical performance component Q, we selected the former 279 records of Wimbledon 2023 final match (2023-Wimbledon-1701) as training group, the remaining records of the match as testing group. MATLAB solution showed that the time series of the technical performance component Q fits ARIMA(1,1,5).

For the mental state component V, we selected the former 320 records of the same match as training group, the rest left as testing group. The time series fits ARIMA(1,1,6).

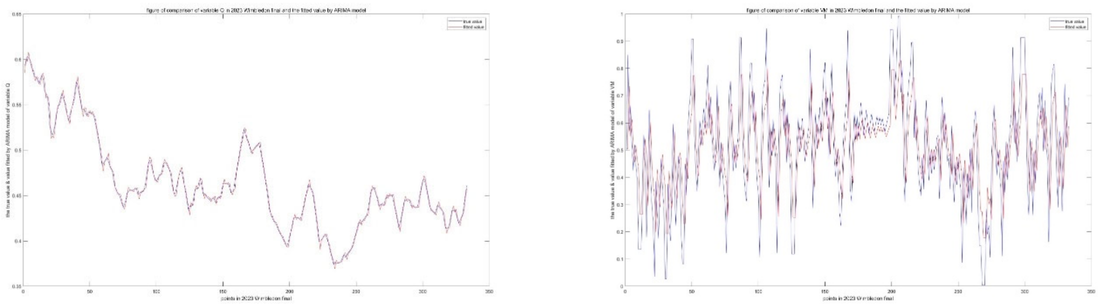


Figure 13. ARIMA regression result of components *Q* and *V*.

The results show ARIMA's ideal regression effect of momentum. This proves that momentum is not random, but is correlated to the time factor.

7 Match Swing Prediction and Affecting Factors

7.1 Kalman Filtering Analysis

Based on our former analysis, if the data was properly measured, the measured result should involve the Gaussian white noise. Therefore, we can regard the momentum system including mental state and technical performance components as a linear dynamic system (LDS).

Based on the assumption, we applied Kalman Filter to our dynamic model to predict occurrence of swings in matches.

It is a recursive Bayesian estimation on the hidden Markov chain. For each moment, the state transition of the previous moment is taken as the priori of the present moment and modeled as a Gaussian distribution. According to the Bayes formula, this priori is modified through the observations modeled as a Gaussian distribution of the present moment, and finally the posterior of the present moment is obtained. This posteriori is the product of the previous two Gaussian distributions.

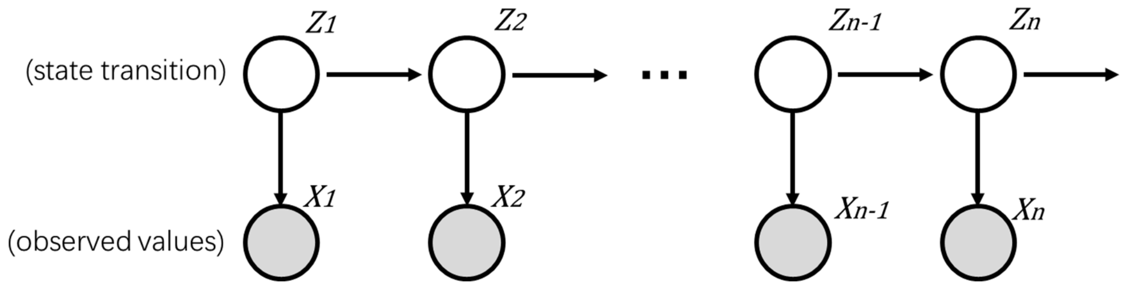


Figure 14. State transition diagram of LDS

A pleasing property is that this posterior is still a Gaussian distribution, and the parameters of the new Gaussian distribution: mean and variance can be derived by comparing the coefficients. This coefficient is called the Kalman gain. This posteriori will act as a priori to the next moment, and so on.

Again, we chose the match 2023-Wimbledon-1701 for model establishment. Because the former 14 records of the match manifested strong randomness, we used the remaining 320 records for Kalman Filter procession. The result was presented in the plots below.

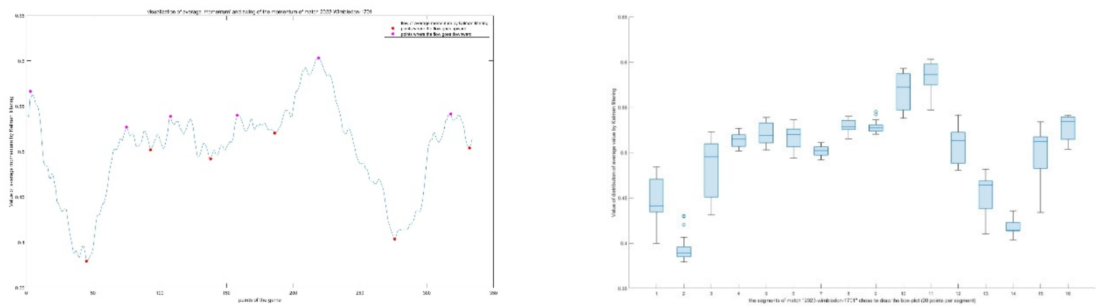


Figure 15. Kalman Filter result of match swing prediction.

Comparing the result with our former modeling and analysis of match 2023-Wimbledon-1701, we could discover that the result reflects the swings of the match well.

7.2 Bayesian Changepoint Detection

To find out possible factors most related to determining swings in matches, we used the Bayesian Online Changepoint Detection to process the technical performance of each applicable match. Note that we chose the technical performance factor as the indicator to avoid the impact of winning or losing scores on the mental state of the player. This is because in higher levels of matches, the mental factor plays minor role in affecting the performance of players. This could help us focus more on the technical and actionable causes of winning or losing points.

Bayesian Changepoint Detection helps detect the position of parameter changepoint caused by systematic factors in the time series. The method predicts the posterior distribution of the run length of current time point since last changepoint occurred, and calculates the posterior distribution of the next point with recursion method. Using the Bayesian formular, the posterior distribution of run length r_t is:

$$p(r_t|x_{1:t}) = \frac{p(r_t,x_{1:t})}{p(x_{1:t})} \tag{13}$$

Then, do a joint probability expansion on $p(r_t,x_{1:t})$:

$$p(r_t,x_{1:t}) = \sum_{r_{t-1}} p(x_t|r_{t-1},x_{t-1}^{(r)})\,p(r_t|r_{t-1})\,p(r_{t-1},x_{1:t-1}) \tag{14}$$

Therefore, $p(r_t,x_{1:t})$ is the total of the product of these three: Underlying Probabilistic Model (UPM), changepoint priori, and its formal state. Thus, the value of $p(r_t,x_{1:t})$ is recursively solvable. Since the beginning of the match is definitely a changepoint, set the initial value $p(r_0=0) = 1$. Apply Bayesian Changepoint Detection to all applicable matches. The following are examples of Bayesian Changepoint Detection results of match 1501, 1502, 1503, 1504, 1601 and 1602.

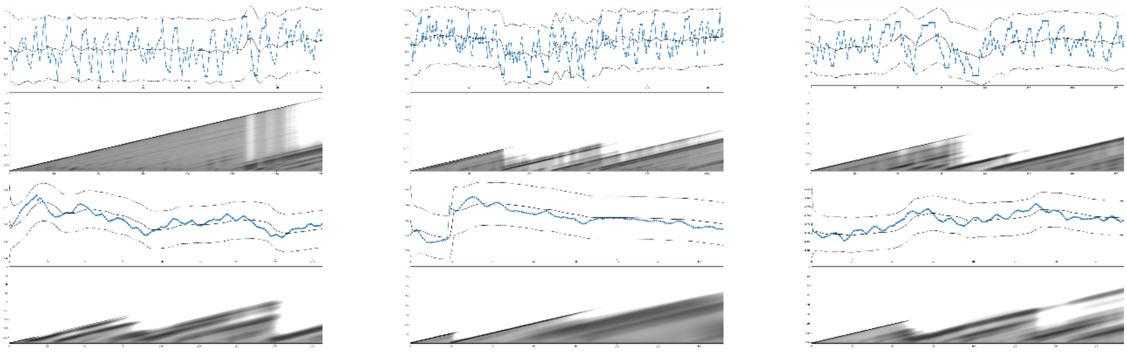


Figure 16. Bayesian Changepoint Detection of match 1501-1602.

129 significant breakpoints were detected in analysis result. We investigated the primitive causes of each point, while the factor of who is the server was omitted. Then, we picked out 5 factors with highest frequency of occurrence:

Table 7. 5 Most related factors affecting match swings.

Factor	*Frequency*
break points	31/129
catching a serve	27/129
error	23/129
physical strength	19/129
ace	18/129

7.3 Advice for players

Based on our research, when a player is going to play a new match against a different player, what can he pay attention to in order to best gather his momentum and get better results?

First, considering the effects of mental state component in our dynamic momentum model, the player's performance is to a considerable extent influenced by his mental state under current match score, and therefore how the score affects his mental state matters a lot. Also, the factor of error appears to play an important role in affecting the swing of the game. Therefore, we suggest players to keep a positive mind when errors or missed points occur, as well as when he falls behind his rival in score. More practically, it might be helpful to take up some training of stress management. This may help minimize possible negative impact of mental state on the player's performance.

Secondly, considering the effects of technical performance component, it emphasizes the importance of training. Good technical performance at the same time boosts the player's spirits, and good performance is always based on continuous training and accumulation. Players should work hard to improve their skills and physical strength. Also, it is valuable to study into new skills and techniques. For example, a good serve will help you win more points and gather more momentum for yourself.

Additionally, if the player could get access to some information about his rival, it is worth the effort to try some special tactics accordingly. For example, if the opponent is physically inferior while the player himself has remarkable physical endurance, he may try running wider to magnify his physical advantage.

8 Model Evaluation and Extension

8.1 Model Evaluation

In former discussion, we compared the Kalman Filter prediction model result to the real situation of the Wimbledon 2023 final match, and the model predicted the real situation well.

For further evaluation of the quality of prediction, we planned to use the concurrency control model.

We modeled the situation of the match as a system. When we use a series a to predict the series b, at first the system is vacant. At this time if an event of series b occurs, the event is regarded as a miss.

Once the system receives an event of *a*, it begins to receive events of *b*. The receiving process continues for an accepting window period of *T_accept*. If the system receives no event from *b* during the window, the event from *a* is regarded as a miss and the system returns to vacant state. If the system receives an event from *a* during the window, it is regarded as an interrupt and the window continues normally. If it receives an event from *b*, the event is regarded as a success and the system enters the vacuum phase. All events during vacuum phase are regarded as interrupt. The length of vacuum phase is not fixed and decreases as the time of accepting period increases, but the minimal length is *T_accept*. After the vacuum period, the system returns to vacant state.

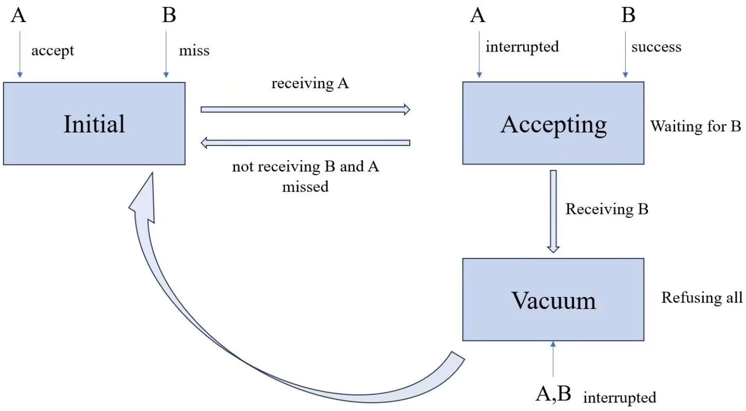


Figure 17. Concurrency control model's control flow diagram.

Table 8. Result of concurrency control model.

Total Events	Miss	Interrupt	Success
204	33	23	148
267	46	73	148

where the item *Interrupt* is the inherent error of the data set and should be excluded.

Quantified analysis with confusion matrix shows that the precision of prediction is 81.77%, and the recall rate of the prediction is 76.29%.

We used the F1 score, which is the harmonic average of precision and recall rate as comprehensive reflection of both indicators. The closer the F1 score is to 1, the higher the quality of model prediction. The F1 score of Kalman Filter prediction is 0.78935, which is relatively ideal.

8.2 Possible Extension

Our model can possibly be applied to other matches. For example, in women's matches, the dynamic momentum model involving mental state and technical performance factors is still applicable. The effect of match score on mental state may well differs little, while the technical factor may differ slightly. Moreover, as for tennis matches on fields with different materials including grass, plastic, etc., the technical factor may also differ slightly, but the mental factor would be relatively stable. The model may also be applicable to other competitive games like table tennis, where both the mental and technical factors might be different.

9 Conclusions

In our work on momentum effects, we first established the Dynamic Match Flow Depiction Model to capture the flow of play. Based on the features of the data set, we determined two main components of momentum: the mental state factor and the technical performance factor, and we modeled momentum of the player as the state vector $Z=[V,Q]$. We used sliding window and Markov chain model for determination of the mental state component V, and used the Principal Component Analysis for determination of the technical performance Q. Then we provided the three-dimensional visualization of the match flow of Wimbledon 2023 final match.

Next, we separately conducted the Swings in Play Analysis and Runs of Success Analysis. We established the coach's random match flow model with Monte Carlo simulation, and took advantage of the Logistic function as evaluation model for the results of the coach's random model and our momentum model. Our model scored 7 and 8.9 for each player in Swings in Play Analysis, 5.6 and 5.7 for each player in Runs of Success Analysis, and the random model scored only correspondingly 1.4 and 0.6, 3.1 and 1.9. Additionally, we conducted ARIMA regression of both components, and found that the result was significantly different from Gaussian white noise, which proved that momentum itself is not simply random series, and has actual effect on swings in play and runs of success.

Based on our former analysis, we assumed that momentum is a linear dynamic system. We used Kalman Filter to predict swings in play, and further found 5 most related factors to the occurrence of swings. Based on these findings, we provided suggestions for tennis players on their future playing.

Finally, we evaluated the model with concurrency control model and confusion matrix. The precision of our model prediction was 81.77% and the recall rate was 76.29%. The F1 score of the model was 0.78935. The model could perform well in most situations.

Additionally, we discussed extended typed of matches where our model could possibly be applicable. Our model may be applied to women's tennis match, matches in fields of different material, or other similar competitive sports like table tennis.

10 Memorandum

To: Coaches
From: MCM Team #2401509
Date: February 5, 2024
Subject: The important role of momentum and how to prepare your players for it

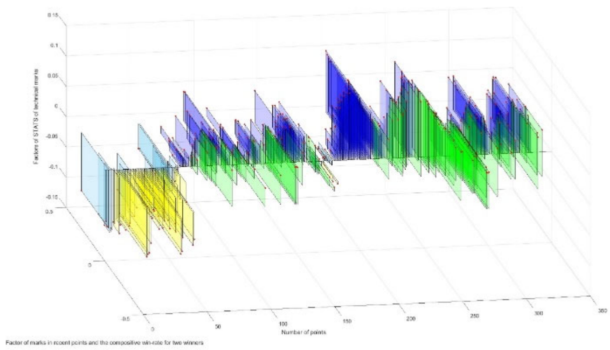
---

Dear Sir or Madam:

The result of competition between tennis players, especially elite players, depends not only on their professional skills and technique, but also to a large extent depends on their mental state, physical adaptation and many other sophisticated factors. Through our research, we have verified the particular role of "momentum" in tennis matches using mathematic model, and we would like to offer you some related advice as we introduce our model.

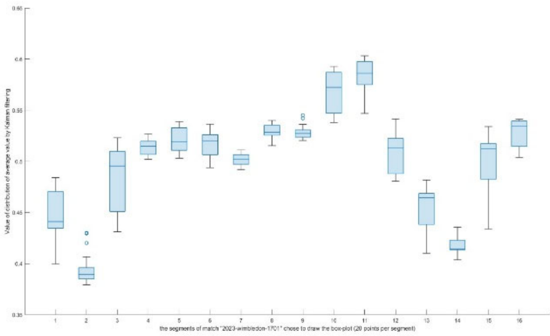
In our research, we first investigated the data set collected from Wimbledon 2023 men's matches. After preprocessing the data, we noticed that the data set contained two main aspects of information: the player's scoring history, and the technical performance of the current point.

Based on this feature of data, we established our dynamic match flow depiction model reflecting player's momentum in match. We described the player's performance with the state vector Z, consisting of two components: the mental state V decided by the player's scoring history, and the technical performance Q. We applied sliding window technique and Markov chain analysis to determine the value of V, and used principal component analysis to decide the value of Q. We applied the model to the Wimbledon 2023 final match. The model showed ideal description ability of the match flow.



A coach once showed us his suspicion about the role of momentum in tennis matches. He thought that swings and success in tennis matches are merely random and momentum could hardly have any effect. However, our research has demonstrated the actual role of momentum in tennis matches. We modeled the coach's random match flow model with Monte Carlo simulation, and evaluates the estimation result of our model and the random match flow model with Logistic evaluation model. The result was that our model scored 7 and 8.9 for each player in swing description, 5.6 and 5.7 for each player in runs of success description, and the random model scored only correspondingly 1.4 and 0.6, 3.1 and 1.9. Our model that took momentum into consideration performed much better than the random model. We further conducted ARIMA regression of the two components V and Q. The regression showed ideal result. This proves that momentum itself is not random as well. Momentum has actual effect on the swings and results of tennis matches.

To provide coaches with feasible advice on preparing their players to better respond to the events that could potentially influence the match flow, we first established the Kalman Filter model to predict the swings in matches. With cross-reference between our former analysis of the Wimbledon 2023 final match, the prediction model proved to be effective.



We further used Bayesian Changepoint Detection to find out the most related factors to swings in matches. We found that break points, catching a serve, error, physical strength and ace were the 5 most related factors.

To prepare your players to respond to the "critical moments" on the match field that could possibly affect the flow of play, we have these suggestions based on our findings:

First, cultivate a strong mind of your players. Mental state is a critical component of the player's momentum. Keeping positive when making errors or missing points could help undermine the negative effects of mental factor. Moderate stress training may help as well.

Second, hard training matters. The player's technical performance and physical ability in match to a large extent decides the player's momentum and the flow of play. Good technical performance and physical capability depend on continuous training and profound accumulation. It's always better to be down-to-earth.

Third, pay attention to the information you have about the rival and make according strategies. Try to fully exert your player's advantage in the game.

Hope our suggestions could be useful to you!

Yours sincerely,

MCM Team #2401509

References

[1]: AELTC. “Home-The Championships, Wimbledon.” Wimbledon, 2 January, 2024. https://www.wimbledon.com/index.html.
[2]: Adams, Ryan P. and David John Cameron MacKay. “Bayesian Online Changepoint Detection.” arXiv: Machine Learning (2007): n. pag.
[3]: Ramsey Faragher. “Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation” IEEE SIGNAL PROCESSING MAGAZINE (2012 SEPTEMBER)
[4]: Jonathon Shlens. “A Tutorial on Principal Component Analysis”, Mountain View. April 7, 2014
[5]: Hamilton, James. Time Series Analysis. (1994) Princeton University Press. ISBN 9780691042893.
[6]: Fawcett, Tom. "An Introduction to ROC Analysis"(2006) . Pattern Recognition Let-ters. 27 (8): 861–874.
[7]: Gagniuc, Paul A. (2017). Markov Chains: From Theory to Implementation and Experimentation. USA, NJ: John Wiley & Sons. pp. 1–235. ISBN 978-1-119-38755-8.

Appendices

Appendix 1

Introduce: Data items involved in technical performance variable.

Appendix Table 1: 29 original technical performance indicators

Variable	Description
who_to_serve	whether the player is the server
consumption_of_strength	physical comsumption of recent 20 points
consumption	total physical consumption
p1_ace	player 1 hit an untouchable winning serve
p2_ace	player 2 hit an untouchable winning serve
p1_winner	player 1 hit an untouchable winning shot
p2_winner	player 2 hit an untouchable winning shot
p1_double_fault	player 1 missed both serves and lost the point
p2_double_fault	player 2 missed both serves and lost the point
p1_unf_err	player 1 made an unforced error
p2_unf_err	player 2 made an unforced error
p1_net_pt	player 1 made it to the net
p2_net_pt	player 2 made it to the net
p1_net_pt_won	player 1 won the point while at the net
p2_net_pt_won	player 2 won the point while at the net
p1_break_pt	player 1 has an opportunity to win a game player 2 is serving
p2_break_pt	player 2 has an opportunity to win a game player 1 is serving
p1_break_pt_won	player 1 won the game player 2 is serving
p2_break_pt_won	player 2 won the game player 1 is serving
p1_break_pt_missed	player 1 missed an opportunity to win a game player 2 is serving
p2_break_pt_missed	player 2 missed an opportunity to win a game player 1 is serving
p1_serve_speed_ave	player 1's average speed when serving
p2_serve_speed_ave	player 2's average speed when serving
width_1	player 1's direction of serve
width_2	player 2's direction of serve
depth_1	player 1's depth of serve
depth_2	player 2's depth of serve
depth_return_1	player 1's depth of return
depth_return_2	player 2's depth of return