WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Research on state volatility of tennis players based on sliding window technique

Yanze Sun *, Zhaoqi Chen, Qiutong Wang

Aulin College, Northeast Forestry University, Harbin, China

* Corresponding author: 2022224838@nefu.edu.com

Abstract. This paper aims to study the state volatility of players in tennis matches and proposes a conceptual model of "momentum" based on sliding window technology. Firstly, Pearson correlation analysis is used to screen out the indicators that are highly relevant to the victory of the match, and the index system including point advantage, service advantage, break point, unforced error, ace, game won and distance run is constructed. Then, a formula for calculating "momentum" is proposed, and the sliding window algorithm is used to calculate the "momentum" value of each athlete. Finally, the change of "momentum" during the game is demonstrated visually, and the correlation between "momentum" and the outcome of the game is verified. In addition, the "momentum" transition point is defined, the "momentum" difference series is calculated by moving average method, and the randomness of the "momentum" transition point is analyzed by Run test. The results show that the change of "momentum" has a significant impact on the outcome of the match, and the change of "momentum" is not random.

Keywords: Tennis match; Player status; Sliding window; Moving average; Momentum.

1. Introduction

Within contemporary sports contests, outcomes can significantly fluctuate, particularly impacting those who are deemed to possess an advantage. This oscillation is referred to as 'momentum'. Momentum is presently conceptualized as: a positive or negative modification in cognition, physiology, emotion, and conduct instigated by a triggering event or series of events leading to a modification in performance[1]. If we view behaviour as reflecting a continuous perception-action cycle, top-down processes, including controlled attention, expectancy and preparedness, play a fundamental role in directing action [2].

This paper mainly solves the following problems:

(1) Quantitative Analysis of athlete state volatility: A conceptual model of "momentum" based on sliding window technology is proposed. By calculating the "momentum" value of each player, the quantitative analysis of tennis players' match state is realized.

(2) Evaluate the correlation between "momentum" and the result of the game: By visualizing the change of "momentum" during the game, the correlation between "momentum" and the outcome of the game was verified.

(3) Explore the randomness of "momentum" transition: Define the "momentum" transition point, use the moving average method to calculate the "momentum" difference series, and use the Run test to analyze the randomness of "momentum" transition point, and find that "momentum" transition is not random.

(4) Provides reference for coaches and athletes: Research results can provide important reference for coaches and athletes to formulate strategies and adjust state.

(5). Explore data science methods to analyze athletes' state volatility: This paper provides a useful exploration for using data science methods to analyze athletes' state volatility.

2. Construction of Momentum Weighted Calculation Model Based on Time-series Sliding Window

2.1. Data Processing

The data preprocessing in this study was conducted in two stages. The initial stage underwent scrutiny for unusual values and missing values, revealing a complete absence of such anomalies. In the subsequent phase, the variable "distance_run" underwent standardisation to enhance the accuracy of subsequent predictive analysis (Figure 1).

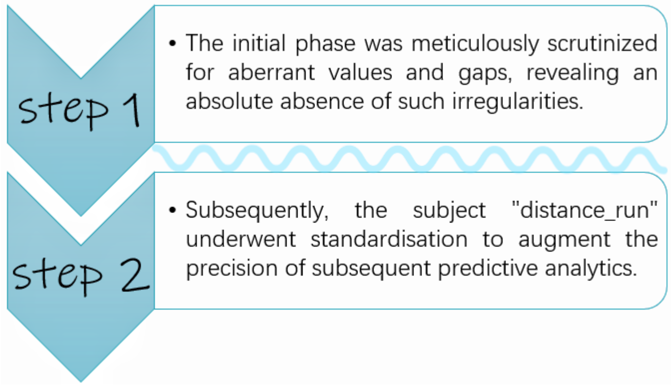

Figure 1. Flow Chart of Data Processing

2.2. Module Overview

The present paper employs Pearson correlation analysis to screen indexes, establishing models for those with high correlation coefficients. Model construction focuses on winning points advantage, won games and sets, serve conditions, break points, unforced errors, ace points, and distance. Subsequently, data visualization processing is conducted (Figure 2).

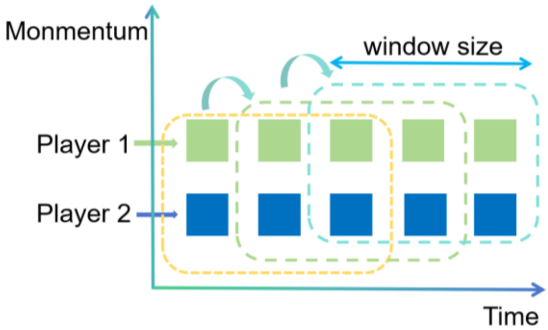

Figure 2. Momentum Time-series Sliding Window Model

2.3. Index Selection

2.3.1. Pearson Correlation Analysis.

We employs the Pearson correlation coefficient analysis on the standard data set to establish relevant indexes through the correlation coefficients [1-4]. The correlation coefficient is calculated as follows:

$$r = \frac{\sum(x-m_x)(y-m_y)}{\sqrt{\sum(x-m_x)^2 \sum(y-m_y)^2}} \tag{1}$$

Where m_x is the mean of the vector x and m_y is the mean of the vector y [5].

The result of the Pearson correlation coefficient is illustrated in Figure 3. Through the analysis of this correlation heatmap, we observe that points advantages exhibit the strongest correlation with other indicators. This suggests that points advantage (PA) is the most crucial index for measuring momentum. Additionally, serve advantage (SA), break points (BP), winners shots (WS) as well as the number of won games (GW) and won sets (SW) also show a high degree of correlation. Therefore, we choose these indicators with strong correlations for the construction of the subsequent momentum calculation model.

2.3.2. Construction of Indicator System.

This paper selected the following indicators based on the value of the Pearson correlation coefficient and literature research and build an Indicator System. By establishing the Indicator System, it is possible to quantify and assess the performance of each player more effectively. Simultaneously, it aids in conducting comparative analysis among different players, identifying strengths and areas for improvement.

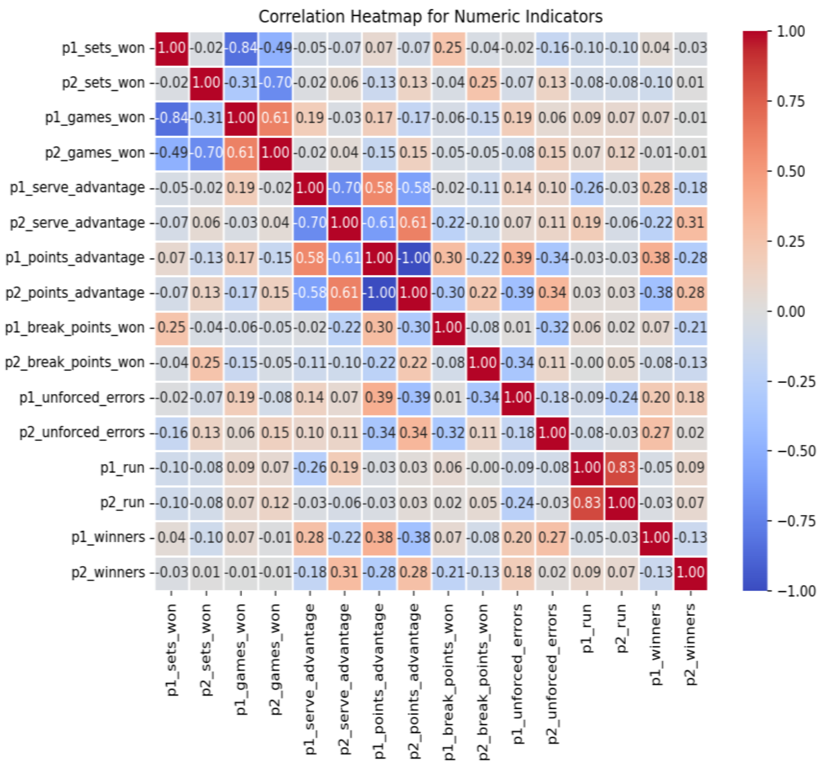

Figure 3. Heatmap of Pearson Correlation

Point Advantage	Perform a comparative analysis of the scoring difference between two players over a selected timeframe.
Serve Advantage	In tennis, the server possesses a substantial advantage. Look at the proportion of points won by the server and their ability to secure service holds.
Break Points Won	The accumulation of break points indicates a player's propensity to win during an opponent's service game.
Winner Shots	The tally of match points reflects a player's offensive prowess and proactive stance in a match.
Sets Won	If a player wins a greater number of sets, they are typically more likely to secure a victory in the match.
Games Won	If a player wins a greater number of sets, they are typically more likely to secure a victory in the match.

Figure 4. Indicator System

2.4. Construction of Momentum Weighted Calculation Model

•Serve Advantage Processing

We've already know that the server has a much higher likelihood of winning points in tennis matches. Therefore, to incorporate this characteristic into our measurement model, we calculate a serving advantage coefficient to provide the serving side with an additional advantage in points.

We define the serve advantage coefficient β as:

$$\beta = \frac{1}{2}\left(\frac{F_1}{S_1} + \frac{F_2}{S_2}\right) = \frac{F_1 S_2 + F_2 S_1}{2 S_1 S_2} \tag{2}$$

Where F1, F2 represents the number of points player 1 and player 2 won in serving games in a single match receptively and S1, S2 represents the total number of serves of player 1 and player 2 in a single match respectively.

•Model Construction

The formulation of the Indicator System is grounded in the specified characteristics. We have established a new indicator——"Momentum", which is employed to assess the performance state of athletes. Similarly, based on the data results from previous correlation analysis, we choose statistical indicators with high significance and practical reference value to form the "Momentum" shown in Figure 5. We defined computational formula of "Momentum" as:

$$M = PA + \beta \cdot SA + BP + R + WS + SW + GW \tag{3}$$

According to this derived equation, we have generated the momentum data for player 1 and player 2 as "p1_momentum" and "p2_momentum".

Figure 5. Composition of Momentum

2.5. Result

We use our model on the match "2023-wimbledon-1701" which is mentioned in the background of the question and visualize the momentum change during the whole match. Figure 6 is a area chart that a line chart that depicts the temporal change in the momentum of player 1 and player 2 and presents an exhaustive account of the fluctuations in the momentum during the match. Here, we utilize momentum as the performance evaluation index for the athletes. The results demonstrate that their change ranges roughly follow counteracting trends. Concurrently, we have selectively extracted set 5 to offer a more detailed depiction of the results shown in Figure 7. Also, we pinpoint the ace point and breaking point of each athlete at precise time points to better interpret the change in curves and mirror the chronology of the entire match.

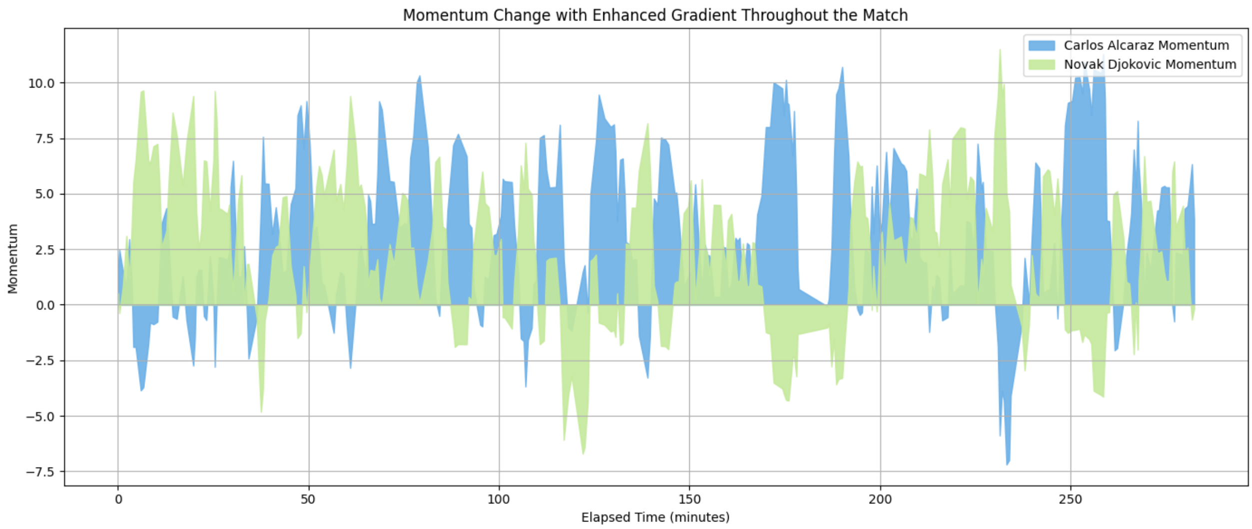

Figure 6. The Momentum Change of Player 1 and Player 2 Throughout the Whole Match

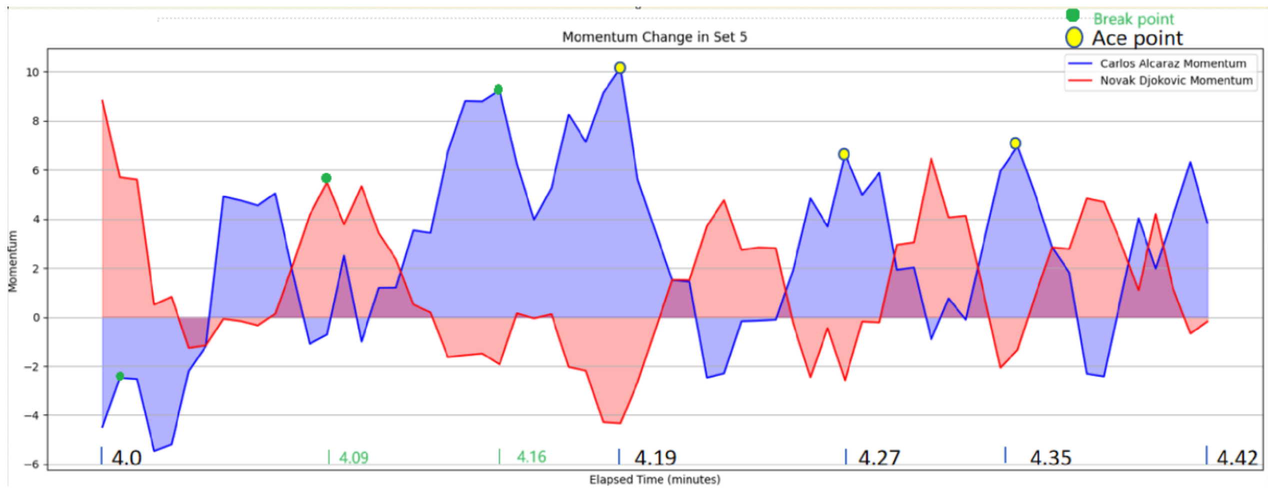

Figure 7. The Momentum Change of Player 1 and Player 2 in Set 5

3. The correlation between swings and momentum

3.1. Model Overview.

To assess the impact of momentum and investigate whether its transitions occur randomly, we employs momentum difference to demarcate turning points. Firstly, we build a Moving Average Algorithm to calculate the difference sequence by subtracting adjacent momentum based on the time series[6-8]. Simultaneously, we identify potential turning points by setting the time movement size and filtering thresholds of the algorithm throughout the game. Ultimately, we utilize the Runs Test to analysis the randomness of the momentum swing and runs of success. Our work in this task is shown below(Figure 8).

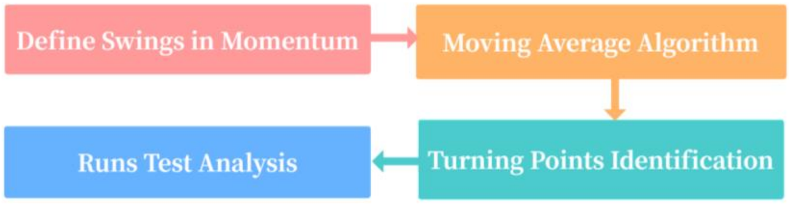

Figure 8. Flow Chart

3.2. Momentum Turning Point Identification Model Based on Moving Average Algorithm

3.2.1. Moving Average Algorithm.

(1) Moving Averages

In the calculation of momentum turning points, we employed the method of moving averages.

This involves determining turning points in a given region by calculating the average difference based on scores within a specific time sequence window [9]. The average of the data points xi within the window size w is defined as follows:

$$\overline{x_i} = \frac{1}{w}\sum_{j=i-\frac{w-1}{2}}^{i+\frac{w-1}{2}} x_j \tag{4}$$

(2) Difference Sequence Calculation

The difference sequence Δx_i can be expressed as the difference between the original sequence and the moving average sequence $\overline{x_i}$:

$$\Delta x_i = x_i - \overline{x_i} \tag{5}$$

(3) Turning Point Identification

Turning points can be identified by checking whether the absolute values of the difference sequence fall within a certain range:

$$|\Delta x_i| \in \ t \tag{6}$$

Where t is the filtering thresholds set parameter of the algorithm used to control the sensitivity of turning point detection[10].

3.2.2. Turning Point Identification and Visualization.

In Figure 9, we counted all the identified momentum turning points of the two players after using the algorithm in the match "2023-wimbledon-1701". Then, we visualized the turning points obtained using the aforementioned method on the momentum change area graph in Figure 10. Through the observation of these results, we observed that the momentum of each player undergoes relatively frequent transformations at different moments during the game, which reflects the player's performance status and scoring situation in a game in a timely manner.

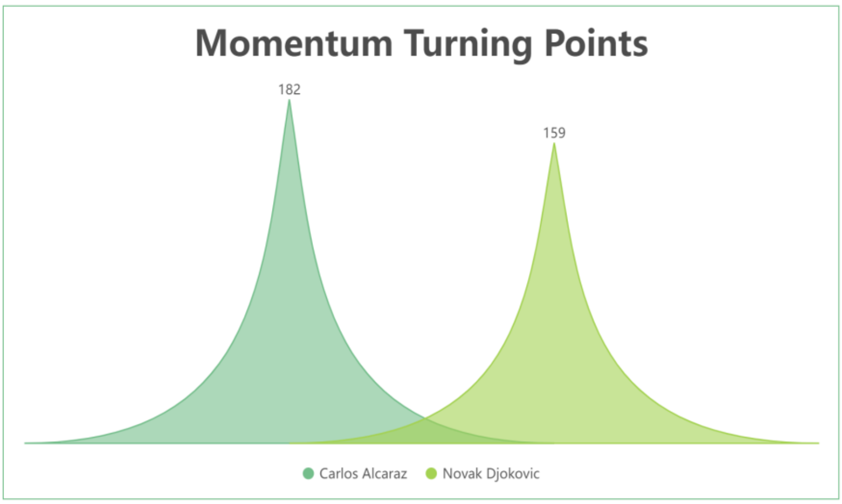

Figure 9. Numbers of Turning Points

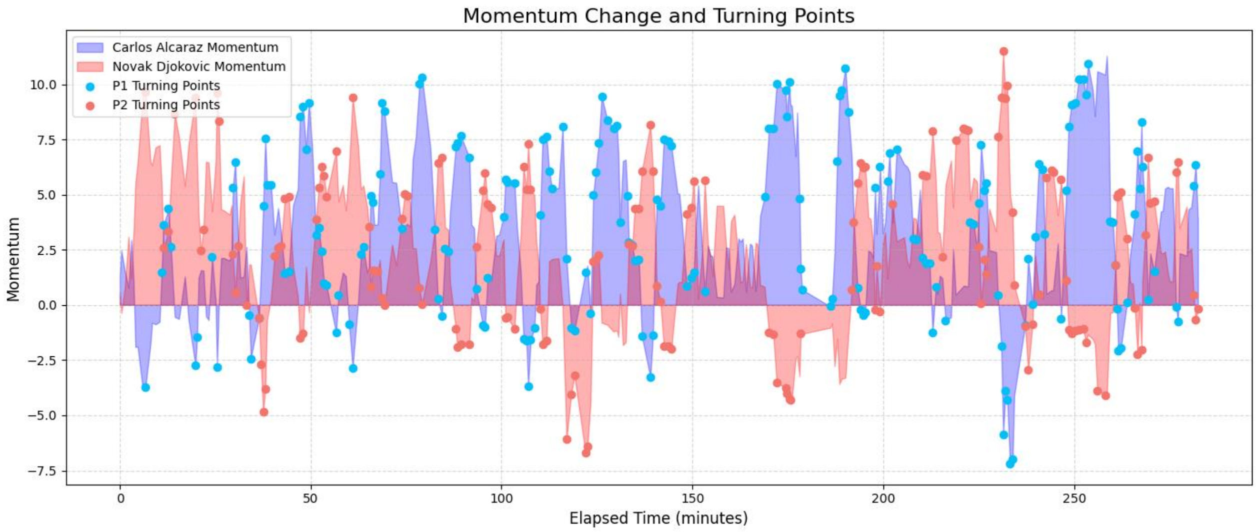

Figure 10. Graph of Sector Area with Turning Points for Dynamic Analysis

3.3. Correlation Between Swings and Momentum

In this research, we conducted a correlation analysis of turning points and momentum using a Runs test. Run test is used to test whether the occurrence of a certain event is random. We analyze whether each analysis item presents significance ($P<0.05$). If it is significant, the null hypothesis is rejected, indicating that the data is non-random; otherwise, it indicates randomness (Table 1).

Table 1. Result of Runs Test

Name	Sample size	z	P
p1_momentum	334	-12.217	0.000***
p2_momentum	334	-11.287	0.000***
p1_turning_points	334	-0.563	0.625
p1_turning_points	334	-0.614	0.548

Note: ***, ** and * represent significance levels of 1%, 5% and 10% respectively.

The table displays the results of the runs test, including the sample size, statistic, and the significance P-value. For p1_momentum and p2_momentum, the P-values are both very close to zero, indicating a very high level of statistical significance, suggesting a significant difference exists statistically. Rejecting the null hypothesis means that there is enough evidence to reject the null hypothesis at a given level of significance, indicating that the data is not random. However, for p1_Turning points and p2_turning points, their significance P-values are higher, at 0.625 and 0.528 respectively, which do not show significant levels and thus do not warrant rejection of the null hypothesis. Consequently, the data is considered to be random. In summary, we conclude that the swings in play are not random, while runs of success by one player may be random and the coach's claim is not correct.

4. Conclusion

In this paper, the concept model of "momentum" is constructed, and the sliding window technique is used to analyze the tennis players' match state. It is found that there is a significant correlation between "momentum" and the outcome of the game, and the change of "momentum" is not random. This provides an important reference for coaches and athletes to develop strategies and adjust their form. In the future, we can further expand the sample size, optimize the model, and try to apply it to other sports to verify the universality and validity of the model. This study provides a useful exploration for analyzing athletes' state volatility by using data science method.

References

[1] Mahato S, Gurao P N, Biswas K. Accelerated prediction of stacking fault energy in FCC medium entropy alloys using multilayer perceptron neural networks: correlation and feature analysis [J]. Modelling and Simulation in Materials Science and Engineering, 2024, 32 (3).

[2] Wang Y L, Li Y, Jiang Q, et al. Clinical characteristics and prognosis of acute myeloid leukemia patients with Runt-related transcription factor 1 mutation: A single-center retrospective analysis. [J]. Hematological oncology, 2024, 42 (2): e3256 - e3256.

[3] Luo X, Ling H, Xing M, et al. A dynamic-static combination risk analysis framework for berthing/unberthing operations of maritime autonomous surface ships considering temporal correlation [J]. Reliability Engineering and System Safety, 2024, 245110015.

[4] Sutherland N G, Cramer L C, III C W P, et al. Association of risk analysis index with 90-day failure to rescue following major abdominal surgery in geriatric patients [J]. Journal of Gastrointestinal Surgery, 2024, 28 (3): 215 - 219.

[5] Liu P, Liu K, Li J, et al. Predictive analysis and correlational exploration of public and academic attention in secondary vocational education [J]. Heliyon, 2024, 10 (4): e25947.

[6] Kovtun V, Giloni A, Hurvich C, et al. Pivot Clustering to Minimize Error in Forecasting Aggregated Demand Streams Each Following an Autoregressive Moving Average Model [J]. Stats, 2023, 6 (4): 1198 - 1225.

[7] Xu L, Ding F. Decomposition and composition modeling algorithms for control systems with colored noises [J]. International Journal of Adaptive Control and Signal Processing, 2023, 38 (1): 255 - 278.

[8] C. A P L. The Predict and Invert Feedback Active Noise and Vibration Control Algorithm [J]. Circuits, Systems, and Signal Processing, 2023, 42 (12): 7640 - 7650.

[9] Li Z, Cai Z, Zheng J, et al. Prediction of Lithium-Ion Battery Remaining Useful Life via Empirical Mode Decomposition- Autoregressive Integrated Moving Average and Regularized Particle Filter Algorithm [J]. Energy Technology, 2023, 11 (9):

[10] S. K, N. J. A Hybrid CNN-LSTM Approach for Monthly Reservoir Inflow Forecasting [J]. Water Resources Management, 2023, 37 (10): 4097 - 4121.