Modeling Tennis Matches Using Monte Carlo Simulations Incorporating Dynamic Parameters

J. Krčadinac, E.M. Maruševec, L. Jerković, I. Kovač, J. Zloić, A. Šarčević, M. Vranić
University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia
jakov.krcadinac@fer.hr

***Abstract* - Although it may seem to be one of the more unpredictable sports, tennis can be rather accurately modelled using the Monte Carlo method. This study aims to evaluate the accuracy of a Monte Carlo simulation that integrates dynamic tennis parameters in forecasting the outcome of a specific match. To predict the outcome of a tennis match, a conventional Monte Carlo simulation based on the identical and independent point distribution assumption requires only two parameters: the probabilities of both players winning a point on their own serve. A more sophisticated method proposed in this paper considers how fatigue affects a player's performance and it analyses and implements the change in the probability of winning a service point after "breaking" an opponent's service game. Calculating the relevant statistics required for player profiling was a critical step in this study. Following that, both previously mentioned variations of the Monte Carlo simulation were implemented to compare their performance. Finally, the method was tested on real-world tennis data.**

Keywords – ***tennis; Monte Carlo simulation; sport match dynamic parameters; prediction; comparing different Monte Carlo simulation versions***

I. INTRODUCTION

Sports analytics has been practiced for decades, but the breakthroughs in data collection have contributed to the popularization of that field in recent years. Sports analytics refers to managing structured data and applying predictive analytics models to analyse diverse aspects of the sports industry. The results of sports analysis are used to assist decision-makers in gaining a competitive advantage in the field of play [1]. Additionally, extensive analysis of historical data can even help in optimizing players' training programs to achieve the best results [2]. In most popular sports, data analysis and the usage of statistics have become commonplace.

Enormous investments in the sports industry are contributing to the immense growth of the sports analytics market in recent years. Consequently, this led to the rapid development of the online gambling and betting industry. Online gambling amounted to 61.5 billion U.S. dollars in 2021 and is expected to rise to 144.4 billion U.S. dollars by 2028 [3].

One of the most popular betting sports is tennis. Tennis is considered one of the hardest sports to play professionally because of its intense physical nature, the need for near-perfect hand-eye coordination during the whole match, and very little financial support for up-and-coming players. The players start their careers by playing in local and regional tournaments, and if they rank highly, they'll get a chance to play in ITF tournaments. If the players are also successful in ITF tournaments, they might get a chance to compete in WTA and ATP tournaments, which are the most important and prestigious tournaments in tennis [4]. Both ATP and WTA tournaments have a sophisticated method of ranking players based on their success in previous tournaments at that level, but most tennis fans are quite aware that those rankings are a very poor indicator when it comes to predicting the winner because lower-ranked players often beat higher ranked players [5].

Despite its seeming unpredictability, tennis is one of the few popular sports that can be rather accurately modelled using methods such as the Monte Carlo simulation, which is the subject of this research. Tennis is very suitable for modelling because a tennis match has no time limit, it is played until one of the players wins and there is no possibility of a draw. It is also an individual sport (in this paper we are not considering the outcome of doubles matches), so there are no complicated team relationships. All these factors contribute to tennis being one of the most popular betting sports [6].

The Monte Carlo method is a mathematical technique used for modelling and simulating complex systems by generating random samples. It generates a set of random numbers according to the data distribution and parameters for each variable to solve mathematical problems and make predictions [7]. This technique is widely used in various fields such as finance, physics, engineering, and computer science. It is particularly useful in problems where an analytical solution is not possible, and instead, the solution is estimated by repeating random trials [8]. Only one parameter for each player is needed to simulate a whole tennis match using the Monte Carlo method, and that is the **probability of a player winning the point when they are serving** [9]. The probability of winning a set, and consequently a match, is independent of which player serves first [10]. By alternating the serving player between games just like in a real tennis match and using

their probability of winning the point whilst simultaneously tracking the result, we can model the whole match. By simulating the match many times and counting how many times each player won, it's possible to estimate which of the two players is more likely to win the match.

The main idea of this paper is to test whether it is possible to improve this simple, already-known approach to predicting the winner of a match by also considering **dynamic parameters** of tennis, such as player **fatigue** and the effect of **one player "breaking" his opponent's service game**.

The research consisted of 3 main stages:

- Data cleaning and extraction of relevant player statistics from the dataset,
- Implementing the calculated statistics in custom versions of the Monte Carlo simulation to predict the winner of tennis matches that already took place in real life,
- Analyzing the results.

The upcoming sections of this paper describe the aforementioned stages of the research and provide further theoretical background on the Monte Carlo simulation. To conclude the research, we compared the winner prediction success rate of different Monte Carlo simulation versions.

II. DATA PREPARATION AND EXTRACTION OF RELEVANT PLAYER STATISTICS

The standard train-test machine learning method was used to test Monte Carlo simulation accuracy. To extract all the needed player statistics and later test the methods' accuracy, a large dataset was used that contains almost all tennis matches that took place in Challenger, ITF, WTA, and ATP tournaments during 2015 and 2016, along with every change in the current score during the match. The dataset was split into two parts – 70% of the data was used for training and 30% of the data was used for testing. The training dataset was also used to calculate the relevant statistics required for player profiling.

In this analysis, a real-life dataset consisting of the results of various tennis matches that were played in 2015 and 2016 was used. The dataset used is the so-called *in-play* dataset, which means that it is supposed to contain all the in-play changes in the score and the corresponding betting odds made by the betting house itself [11]. Apart from the in-play datasets, there are also so-called *play-by-play* datasets. Play-by-play datasets are the most insightful types of datasets in the world of sports. They contain much more information about the within-match events than in-play datasets, but appropriate tools for modelling and simulating that use play-by-play data have not yet been developed, as opposed to the ones that use in-play datasets [12]. Play-by-play is also better for analyzing sport events with more interaction between various of its components, e.g., football matches, where a play-by-play analysis is needed to reflect actual interplay and to qualify different interaction phases [13]. On the other hand, the in-play method is suitable for sport events with more discrete nature of the game, such as tennis, or cricket [14].

Before extracting the relevant statistics from the real-life dataset, the dataset must be thoroughly examined. Any matches that don't have the final result should be discarded, along with doubles matches because only singles matches are relevant for this research. Rows in which the current result is the same as in the previous row were also removed because they are redundant. Since there is much more available data on well-established players who regularly take part in WTA and ATP tournaments, the decision was made to also remove all matches from tournaments that are not of WTA or ATP level. The distribution of the number of matches played for all players in the dataset is shown on Figure 1. To be able to model the performance of individual players well, players that have fewer than 10 matches in ATP or WTA tournaments were removed from the dataset along with all the matches they participated in. Only matches where both players have played more than 10 matches in ATP or WTA tournaments were modelled using the Monte Carlo methods. After this step, 55% of the matches and 28% of the players remained in the training part of the dataset with an average of 11.4 matches per player.

One other problem arose when we tried to analyze the data, and that are the names of players containing non-ASCII characters. There were a few different versions of those players' names in the dataset. For example, there were 4 versions of the name "Marin Čilić", each containing different non-ASCII characters. After identifying all players with non-ASCII characters in their names, those names were lemmatized and standardized (all variations of "Marin Čilić" were changed to "Marin Cilic").

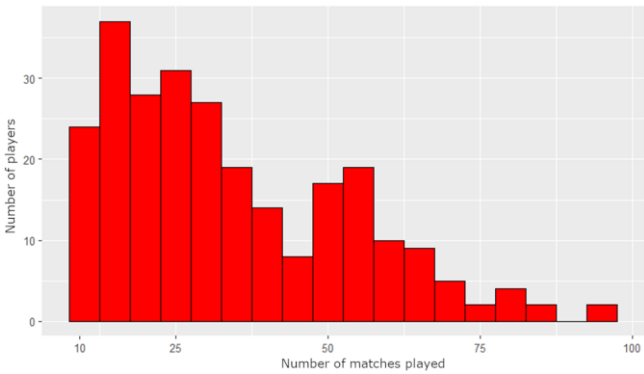

Figure 1. Distribution of the number of matches played for all players in the dataset

After cleaning the data, it was possible to calculate the probability of winning a point when serving for each player in the dataset. That statistic is later used in the basic Monte Carlo prediction model that doesn't take the dynamic parameters of tennis into consideration. One of the dynamic parameters is fatigue, and to test how players perform under fatigue it was decided to differentiate between the probability of winning a point when serving during the first 2 sets and if the game goes past the second set, when almost all players will start to feel fatigued. After calculating the probability of winning a point when serving in the first two sets and the probability of winning a point after the second set, it was clear that the top players on the tour like Djokovic, Nadal, and Federer

manage to either keep or even slightly improve their probability of winning the point when serving during the whole match. When taking into consideration the effect of one player "breaking" his opponent's service game, which is the second dynamic parameter that we propose to be used in the Monte Carlo method, the probability of a player winning the point after breaking their opponent was calculated, along with the probability of winning when the opponent breaks them.

III. Modelling the matches using Monte Carlo methods

A. Monte Carlo method

The term *Monte Carlo methods* refers to a class of computational algorithms that use random sampling from a probability distribution to solve a problem that involves a large number of degrees of freedom and is usually of mathematical or statistical nature. This research uses the Monte Carlo method to simulate real-life tennis matches many times and determine the more likely winner. Monte Carlo methods have a wide range of applications in many different fields of science such as radiological sciences [15], as well as in engineering [16] and finance [17].

One of the earliest applications of Monte Carlo methods was in late 1940s, when physicists Nicholas Metropolis and Stanislaw Ulam developed a method to solve complex integrals that were not analytically computable [18]. Monte Carlo methods have become widespread ever since, finding applications in numerous fields of scientific computing [19].

A variety of problems, from simple integrals to complex simulations of advanced multivariable systems, can be solved by using Monte Carlo methods. Handling systems with uncertain or random inputs is one of the crucial strengths of these methods, which allows researchers to precisely estimate the probability of rare events, as well as to measure the quantity of their results [20].

There are many variations on the basic Monte Carlo method, including importance sampling [21], Markov chain Monte Carlo [22], and quasi-Monte Carlo methods [23]. Importance sampling can improve the efficiency of Monte Carlo simulations by focusing on the most important parts of the probability distribution. Markov chain Monte Carlo methods use a series of dependent random samples to explore complex distributions, while quasi-Monte Carlo methods use low-discrepancy sequences of points to generate more representative samples [24].

One of most widely known applications of Monte Carlo methods is in the financial domain. Experts use them to simulate the behaviour of financial markets [25] and to find the oprimal way of trading financial instruments such as stocks or ETFs. They can be used to assess the probability of a stock portfolio losing certain amount of value over a specific interval, or to appraise an option that depends on the value of an underlying asset [26].

Along with numerous advantages, Monte Carlo methods also have some disadvantages. They can require extremely high computational resources, especially for problems with multidimensional data space or multivariate dependencies, which can be difficult to implement [27]. However, as the system size grows, the problem converges to a simpler, deterministic limit that is cheap to solve [28]. To ensure the accuracy and reliability of Monte Carlo simulations, researchers must pay close attention to the quality and representativeness of the random samples used in the simulation.

In general, Monte Carlo methods are an efficient, multifunctional and very flexible tool used for predicting the most likely outcome and simulating complex systems. That is why researchers in a wide range of scientific fields tend to use them for gaining insight into a variety of problems.

B. Standard Monte Carlo approach

The basic Monte Carlo model used in this research is a point-based model because it only needs the probability of Player A winning any given point versus Player B for predicting the winner of the match [29]. Monte Carlo tennis simulation starts from a random number generator that generates values for the success or failure of both players p_a and p_b sampled from a uniform distribution on the interval [0, 1]. When player A is serving, a value on the unit interval is sampled and if it lies in the range [0, p_a], player A wins the point. If not, player B wins that point. When player B is serving, if the value on the unit interval is in the range [0, p_b], player B gets the point. If not, player A gets that point. [30]. By keeping score after simulating each point and alternating the serve between players just like in a real tennis match, the model predicts the possible result and match winner. Figure 2 shows this scoring system with probabilities used when the player A is serving.

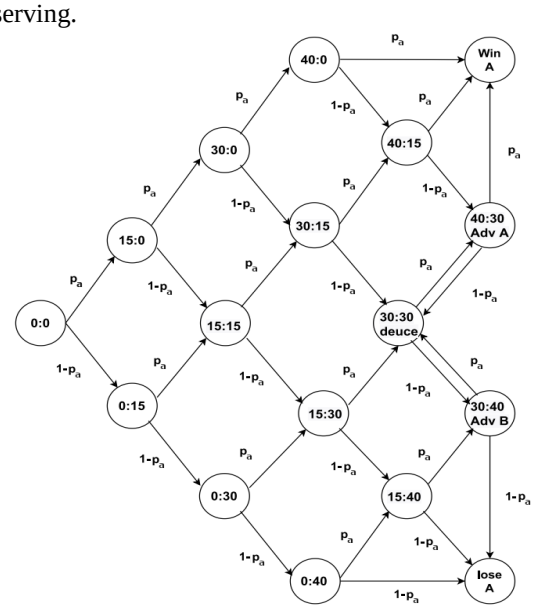

Figure 2. Markov chain for a game in which Player A is serving [37]

Of course, the basis of a Monte Carlo method is not to simulate the whole match just once, but many times and reach the conclusion of who is the more likely winner

[31]. Each match in our test dataset was simulated 400 times. Two other versions of the Monte Carlo model that incorporate dynamic parameters of tennis were also used to predict the outcome of tennis matches, and the accuracy of all three Monte Carlo simulations was compared.

C. *Monte Carlo simulation incorporating dynamic parameters*

As in any sport, fatigue is one of the biggest factors when it comes to an athlete's performance. In tennis, the players need to be very explosive and precise when returning or serving the ball whilst being able to keep that level of explosive strength during the entire match. Because of that, fatigue seems to be one of the most important dynamic parameters to consider when predicting tennis matches. Of course, almost all of the players in ATP and WTA tournaments, especially those at the highest level, try to adjust their style of play during the match to better deal with fatigue [32]. Despite that, in this paper it was assumed that almost all players will start to feel significantly fatigued after 2 sets of play, which equates to approximately an hour and 20 minutes, not taking rest breaks into consideration [33]. That's why it was chosen to differentiate between the percentage of points won when serving during the first two sets, and the percentage of points won after the game goes past the second set. Both percentages were calculated by looking at the in-play data from all the matches in the training part of the dataset. The results from the first two sets were then used when calculating the percentage of points won during them, and when calculating the percentage of points won after the game goes past the second set only results from the third, fourth, or fifth set were considered. Both percentages were used accordingly in the Monte Carlo simulation that incorporates dynamic parameters. This is shown in Figure 3, where it is visible that after the game goes past the second set, the percentage of points won when tired is used instead of the percentage of points won during the first two sets.

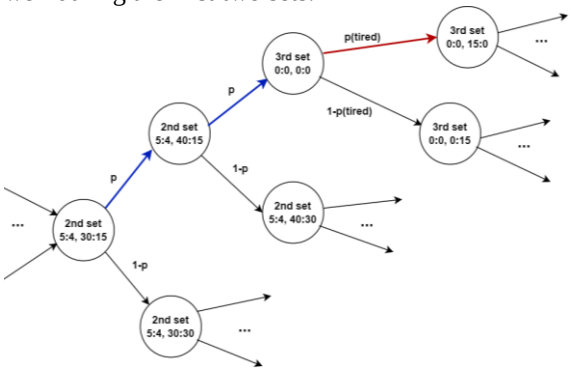

*Figure 3. When the simulated match enters the third set, the Monte Carlo simualtion starts using the calculated percentage of points won after the game goes past the second set (**p(tired)**) instead of the the percentage of points won during the first two sets (**p**)*

In the world of sports, another widely researched parameter of sport dynamic is momentum or the so-called *"hot hand"*. "Hot hand" is a situation where one player exceeds their normal expectations because of a previous victory [34]. This effect is considered very important in table tennis [35] as well as in regular tennis matches [36]. In this paper, we considered the momentum in break situations. In tennis, a break is when a player wins a game as the receiving player, thereby they break the opponent's service game. To account for break situations in matches, we calculated the percentage of points won only in the moments when one player broke more of their opponent's serves in a set than the opponent because in those situations one player has the advantage over the other. Those situations were registered by observing results in breaks whilst considering is serving. If one player is serving and he has more break points than the opponent, that indicates he/she broke at least one more of opponent's serves than his opponent did, and from that point on his/her percentage of points won was recorded. This observed situation lasted until the set was over or when the number of broken opponent's serves from both players become equal. That situation was registered by checking when the player who is serving will have equal or just one break point less than his opponent. Percentages calculated by this method were used in the Monte Carlo simulation that incorporates dynamic parameters in the previously described moments. If there was both a situation of advantages in broken serves and long match duration at the same time, the mean of both percentages was used.

Three versions of the Monte Carlo simulation were implemented in this research. First was standard Monte Carlo. Second was Monte Carlo simulation which only considers fatigue and the third that considers both fatigue and the effect of one player breaking the other.

IV. RESULTS

The accuracy of all the aforementioned Monte Carlo simulations was tested on a testing part of the dataset which consisted of 1232 matches that took place in 2016 on ATP or WTA tournaments which were not used to calculate relevant player statistics. Each match was simulated 400 times to determine the more likely winner, after which the simulation winner predictions were compared to the real winner of each match. For example, in the match between Novak Djokovic and Kei Nishikori that took place at the Australian Open on 26 January 2016, which Djokovic won, the basic version of the Monte Carlo simulation calculated that the chance of Djokovic winning is 58, 50%, the version of Monte Carlo that considers only fatigue calculated the chance of Djokovic winning to be 61,25% and the third version the Monte Carlo method that considers both fatigue and the effect of one player breaking the other calculated the chance of Djokovic winning to be 57,25%. The likelihood of a player winning the match is a percentage calculated by dividing the number of simulated matches in which they won and the number of simulated matches.

Table I presents the number and the percentage of correct winner predictions using each Monte Carlo simulation variations. The accuracy of all methods ended up being almost identical, with all having around 60% correct guesses.

TABLE I
PREDICTION ACCURACY OF DIFFERENT MONTE CARLO METHODS

Monte Carlo simulation	Number of correct winner predictions (out of 1232)	Percentage
Basic	756	61,36%
Including player fatigue	731	59,33%
Including both player fatigue and the probability of winning after a break	725	58,85%

If the Monte Carlo methods calculate that the chance of one player winning the match is just slightly higher than 50%, that means the players are very evenly matched in terms of their skill, and the prediction is highly uncertain. By considering only the matches where the Monte Carlo method is significantly more than 50% certain of the win of one player, each version becomes more accurate in predicting the winner. Table II shows the number and the percentage of correct predictions when one player is more than 70% likely to win. In these circumstances, the percentage of correct guesses increases to 67,21%. Figure 4. shows how accuracy of the Monte Carlo simulation improves when one player has a 70% chance of winning rather than both players having almost equals chance of winning.

TABLE II

PREDICTION ACCURACY OF DIFFERENT MONTE CARLO METHODS WHEN ONE PLAYER IS **MORE THAN 70% LIKELY TO WIN**

Monte Carlo method	Correct winner predictions	Percentage
Basic	378 out of 572	66,08%
Including player fatigue	295 out of 439	67,20%
Including both player fatigue and the probability of winning after a break	291 out of 433	67,21%

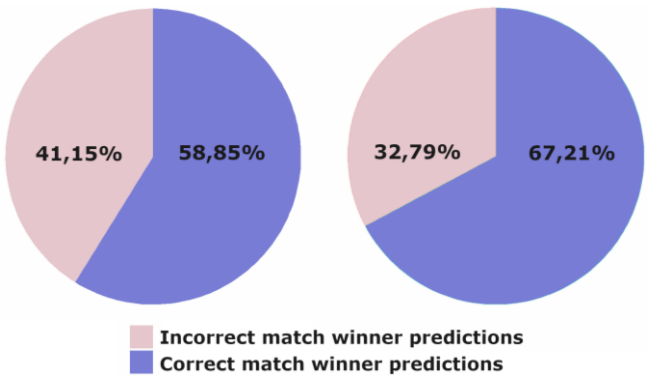

*Figure 4. Accuracy of the Monte Carlo simulation that consideres both player fatigue and the probability of winning after a break when the chance of one player winning is **50%** or more (**left pie chart**) and when the chance of one player winning is **70%** or more (**right pie chart**)*

V. CONCLUSION

The goal of this study was to compare the accuracy of predicting the match winner between different versions of the Monte Carlo simulation. One version uses only the percentage of points won, another version considers both the percentage of points won and player fatigue during the match and the last version uses all of the previously mentioned factors as well as the effect of one player breaking his opponent's service game. The dataset used was a real-life in-play dataset. It contained a lot of irrelevant data that needed to be removed, such as matches without the final result, doubles matches etc. Afterwards, the impact of fatigue on players' performance and the change in the probability of winning a service point after "breaking" an opponent's serve could be calculated. Lastly, all of the previously mentioned Monte Carlo simulation versions were implemented using the calculated player statistics and tested on a different dataset containing matches from 2016.

Even when all the tested versions of Monte Carlo simulations are more than 70% sure of the win of one player, the accuracy of their predictions is about 67%. By using different, perhaps more theoretical models instead of the Monte Carlo simulation such as point-based models, paired comparison models, and machine learning models, the prediction accuracy still doesn't surpass 70% [37]. Wilkens [38] also concludes that the average prediction accuracy cannot be increased to more than about 70% regardless of the model used. This further proves that tennis is a highly unpredictable and dynamic sport, even though it can be easily and accurately modelled. It is impossible to quantify dozens of factors such as emotional state, injuries, and fan support that can influence the outcome of tennis matches [39]. Incorporating dynamic parameters of tennis and building a more sophisticated prediction model based on the Monte Carlo method does not necessarily result in a significant increase in correctly predicted match winners. The results could perhaps be improved by testing different ways of using the calculated player statistics in the simulation that considers dynamic parameters, but there is little hope that the increase in accuracy would be significant.

ACKNOWLEDGMENT

This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

REFERENCES

[1] E. Morgulev, O. H. Azar, and R. Lidor, "Sports analytics and the big-data era", International Journal of Data Science and Analytics, 2018, pp. 213-222.

[2] A. Knobbe, J. Orie, N. Hofman, B. Burgh, and R. Cachuco, "Sports analytics for professional speed skating", Data Mining and Knowledge Discovery, 31, 2017, pp. 1872-1902.

[3] Statista, Market size of the online gambling and betting industry worldwide in 2021, with a forecast for 2028, viewed 16 January 2022, <https://www.statista.com/statistics/270728/market-volume-of-online-gaming-worldwide/>.

[4] TennisGate LTD, You do not choose to be a professional tennis player – it chooses you, viewed 6 February 2023, <https://www.tennisgate.com/you-do-not-choose-to-be-a-professional-tennis-player-it-chooses-you/>.

[5] Tennis rankings: How they work and difference between ATP and WTA systems (ATP Rankings – men's singles and doubles tennis rankings), viewed 6 February 2023 <https://olympics.com/en/news/tennis-rankings-atp-wta-men-women-doubles-singles-system-grand-slam-olympics>.

[6] Top 7 biggest betting sports in the world, viewed 19 February 2023 <https://mwwire.com/2022/11/02/top-7-biggest-betting-sports-in-the-world/>.

[7] K. R. Koch, "Monte Carlo methods", GEM - International Journal on Geomathematics, 9, pp. 117-143, 2018

[8] D. Luengo, L. Martino, M. Bugallo, V. Elvira, and S. Simo Särkkä, "A survey of Monte Carlo methods for parameter estimation", EURASIP Journal on Advances in Signal Processing, 25, 2020

[9] P. K. Newton, and K. Aslam, "Monte Carlo tennis", SIAM Review, 48(4), 2006, pp. 722-742.

[10] P. K. Newton, and J. B. Keller, "Probability of Winning at Tennis I. Theory and Data," Studies in applied Mathematics, 114(3), 2005, pp.241-269.

[11] A. Šarčević, D. Pintar, M. Vranić, and A. Gojsalić, "Modeling in-match sports dynamics using the evolving probability method", Applied Sciences-Basel, 11(10), 2021

[12] P. Vračar, E. Štrumbelj, and I. Kononenko, „Modeling basketball play-by-play data", Expert Systems With Applications, 44, 2016, pp. 58-66

[13] F. Korte, D. Link, J. Groll, and M. Lames, 2019., „Play-by-Play network analysis in football", Frontiers in Psychology, 10, 2019.

[14] M. Asif, and I.. G. McHale, „In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model", International Journal of Forecasting, 32(1), 2016, pp. 34-43

[15] R. L. Morin, Monte Carlo Simulation in the Radiological Sciences, CRC Press, Boca Raton, 2019

[16] X. Zhang, Z. Lu, K. Cheng, and Y. Wang, "A novel reliability sensitivity analysis method based on directional sampling and Monte Carlo simulation", Proceedings of the Institution of Mechanical Engineers Part O Journal of Risk and Reliability, 234(19), pp. 622-635, 2020

[17] M. S. Shahriar, K. B. M. R. Hasan, T. Hossain, T. H. Beg, K. M. A. Islam, and N. M. Zayed, "Financial decision making and forecasting techniques on project evaluation: A planning, development and entrepreneurial perspective", Academy of Entrepreneurship Journal, 27(4), pp. 1-7, 2021

[18] M. Betancourt, "The Convergence of Markov Chain Monte Carlo Methods: From the Metropolis Method to Hamiltonian Monte Carlo", Annalen der Physik, 531(3), 2017

[19] A. Mücke, R. Engel, J. P. Rachen, R. J. Protheroe, and T. Stanev, "Monte Carlo simulations of photohadronic processes in astrophysics", Computer Physics Communications, 124(2-3), 2000, pp. 290-314

[20] D. Xiu, and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs", SIAM Journal of Scientific Computing, 27(3), 2005, pp. 1118-1139

[21] T. Müller, B. Mcwilliams, F. Rousselle, M. Gross, and J. Novák, "Neural Importance Sampling", ACM Transactions on Graphics, 38(5), pp. 1-19, 2019

[22] C. Nemeth, and P. Fearnhead, "Stochastic Gradient Markov Chain Monte Carlo", Journal of the American Statistical Association, 116(533), pp. 433-450, 2021

[23] G. Ökten, and Y. Liu, "Randomized quasi-Monte Carlo methods in global sensitivity analysis", Reliability Engineering & System Safety, 210, 2021

[24] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Markov chain Monte Carlo in practice, Chapman & Hall/CRC, Boca Raton, 1996

[25] L. Gan, H. Wang, and Z. Yang, "Machine learning solutions to challenges in finance: An application to the pricing of financial products", Technological Forecasting and Social Change, 153, 2020

[26] D. W. H. Alrabadi, N. I. A. Aljarayesh, "Forecasting stock market returns via Monte Carlo simulation: The case of Amman stock exchange", Jordan Journal of Business Administration, 11(3), 2015.

[27] R. Bardenet, A. Doucet, and C. Holmes, "On Markov chain Monte Carlo methods for tall data", Journal of Machine Learning Research, 18, pp. 1-43, 2017

[28] D. F. Anderson, D. J. Higham, and Y. Sun, "Computational complexity analysis for Monte Carlo approximations of classically scaled population processes", Society for Industrial and Applied Mathematics, 16(3), pp. 1206-1226, 2018

[29] Chew the Stat, MATHS POINT: Finding Tennis' Winning Formula, viewed 12 February 2023, <https://chewthestat.com/finding-the-winning-formula/#:~:text=Points%2Dbased%20modelling%20operates%20on,given%20point%20versus%20Player%20B.>

[30] K. Aslam, "A stochastic Markov chain approach for tennis: Monte Carlo simulation and modeling", Thesis(Ph. D.), University of Southern California, 2012.

[31] E. Koehler, E. Brown, and S. J.-P. A. Haneuse, "On the assessment of Monte Carlo error in simulation-based statistical analyses", The American Statistician, 63(2), 2009, pp. 155-162.

[32] M. Reid, and R. Duffield, "The development of fatigue during match-play tennis", Br J Sports Med., 48, 2014.

[33] M. Ferrante, G. Fonseca, S. Pontarollo, 2017, "How long does a tennis game last?", in Proceedings of MathSport International 2017 Conference, 2017.

[34] M. Bar-Eli, S. Avugos, and M. Raab, "Twenty years of "hot hand" research: Review and critique", Psychology of Sport and Exercise, 7(6), 2006, pp. 525-553

[35] R. J. R. Den Hartigh, and C. Gernigon, "Time-out! How psychological momentum builds up and breaks down in table tennis", Journal of Sports Sciences, 36(23), 2018, pp. 2732-2737

[36] H. Dietl, and C. Nesseler, "Momentum in tennis: controlling the match", UZH Business Working Paper Series 365, University of Zurich, 2017.

[37] Ana Šarčević, Mihaela Vranić, Damir Pintar, Agneza Krajna, 2022., "Predictive modeling of tennis matches: a review", Point based models

[38] S. Wilkens, "Sports prediction and betting models in the machine learning age: The case of tennis", Journal of Sports Analytics, vol. 7, no. 2, pp. 99-117, 2021.

[39] M. de Araujo Fernandes, "Using Soft Computing Techniques for Prediction of Winners in Tennis Matches", Machine Learning Research, 2(3), 2017, 86–98.