Applying Coverage Control Models to the Position of Baseball Players

Kerry Johnson and Erick J. Rodríguez-Seda
Department of Weapons, Robotics, and Control Engineering
United States Naval Academy
Annapolis, MD
m263156@usna.edu, rodrigue@usna.edu

***Abstract*—Coverage control models are widely applied in areas such as data compression, resource allocation, and the modeling of animal behavior. These models have facilitated the optimization of complex systems. Recent research has demonstrated that coverage control methods can also be used to analyze the spatial interactions of athletes. This article applies coverage control techniques to study the positioning of baseball players, incorporating player history as a key variable. The study explores various approaches to modeling player range and ability while adhering to the positional rules of Major League Baseball (MLB). The model's outcomes are compared to MLB statistics, including actual player positioning and field coverage efficiency. By leveraging MLB's advanced camera systems and statistical archives, the model was refined to accurately reflect real-world conditions. The model created increases the ability for the players to cover the playing field with a relatively small change in positioning. The development of Voronoi cells in this context revealed spatial patterns and trends in sports, demonstrating broader applications in fields such as military personnel allocation and multi-agent systems for situational awareness.**

***Index Terms*—Coverage control, Multi-agent systems, Lloyd's algorithm, Voronoi cells**

I. Introduction

The MLB has undergone a significant evolution in data analysis, primarily driven by the advent of Statcast [1]. This high-tech tracking system provides coaches, players, fans, and broadcasters with unprecedented insight into on-field events. By combining radar, cameras, and optical tracking technologies, Statcast tracks virtually every aspect of the game, including metrics such as pitch velocity, exit velocity, and launch angle [2].

MLB began its transition toward this data-rich environment by installing pitch-tracking systems in every stadium, initially to better understand the game's complexities. Statcast was officially introduced in 2015 following a year-long pilot program in 2014, with installations in all 30 MLB stadiums [3]. This expansion allowed for the tracking of dynamics beyond pitching, including player positioning, batting mechanics, and sprint speed. In 2020, MLB integrated Hawk-Eye, a high-speed camera system renowned for its precision in tennis [4]. Currently, every stadium utilizes 12 Hawk-Eye cameras dedicated to tracking player interactions, providing a comprehensive dataset for analysis [2].

Crucially, MLB publishes these data publicly, democratizing analytics for fans and researchers. The Baseball Savant platform provides access to data for every pitch thrown since 2015, resulting in a heavier emphasis on statistical analysis in baseball than ever before [2].

Sabermetrics (originally SABRmetrics) serves as the foundation for modern sports analytics in the United States [5]. Defined by the University of Syracuse as "*the empirical analysis of baseball through statistics, used to predict the performance of players, giving teams a winning edge*," the term was coined by baseball analyst Bill James in 1980 [6].

The practical application of Sabermetrics gained mainstream recognition through Michael Lewis's book, *Moneyball: The Art of Winning an Unfair Game*, and its subsequent film adaptation. The narrative depicts Oakland Athletics manager Billy Beane's strategy of leveraging statistical analysis to overcome financial constraints. By identifying undervalued players through statistics, the Athletics demonstrated that Sabermetrics could level the playing field against teams with significantly larger payrolls.

To analyze MLB data effectively, this study proposed the use of Voronoi diagrams and geometric computational methods [7]. A Voronoi diagram assumes a set of N agents scattered across a region. The region is subdivided into an equal number of cells, with every boundary equidistant between two or more points, as shown in Fig. 1. Consequently, any location within a specific cell is closer to the generating point of that cell than to any other point in the region.

Voronoi patterns are prevalent in nature, architecture, and art due to their spatial efficiency. Growth that occurs at a uniform rate from separate points—such as the melanin-secreting cells determining the spots on a giraffe—naturally exhibits Voronoi patterns [8].

Optimization of these cells is often achieved using Lloyd's algorithm [9]. This iterative process constructs Voronoi diagrams and calculates the centroid (geometric center) of each cell. After a cell is constructed, the generating point is moved toward the centroid, and the process is repeated. This algorithm ultimately drives points toward locations that distribute the cells more evenly. Applications of Lloyd's algorithm range

The views expressed in this document are those of the author(s) and do not reflect the official policy or position of the U.S. Naval Academy, Department of the Navy, the Department of War, or the U.S. Government.

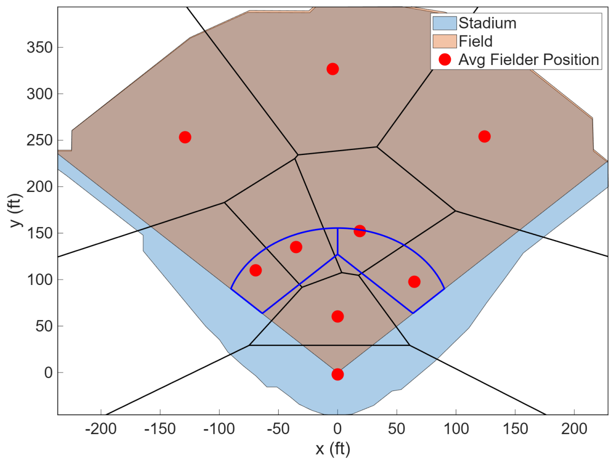

Fig. 1: A Voronoi diagram of players on a baseball field using euclidean distance to define the boundaries of the cells.

from quantization to lossy data compression, making it highly effective for optimizing shared space.

This project explored the application of Voronoi diagrams to optimize the positioning of baseball players. By utilizing Voronoi diagrams, the study identified spatial patterns and developed methodologies for constructing cells that characterize player interactions. As the MLB continues to pioneer the adoption of Sabermetrics, this research leveraged Voronoi techniques to enhance the understanding of human spatial dynamics in sports. These findings regarding human spatial interaction suggest broader applicability, specifically for military applications regarding coverage control, battlespace awareness, and security detail personnel allocation.

II. Problem Statement

This study addresses the following question: **"Can coverage control models based on player data be used to position defensive players before a pitch is thrown?"**

The data implemented in this research included the historical distribution of a batter's hit locations and the average recorded MLB fielder positions. The resulting model was compared to the positioning strategies of modern baseball teams, evaluating the divergence between the model's calculated optimal positions and the actual positions taken by players.

The analysis relied on the assumption that players make optimal decisions and execute performance successfully once the pitch is thrown. To achieve this, the model demonstrated the use of heterogeneous Voronoi diagrams, constructed using a density function derived from batter data and weighted by the relative distance from home plate.

III. Literature Review

Centroidal Voronoi Tessellations (CVTs) are utilized across a wide range of disciplines, including image processing, quadrature finite difference methods, resource distribution, cellular biology, and statistics [7]. The versatility of CVTs stems from their adaptability to optimization problems. In a standard Voronoi diagram, every region has a center of mass that is closer to any point within that region than the center of any other region. This geometric property makes Voronoi diagrams particularly effective for coverage control and the positioning of agents relative to one another and their environment.

Recent research has extended these concepts to athletics. For instance, a study in [10] examined the relationship between coverage control methods and the movement of Futsal players. By analyzing camera recordings of Futsal games, the researchers identified spatial patterns between players using Voronoi diagrams, concluding that player behavior could be reliably predicted using these geometric models.

However, typical centroidal Voronoi diagrams rely on the assumption of a homogeneous environment, which leads to inefficiencies when environmental factors vary. Research by [11] highlighted this issue in the context of drone coverage. While drones can position themselves based on a static environment, assuming homogeneity implies equal travel time in all directions. The introduction of wind alters the effective distance a drone can cover, depending on direction. Therefore, accurate coverage control methods must account for dynamic environmental variables.

Heterogeneity also applies to the coverage capabilities of individual agents. A control strategy developed by [12] addressed robots with varying sensing ranges. In this model, each agent possesses a specific *region of dominance* that it can cover within a set period. By adjusting boundary definitions based on agent characteristics, this strategy capitalizes on the extended range of high-performing agents while reducing the load on lower-performing ones.

In the context of baseball, the region a player can defend has been modeled as an ellipse by [13]. This approach used player data to determine the range within which a player can successfully field a ball. This ellipse effectively represents a player's region of dominance, which may expand or contract based on the distance from a hit ball and the time available before the ball lands.

Drawing upon this literature, this project integrated environmental factors that dictate where players can and cannot be positioned and differences in ball park dimension. The environment was defined by the specific dimensions of the field—accounting for variations in ballpark size—and by batter data, where a density function represented the number of hits to a location.

IV. Methodology

This study utilized player data from the Baseball Savant database, hosted by the MLB, encompassing multiple seasons. Data were extracted on a player-by-player basis, documenting every pitch a hitter saw over the specified period. Key variables included hit coordinates (x, y) and the actual positioning of defensive players, as seen in Fig. 2.

To construct the simulation environment, the study processed the hit locations for specific batters, removing data points landing outside the field of play. These data were used

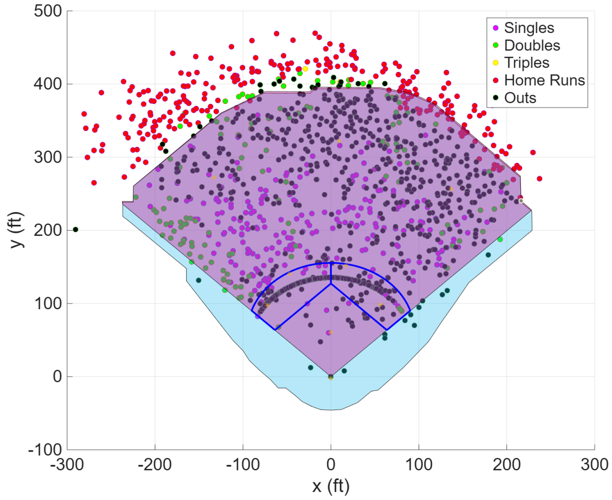

Fig. 2: A spray chart based on player data.

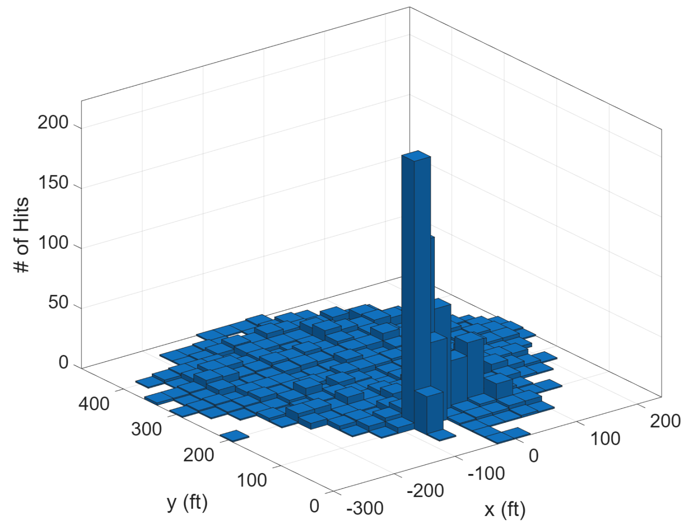

Fig. 3: A discrete density function representing hit frequency and location.

to generate a density function (DF), $p(x, y)$, representing the weighted number of balls landing in a specific area. Ground balls were moved to the midpoint of the infield region collinear with the direction of the hit:

$$\theta_{spray} = -\arctan\left(\frac{x - 130}{213 - y}\right) \tag{1}$$

$$D_{new} = 60.5\cos\theta_{spray} + \sqrt{75^2 - (60.5\sin\theta_{spray})} \tag{2}$$

where θ_{spray} is the direction of the hit and D_{new} is the midpoint of the infield in the direction of the hit. This represents the ability of the fielders to play ground balls and throw the runner out, effectively moving the ground balls into the infielders' regions.

The density function was weighted according to the distance from home plate where $w(x, y)$ represents the weight for each hit. This increases the density in certain regions to reflect the severity of deeper hit balls as compared to balls hit into the infield. The distance from home plate is divided by 1000 so that the minimum value will be 1 when located at home plate and the maximum value will be between 1.3 and 1.4 when located at the furthest point, typically between 350 and 400 feet, depending on the ball park.

$$w(x, y) = 1 + \frac{\sqrt{x^2 + y^2}}{1000} \tag{3}$$

The DF can be visualized as the frequency of hits to different regions, as illustrated in Fig. 3.

The core of the simulation involved constructing Voronoi diagrams to optimize defensive positioning. The field, W, was partitioned into N regions (Voronoi cells), W_i, where each cell represented the area of responsibility for a specific player (agent). The efficiency of the field coverage was evaluated using a Locational Cost function, H_i, defined as [9]:

$$H_i = \iint_{(x,y)\in W_i} ((x_i - x)^2 + (y_i - y)^2) \cdot \rho(x, y)\, dx\, dy \tag{4}$$

where (x_i, y_i) represents the position of the ith player. The objective is to minimize the total cost function, $H = \sum_{i=1}^{N} H_i$, thereby maximizing coverage efficiency relative to the batter's history.

The proposed approach uses Lloyd's algorithm to minimize the locational cost. Lloyd's algorithm is an iterative control law that drives agents toward the centroid of their respective Voronoi cells. Since the probability density function $p(x, y)$ remained static for a given batter, Lloyd's algorithm was a suitable choice for convergence.

In this approach, the position of each player is modeled as single integrator:

$$\dot{q}_i = u_1 = \begin{bmatrix} u_{i,x} \\ u_{i,y} \end{bmatrix} \tag{5}$$

where q_i are the Cartesian coordinates of ith defensive player. The control law u_i is based on Lloyd's algorithm:

$$u_i = K_p(c_i - q_i) \tag{6}$$

where $K_p > 0$ is a proportional gain and $c_i = [c_{i,x}, c_{i,y}]^T$ is the center of mass, or centroid, of the Voronoi cell, computed according to:

$$m_i = \iint_{W_i} \rho(x, y)\, dx\, dy \tag{7a}$$

$$c_{i,x} = \frac{1}{m_i} \iint_{W_i} x \cdot \rho(x, y)\, dx\, dy \tag{7b}$$

$$c_{i,y} = \frac{1}{m_i} \iint_{W_i} y \cdot \rho(x, y)\, dx\, dy \tag{7c}$$

where m_i represents the ith cell's mass.

The application of Lloyd's algorithm was subject to specific MLB positional rules and game states:

- Fixed Agents: The pitcher and catcher were constrained to the mound and home plate, respectively.

Algorithm 1 Main Loop

for each time step k **do**
  for each agent i **do**
    Draw Voronoi Cells, W_i
    Calculate Centroid, c_i
    Move Players with Proportional Control, u_i
    if agent i is positioned outside their allowable region **then**
      Reassign position inside their region
    end if
    Calculate Locational Cost
  end for
end for

- Infield Constraints: Four infielders were required to remain within the infield boundary, with two positioned on either side of second base.
- Game State: The simulation modeled "bases empty" scenarios only, prioritizing ball interception over base coverage.

The control law for the agent positioned around first base was changed to incorporate an emphasis on proximity to first base. A proportional attraction to the location of first base was added to the proportional attraction to the centroid:

$$u_{ix} = K_p(c_{ix} - q_{ix}) + K_1(90 \cos 45 - x) \quad (8)$$

$$u_{iy} = K_p(c_{iy} - q_{iy}) + K_1(90 \sin 45 - y) \quad (9)$$

where K_1 is the proportional gain for the attraction to first base. This facilitates other players fielding ground balls and throwing runners out.

V. Simulated Results

The simulations were conducted within the bounds of the San Diego Padres stadium, Petco Park. The stadium limits, in-play area, left infield, and right infield were all defined as polygons.

The player data with which this simulation was completed was from the five right handed players who saw the most pitches from 2021 to 2025. These players are Randy Arozarena, Marcus Semien, Paul Goldschmidt, Eugino Suárez, and Aaron Judge. The larger sample size creates a better discrete density function. The data points used for Aaron Judge are shown in Fig. 2 as a spray chart. Each ball batted into play is charted against the stadium and field polygons. For the simulation, each point outside the limits of the stadium is not included in the creation of the density function. A visual representation of the density function created for the simulation is displayed in Fig. 3.

The initial placement of the players in the simulation is the MLB average fielder position against right handed batters. This is an appropriate location to start so that the model can be applied to various stadiums and right handed players. The results of the simulation will be compared to the MLB average player location to assess effectiveness.

Algorithm 1 presents the main loop of the simulation. The simulation was run for 10 seconds with a step time of 0.02 seconds, K_p of 1, and K_1 of 0.5. First, a Voronoi diagram is drawn with each player as the agents and the density function as the environment. After each player has a region defined around them, as depicted by the black lines in Fig. 4, the centroid for their region is calculated according to (7).

Each player is then driven towards their centroid using Lloyd's algorithm (6). The infielders were restricted to their side of the infield. If the movement of a player put them outside of their respective infield polygon, they were placed back inside the polygon at the nearest location to that new position. The pitcher and catcher remain in fixed positions for the duration of the simulation. The MLB average fielder position against right-handed hitters, and the final position of the simulation for Aaron Judge is displayed in Fig. 4. The model moves the players from the MLB average locations using Lloyd's algorithm to their final positions.

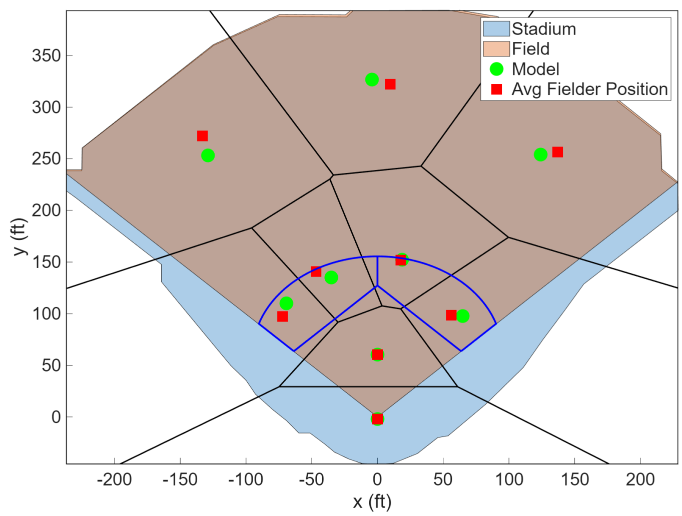

Fig. 4: The initial and final positions of the players during a simulation.

The total locational cost of all players is shown in Fig. 5. Locational cost is a metric used to evaluate the efficiency of the players covering the region. The locational cost decreases rapidly and eventually settles. This indicates that the simulation has found an equilibrium point and the players have reached their final positions.

The difference between the final and initial positions is displayed in Table 1. This table is the average across all five players in the study. Visually comparing the spray chart in Fig. 2 and the shift occurring reveals that the players generally move towards regions with a higher frequency of hits, creating a lower locational cost. This is the desired effect of the basic coverage control model created.

The reduction of locational cost indicates a more even distribution of players across the baseball field, reducing the total distance to each hit. All players saw a significant reduction in locational cost. The reduction in locational cost is displayed in Table 2 with an average reduction, across all simulations, of 6.15%.

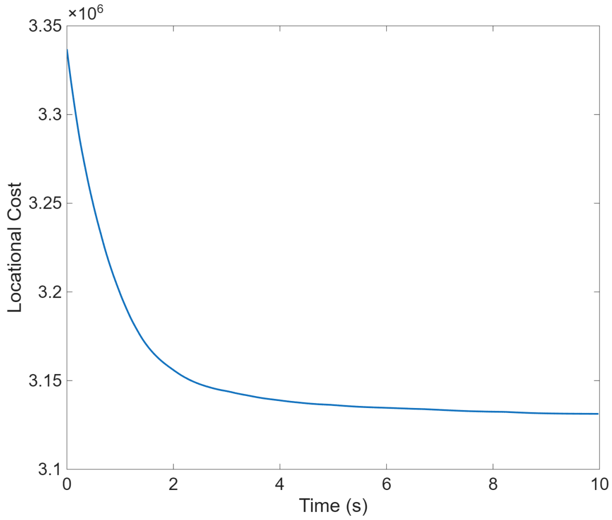

Fig. 5: Locational cost for the duration of a simulation.

TABLE I: Positioning Difference

Position	ΔX (ft)	ΔY (ft)	$\Delta\theta$ (deg)	$\Delta Depth$ (ft)
1st Base	16.47	-0.91	-6.91	15.57
2nd Base	8.61	-0.54	-3.22	8.07
Shortstop	13.28	1.23	-5.11	14.50
3rd Base	3.94	15.58	-5.41	19.51
Left Field	3.75	-15.21	0.67	13.20
Center Field	-10.17	5.14	1.83	15.99
Right Field	-14.58	-1.63	2.46	16.20
Average	11.18	6.65	3.88	14.72

The MLB provides statistics based on hang time, H, and distance, D, from the nearest player. This data represents the statistical likelihood of a player catching a ball. This study used this catch probabilty data to estimate the effectiveness of the model's position as compared to the league average positioning. The catch probability, p, can be represented by:

$$p = \frac{1}{1 + e^{-(3.5*H - 0.1*D - 8)}} \tag{10}$$

This catch probability is visualized in Fig. 7. This function was used to assign a catch probability to every hit location in the data set across all simulations. An overall catch probability was achieved by taking the average of all catch probabilities. This represents the ability for players to catch any given ball hit into play. Comparing the overall catch probability from the model positions and the average MLB positions revealed that the coverage control model had a lower overall catch probability, with a reduction of 0.054%.

VI. Conclusions and Future Work

The results of the five simulations show that the model can position baseball players within MLB rules while reducing the locational cost of the fielders. Locational cost saw a reduction of 6.15% as a result of players being positioned a mean distance of 13.00 ft from MLB league average positioning. This shows that a relatively small change in distance can potentially produce more effective coverage of the baseball

TABLE II: Reduction in Locational Cost

$Player$	$\%\Delta LocationalCost$
Randy Arozarena	-5.33
Marcus Semien	-7.03
Paul Goldschmidt	-4.58
Eugino Suárez	-7.66
Aaron Judge	-6.16
Average	-6.15

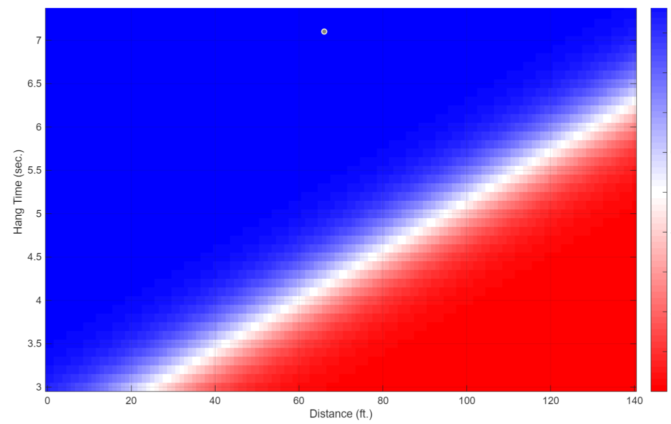

Fig. 6: Catch probability as a function of hang time and distance.

field. However, the model has a relatively small reduction in catch probability.

The improvement of locational cost with a very small change in catch probability is expected when examining the mechanism in which the model finds its final positioning. Lloyd's algorithm is used to lower location cost by incrementally moving the agents to the centroid. This relationship is depicted in Fig. 7. Both position 1 and 2 have an average catch probability of 50%. By moving from position 1 to position 2, the overall catch probability does not change, but the locational cost would be greatly reduced because the distance is more evenly spread out. This way of positioning may lend itself well to baseball through the reduced severity of balls hit into gaps, but the catch probability did not see any significant improvement.

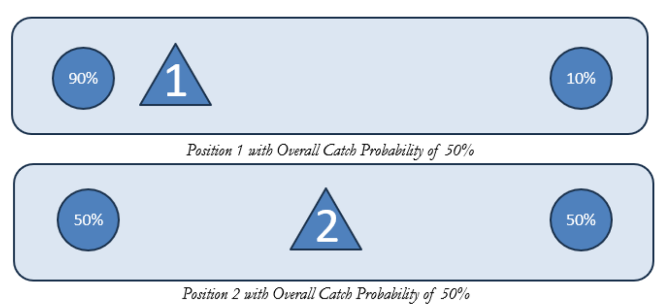

Fig. 7: Two methods of positioning that result in different locational cost values, but the same catch probability.

This model successfully uses coverage control techniques to position baseball players, but is limited by the complexity with which the model draws Voronoi cells and the robustness of the input data. Unlike standard Voronoi models that assume identical agents, in the future this model could account for the heterogeneous capabilities of baseball players:

- The range of players will be modeled as an elliptical region of dominance rather than a circular one. This ellipse will account for variations in speed and the physics of the ball's trajectory (hang time).
- Adjust ground ball data to be more accurately represented in the model.

These constraints will modify the boundary definitions of the Voronoi cells, ensuring that the resulting partitions reflect the actual physical limits of the players.

Improvements in the model can be made by introducing curved Voronoi cells [14] according to player ability as well as hit trajectory. In order to make the model more accurate, the players' physical limitations need to be considered as input into the model. Factors such as relative position to home plate, ball flight time, and player speed may be incorporated.

References

[1] J. Mizels, B. Erickson, and P. Chalmers, "Current state of data and analytics research in baseball," *Current Reviews in Musculoskeletal Medicine*, vol. 15, no. 4, pp. 283–290, 2022.

[2] Statcast glossary. Major League Baseball. [Online]. Available: https://www.mlb.com/glossary/statcast.

[3] M. Lage, J. P. Ono, D. Cervone, J. Chiang, C. Dietrich, and C. T. Silva, "Statcast dashboard: Exploration of spatiotemporal baseball data," *IEEE Computer Graphics and Applications*, vol. 36, no. 5, pp. 28–37, 2016.

[4] P.-Y. An, "Sports broadcasting: How big data technology impacts the viewer experience in baseball broadcasting," *Engineering Proceedings*, vol. 74, no. 1, p. 60, 2024.

[5] P. Beneventano, P. D. Berger, and B. D. Weinberg, "Predicting run production and run prevention in baseball: The impact of Sabermetrics," *International Journal of Business, Humanities and Technology*, vol. 2, no. 4, pp. 67–75, 2012.

[6] Sabermetrics: Baseball analytics and the science of winning. Syracuse University. [Online]. Available: https://onlinegrad.syracuse.edu/blog/sabermetrics-baseball-analytics-the-science-of-winning/.

[7] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: Applications and algorithms," *SIAM review*, vol. 41, no. 4, pp. 637–676, 1999.

[8] J. S. F. Garcia, M. J. M. Olivares, A. Gomez, and J. D. V. Morales, "A genetic algorithm-based methodology for the structural optimization of voronoi flat roofs," *Applied Soft Computing*, vol. 171, p. 112742, 2025.

[9] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[10] S. Fonseca, J. Milho, B. Travassos, and D. Araújo, "Spatial dynamics of team sports exposed by voronoi diagrams," *Human Movement Science*, vol. 31, no. 6, pp. 1652–1659, 2012.

[11] X. Feng and A. T. Murray, "Allocation using a heterogeneous space voronoi diagram," *Journal of Geographical Systems*, vol. 20, no. 3, pp. 207–226, 2018.

[12] R. Lin and M. Egerstedt, "Dynamic multi-target tracking using heterogeneous coverage control," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2023, pp. 11 103–11 110.

[13] J. Albert, *Visualizing baseball*. Chapman and Hall/CRC, 2017.

[14] T. Leung and E. J. Rodríguez-Seda, "Persistent coverage control of large-scale domains using multi-agent systems with anisotropic sensors," in *IEEE SoutheastCon*. IEEE, 2024, pp. 693–698.