Player Movement

Quantifying Off-Puck Player Positioning: A Physics-Driven Model Using Voronoi Tessellations

Sean Boyd

1 Questions Asked

The inherent complexity of hockey makes it challenging to objectively define what constitutes “good positioning,” let alone establish a reliable, quantifiable metric to evaluate a player’s ability to be effectively positioned during gameplay. Positioning is often linked to ubiquitous yet abstract and subjective qualities such as “hockey IQ” and “awareness,” which are frequently used in player evaluations but lack standardized definitions or consistency among evaluators. Despite this subjectivity, strong positioning is widely recognized as a critical factor in a player’s overall performance.

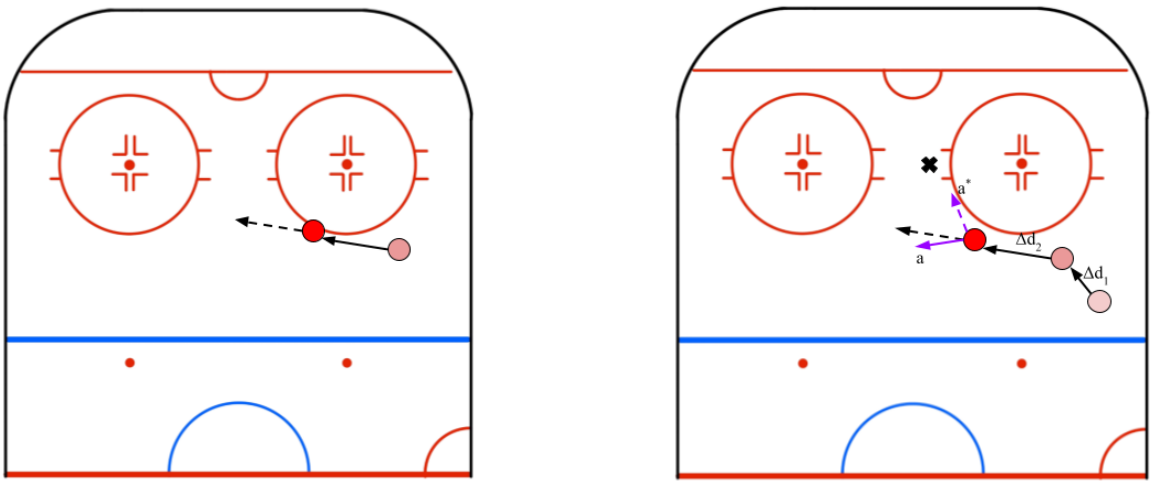
Off-puck positioning, in particular, is integral to both offensive and defensive strategies. Offensively, a player’s ability to find open space in high-danger areas significantly enhances scoring opportunities and team effectiveness. Defensively, maintaining control of dangerous areas, such as the slot and high-traffic zones near the net, is essential for minimizing scoring chances against. These nuanced aspects of the game often elude traditional statistical measures, leaving room for significant advancements in quantifying and analyzing these critical components of play.

This project addresses these gaps by developing a framework to quantitatively evaluate two key aspects of off-puck positioning: (1) a player’s ability to create opportunities by positioning themselves in high-danger areas of the offensive zone, and (2) their ability to control and protect dangerous areas in the defensive zone. By tackling these objectives, this analysis aims to answer the following core questions:

- Is it possible to quantitatively measure effective positioning in hockey?
- How can we define the area that a player controls, considering the complexity and movement involved in a game?
- What methods can be used to identify optimal areas of the ice for players to control?
- How does movement factor into static positions in a given frame?
- Does off-puck positioning have any correlation with outcomes and events in the game?

2 Approach

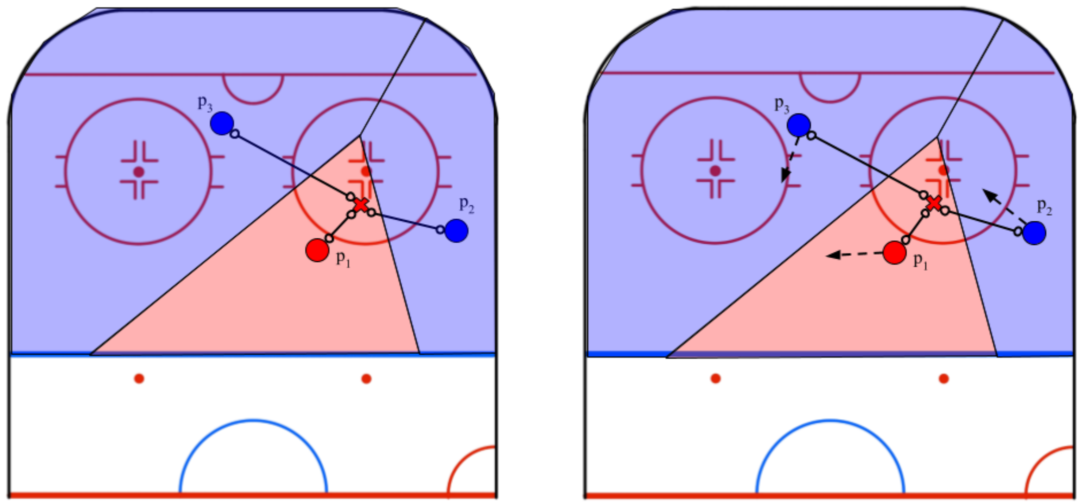
This model is constructed frame by frame, analyzing player positioning in sequential frames to predict the collection of points in the current zone that the player is expected to reach in the shortest time of any player present on the ice. Figure 1 (a) shows an example of the method used to calculate a player’s velocity at any given frame. Velocity is found by taking the player’s position in the previous frame, if accessible, and subtracting their current position from their previous in both the x and y axis to construct a velocity vector. Figure 1 (b) depicts an expanded visual of the previous example, showing the method used to calculate the acceleration of a player at any given frame. Found by taking the change in the previous two velocity vectors, in both the x and y direction, to construct an acceleration vector. However, in this model, since we're attempting to predict the time it will take for a given player to reach any given point in this zone, we maintain the current magnitude of the acceleration vector, directing it toward the current point of interest. Given a larger, seasonal dataset, the magnitude of acceleration could be found by taking a weighted average of the highest percentiles of acceleration recorded for the specific player. This way, the model will additionally reward players who have the ability to change direction and reach high speeds in significantly less time than their peers, even if in the directly previous frames, they did not display such capabilities. In light of the dataset provided for this project, assessing the player's current magnitude of acceleration presents a suitable and effective approach for this analysis.



(a) Velocity calculation (b) Acceleration calculation

Fig. 1 Visuals for Vector Calculations

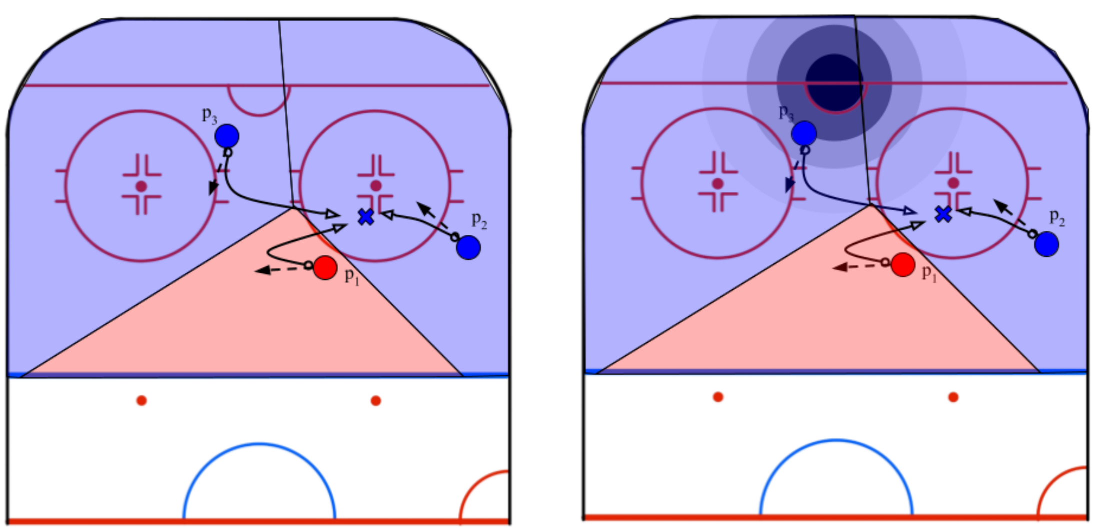
Now that we've defined a method for deciding which points on the ice are "controlled", or directly accessible for a specific player if the puck were to reach that point, we can construct a Voronoi tessellation. As we'll see shortly, the naive approach of using Euclidean distance is inferior to using a physics-based approach. In Figure 2 (a) we can see that Player 1 (denoted p_1) is the nearest player by Euclidean distance to the point marked "X". However, given a specific game scenario, such as the one displayed in Figure 2 (b), which includes the player's velocity vectors in this frame (depicted using black dashed arrows), it becomes less clear which of the three players should have "control" of the point.



(a) Controlled Regions (b) Velocity vectors

Fig. 2 Euclidean Voronoi Tessellation

Figure 3 (a) displays our approach of physics-based control as well as the corresponding updated controlled regions. A loose depiction of the path the player might take in their minimal time to reach the point marked, shown with solid black arrows. We can now see how the physics-based approach determines Player 2 (denoted p_2) as the player in control of the region containing the marked point. This is a more accurate denotation since Player 2 is moving in a direction far more aligned to the direction of the puck and is in a position of similar distance to Player 1. If a player's previous frames were inaccessible, the Euclidean distance approach was used to define control for the frame as velocity and acceleration were unavailable.



(a) Controlled regions (b) Expected Goals (xG)

Fig. 3 Time Voronoi Tessellation

With a clearly defined and appropriate model for controlled regions within a zone, we can define "dangerous areas" on the ice using an approximation for expected goals based entirely on shot location. This way, we can identify points on the ice where defensemen are expected to maintain control within their own zone and where opposing forwards are expected to attack and attempt to gain possession. Through testing the following formula was found to be most appropriate and best aligned with more complex expected goals calculations.

$$Expected\ Goals\ =\ 1-\frac{Distance\ from\ goal}{Distance\ from\ goal+1.5}$$

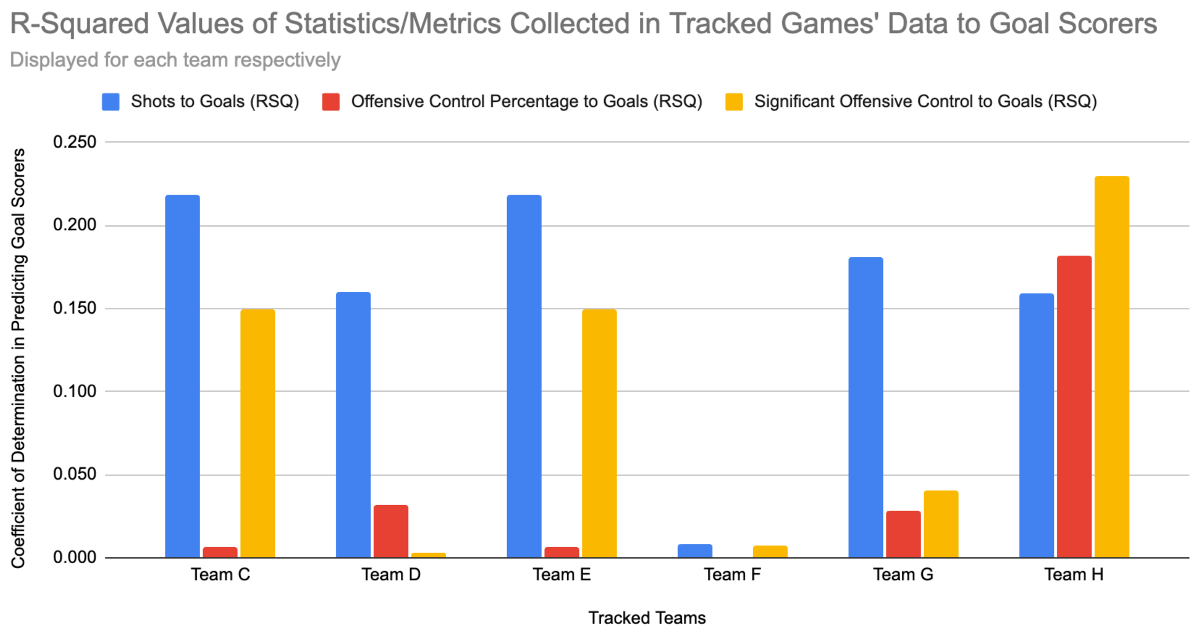
Figure 3 (b) displays this expected goals model overlaid onto our previous example and its Voronoi tessellation, with darker regions indicating locations with higher expected goals. This, in conjunction with our physics-based model for control, will reward players who attack the net and take advantage of open dangerous space, like Player 2 in this example. Additionally, defensemen will be rewarded for limiting the speed and acceleration of opposing forecheckers as they move toward dangerous areas and also for physically preventing forecheckers from entering positions on the ice where they could have access to dangerous regions.

For the construction of the metric used to quantify which players perform strongly in this model, we take the sum of the expected goals for each square foot controlled by a player and divide by the maximum sum acquirable for the given zone. Taking the average of this metric for a player in a given zone for each frame of a game produces “Offensive Control Percentage” and “Defensive Control Percentage” respectively. Additionally, we can define “Significant Offensive Control” and “Significant Defensive Control” as countable statistics recorded for each frame in which a player crosses a specific threshold for control percentage in a frame. A 15% threshold was proposed, as it would represent fifty percent more dangerous area control than the expected 10.0% (in any given frame the total control percentage is 100.0% across all ten players in 5v5 hockey).

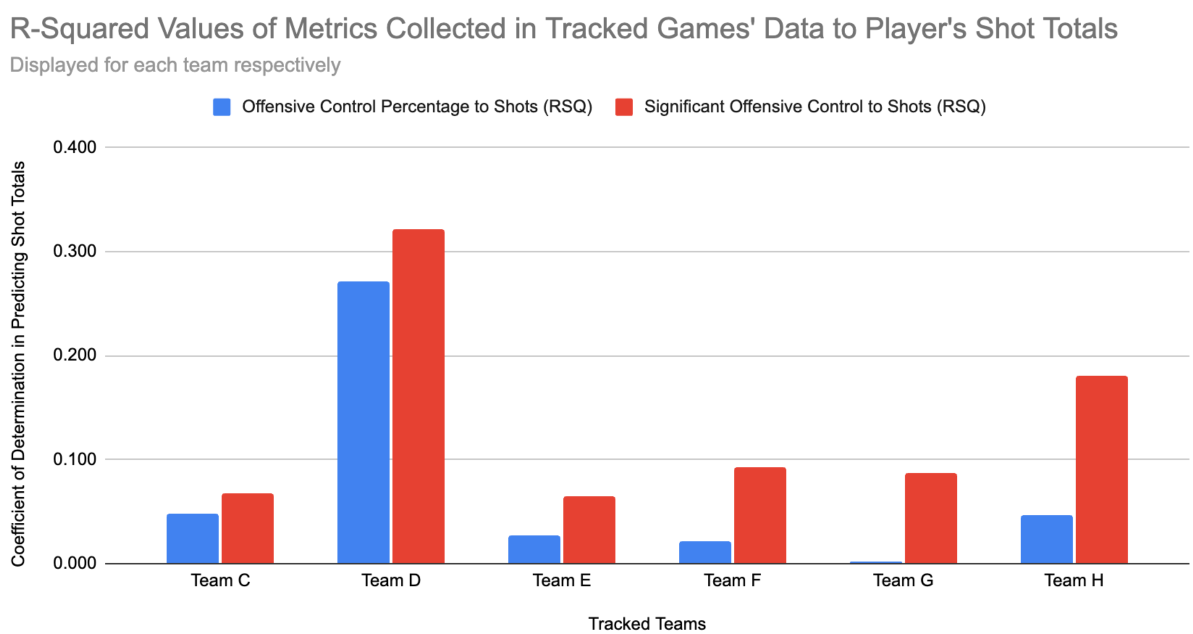
3 Findings

In the collection of data from the three provided games of tracking data, the Significant Offensive Control (SOC) metric had a higher correlation than shots when predicting goal scorers for a team 16.67% of the time. To further emphasize the significance of this performance, this metric has no access to the players’ locations when the puck is in their possession, and these values are recorded purely based on off-puck movement. Any substantial correlation, relative to shots (which are directly tied to scoring as they are a prerequisite for goals), is a success as this model is developed independently of all “On-Puck” statistics.

Data used in the following graph and ensuing analysis was retrieved every 30th frame for efficiency purposes and since increased specificity yielded minimal difference in resulting metrics for each player.



Additionally, Significant Offensive Control (SOC) had a higher correlation than Offensive Control Percentage (OCP) when predicting shots. This is expected as SOC is a countable version of OCP, recorded only once a significant threshold is surpassed.



4 Key Action Points from Analysis

Offensive Control Percentage (OCP) in unison with Significant Offensive Control (SOC) can be used to differentiate between players who display consistently higher average values of control (likely generally indicating players who are consistently well-positioned and more calculated), as opposed to players with short spurts of high control percentages (likely generally out-of-position players or sub-optimally positioned players who are more likely to take advantage of high-danger chances).

Since this dataset is limited to only a single game for each team it is difficult to extrapolate and prove that a strong average control in either zone is successful at predicting players who will consistently be impactful in games, independent of whether or not they register goals or shots. However, given the methods used to construct control percentages are nearly entirely independent of the events they showed correlation to, it is reasonable to assume their consistency would remain when tested across a larger sample size. This would most likely indicate a strong correlation to performance-based countable statistics like shots, goals, assists, and points as well as differential statistics like plus/minus, Fenwick, and Corsi with the inclusion of Defensive Control Percentage (DCP).

In future work, I plan to investigate defensive control, defined in contrast by decreasing the OCP of the nearest opponent, rather than maximizing their personal DCP. This could potentially more aptly reward defensemen for limiting forecheckers' ability to get into dangerous areas by inhibiting their ability to build up speed, accelerate, or maintain strong positioning in open areas.

Additionally, the introduction of shot angles into the expected goals model could improve the definition of how dangerous a specific point is. As well as introducing season-long highest percentiles for acceleration, rewarding players who play faster games because have the ability to change direction and gain momentum faster than their peers, rather than discrediting them for their high velocity in a specific direction. In a similar vein, probability mapping for the places a puck is most likely to be next would reward players who are moving with speed toward locations where the puck is more likely to be. This could be done via distance from the puck as a simplified approach, or via a neural network developed using the positions of players and the puck in frames from an entire season of data as a more robust approach.

5 Appendix

- Big Data Cup - Data Retrieval and Processing
- Big Data Cup - Findings
- Big Data Cup - CvD
- Big Data Cup - EvF
- Big Data Cup - GvH