Evaluating Space Creation in the National Hockey League using Puck and Player Tracking Data

Hassaan Inayatali and Timothy Chan

University of Toronto

Abstract. Star ice hockey players are often described as having a magnetic pull, with the ability to draw out opponents and generate dangerous opportunities for their linemates in the space left vacant by defenders. Using spatiotemporal Puck and Player Tracking (PPT) data, we develop a quantitative approach to measure how players create space while in possession of the puck, termed On-Puck Space Generation (OPSG). The benefits of our model's approach include its decomposition into three components: 1) Rink Control, the probability of controlling the puck at a given location; 2) Rink Value, the probability of scoring from a given location; and 3) Transition Probability, the probability that the next on-puck event will occur at a given location. Preliminary results of our metric show that players who achieve high levels of OPSG are more likely to lead their team in goals, assists and points. Our model can be used to analyze which players are in positions of danger, identify instances in which an individual created valuable space for their teammates, and understand which teams are best at generating space.

1 Introduction

While much of the information used to construct ice hockey teams and evaluate players is limited to the contributions of an individual, success in ice hockey requires high degrees of coordination among teammates. A common point of discussion with regards to play-making is space creation, movement which enables fellow teammates to position themselves in areas of high value. The work in this paper aims to address the following research question: How can we quantify the value of player movement with respect to influencing defender actions and creating scoring opportunities? In this work, we develop a model to quantify space creation by players in possession of the puck.

In the past, ice hockey analytics have been limited to event and stint data, which includes actions on the puck, the players involved and which players were on the ice at the time of a given event. Building off of the work by Sam Green [5] in soccer, expected goals (xG) models were developed to better understand the quality of shot opportunities in ice hockey [7, 3]. More comprehensive models such as Regularized Adjusted Plus-Minus (RAPM) models [4] have been built to better understand an individual's offensive and defensive impact. However, these models were unable to include information on teammate and opponent positioning on the ice.

With the introduction of Puck and Player Tracking (PPT) Data during the 2019-2020 NHL Playoffs, we can obtain more context into the game state when an event occurs and answer more complex questions relating to space creation. With other sports like soccer having had access to spatiotemporal data for nearly a decade [6], researchers have developed methodologies which focus specifically on inter-player dynamics. Thus, we have the opportunity to adapt existing models to ice hockey while accounting for the differences in these sports.

From this literature, we develop a model to determine the probability of scoring on the next on-puck event given the state of the game conditioned on an instantaneous pass event. Our model consists of three components:

1. **Rink Control**: The probability of controlling the puck at a given location.
2. **Rink Value**: The probability of scoring from a given location.
3. **Transition Probability**: The probability the next on-puck event will occur at a given location.

We calculate the change in the probability of scoring from the start of the possession to the end and aggregate across all possessions. This yields our final metric, On-Puck Space Generation (OPSG). The results of our model can help in understanding specific instances of space creation, evaluate player movements, and discern which teams are able to generate space consistently against their opponents. The contributions of this work are as follows:

- We develop a novel transition probability model for ice hockey conditioned on the locations and movements of all players on the ice.
- We propose a model to measure space creation by players in possession of the puck. Our model is composed of three sub-models which increase interpretability of the model's predictions.
- We perform an evaluation over 35 NHL games from the 2023-2024 season. We aggregate OPSG for individual players and teams. Our results show that OPSG has the strongest correlations with forward assists, defensemen goals, and team shot attempt differential.

2 Related Work

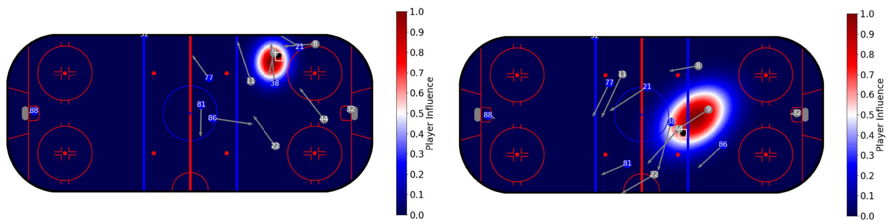
Pitch control refers to the "the probability that a player or team will be able to control the ball if it were at that location" [11]. Pitch control models have been developed in various forms, through the use of Voronoi Diagrams [12], Player Influence Models [1], and Poisson Point Processes [10]. While Pitch Control provides insight into spatial ownership, the value of this space is not considered. Pitch Value models aim to learn the value of space in different areas of a playing surface. These models can apply defensive positioning [1], distance from the net [10], or other models to decompose possession value into various actions [2]. These models are combined to better understand the quality of space controlled by each team [1, 10]. These can be used to create off-ball space creation metrics including Off-Ball Scoring Opportunity (OBSO) [10] and Space Generation Gain (SGG) [1]. In ice hockey, past research with PPT Data have focused on passing lanes [8] and passing value [9].

In this paper, we adapt Pitch Control and Pitch Value as well as expected pass speed from the aforementioned passing lane literature [8] to ice hockey to construct each of our Rink Control, Rink Value, and Transition Probability models. Our final result is a novel metric, entitled On-Puck Space Generation (OPSG), which examines a player's creation of space while in possession of the puck. We analyze how OPSG correlates with a player's cumulative production in terms of goals, assists, and points. Furthermore, we delve into a team's game-level space creation and how it relates to their performance in terms of shot attempts, shots on goal (SOG) and goals scored. Unlike previous work that focuses explicitly on puck transitions between players, our model provides insight into how players generate value when in possession of the puck. To the best of our knowledge, this is the first such model in the domain of ice hockey.

3 Methodology

3.1 Rink Control

We develop our Rink Control model using a bivariate normal distribution, in a similar fashion to Bornn and Fernandez [1]. To account for the speed of ice hockey, we increase the denominator in calculating a player's normalized speed ratio to 1500 ft/s. Additionally, we set the range of influence to be a minimum of 12 ft, which increases with distance from the puck up to a maximum of 30 ft. Distance from the puck affects the range of influence, aligning with the premise of Bornn and Fernandez, that "if the ball moves toward the player he would have more time to reach the ball within a larger space" [1]. Figure 1 shows two examples of player influence for a single player, both with the puck (Figure 1a) and without the puck (Figure 1b), ranging from 0 to 1. We focus on #9 in grey, where darker shades of red indicate higher levels of influence. For this and all future figures, the white square highlights the puck's location.



(a) Player Influence with the Puck (b) Player Influence without the Puck

Fig. 1. Visualizations for Continuous Player Influence

The influence of each player is aggregated and the home team's influence is subtracted from the visitor's at each location. The logistic function is applied to obtain a measure of Rink Control for each team, in the range of 0 to 1. A sample of this can be seen in Figure 2. Darker shades of red represent higher levels of influence for the team in gray, whereas darker shades of blue represent higher levels of influence for the blue team.

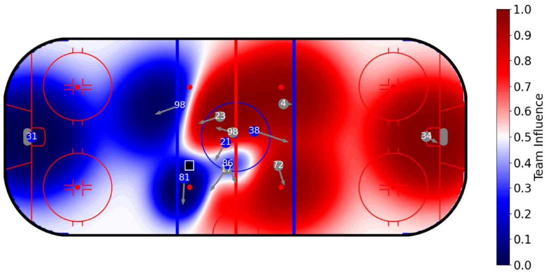


Fig. 2. Team Influence Model

3.2 Rink Value

To determine the value of a region on the ice, we develop an expected goals (xG) model to predict the probability a goal will be scored from a given location. This model is a logistic regression which predicts the probability of scoring based on the distance and angle of a shot. Our model is trained using NHL play-by-play (PBP) data from the 2015-2016 season, and achieves a cross-validated AUC of 0.731. This model can be seen in Figure 3. Darker shades of red represent a higher probability of scoring from the perspective of the team in grey, from 0 to 0.4.

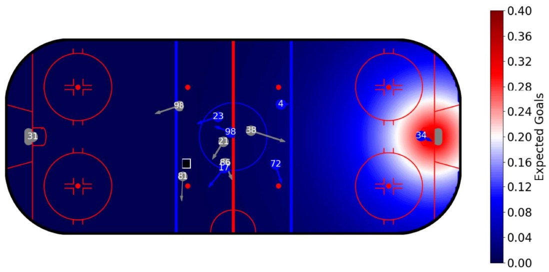


Fig. 3. Expected Goals (xG) Model

3.3 Transition Probability Model

With our Rink Control and Rink Value models, we determine the probability the next on-puck event occurs at a given location. using the following procedure:

1. Obtain a dataset of intended passes (successful and failed).
2. Model the probability a given pass will be successful.
3. Calculate the probability a pass will be successful to any location.
4. Normalize Pass Probability such that the sum over the rink is equal to 1

Thus, we assume Transition Probability is proportional to the probability a pass will be successful to a given location.

Possession Model Given that we do not have access to event data/passes, it is assumed that passes are transfers of possession between teammates. To obtain passing instances, we develop a possession model that produces a binary variable indicating the team in possession of the puck. The rule-based model is as follows:

1. The player is closest to the puck AND
2. The player is within six feet of the puck (one stick length) AND
3. The puck is traveling slower than 35 ft/s (max skater speed) AND
4. (a) The player was the previous player in possession of the puck OR
   (b) The player has an additional six feet between themselves, the puck, and the nearest defender

Pass Regression Using successful passes, we apply the linear regression methodology employed by Radke et al. [8] to predict the time until a pass will arrive at a receiver given their the distance from the passer. Transfers of possession were limited to less than two seconds to ensure only intended passes were analyzed as opposed to dump-ins or puck recoveries. This model was developed using 697 successful passes taken over the course of an NHL game, shown in Figure 4.

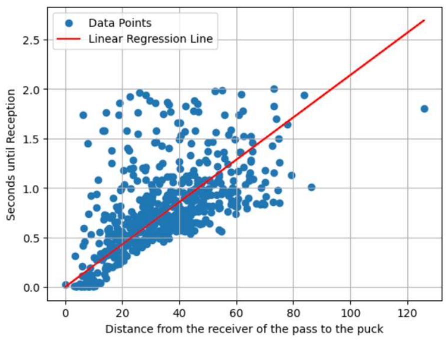


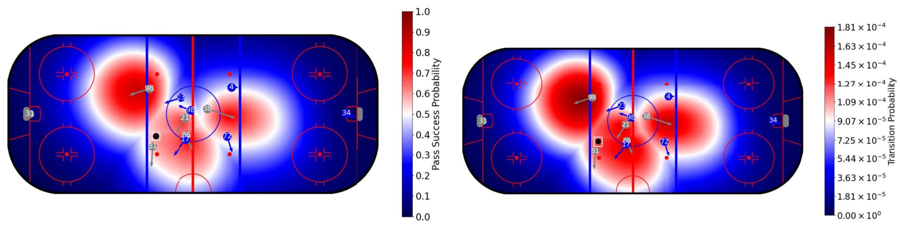
Fig. 4. Passing Linear Regression Model

The slope of this regression is 0.021 s/ft, meaning that for every additional foot a player is away from the puck carrier, it will take an additional 0.021 s for the puck to reach this player.

Pass Prediction To predict passes, we take the velocity vector of the puck at the moment the player no longer has possession, decided by our rule-based model presented in Section 3.3. For each teammate, the puck's velocity is projected according to how far they are from the puck carrier, using the linear regression in Figure 4. Each teammate's velocity is projected forward to estimate where they will be at the puck's arrival [8]. The distance between the puck's projection and teammate's projection is calculated. The player with the smallest distance is predicted to be the receiver of the pass. We exclude passes which hit the boards (dump-ins, bank passes, rims) as well as those which would be further from their intended receiver than 10 feet, to ensure we only analyze passes in which there was the direct intention to be received by a given player. A pass is successful if the following player in possession of the puck is the passer's intended target.

We fit a logistic regression model to predict pass success probability using five variables: defensive influence at the origin, midpoint and destination of the pass, projected distance between reception and receiver, and pass distance. The model was trained on 7000 passes and achieved a cross-validated AUC of 0.751.

Transition Probability We assume that a pass to a given location is intended for the player with the highest probability of receiving it. Because Transition Probability is proportional to the probability a pass would be successful to a given location, we normalize team pass probabilities across the rink surface to sum to 1. Team Pass Probability and Transition Probability can be seen in Figure 5. In Figure 5a, the right-defensemen of the grey team is in possession of the puck and the highest probability of a pass being successful is to his defensive partner. However, there are also passing lanes to each forward, and this can be seen by the darker shades of red in the direction each is travelling. Figure 5b takes the results from Figure 5a and normalizes across the rink surface by dividing the probability at each location by the sum of all probabilities.



(a) Team Pass Probability Model (b) Transition Probability Model

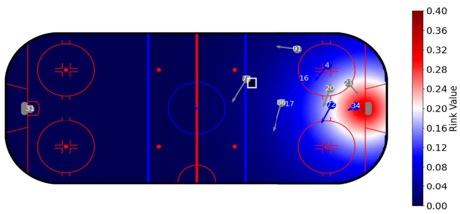
Fig. 5. Visualizations for Continuous Player Influence

3.4 Combined Model

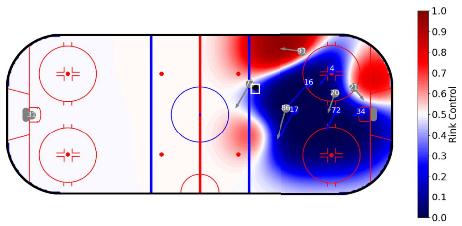
Using the framework presented by Spearman [10], we predict the probability of scoring on the next on-puck event for the attacking team conditioned on the puck possessor passing the puck. Let G_r denote the probability of scoring from location r, C_r represent the probability of controlling the puck at location r, and T_r signify the probability of passing the puck to location r. D and M represent the state of the game and a boolean representing an instantaneous passing event, respectively. Equation 1 shows how these models are combined to calculate the probability of scoring on the next on-puck event for the attacking team given the state of the game conditioned on an instantaneous pass event, $P(G|D,M)$.

$$P(G|D,M) = \sum_{r \in RXR} P(G_r|C_r,T_r,D,M)P(C_r|T_r,D,M)P(T_r|D,M) \tag{1}$$

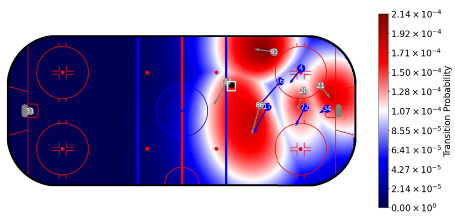
Figure 6 illustrates each component of the model, with Figure 6a showing Rink Value, Figure 6b displaying Rink Control, Figure 6c demonstrates Transition Probability and Figure 6d presents the combined model, with darker shades of red representing higher values. The sum of the combined model across the rink surface is $P(G|D,M)$. In Figure 6d, the probability of scoring is concentrated on the top right of the rink surface, driven by there being a player who is in a position of high value (Figure 6a), able to control the puck if it were to reach them (Figure 6b) and receive a pass at that location (Figure 6c).



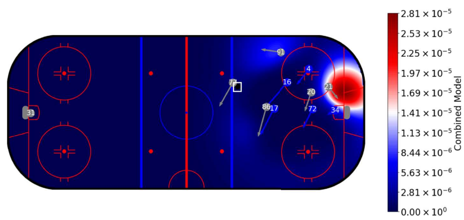
(a) Rink Value Model



(b) Rink Control Model



(c) Transition Probability Model



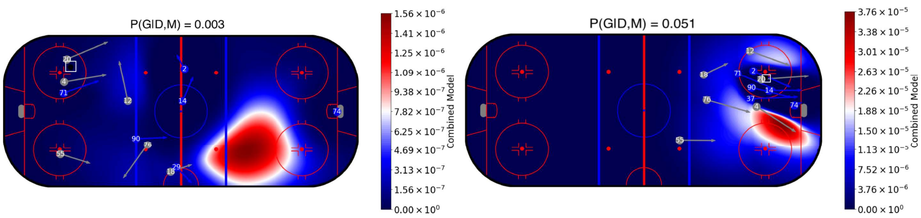
(d) Combined Model

Fig. 6. Visualizations for Continuous Player Influence

4 Results

4.1 Tactical Analysis

Using our model presented in Section 3, we measure how players create valuable space for their teammates. Figure 7 represents one of these situations, when the puck carrier starts with the puck in their own zone and proceeds to carry the puck from end-to-end and creates a direct passing lane to their teammate on a 2-on-1. At the beginning of the possession (Figure 7a), the probability of scoring on the next on-puck event is concentrated for #18 (bottom), which would depend on his being able to beat his defender to receive the pass from #20 (top left). By the end of the possession (Figure 7b), the teammate closest to the puck carrier (#4) on the 2-on-1 is occupying a high value area with a passing lane to receive the puck. This process can support coaches in identifying moments where an individual creates valuable space for their teammates in opposition scouting as well as better understanding how their players create space for one another.

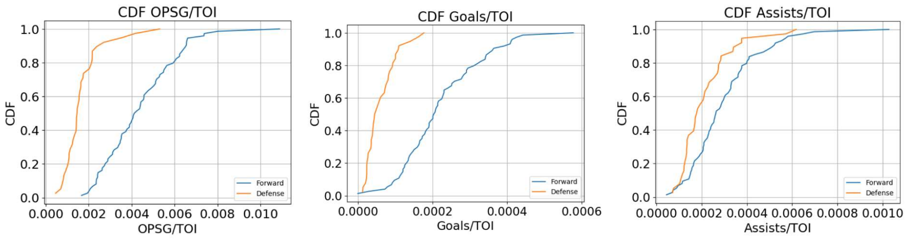


(a) Start of Possession (b) End of Possession

Fig. 7. Visualizations for Continuous Player Influence

4.2 Player Evaluation

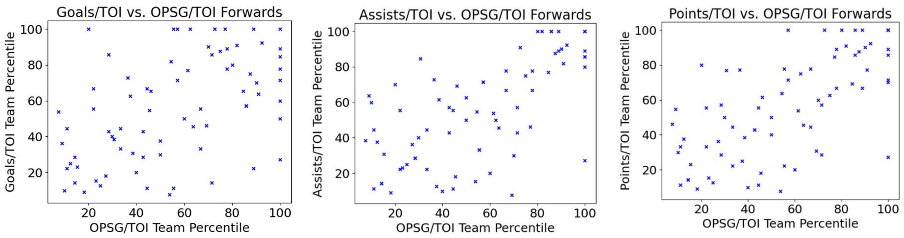
We aggregate the change in probability of scoring on the next on-puck event over a player's possessions to measure their On-Puck Space Generation (OPSG). We complete this process across 35 NHL games from the 2023-2024 season. Matchups were selected to maximize games played (GP) by a subset of teams to gather a representative sample for each player given compute constraints. Our evaluation examines players with five or more GP. This includes 74 forwards and 38 defensemen, with their mean time on ice (TOI) being 16.2 and 20.7 minutes per game, respectively. We normalize metrics using TOI to ensure usage does not affect player comparisons. In Figure 8 we plot CDF's for OPSG/TOI, Goals/TOI and Assists/TOI. Forwards generate more OPSG/TOI than defensemen (Figure 8a), with the 80^{th} percentile forward generating three times more than the 80^{th} percentile defensemen. This is comparable for Goals/TOI (Figure 8b); however the difference for Assists/TOI is considerably smaller (Figure 8c).



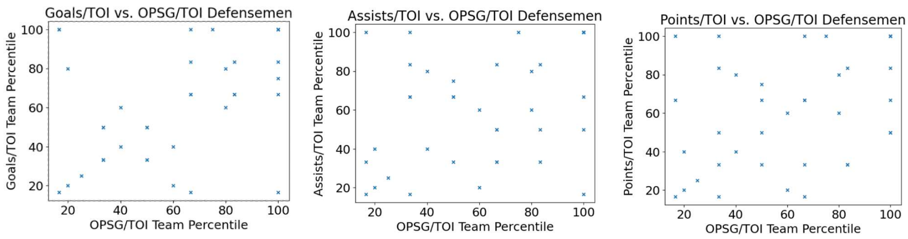
(a) OPSG/TOI (b) Goals/TOI (c) Assists/TOI

Fig. 8. CDF Plots by Position Category

Figure 9 demonstrates the relationship between OPSG/TOI and Goals/TOI, Assists/TOI and Points/TOI using their team percentile within their position category. Figures 9a, 9b and 9c show forward results and Figures 9d, 9e and 9f show defensemen results. While these scoring metrics generally increase as OPSG/TOI rises, the results are more pronounced for forwards compared to defensemen. Team percentiles are presented to maintain player/team anonymity. It should be noted OPSG focuses on player movements while the puck is possessed; it does not evaluate a player's eventual decision with the puck.



(a) Forward Goals (b) Forward Assists (c) Forward Points



(d) Defensemen Goals (e) Defensemen Assists (f) Defensemen Points

Fig. 9. Relationship between OPSG/TOI and Scoring Statistics by Position Category

Table 1. Average OPSG/TOI and Correlation with OPSG/TOI by Position Category

Position Category	Avg. OPSG/TOI	Goals/TOI	Assists/TOI	Points/TOI
Forwards	0.0048	0.504	0.654	0.660
Defensemen	0.0017	0.419	0.344	0.363

Table 1 shows the average OPSG/TOI and the correlation between OPSG/TOI and each of the aforementioned scoring metrics by position category. OPSG/TOI's correlation with Goals/TOI, Assists/TOI, and Points/TOI is higher for forwards compared to defensemen. Because defensemen are usually last to lead the rush and position themselves from the point in the offensive zone, we hypothesize they are less likely to move the puck into a space that will generate dangerous opportunities for their teammates relative to forwards. As seen in Table 1, defensemen generate nearly three times less valuable space for their teammates compared to forwards. Thus, the relationship between OPSG/TOI and Assists/TOI is weaker than forwards with a correlation of 0.344 compared to 0.654. This follows intuition given that this metric is designed to measure how well players create space for others, as opposed to themselves, and forwards are consistently moving the puck within the offensive zone to create scoring opportunities for their teammates.

Additionally, the correlation for defensemen is higher in goals compared to assists. One reason this might occur is that offensive defensemen that generate goals are more likely to carry the puck in more valuable areas in the offensive zone. Considering there are two assists for each goal, defensemen do not need to be in these areas to generate assists. Deeper analysis is required to confirm the underlying cause for these relationships. We leave this for future work.

We can also analyze the OPSG/TOI breakdown by position, shown in Table 2. While left-wingers and centers seem to generate more OPSG/TOI compared to right-wingers, these are subject to the composition of the rosters analyzed, and thus, a more comprehensive analysis of this area is needed. A similar statement could be made on left and right defensemen.

Table 2. OPSG by Position

Position	OPSG/TOI
Right Wing	0.0044
Center	0.0048
Left Wing	0.0050
Left Defensemen	0.0018
Right Defensemen	0.0016

4.3 Team Evaluation

We calculate the OPSG Differential between the Home and Away team, along with their shot attempts, SOG, and goal differential for each game in the dataset. The correlations between these measures are shown in Table 3. OPSG has a stronger relationship with shot attempts and SOG compared to Goal Differential. As various pieces on expected goals have noted, goals occur more randomly in comparison to shot attempts and SOG [3, 5]. Exploring the slightly negative correlation between OPSG and Goal Differential is left for future work with a larger set of games.

Table 3. OPSG Differential Correlation

Metric (H)	Correlation
Shot Attempt Differential	0.647
SOG Differential	0.603
Goal Differential	-0.055

5 Limitations and Future Work

While the model we propose in this work helps characterize space creation, it has several limitations. A limitation of our Rink Control Model is that it is a descriptive model and is not calibrated to event outcomes e.g. passes/puck recoveries; thus the model may not be indicative of actual values of control beyond player orientation to a specified location. Additionally, our Rink Value Model assumes the receiver of a pass is able to shoot. In reality, the value of possession in different areas on the ice should incorporate all possible decisions available to a player and their expected outcome. With regards to our Transition Probability Model, we have not incorporated the NHL offside rule into our models; thus receivers may not be in positions to receive a pass despite having a high transition probability. Finally, our combined model does not value passing lanes differently from one another. While low-high and seam passes increase the probability of a shot being scored relative to other shots at the same location, our current model does not incorporate features to differentiate these types of plays. Each of these limitations is left for future work.

Furthermore, given OPSG results are limited to a 35-game sample, an immediate direction of future work is to run this process across a full season to better understand the consistency in a player's ability to create space with the puck and our metrics repeatability.

A puck carrier's teammates also have the responsibility of moving into positions where they can receive the puck and create scoring opportunities. To understand a player's off-puck space creation for teammates, we could follow the framework presented by Bornn and Fernandez [1] for Space Generation Gain (SGG). To understand a player's off-puck space creation for themselves, we could isolate a player's contribution to Transition Probability through their Pass Probability Fabric, their contribution to Rink Control through their individual influence as seen in Figure 1, and Rink Value which does not require any adjustments. We could then measure the changes in their probability of scoring on the next on-puck event assuming the possessor of the puck were to pass it to them. This could then be aggregated across all possessions.

Another direction of future work is to measure how well teams perform in blocking passing lanes using our pass probability model. Given a set of intended passes, we can calculate the probability of the pass being successful, and measure how teams block opponent passes compared to the probability of the pass being successful. This would require an accurate dataset of incomplete and blocked passes, and thus, we leave this study for future work.

An additional area of future work would be to better understand which types of players play best together, known as team formation problems in the broader artificial intelligence (AI) literature. Teams may not want to have three players who excel at on-puck space creation but are poor in off-puck space creation on the same line. These types of traits could be applied further into team formation algorithms to better predict team success.

6 Conclusion

The presence of spatiotemporal data in ice hockey shifts the types of questions that can be addressed through analytics and the methodologies employed to approach them. Research developed in other sports, most specifically soccer, lends itself well to ice hockey and allows for more complex approaches to these problems. In this paper, we present a new metric for ice hockey, On-Puck Space Generation (OPSG). This metric can be used to better understand play-making through a quantitative approach to space creation using Puck and Player Tracking Data. Our models can be applied at the game-level to understand which players are creating space and where it is most often being generated. Furthermore, we can learn which teams are best at collaborating to generate space for one another, and how this relates to collective success. We believe this framework represents the first step toward better understanding how players create space in valuable areas in ice hockey.

Acknowledgments. We would like to give special thanks to Dr. David Radke from the Chicago Blackhawks for his invaluable suggestions throughout this process. Additionally, we would like to thank Ryan Kruse, Sam Forstner, Mark Weinstein, Albert Lyu, Daniel Codos, Zack Brusso, and Chris McCorkle from the Chicago Blackhawks for their feedback. Finally, we would like to thank Brant Berglund and Keith Horstman at the NHL, as this project would not be possible without the use of PPT data.

References

1. Bornn, L., Fernandez, J.: Wide open spaces: A statistical technique for measuring space creation in professional soccer. MIT Sloan Sports Analytics Conference (2018)

2. Bornn, L., Fernandez, J., Cervone, D.: Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. MIT Sloan Sports Analytics Conference (2019)
3. EvolvingHockey: A new expected goals model for predicting goals in the nhl (2018)
4. EvolvingHockey: Regularized adjusted plus-minus. https://evolving-hockey.com/glossary/regularized-adjusted-plus-minus/ (2021)
5. Green, S.: Assessing the performance of premier league goalscorers. Stats Perform (2020)
6. Lacey, P., Bialkowki, A., Carr, P., Matthews, I.: Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. MIT Sloan Sports Analytics Research Paper Competition (2015)
7. Moneypuck: Shot prediction expected goals model
8. Radke, D., Brecht, T., Radke, D.: Analyzing passing metrics in ice hockey using puck and player tracking data. Linköping Hockey Analytics Conference (2022)
9. Ritchie, R., Harell, A., Shreeves, P.: Pass evaluation in women's olympic ice hockey. MMSports '22: Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports (2022)
10. Spearman, W.: Beyond expected goals. MIT Sloan Sports Analytics Conference (2018)
11. Spearman, W., Basye, A., Dick, G., Hotovy, R., Pop, P.: Physics-based modeling of pass probabilities in soccer. MIT Sloan Sports Analytics Conference (2017)
12. Swartz, T., Wu, L.: A new metric for pitch control based on an intuitive motion model (2023)