Puck Possession and Net Traffic Metrics in Ice Hockey

by

Miles Pitassi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2025

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis builds upon a research paper published in the 2024 Linköping Hockey Analytics Conference (LINHAC), co-authored with Prof. Tim Brecht and Mingyue Xie [47]. It was then expanded to become Chapter 3 of this thesis. Additional contributions to this thesis for Chapter 4 stem from collaborations with Evan Iaboni, Sebastian Negulescu, and Fauzan Lodhi, under the supervision of Prof. Tim Brecht.

Abstract

This thesis investigates two elements of hockey widely believed to be critical to success: puck possession and traffic (skaters who are in or near the triangular area formed between the puck and the posts during a shot attempt). Our analysis draws on puck and player tracking (PPT) data from the 2023–2024 and 2024–2025 NHL regular seasons. We determine average team puck possession percentage, defined as the average proportion of total game time that each team has possession of the puck. We find that this metric has only a moderate correlation with average goal differential (r=0.56). To further explore how different aspects of possession relate to team success, we compute additional metrics, including Average Offensive Zone Possession Time Differential (Avg. OZPTD). This captures the difference between the time a team spends with possession in the offensive zone and the time their opponents spend with possession in their offensive zone. We find a strong correlation (r=0.77) between Avg. OZPTD and average goal differential. Further analysis shows that Avg. OZPTD is stable across games, effectively distinguishes between teams, and, despite being correlated with existing metrics like Shot Attempt Percentage (SAT%), offers additional predictive value for goal differential. SAT% (also known as Corsi) refers to the percentage of total shot attempts that each team takes.

We also study the relationship between the amount of skaters creating traffic during shot attempts and shot outcomes. Our findings show that increased levels of traffic significantly increase the percentage of shot attempts that are blocked and reduce the chance of a shot attempt resulting in a shot on goal. Overall, we find that 29% of all shot attempts are blocked and that the highest goal rates occur from the center of the ice on short-to-mid-range attempts with no traffic present. For long-distance shot attempts that reach the goaltender, scoring probability increases with traffic. We also show that defensive skaters primarily reduce shot-on-goal rates but can inadvertently increase goal likelihood on mid-range shot attempts (23-45 feet) presumably due to screening their own goaltender.

Together, the findings in this thesis offer valuable insights into how puck possession contributes to team success and how traffic influences shot outcomes. In addition to these empirical results, we contribute a suite of methodological techniques that can support future analysis of possession and traffic. We present a comprehensive pipeline for cleaning, filtering, and processing individual possession data sourced from the NHL's puck and player tracking system which is an essential foundation for our findings and a resource for future research. We also describe how we assemble a set of shot events by aligning official NHL API shot data with shot attempts in the PPT data. This involves adapting the NHL's own inference algorithms to identify undetected shot attempts and applying custom techniques to improve timestamp accuracy.

Acknowledgements

This thesis marks both the completion of my degree and a turning point in my career. With the support of my family and my supervisor, Tim Brecht, I was able to pursue a path where I can combine my practical experience in coding, data analysis, and problem solving with my lifelong love for sports. For that, I am immensely grateful.

First and foremost, I want to thank my family for their endless support and enthusiasm throughout my life and during my pursuit of this degree. They are my favourite people to hang out with and the people I look up to the most. My best linemates will always be my brother, Cole, and my dad, Rich. Thank you for teaching me how to play and love hockey. I also want to thank my mom, Toni. When I was six years old and couldn't even raise the puck, I spent hours teaching her how to take slap shots in our basement until she got it. She attended my games while reading research papers and loved watching me play baseball (her sport) because it fueled her passion for analytics just like it did mine. I used to attend and score Blue Jays games by myself just so I could share the scorecards with her later. Thank you to my sister, Ella, for being my biggest supporter in more ways than one. When she saw me get bodychecked in my hockey games, she would bang on the glass and yell for me to hit them back. It's a beautiful reminder that there is something in sports for everyone.

I'm incredibly grateful to my supervisor, Tim Brecht, for his advice, mentorship, and for giving me the opportunity to pursue this work. His guidance has helped me become more thoughtful and critical in how I approach problems. Instead of looking for the data to support my ideas, I now focus on forming insights based on the data itself (and comparing those insights to my initial hypotheses often leads to more meaningful discoveries). He also helped me improve as a writer, as a presenter, and as someone who enjoys research and discussion. I am certain I would not have been able to publish several papers and complete this thesis without his mentorship.

Next, I want to thank my partner, Rayah. She is the person I share everything with and my number one supporter. She has a remarkable ability to understand the core of what I'm working on, even with little context, and her feedback always makes a difference. She brings light into my days (literally, she opens blinds when I have been holed up all day and forces me to go for walks) and adds joy and balance to my life. I'd also like to thank her whole family including Armaan, Alida, Amin, Nani, Nana, and all their families for welcoming me as part of their own and for being like a second family to me. Their support has given me confidence in my work and more role models to look up to. I once tried to dismiss the importance of working in sports, but I'll never forget Alida reminding me how much sports means to people, how improving a player's or team's success matters deeply, and how my work can enhance a fan's appreciation of the very sports that have such a profound impact on their lives, just as they do on mine. Thank you for all your love (and food).

I also want to thank my extended family and especially my two grandmothers, Madeleine Pitassi and Aida Zemel, and my grandfather, Reuben Zemel, for being the strongest people I know. Madeleine (or Nunnee) worked for the Pittsburgh Penguins, Pirates, and Steelers for over 40 years. I'll never forget going to Pirates games together even though she was forced to watch 81+ of them every year. And to my late grandma Aida (Muck), she couldn't care less about hockey even though her husband Reub was a McGill Redman and all three of her boys and their kids were obsessed with it, but I know she would have loved to hear me talk about what I'm doing and to hear how happy I am in pursuing my passion.

Next, I'd like to thank my collaborators Mingyue Xie, Evan Iaboni, Sebastian Negulescu, and Fauzan Lodhi for their invaluable contributions to making these research projects possible. I wish you all continued success in your chosen fields (however I believe hockey and sports would greatly benefit from your involvement). I am also grateful to my collaborators at the National Hockey League, especially Ben Resnick. I did not anticipate the partnership would be so rewarding. Your experience working with teams and the league, combined with your generosity in sharing your time and insights, was instrumental in helping me navigate the data and deepen my understanding of hockey analysis. Thank you as well to Jonah Eisen at Rogers for your thoughtful feedback throughout my degree. I also want to acknowledge Dave Radke and his team for winning the Rogers hackathon and partnering with Tim Brecht to establish the UW / Rogers / NHL collaboration that gave me access to this valuable data. And thank you to Jesse Hoey and George Labahn for serving on my examining committee and for your thoughtful comments and questions.

Finally, I want to thank all my friends, teammates, and coaches over the years that I played with and spent my childhood with. My time at Waterloo has been a culmination of everything that has come before, and it wouldn't have meant nearly as much without the people who supported me throughout.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis proposes new methods for extracting and analyzing information about team puck possession and traffic (skaters that are in or near the triangular area between the puck and the posts during a shot attempt) using puck and player tracking (PPT) data from the National Hockey League (NHL).

1.1 Motivation

During the 2021–22 season, the NHL introduced PPT technology across all arenas, using light-emitting infrared diodes (LEDs) embedded in the puck and player sweaters to collect high-frequency, high-resolution spatiotemporal data. Through a partnership between Rogers, the NHL, and the University of Waterloo, we obtained access to this data. In contrast to other major sports like soccer and basketball, where tracking data has already transformed player and team evaluation and analysis (see Section 2), hockey is only beginning to explore the potential of this rich data source.

This thesis is motivated by the opportunity to explore what PPT data can reveal about two core but historically under-analyzed elements of hockey: team puck possession and traffic during shot attempts. Both are frequently discussed by coaches, players, and broadcasters as key factors in team success, but their impact has rarely been quantified rigorously. In the early 2000s, analysts began to estimate possession by using proxies like the percentage of shot attempts (SAT%, also known as Corsi) [63]. Then in 2013, Tulsky et al. [64] demonstrated that entering the offensive zone with possession leads to more than twice as many shots and goals as dump-ins, helping to drive a broader strategic shift toward possession-based play (Eric Tulsky is now the General Manager of the Carolina Hurricanes). Despite this previous work using possession proxies and zone entry data, there remains a gap in fully leveraging tracking data to measure and analyze possession directly. With access to league-wide PPT data, this thesis examines possible connections with success across zones and situations.

Traffic during shot attempts is also widely believed to play a critical role in scoring. Coaches often emphasize the importance of generating traffic to obstruct the goaltender's view, create deflections, or capitalize on rebounds, while defenders are tasked with blocking shot attempts, clearing shooting lanes, and tying up sticks. Despite this attention, traffic has, to our knowledge, not been systematically defined and studied. With PPT data, we analyze traffic in a spatial and temporal context, identifying how many skaters are in or near the shooting lane, what roles they play, and how their presence affects shot outcomes. This creates an opportunity to rigorously evaluate a long-standing assumption in hockey strategy and open new avenues for understanding how skaters' off-puck interactions affect shot attempts.

1.2 Objectives

The primary goal of this thesis is to develop new methods for analyzing team puck possessions and traffic using NHL puck and player tracking (PPT) data to obtain insights into how these factors influence team performance and shot outcomes. Our objectives in this thesis are:

- Develop a methodology for cleaning and filtering raw, individual possession data and integrating that data with game context (e.g., score, strength, location) to construct detailed team-level possession sequences.
- Create metrics to characterize team puck possession (e.g., total possession time, frequency) and assess their stability, independence from existing statistics, and ability to distinguish between teams. Then evaluate the relationship between these metrics and team success.
- Design and implement a methodology for detecting and augmenting shot annotations in the PPT data and use PPT location information to determine who is creating traffic for each shot attempt.
- Create metrics that quantify traffic, including the difference between offensive and defensive traffic, and evaluate their relationship to shot success.

1.3 Contributions

This thesis contributes new methods and empirical findings for analyzing team puck possession and traffic in the NHL using puck and player tracking (PPT) data. It offers practical techniques for analysts working with PPT data, particularly to support future research on possession and traffic. Additionally, it provides empirical insights for NHL front office staff, coaches, players, and fans into the relationships between team puck possessions and success, as well as the relationship between traffic and shot outcomes.

Methodological Contributions:

- We design a robust pipeline for cleaning and filtering the NHL's player possession data including: removing duplicate possessions, resolving overlapping possessions, improving end of possession detection, incorporating puck location for zone information, integrating game context such as strength and score state, and aggregating player-level possession segments to determine team-level possession.
- We develop a process to combine shot attempts with precise timestamps by matching official NHL shot data from the NHL API to shot attempts in the PPT data. Shot attempts in the PPT data are found by leveraging the NHL's shot detection system, adapting their inference algorithm to recover undetected shot attempts, and applying our own augmentation methods to improve timestamp accuracy.
- During our initial analysis of the impact of traffic on shot outcomes, we found that traffic is strongly correlated with shot distance (longer shot attempts are typically subject to more traffic) and moderately negatively correlated with shot angle (shot attempts from sharper angles typically have less traffic). To isolate the effect of traffic from these confounding spatial factors, we develop an algorithm that divides the offensive zone into regions that minimize within-region variation in distance and angle, allowing us to study the impact of traffic separately from distance and angle.

Empirical Contributions:

- We introduce the Average Offensive Zone Possession Time Differential (Avg. OZPTD) metric, defined as the average difference between a team's offensive zone possession time and that of their opponents. A positive Avg. OZPTD indicates that, on average, the team spends more time in the offensive zone with possession than their opponents do in theirs (i.e., the team's defensive zone). We find that Avg. OZPTD is highly correlated (r=0.77) with Average Goal Differential. This highlights its potential for enhancing our understanding of team success in the NHL.

- We show that Avg. OZPTD's is stable across two halves of our dataset (i.e., useful for prediction), is able to differentiate between teams, and is independent from existing metrics which demonstrate its potential as a useful new metric.

- When examining the impact of traffic, we find that for all regions, increased levels of traffic significantly increases the percentage of shot attempts that are blocked and reduces the chance of a shot attempt resulting in a shot on goal.

- We find that in all short-range regions, goals with no traffic exceed the combined total of all other traffic levels. Across all regions, the highest number of goals occurs in the 9–23 ft center region with zero traffic, highlighting the value of shot attempts in this region.

- For long-range shot attempts (45-90 ft), 38% of shot attempts are blocked, compared to 29% for all shot attempts in our dataset. However, long-range shot attempts through higher levels of traffic that reach the goaltender are associated with an increased likelihood of a goal.

- For mid-range shot attempts (23-45 ft), the effect of traffic on outcomes is generally nuanced, with no statistically significant trends. However, when there are more defenders than attackers in the shooting lane, shot attempts that reach the goaltender are significantly more likely to result in goals, which suggests that defensive traffic may unintentionally screen the goaltender or alter shot trajectories in favour of the shooter.

- To our surprise, we find that the percentage of missed shot attempts is consistent across traffic levels and locations, averaging 22.4% with variation of just 1.5% and 2.5%, respectively.

1.4 Thesis Outline

In Chapter 2, we present background information and related work. We describe the puck and player tracking (PPT) dataset used throughout the thesis and outline conventions applied in our analyses. We also review prior research on possession and traffic in hockey and other sports. In Chapter 3, we explore team puck possession in the NHL. We define and process team-level possession data and introduce a new metric, Average Offensive Zone Possession Time Differential (Avg. OZPTD). We evaluate its correlation with team success, as well as its stability, discriminatory power, and independence from other metrics. In Chapter 4, we analyze the impact of traffic on shot outcomes. We present methods for determining shot timing, quantifying traffic using PPT data, and isolating the impact of traffic from that of shot location on shot outcomes. We then evaluate how traffic affects goal likelihood, with further analysis separating traffic caused by attacking versus defending skaters. Finally, in Chapter 5, we summarize our findings and contributions, discuss the implications for hockey analytics and future work, and offer concluding remarks.

Chapter 2

Background and Related Work

2.1 Dataset Overview

For over a century after the NHL's founding in 1917, skaters were measured using traditional box score statistics such as goals, assists, and saves. In the 1967-68 season, the controversial plus-minus statistic was introduced, which is a team's goal differential while a particular skater is on the ice, excluding power play goals for and against but including empty net situations. In the early 2000s, "advanced stats" began to emerge, introducing metrics such as the percentage of shot attempts (SAT%, also known as Corsi), which served as a proxy for puck possession. However, as many have noted, including an excellent explanation in the opening chapter of Hockey Analytics: A Game-Changing Perspective [56], hockey is particularly resistant to statistical simplification. It is a low-scoring, fluid, and highly interdependent sport where critical plays often go unrecorded in the stat sheet. That potentially changed during the 2021-22 season, when the NHL made a major technological leap by introducing a puck and player tracking (PPT) system. This thesis leverages the NHL's PPT dataset as a core component of its analysis.

In this thesis, we use both PPT and official data made available through the NHL's public Application Programming Interface (API) [38]. This API provides access to a broad range of game-related data, including player statistics (e.g., goals, assists) and team performance metrics (e.g., wins, losses). Integrating this data with the PPT data allows for a more comprehensive analysis by correlating detailed movement and positioning data with game outcomes and individual performance metrics. The API has proved especially valuable in cases where PPT data is incomplete, such as games with missing tracking information or players excluded due to small sample sizes. Rather than relying solely on season-level aggregates, the API allows us to dynamically retrieve game-specific or player-specific data on demand which enables more targeted filtering, validation, and comparison.

2.1.1 Puck and Player Tracking Overview

The NHL's PPT system relies on light-emitting infrared diodes (LEDs) embedded in both the puck and in players' sweaters. Each player's sweater contains an LED located near the right shoulder blade and the puck contains LEDs positioned on its top and bottom surfaces (see Figure 2.1 for LED locations). Notably, the system was updated in the 2022–23 season to move the LEDs closer to the puck's edges, increasing the dispersion angle and improving detection speed and accuracy, as well as resolving concerns over the puck's gliding ability [13]. To capture location data, up to 20 infrared cameras are mounted above the ice in each arena. These cameras are calibrated to triangulate emitter positions with high accuracy in three dimensions. The result is a continuous stream of high-frequency data. Positional data, including x, y, and z coordinates, is recorded at a frequency of 60 times per second for the puck and 12 times per second for each player on the ice. An additional update is provided once per second for players on the bench, resulting in around 734,400 location points in a typical 60-minute game.

Figure 2.1: LEDs embedded in a player's sweater (left) and in the puck (right)

The NHL's choice of infrared-based tracking offers several key advantages: it provides high positional accuracy in the x and y directions, it supports real-time data streaming with low latency, and it performs reliably in indoor arenas where a Global Positioning System (GPS) is ineffective. Infrared systems excel at tracking small, fast-moving objects like the puck due to its high sampling rate (often 60 Hz or more). However, the system also has limitations. Infrared tracking requires a line of sight between infrared emitters and cameras, meaning occlusion can lead to data loss or inaccuracies. The system's z-axis tracking (in this case, the distance from the ice surface) is generally less reliable due to overhead camera placement and limited vertical triangulation. Additionally, the PPT system depends on specialized hardware (emitters in pucks and sweaters and up to 20 cameras in each rink) which increases operational complexity and cost. Other sports leagues and data providers have explored a variety of tracking technologies, each offering different advantages and disadvantages. Below, we briefly review the main alternatives relevant to ice hockey:

- **Radio Frequency Identification (RFID):** The NHL initially introduced puck and player tracking using Radio-Frequency Identification (RFID) technology rather than infrared. This debut occurred during the 2015 All-Star Game and relied on a system developed by Sportvision, which has since been acquired by SMT (the NHL's current tracking technology provider) [16] [25]. While the NHL did not publish accuracy figures for this system, similar RFID-based tracking has been used in other sports, such as American football [36]. In 2021, NFL Vice President of Emerging Products and Technology Matt Swensson reported that the system's positional accuracy was approximately six inches [9]. This level of precision falls short of the accuracy required for tracking small, fast-moving objects like a hockey puck in real time.

- **Bluetooth:** Finland's top professional ice hockey league, Liiga, uses the Wisehockey system developed by Bitwise [41], which is built on Quuppa's Bluetooth Low Energy (BLE) positioning technology [48]. Each player and the puck are equipped with BLE transmitters that communicate with approximately 20 locators placed around the rink. This approach offers a relatively low-cost, scalable solution that is effective in indoor environments and provides moderate positional accuracy. It is also more robust to occlusion than computer vision or infrared systems since it relies on signal trilateration rather than line-of-sight detection. However, bluetooth-based tracking typically has a slightly higher error margin than infrared systems. For example, the Wisehockey system is reported to track positions with an accuracy of approximately four inches (10.16 cm) [41]. While the NHL does not publish accuracy figures for its infrared-based tracking system, other infrared technologies used in optical tracking have demonstrated mean square errors of less than 0.3 mm when measuring distances between light-emitting points [8].

- **Computer Vision:** Puck and player locations can also be determined using computer vision, a field of artificial intelligence that enables computers to interpret and extract meaningful information from visual inputs like images or video. In sports, computer vision systems are used to automatically detect, track, and identify players and objects throughout a game [65] [25] [9]. Computer vision has the benefit of being unobtrusive and hardware-free for players, making it easy to deploy without modifying equipment. It is also scalable across venues and capable of capturing visual features such as body posture. However, it faces significant challenges in hockey as it is particularly vulnerable to occlusion when the puck or players are behind a player, the boards, or are out of the camera's field of view, and because it can be challenging to distinguish between players.

While each tracking technology has advantages and disadvantages, the NHL's infrared-based system offers the precision needed to capture the fast, complex movement in hockey. However, its limitations such as inaccuracies in the z-dimension and occasional data gaps from occlusion means it supplements rather than replaces traditional tools like video review and manual scoring for critical decisions such as goal-line detection. As well, its application is subject to specific limitations. According to the agreement between the NHL and the NHL Players' Association (NHLPA), tracking data cannot be used in salary arbitration hearings or contract negotiations [32]. This restriction is designed to protect players from potential misuse of performance data in financial discussions.

2.1.2 Applications and Research Using Puck and Player Tracking Data

The NHL makes PPT data available to every team, a small number of external partners, and our research group. While teams use it in a variety of ways that are often kept private for competitive reasons, a limited but growing body of public applications and research has also emerged. The NHL has used PPT data to release some fan-facing tools, such as the NHL EDGE website, which includes leaderboards and visualizations for player and team metrics like skating speed and shot locations. For shot locations, they have produced heatmaps, specific to each goaltender, showing which areas of the ice opposing shooters score the most goals from [43].

Internally, and in collaboration with their external partners such as Amazon Web Services (AWS), the NHL has also developed models such as opportunity analysis to better evaluate scoring chances [37]. In academic research, our group led by Tim Brecht and

David Radke has explored various applications of PPT data. Early work in 2021 introduced, among other things, pressure and passing metrics [49]. Follow-up studies expanded on this by identifying pass types and improving passing lane models [49], and by analyzing passing metrics and player openness [50].

There are also instances of research using puck and player tracking data that is not provided by the NHL. A notable example is the Big Data Cup, an annual competition launched by Stathletes in collaboration with the University of Toronto's Rotman School of Management and the University of Toronto Sports Analytics Student Group (UTSPAN) [61]. The event provides anonymized tracking data, generated from broadcast video of NHL games using computer vision, to students and researchers exploring a wide range of hockey analytics questions. This thesis builds on that growing foundation by demonstrating the capabilities of full-resolution PPT data and providing transparent methods and insights that can support future research in hockey analytics.

2.2 Conventions Used for Correlation Analysis

In this thesis, we frequently use Pearson's correlation coefficient (r) to measure relationships between variables. Values closer to 1.0 or -1.0 indicate stronger linear relationships, whereas values near 0 suggest little to no linear association. Positive values indicate a direct relationship (i.e., as one variable increases, the other tends to increase), while negative values indicate an inverse relationship (i.e., as one variable increases, the other tends to decrease). Throughout the thesis, we apply the thresholds outlined in Table 2.1 to maintain uniformity in our interpretations. It is important to understand that interpretations of correlation values can vary depending on the context; by consistently applying the thresholds shown in Table 2.1, we aim to provide a clear basis for understanding how we derive our conclusions from the data.

Correlation (r-value)	**Interpretation**
± 0.9 to ± 1.0	Very strong
± 0.7 to ± 0.9	Strong
± 0.5 to ± 0.7	Moderate
± 0.3 to ± 0.5	Weak
± 0.0 to ± 0.3	None to Negligible

Table 2.1: Interpretation of Pearson correlation coefficients (r) used in this thesis

2.3 Related Work

There are two sections of related work, one for each of the two main chapters of the thesis.

2.3.1 Puck Possessions and Team Success

Previous studies on the importance of possession across sports provide context for this section of the thesis, particularly highlighting the research gap in hockey analytics.

In football (soccer), the relationship between ball possession and team success has been extensively studied using various research methods. Some of these studies find a positive relationship. For instance, researchers studied the 2016 UEFA Euro and found that the average possession time for a leading team was 20.3 minutes with a standard deviation (SD) of 16.0 minutes, compared to 18.2 minutes with a SD of 16.8 minutes for teams when the score was tied, and 13.7 minutes with a SD of 12.3 minutes for a trailing team [10]. The authors explained that the p-value, which assesses the likelihood that these differences occurred by chance, was less than 0.01, indicating a statistically significant difference. Additionally, researchers studying the 2006 FIFA World Cup found that the percentage of ball possession, analyzed using principal component analysis, had the greatest influence on match outcomes with a coefficient with an absolute value of 0.72. This indicates that it is an important variable for discriminating winning teams from those that lose or draw [35]. Another study found that ball possession had a positive effect on winning in the 2014 FIFA World Cup, with an 11% increase in the probability of winning for all matches and a 14% increase for close matches when ball possession increased by two standard deviations [27]. Also, a study covering the 2017-18 and 2018-19 season in the German Bundesliga showed a positive correlation (r=0.75) between team possession and overall points earned [3].

However, other studies have found that possession may not correlate with or may even negatively impact team success. For example, researchers for the FIFA Training centre studied the 2022 FIFA World Cup and found that, for the men's tournament, teams with less possession than their opponents won slightly more games (26 wins versus 23) [1]. Additionally, a study of the 2010-11 season in the Portuguese Premier League found that the amount of ball possession had a very weak negative correlation (r = -0.192) with the match result [18]. In a study analyzing elite leagues in Europe, researchers found that a significant difference in ball possession percentages between winning and losing teams only occurred in matches with wide result margins (3 or more goals). In the other, closely contested matches, the difference in possession between winning teams (51.48% with a SD of 13.05%) and losing teams (48.52% with a SD of 13.05%) was not statistically significant [19]. Similarly, researchers studying the World Cups of 2002, 2006 and 2010 found that ball possession was slightly higher for winning teams (51.6% with a SD of 6.8%) compared to those that drew (49.9% with a SD of 5.8%) or lost (48.5% with a SD of 6.8%), though the differences were not statistically significant [11]. In another study using data from five European leagues, UEFA, and FIFA tournaments, researchers found that possession time was a poor predictor of team success once team quality and home advantage were accounted for [14].

In basketball, intuition might lead one to believe that possession is less important due to the shot clock, which mandates a field goal attempt within 24 seconds in the NBA and most European leagues. Previous studies have shown a positive but insignificant correlation between longer possessions and success. Research on the Spanish Basketball Playoffs from the 2004-05 season investigated the possession durations of winning and losing teams against various defensive systems. They found that, when averaged across all defenses, winning teams had an average possession duration of 13.1 seconds with a SD of 6 seconds, compared to 12.32 seconds with a SD of 5.88 seconds for losing teams [67]. Although significant differences were observed depending on the defensive system faced, these differences did not translate into statistically significant overall differences in possession durations between winning and losing teams.

In American football, significant value is placed on time of possession, notably because it allows the defense to rest, enhancing both offensive and defensive performance. Time of possession refers to the amount of game time an NFL offense has the ball. Researchers studied the 2003-04, 2004-05, and 2005-06 NFL seasons and found that 67% of teams with greater time of possession than their opponents won their games [4]. However, the research recognized potential biases; leading teams often prolong their possessions near the end of the game to conserve their lead. To avoid this bias, the analysis was confined to first-half data. In this analysis, a logisitic regression model was applied to predict the halftime score. The model revealed a negative coefficient for time of possession (β = -0.126), indicating that for each additional minute of possession in the first half, the log-odds of winning at halftime decrease by 0.126. This indicates that more possession, with biases removed, does not contribute positively to winning.

Hockey's analysis of puck possession has comparatively been less robust as it relies on manual tracking [29] or metrics approximating possession such as SAT% [63]. Some studies using manual tracking or SAT% have found a positive correlation between possession and team success. For instance, a study of 243 NHL overtime periods from 2015 to 2021 in which possessions were manually tracked revealed that victorious teams in 3-on-3 overtime generally have a higher count of individual possessions (53 percent of the total number of individual possessions of both teams), a higher duration of individual possession (54 percent of the total duration of individual possession of both teams), and more offensive zone time (57 percent of the total offensive zone time of both teams) compared to teams that lost [29]. Additionally, a study of the 2007-08, 2008-09, and 2009-10 NHL regular seasons revealed that SAT% Tied (even strength SAT% with the score tied) is more predictive of how well a team will perform (r=0.47) than goal ratio (r=0.35) or winning percentage (r=0.34) [63]. This correlation is relatively low compared to our findings, where higher correlations emerge from utilizing PPT data to measure various metrics of puck possession, most notably for Average Offensive Zone Possession Time Differential (Avg. OZPTD).

Although previous hockey analytics research shows a positive correlation between possession and team success, there are challenges to manually tracking possession. As well, SAT% has its limitations, as it does not account for possession in the defensive or neutral zones and may not reflect the strategy of teams that prioritize high-quality shots over quantity.

In recent years, expected goal (xG) models have gained popularity. Originating in football (soccer), xG represents the probability that a scoring opportunity will result in a goal. It addresses some issues with SAT% as it includes weighting shot attempts based on quality, recognizing that certain shots have a higher probability of resulting in a goal. In hockey, efforts to evaluate shot quality began in 2004 [54] [23] [24]. This foundational work led to the first explicit mention of xG in hockey in a 2012 study, which used ordinary least squares (OLS) regression and ridge regression to predict goals, incorporating variables such as goals, shots, missed shots, blocked shots, faceoffs, hits, turnovers, and zone starts [31]. Since 2012, numerous xG models have emerged, each aiming to capture the best set of predictive variables, often including more than ten variables weighted during model training [60] [58] [66] [59]. These models generally outperform SAT% and other metrics in predictive accuracy [60] [59].

However, there are drawbacks to xG models. First, there are many different xG models, which can have varying parameters, potentially leading to inconsistencies when advising a team on how to improve their xG to win more games. Additionally, to our knowledge, there has been limited work on testing the stability of these models, meaning the parameters and weights might not remain consistent from one season to the next. Lastly, because these models have several parameters, determining the specific actions a team can take to improve their xG may not be straightforward.

In this thesis, we utilize PPT data to conduct a detailed investigation into measures of puck possession and their correlation with NHL team success. Our findings indicate that a single metric of possession can be as effective, if not more so, than existing, publicly available xG models in predicting team success.

2.3.2 Examining the Impact of Traffic on Shot Attempts

Traffic Research in Football (Soccer): Traffic-related research in football provides relevant context for this thesis. For example, a 2015 study using tracking data from a professional football league (via Prozone, now Stats Perform [62]), found that defender presence in the shooting lane significantly reduced goal likelihood [28]. However, its contribution to the predictive model used in the paper was limited, suggesting that its effect may be confounded by factors like shot location. A 2016 study computed *dangerousity*, the probability of a goal during possession, for 64 Bundesliga games (2014-15). As an input to their model, they proposed a more detailed traffic metric, *shot density*, which incorporates not only the number of players in the lane but also their proximity to the shooter and whether they are attackers or defenders [26]. While the authors did not isolate the effect of shot density in their model of dangerousity, the study provides a valuable framework for analyzing player presence and placement in traffic. However, we currently do not focus on specific placement within the shooting lane as we recognize that different locations in the shooting lane in ice hockey have varying impacts such as being close to the shooter potentially increasing the chance of blocking the shot attempt, while being near the goaltender may enhance the likelihood of obsructing the goaltender's view. These studies highlight how to account for the number and type of players in traffic. However, ice hockey poses unique challenges not present in football, such as faster puck and player movement, the difficulty of tracking and blocking a small, fast-moving puck, a higher degree of physical contact, and a smaller, enclosed playing surface. These factors underscore the necessity for tailored approaches to fully understand how traffic influences shot outcomes in ice hockey.

Related work in football has also examined how defender positioning can unintentionally impair goalkeeping performance. Using virtual reality to simulate free-kick scenarios, López-Valenciano *et al.* [30] found that defensive walls, though intended to reduce scoring, can actually hinder a goalkeeper's reaction by occluding their view of the ball. This raises an interesting parallel for ice hockey, where skaters in the shooting lane may likewise reduce goaltender effectiveness by unintentionally obstructing the goaltender's view, a possibility we explore later in our analysis.

Traffic-Related Research in Ice Hockey: A common proxy for traffic in ice hockey is the use of shot types typically associated with traffic such as *tips* and *deflections*. Tips occur when a puck traveling towards the net is redirected via the stick with the goal of changing the puck's direction while not adding momentum to the puck. Deflections occur when a puck traveling towards the net is redirected via the body or skates. While these shot types can signal the presence of traffic, they offer only an indirect measure of its broader impact. One study analyzing NHL power plays during the 2015–16 season took a more expansive approach by manually identifying *screens* through video review, defining them as shot attempts where the goaltender's view was obstructed [45]. The study found that screened shot attempts made up 24.9% of total shot attempts and 21.3% of goals. Although this manual approach enables detailed insights, it lacks scalability and is difficult to evaluate further as no comparable data point for screens currently exists in league-wide datasets. More recently, the NHL introduced an "Opportunity Analysis" model designed to evaluate the quality and context of each shot attempt [2]. This model incorporates variables such as player positioning, shot type, and game situation, as well as counts of attacking and defending skaters within fixed distances of the shot cone (the area between the puck and the net), and a flag for "Possible Goalie Vision Block". However, the impact of these traffic-related variables are not publicly available on a per-shot basis, making it difficult to assess their influence on shot outcomes.

Chapter 3

Puck Possessions and Team Success

3.1 Introduction

The margin for victory in the NHL is small, and teams are in constant pursuit of advantages to improve their chances of success (i.e., winning). The unpredictable nature of hockey compounds the difficulty of identifying and quantifying metrics that genuinely influence outcomes. In this chapter, we analyze puck possession and its use as a potential indicator of success. This work is further motivated by the premise shared across sports that possessing the ball or puck for significantly more time than the opponent increases your chance of winning.

Research in other sports present mixed results regarding the correlation between possession and team success; some studies affirm a strong correlation [35] [27] [67] [10] [3], while others find no significant relationship [4] [11] [14] [19] [18] [1]. In the NHL, prior investigations into puck possession have mainly relied on manual tracking [29] or metrics approximating possession such as Shot Attempt Percentage (SAT%), also known as Corsi [63]. SAT% measures a team's share of the total shot attempts in a game. The rationale behind SAT% is that a higher number of shot attempts, which can only be credited during a possession, indicates superior puck control.

Despite these indirect methods for measuring possession, the idea that puck possession is of strategic importance in the NHL has gained traction, particularly for the overtime period since the introduction of 3-on-3 overtime in 2015. This growing emphasis is underscored by the NHL general managers convening this season to discuss potential regulatory changes, such as implementing a shot clock during overtime [53]. This consideration directly reflects concerns about extended possessions during overtime, highlighting the central role puck control has come to play in modern NHL strategies. With the recent introduction of puck and player tracking (PPT) technologies in the 2021-22 NHL season and the established significance of puck possession, this chapter investigates whether teams with more puck possession have greater success.

3.2 Individual and Team Puck Possession

Before we delve into the dataset and analysis, we define the concepts of individual and team puck possessions as utilized in this chapter.

According to the NHL definition for the model that produces the individual possession data we employ, a player is considered to have possession and control of the puck, and thus in individual possession, when they make two or more consecutive touches with the puck. The start of the individual possession is marked by the first touch, which is confirmed upon a second touch. Individual possessions also includes brief moments during one-touch actions, like shots, passes, or area plays (e.g., dump-ins). An individual possession ends when the player is separated from the puck or when another player gains possession. We delineate these episodes to identify windows of time with "no individual possession", representing segments of active gameplay where the puck is not under direct control by any player. This includes scenarios ranging from face-offs, puck battles, and loose pucks to passes, shots, and "area plays" (e.g., dump-ins and dump-outs). The top line in Figure 3.1 shows examples of individual possessions by members of different teams (red and blue lines) and "no possession" (orange dotted lines).

We define team puck possession as the aggregate of individual possessions with continuous possession by members of the same team, interrupted only by game stoppages or a change in possession to the opposing team. Consequently, "no team possession" intervals are distinct from "no individual possession" intervals. The bottom line in Figure 3.1 shows examples of team possession (red and blue lines) and "no team possession" (green dashed lines). As shown in the figure, team possessions end when the puck is last touched by one team, prior to the opposing team gaining possession. Our use of team possession differs slightly from the official NHL definition as the details required to implement the NHL's definition aren't available in the current dataset.

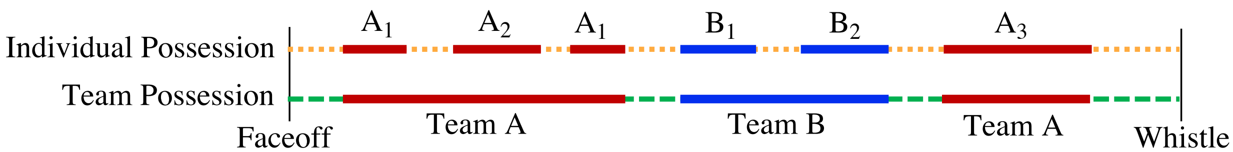

Figure 3.1: Differentiating individual and team possessions

3.3 Dataset Details

In March 2023, a significant advancement was made with the introduction of an individual player possession model into the "DISH" data stream, which features Delayed, Interpolated, Smoothed and Hundred-Hertz enhancements. This dataset is considered unofficial by the NHL and may differ from other datasets that track possession information (e.g., a hand-labeled dataset). This thesis uses the DISH data to compute team possessions, which form the basis of the analysis in this chapter. Since this data only became available in March 2023, the dataset for the 2022-2023 season is limited. Consequently, our analysis focuses on the 2023-2024 NHL season, using data from 780 games played up to January 31, 2024 (the All-Star break). After excluding games with significant data issues, as detailed in the following section, or those with no tracking data, such as the Heritage Classic and the games played in Europe, 708 games remained for analysis.

3.4 Dataset Cleaning and Filtering

The player possession data provided in the NHL's DISH data stream, indicating who held the puck and for how long, lacks broader game context such as powerplay situations and player locations. To address this, we merge it with data from a detailed game information file, enriching player possessions with relevant game context, and then compute team possessions based on this integrated data. Through this process, we encounter challenges that necessitate extensive cleaning and preprocessing to ensure data integrity. Cleaning refers to the process of correcting or removing inaccuracies within the data that can be rectified, such as adjusting timestamps or eliminating duplicates. Conversely, filtering is our strategy for dealing with more complex issues that cannot be directly corrected; it involves the exclusion of entire games from the dataset.

3.4.1 Dataset Cleaning

Possessions Occurring During Stoppages: One issue with the data is that some possessions occur either entirely during stoppages or span active and stopped intervals. These erroneous possessions are identified after merging the player possession file with the game information file and computing active gameplay intervals. To resolve this issue, we eliminate portions of the possession that occurred during stoppages, ensuring accuracy in active play representation.

Abnormal Timestamps and Non-chronological Data Entries: The game information file contains updates every hundredth of a second, but some of these updates display additional digits of precision and are out of sequence. These extraneous updates, found to be non-essential, are removed to maintain dataset integrity. After their removal, the data is re-sequenced to reflect the actual gameplay order.

Clock Resets: Another issue encountered in the dataset are unexpected time jumps, with the time remaining on the scoreboard clock suddenly increasing, leading to duplicated timestamps. These time jumps primarily occur after video reviews where time is added back to the clock, such as when a play is subsequently ruled as offside. Smaller adjustments may also result from false face-offs or if the clock inadvertently continues running briefly after a whistle. The NHL addresses these situations by eliminating all recorded statistics and events that transpired during the time that is later nullified. Our approach mirrors this; upon identification of such a clock reset, we disregard stats and possessions recorded during the time frame subject to the reset.

Overlapping and Duplicate Player Possessions: The last challenge rectified through cleaning is the presence of duplicate or overlapping possessions. Duplicates are resolved by retaining a single entry. For overlapping possessions, we evenly distribute the overlapping time (i.e., the period during which the data indicates two players simultaneously possess the puck) among the involved players.

3.4.2 Dataset Filtering

There are cases where the above cleaning methods are insufficient to repair the data and preserve the integrity of the dataset. Consequently, we establish exclusion criteria based on the severity of data corruption: if the data is compromised for either more than 4% of a game's duration or more than 4% of a team's possession time, the game is excluded from the analysis. This filtering process results in the exclusion of 68 games, leaving 91% of the games for which we have data available for use in the analysis. The distribution of team appearances in the excluded games varied, with an average of 4.5 games per team, a standard deviation of 2.2 games, a minimum of one game for the Tampa Bay Lightning (TBL), and a maximum of 11 games for the Vancouver Canucks (VAN), constituting 22.4% of their total games. In Section 3.4.1, we show that robust analysis can be achieved with just 20% of a team's games in the dataset, as the correlation between early game data and the rest of the season stabilizes after this. Table 3.1 shows the number and percentage of games impacted by each filter; note that the sum of games exceeds 67 and the sum of percentages exceeds 9% since 20 games were subject to more than one filter criterion.

Filter Criteria	**Games Impacted**	**Percent Impacted**
Irregular Possession Lengths	34	4.4%
Clock Gaps	30	3.8%
Irregular Period Lengths	26	3.3%
Possessions with Missing Data	5	0.6%
Excessive Distance Between Puck and Possessor	2	0.3%

Table 3.1: Impact of various filters on the game dataset

Irregular Possession Lengths: Games are flagged for exclusion when the duration without any possession or the length of specific possessions significantly exceeds normal expectations. For total "no possession" time, we apply the statistical outlier definition of mean plus three standard deviations. Given the mean of 62.8% and the standard deviation of 4.8%, this led to the exclusion of any game exceeding 77.2%. Additionally, games with a no possession duration longer than 144 seconds, or any individual possession lasting more than 48 seconds, are excluded, impacting 4.4% of the total games. The limit of 144 seconds corresponds to 4% of a 60-minute game and 48 seconds represents 4% of the average of the per game sums of individual possession times (20 minutes).

Clock Gaps: We identify games with significant windows of time missing in scoreboard data timestamps, indicating lost data segments affecting puck locations, player locations, or possession details. We set a 144-second threshold for these gaps, equivalent to 4% of a 60-minute game. Games exceeding this limit due to missing data are excluded, affecting 3.8% of the dataset.

Irregular Period Lengths: We identify games with periods deviating significantly from the standard 20-minute length in order to filter games with extensive data loss or situations where the data cleaning techniques may be ineffective. We exclude games with periods exceeding or falling short of the expected duration by more than 48 seconds, equivalent to 4% of a 20-minute period, impacting 3.3% of the total games.

Possessions with Missing Data: Games are flagged when they contain missing player data, or missing possession start or end times. This is likely due to tracking failures in the puck or sweaters, or instances where a player does not have a tracking device in their sweater. Games with more than two instances of missing data related to possessions are removed from the dataset; impacting 0.6% of the games.

Excessive Distance Between the Puck and Possessor: We considered possessions where the distance between the puck and its possessor is too large. We focus on possessions where the puck is over 16 feet from the possessor continuously for more than 2 seconds, indicating potential data inaccuracies. Games with a total "excessive distance duration" exceeding 48 seconds, equivalent to 4% of a team's average possession time of 20 minutes, are excluded, affecting 0.3% of the total games. In previous work, we adjusted the timestamps for events like shots and passes to try to more accurately capture the point of release [49]. We considered a similar approach in this work but the problem proved more difficult because we found instances where the distance between the puck and possessor is large in the middle of the possession. Adjusting such possessions would amount to building a new model, which is currently the domain of the NHL.

3.5 Analysis of Team Possessions

In this section, we explore the relationship between team success and possession metrics, focusing on team possession percentage, aggregate individual possession count differential, and offensive zone team possession time differential. Team success is measured primarily by goal differential because it is adaptable across game situations, unlike points per game, which is less flexible. Additionally, for the games in the dataset, average goal differential exhibits a very strong correlation with average points per game ($r = 0.95$). Note that, unless stated otherwise, the analysis includes all strengths (i.e., even-strength and powerplays) and pertains exclusively to regulation time. This means that for our analysis, each team is awarded one point if a game goes to overtime.

3.5.1 Team Possession Percentage Versus Team Success

Team possession percentage is calculated by dividing the total duration of team A's possession by the combined possession duration of team A and the opposing team. Team possession percentage is calculated for each team in every game and subsequently averaged across all games played. We compute the correlation between average team possession percentage and average goal differential (Average GoalDiff), as well as average goals for (Average GF) and average goals against (Average GA), aiming to delineate the correlations of possession with offensive and defensive metrics.

As shown in Table 3.2, average team possession percentage is moderately correlated with Average GF, suggesting that teams with higher possession tend to score more goals. In contrast, the correlation between average team possession percentage and Average GA is weaker, implying that while possession might play a role in limiting opposition goals, its effect is not as strong.

Furthermore, the analysis reveals a nonlinear relationship among the correlations of average team possession percentage with Average GF, Average GA, and Average GoalDiff. Intuition might lead one to expect these correlations to sum linearly; for example, given the correlation between average team possession percentage and Average GF is +0.56, and between average team possession percentage and Average GA is -0.38, one might anticipate the correlation between average team possession percentage and Average GoalDiff to be the difference, equating to +0.94. This is not true and can be explained by understanding the correlation formula's normalization process. The Pearson correlation coefficient is:

$$r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{3.1}$$

where $\text{cov}(X, Y)$ is the covariance between X and Y, and σ_X and σ_Y are the standard deviations of X and Y, respectively. The denominator normalizes the covariance by dividing it by the product of the standard deviations of X and Y, ensuring the correlation values fall within the range of -1 to +1. Given the distinct standard deviations for Average GF (0.44), Average GA (0.39), and Average GoalDiff (0.70), this normalization introduces nonlinearity to the relationships.

Possession Metric	**Success Metric**	**r-value**
Average Team Possession Percentage	Average GF	0.56
Average Team Possession Percentage	Average GA	-0.38
Average Team Possession Percentage	Average GoalDiff	0.56

Table 3.2: Correlations between average team possession percentage and team success metrics

3.5.2 Possession Count Differential Versus Team Success

Shifting the analysis from the percentage of team possession to the aggregate quantity of individual possession instances can potentially offer new insights by capturing both the totality of possessions gained through turnovers or puck battles and the extent of puck movement within team possessions. To assess which teams excel in managing aggregate individual possession quantity, we introduce a metric called average possession count differential.

For team A, the possession count differential is defined as the count of team A's individual possessions, minus the count of the opposing team's individual possessions. We compute this metric for each game and subsequently determine the average across all games played by each team. Utilizing this metric reveals a slightly enhanced correlation with Average GoalDiff ($r = 0.63$) compared to the correlation between average team possession percentage and Average GoalDiff ($r = 0.56$). This improved correlation may indicate the potential impact of frequent and dynamic possession changes to outscoring opponents, suggesting a strategy centered on maximizing possession instances correlates positively with achieving a better goal differential.

3.5.3 Offensive Zone Possession Time Differential Versus Team Success

We now examine the significance of possession within the offensive zone. The rationale for this approach is that possessions in the defensive or neutral zones can serve to facilitate transitions, whereas offensive zone possessions might contribute more directly to scoring goals and outscoring the opponent. In this refined analysis, we introduce a new metric, Offensive Zone Possession Time Differential (OZPTD), which is defined as the sum of the duration of team A's individual possessions in the offensive zone, minus the sum of the duration of the opposing team's individual possessions in their offensive zone (team A's defensive zone). For possessions that span multiple zones, the duration is allocated proportionally based on the time spent in each zone.

In Figure 3.2, the zones of offensive possession for Team A and the opposing team are delineated. To calculate OZPTD, we subtract the offensive zone possession time of the opposing team from that of Team A. For example, if Team A controls the puck in their offensive zone for 10 minutes and the opposing team controls the puck in their offensive zone for 8 minutes, then Team A's OZPTD is calculated as 10 - 8 = +2, while the opposing team's OZPTD is -2.

Similar to the previously examined metrics, OZPTD is computed for each game and subsequently averaged across all games played by each team.

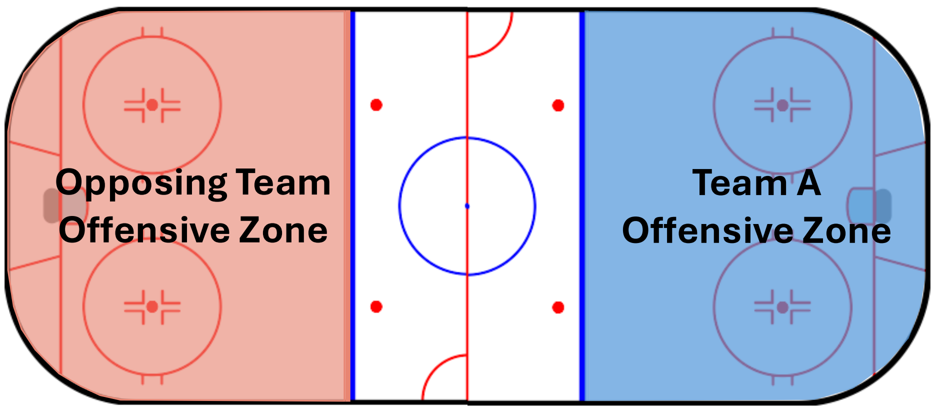

Figure 3.2: Visual representation of offensive zone possession time differential between Team A and the opposing team

As shown in Figure 3.3, the analysis reveals a strong correlation of 0.77 between Average OZPTD and Average GoalDiff. This finding highlights the importance of not just possessing the puck more than the opponent, but doing so in the offensive zone where it more strongly correlates with outscoring the opponent. Teams such as the Colorado Avalanche (COL) and Florida Panthers (FLA) who, on average, maintain offensive zone possession longer than their opponents, typically see positive goal differentials. Interestingly, the Winnipeg Jets (WPG), Boston Bruins (BOS) and Vancouver Canucks (VAN) achieved the highest Average GoalDiff values despite having values of Average OZPTD near the league average of 0. In contrast, the San Jose Sharks (SJS) and Chicago Blackhawks (CHI) exhibit negative Average OZPTD values and, correspondingly, negative Average GoalDiff values. Recognizing that SJS and CHI may contribute significantly to the strong correlation, we compute the correlation coefficient without those two teams and observe an r-value of 0.63. In future work we plan to examine if offensive zone possession counts and differential are also correlated with success.

As Average GoalDiff is very strongly correlated ($r = 0.95$) with average points per game for the games in the dataset, Figure 3.3, which arranges teams from left to right based on Average GoalDiff, provides a useful reference for readers to assess team standings, offering

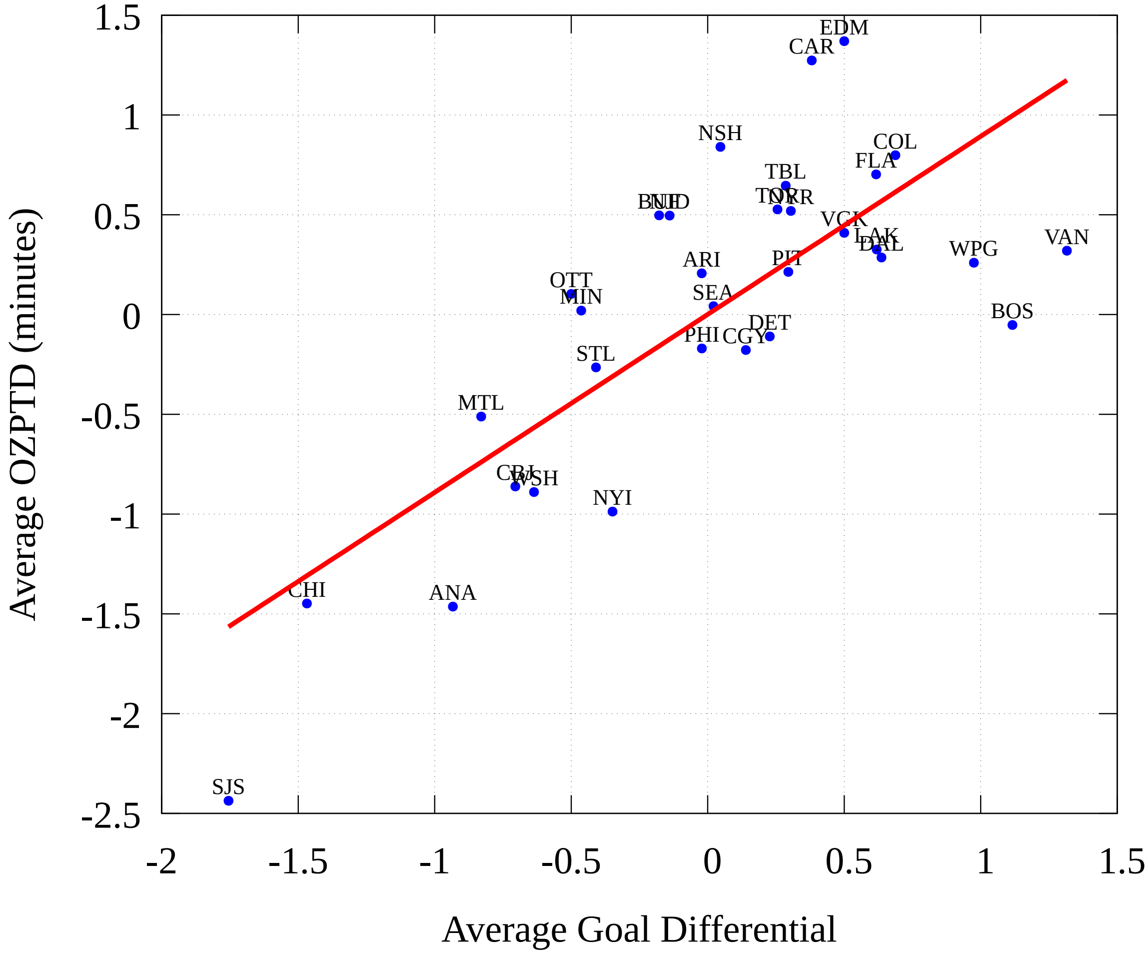

Figure 3.3: Average offensive zone possession time differential versus average goal differential ($r = 0.77$)

a more accurate perspective than actual standings that include games outside our analysis.

Recognizing the significance of the correlation, we conduct a deeper examination of its components, focusing exclusively on even-strength play. The correlation remains strong at 0.73, indicating that the initial correlation is not simply a byproduct of power plays but is also prevalent during even-strength play, reinforcing the importance of offensive zone control throughout the game.

Given Average OZPTD's strong correlation with Average GoalDiff, we also analyzed it on a per-game basis, as depicted in Figure 3.4. This per-game analysis shows an r-value of 0.00, indicating no correlation. This finding suggests that, despite the correlation between Average OZPTD and Average GoalDiff across many games, individual games show high variability. Thus, while superior offensive zone possession doesn't guarantee game victories, teams with consistently higher offensive zone time may outscore their opponents over the course of a season.

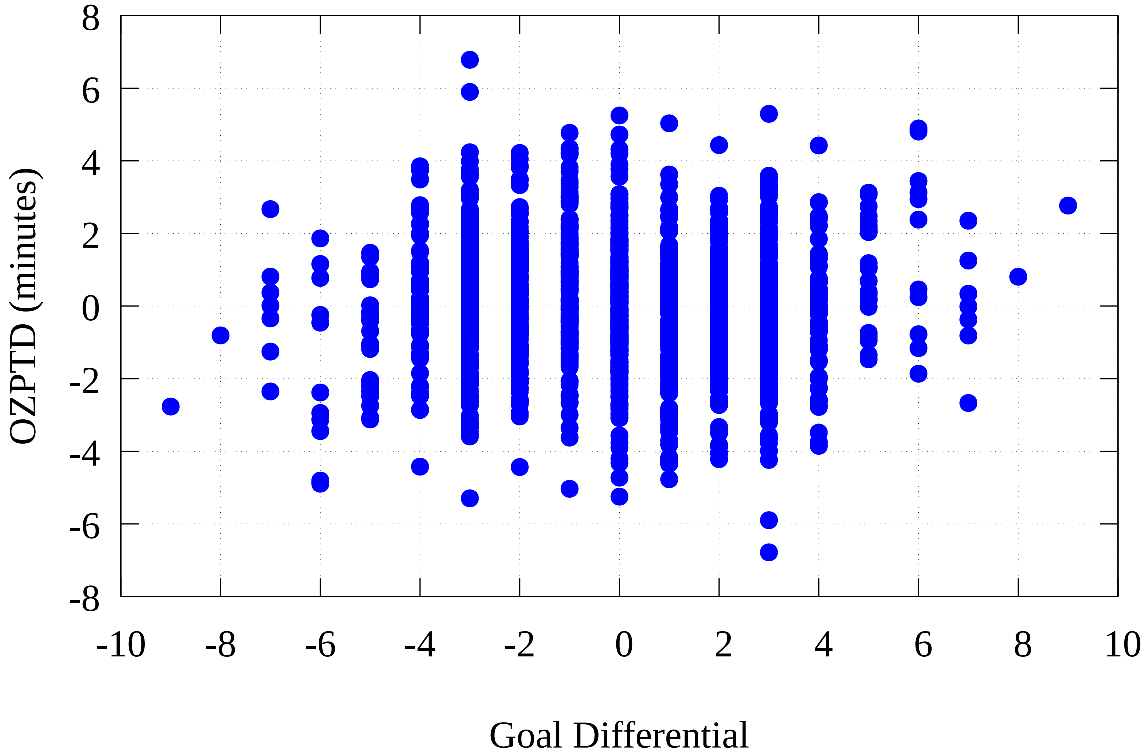

Figure 3.4: Game-level offensive zone possession time differential versus goal differential (r = 0.00

3.5.4 Possession Across Different Strengths

Building on the earlier findings, this section delves deeper into possession metrics across different strength scenarios, as shown in Figure 3.5. We observe that at even strength, possession is typically balanced between teams. However, with a plus-1 strength advantage, teams dominate possession. In contrast, a minus-1 strength differential leads to a substantial decrease in possession percentage for the disadvantaged team.

The variance in average team possession percentages is notably higher in even strength scenarios than in situations of plus-1 or minus-1. Specifically, the Chicago Blackhawks (CHI) and San Jose Sharks (SJS) show lower possession percentages at even strength, yet they are near the league average in plus-1 and minus-1 situations.

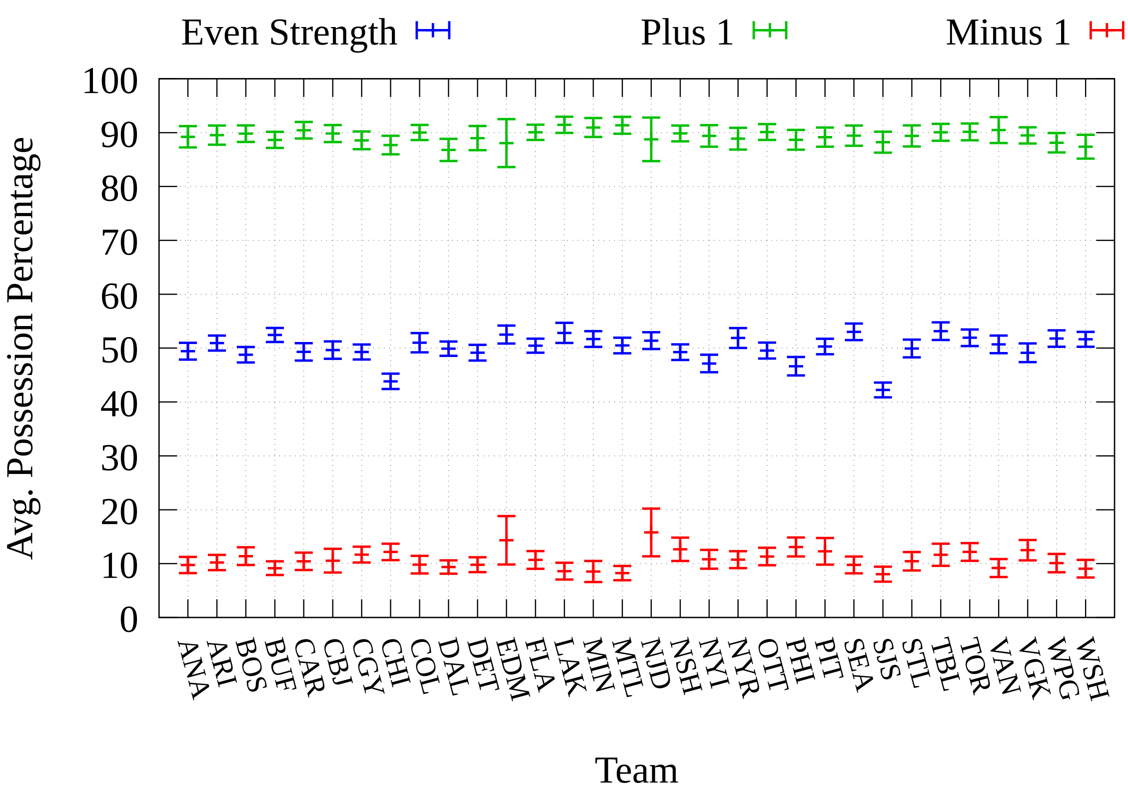

Figure 3.5: Average team possession percentage: even strength, +1 and -1

3.6 Meta Metrics: Evaluating Average OZPTD

Due to Average OZPTD's strong correlation with Average GoalDiff, and thus its potential to offer insights, it is imperative to evaluate this new metric. We utilize the notions introduced by Franks et al. [17], which emphasizes three key properties: stability, discrimination, and independence. While some of our tests of these properties differ slightly from those suggested in their paper, we maintain the spirit of each property. Stability measures the consistency of a metric across seasons or portions of a season (e.g., the value of using the metrics in predictions), discrimination measures its ability to distinguish between players or teams, and independence assesses whether it provides unique insights when compared with existing metrics.

3.6.1 Stability

To assess the stability of Average OZPTD and determine its potential for predictive use, we calculate Average OZPTD separately for the first and second halves of the dataset. The observed strong correlation ($r = 0.84$) between Average OZPTD in the two halves, depicted in Figure 3.6, validates the metric's consistency. To further our understanding of the metric's stability, we conduct a rolling correlation analysis where the Average OZPTD is calculated for each team across incremental segments of the dataset, ranging from 5% to 50% and then these values are compared with Average OZPTD for the remaining games. The correlation starts at 0.68 when using the first 5% of the games to predict the Average OZPTD of the remaining 95% of the games and stabilizes above 0.80 when using the first 20% of the games to predict the remaining 80% of the games.

Predictive Power:

To evaluate the predictive accuracy of Average OZPTD, we divide the dataset into two halves. Using data from the first half of our dataset, we build a linear regression model to establish the relationship between Average OZPTD and Average GoalDiff. We then test this model with data from the second half of the dataset, using Average OZPTD to predict Average GoalDiff for each team. The comparison between the predictions and the actual outcomes results in an R^2 value of 0.49, and a correlation coefficient of 0.73. This correlation indicates a relatively strong correlation between the predicted and actual values.

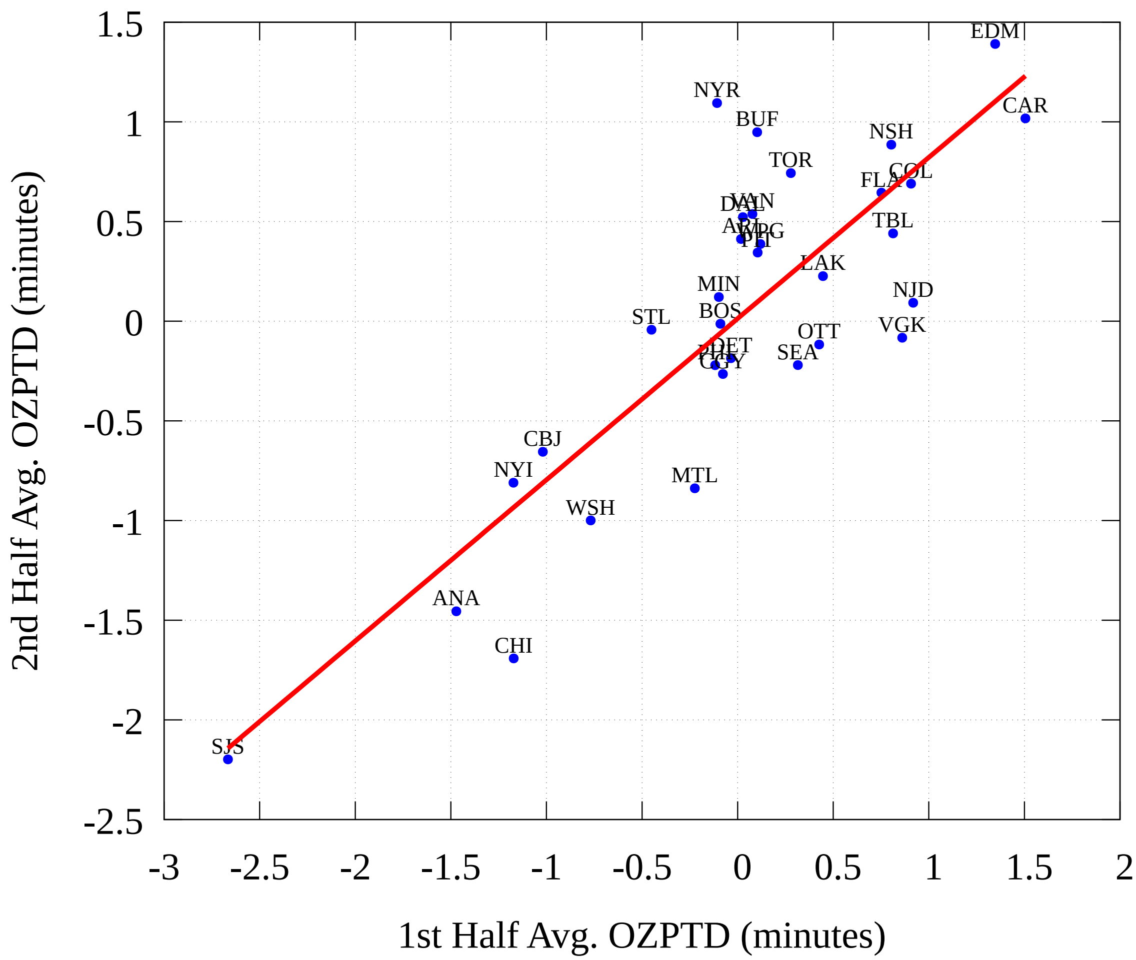

Figure 3.6: Average offensive zone possession time differential across dataset halves ($r = 0.84$)

To compare results obtained when using other metrics for predicting team success, we found that the 2012 study by Macdonald yielded a correlation between actual and predicted goals of 0.69 using his ridge regression model [31]. Next, we found that a 2022 study, described on the Hockey-Statistics website [59], that reports that using their xG model and the xG model from Evolving-Hockey [66] to predict the expected goals for percentage (xGF%) yielded an R^2 value of 0.49 in both cases. This prediction was based on the xGF% from the first 41 games of a season for each team to forecast the xGF% for the last 41 games. Note that xGF% is the ratio of a team's expected goals for compared to their opposition. Another study from 2015 considers a different expected goals model, and found that when using their model built using the first 40 games to predict GF for each team at the end of the season, they obtained an R^2 value of 0.51 [60].

While a more in-depth evaluation needs to be done using larger sample sizes with a direct comparison between metrics, this preliminary investigation indicates that our fairly simple Average OZPTD metric performs on par with existing, relatively complex models (because they typically use a large number of parameters that appear to require tuning) for predicting team success.

3.6.2 Discrimination

Our evaluation of Average OZPTD's discriminatory power, depicted in Figure 3.7, shows the Average OZPTD for each team, including 95% confidence intervals. There are statistically significant differences between some teams, however the overlap in confidence intervals for many teams indicates that the metric might have moderate discriminatory power.

3.6.3 Independence

In assessing the independence of Average OZPTD, we revisit SAT% and expected goals (xG). SAT% has traditionally been used to approximate possession by measuring the ratio of a team's shot attempts (goals, shots on net, shots that miss the net, and blocked shots) to the total shot attempts in the game. xG models attempt to improve on the predictive power of SAT% by including several variables related to the shot to better describe the context around the shot. To analyze the independence of Average OZPTD from these two metrics, we show the correlation between them, as well as each metric's correlation with team success (Average GoalDiff) as shown in Table 3.3. Note that the data used for the xGF% model is from Natural Stat Trick [40].

The results indicate that Average OZPTD is strongly correlated to SAT% but shows a stronger correlation to Average GoalDiff compared to the correlation between SAT% and Average GoalDiff. This stronger correlation for Average OZPTD implies it provides additional insights beyond SAT%, especially in relation to game outcomes. The results also indicate that Average OZPTD is strongly correlated to xGF%, with both metrics having the same correlation to Average GoalDiff. However, as mentioned previously, there are

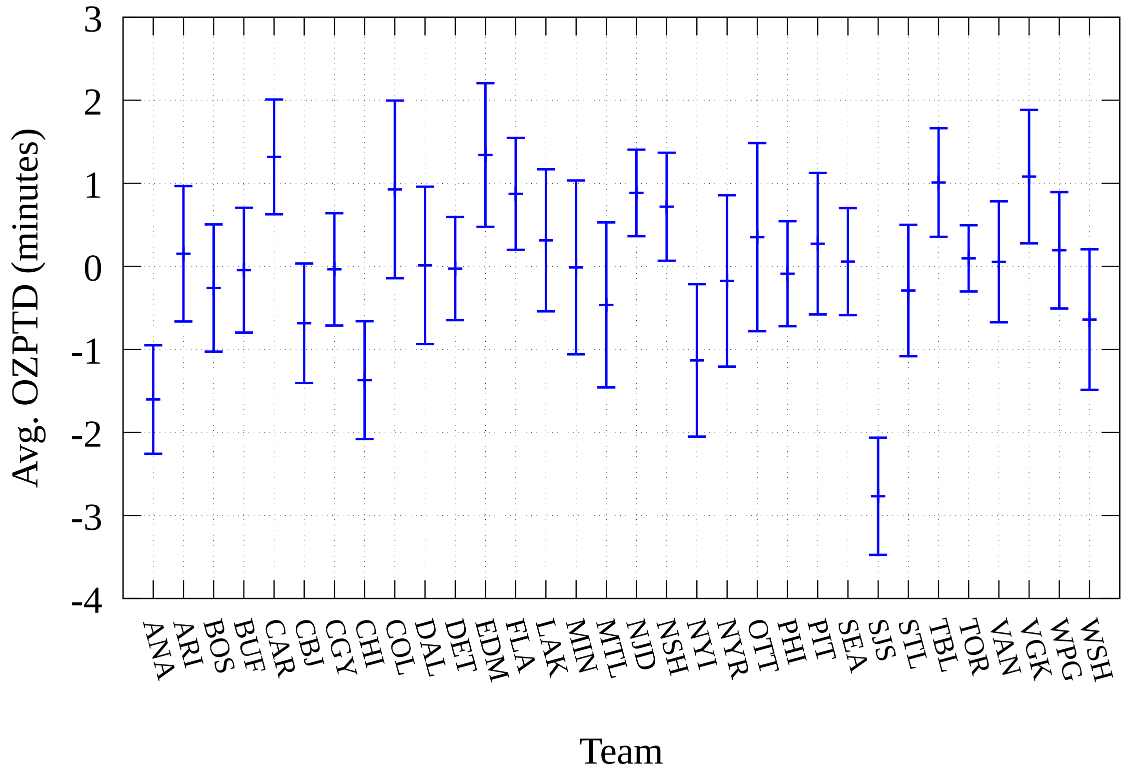

Figure 3.7: Average team offensive zone possession time differential

potential drawbacks to xG models, such as possible inconsistencies in parameters across different models and the complexity of determining specific actions to improve xG.

Metric	**Correlation with Average OZPTD**	**Correlation with Average GoalDiff**
Average OZPTD	1.00	0.77
SAT%	0.83	0.62
xGF%	0.88	0.77

Table 3.3: Correlation of metrics with average offensive zone possession time differential and average goal differential

In terms of evaluating established metrics, the work by Franks et al. [17] does evaluate some NHL metrics for individual players but does not include team metrics. As described in Section 3.6.1, some previous studies have examined the predictive power of various expected goal (xG) models. However, there is a lack of work in evaluating those metrics in terms of stability, discrimination and independence. In the future, we hope to evaluate established team performance metrics alongside the metrics posited in this thesis.

Chapter 4

Examining the Impact of Traffic on Shot Attempts

4.1 Introduction

Shot quality in ice hockey is influenced by many factors including distance, angle, and preceding events [60] [58] [66]. One often-discussed but less precisely quantified factor is *traffic*, which we define as the presence of skaters in or near the *shooting lane*: a triangular area between the puck and the posts ("near" is defined precisely in Section 4.3). Traffic is widely believed to determine the quality of scoring chances but its effects are nuanced and to this point have not been studied in detail. In this thesis, using puck and player tracking (PPT) data from the National Hockey League (NHL), we examine the impact of traffic on shot outcomes.

In our initial exploration of traffic and its relationship to shot outcomes for this thesis, we observed that traffic levels were strongly correlated with shot location. Specifically, when grouping shot attempts by traffic level (N) and computing average distance and angle within each group, we found a strong positive correlation between N and shot distance ($r = 0.94$) and a moderate negative correlation between N and shot angle ($r = -0.64$). This indicates that shot attempts with more traffic tend to be taken from farther away and towards the center of the ice. Because traffic is correlated with shot location, any attempt to measure the impact of traffic must first control for the confounding effects of shot location. To do this, we group similar shot attempts together by dividing the offensive zone into regions based on distance and angle, as shown in Figure 4.1 (see Section 4.4 for details on how these regions are defined). We then examine how traffic impacts shot outcomes within each region.

Each region is labeled with their distance range in feet and their orientation: "c" represents center-angle shot attempts, "o" represents off-center shot attempts (which are only included in the larger distance regions), and "w" represents wide-angle shot attempts. Figure 4.2 shows the relationship between traffic and goals for each region, with the x-axis representing traffic levels within each of the ten distance-angle regions (ordered by distance and then angle), and the y-axis showing the number of goals. This figure highlights how analyzing traffic can reveal new insights about shot attempts. For example, in all close-range regions (within 23 ft), the number of goals scored with no traffic exceeds the combined total of goals scored with traffic. While this figure doesn't show how many shot attempts are in each traffic-region group (we explore that in Section 4.5), it provides a preview of the types of patterns we examine. One reason for the high number of goals from close range may be the limits of human reaction time. Prior research suggests that the fastest male humans have an average reaction time of 0.22 seconds [7], which means a 70 mph shot attempt would need to be at least 23 feet away for the goaltender to react in time (before even accounting for the time needed to move their body). Across all regions, the highest number of goals occurs in the 9–23c region with zero traffic. For long-range shot attempts (23+ ft), the relationship becomes more nuanced. Section 4.5 explores this further, including the impact of traffic based on whether it is dominated by attacking or defending skaters.

4.2 Determining Shot Timing and Duration

In this section, we describe our process for identifying and augmenting shot attempts, and then determining a shot start and end time in which to measure traffic.

4.2.1 Identifying Shot Attempts

Each official shot attempt is available from the NHL via their Application Programming Interface (API) [38], but this data is only recorded in whole second granularity using scoreboard time. In contrast, PPT data provides locations every hundredth of a second. Since the puck travels about 103 feet per second (for a 70 mph shot), the coarse resolution of scoreboard time is insufficient for aligning tracking data with shot attempts. As such, we construct our shot dataset using a combination of *detected shot attempts* based on the

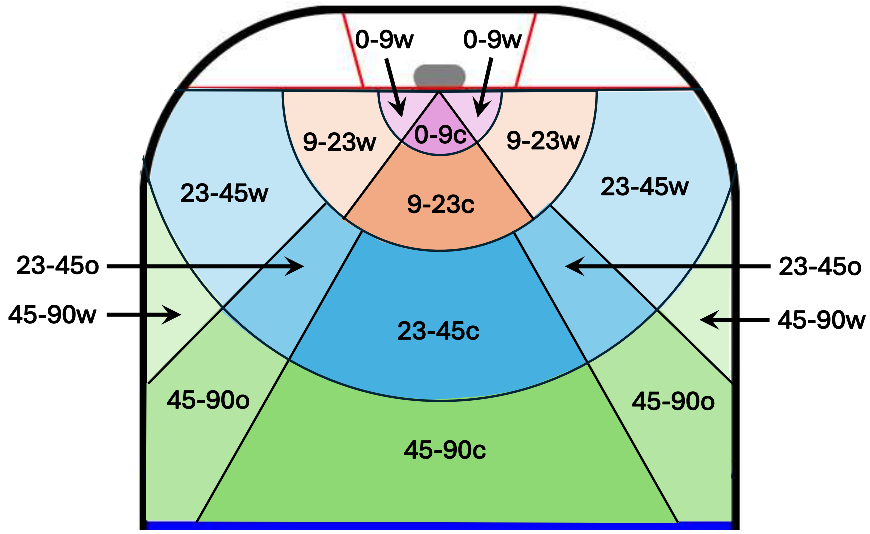

Figure 4.1: Shot location regions used to control for distance and angle in traffic analysis.

NHL's physics-based shot event model applied to PPT data, and *inferred shot attempts* which we derive from puck touch data, which represents the NHL's effort to determine every instance of a player making contact with the puck. Combining detected and inferred shot attempts allows us to reconstruct all official shot attempts for which valid tracking data exists.

Detected Shot Attempts: The NHL detects shot attempts using the PPT data and an automated, physics-based event classification algorithm that identifies when a player directs the puck toward the net. Detection is extremely difficult and consequently, errors may occur such as passes being misclassified as shot attempts. To improve accuracy, the NHL incorporates a manual shot reviewing process where human reviewers verify and correct PPT shot event data. Unfortunately, even after manual review, discrepancies remain where an official shot attempt is not recognized by the shot detection system or where attributes such as the shooter or shot location differ between sources. To address this, we compare the PPT data with the official NHL API data and infer the timing of undetected shot attempts using puck touch data, as described in the following section.

Inferred Shot Attempts: We identify the timestamps (or release times) of undetected shot attempts in the PPT data by comparing NHL API data and puck touch data. From the NHL API data, we can obtain the scoreboard time, shooter, and shot outcome. To match it with a puck touch, we implemented a matching algorithm, adapted from a method

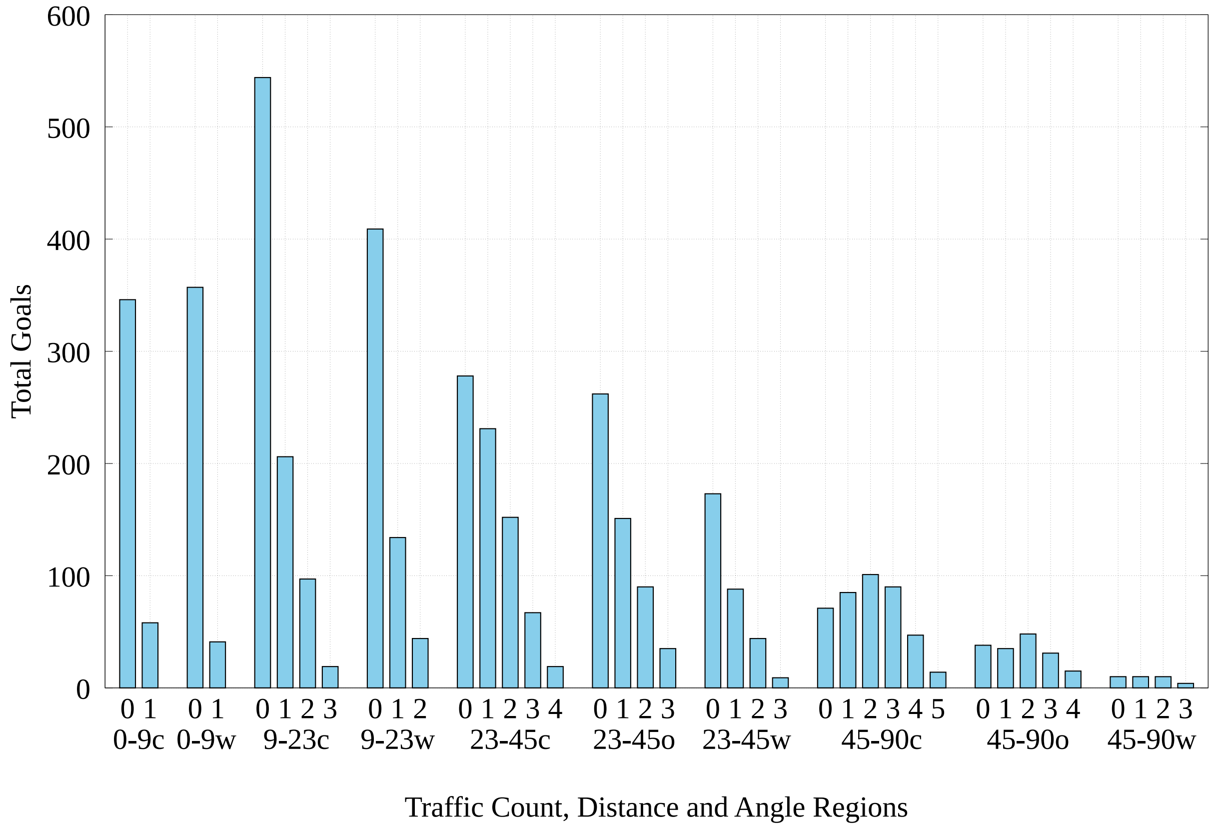

Figure 4.2: Goals scored across traffic levels and location regions.

originally developed by the NHL's Research and Development team [52]. This algorithm attempts to match official shot attempts and puck touches by considering the touch's timing, location, and the puck's incoming and outgoing direction and velocity.

4.2.2 Augmenting Shot Attempts

Potential Start Time Inaccuracies: After completing these steps, we obtain a dataset of shot attempts with estimated release timestamps. Some timestamps originate from the NHL's physics-based shot detection model, while others are derived from the end of a puck touch. In an attempt to verify and possibly improve shot release times, we use techniques from previous work on pass and shot timing corrections [50]. Specifically, we verify that the puck is within four feet of the shooter's location at the moment of release. If not, we attempt to adjust the timestamp accordingly (adjusting 20.8% of all shot attempts).

Handling Compound Shot Attempts: Tips occur when a puck traveling towards the net is redirected via the stick with the goal of changing the puck's direction while not adding momentum to the puck. Deflections occur when a puck traveling towards the net is redirected via the body or skates. We refer to both the initial play action and the tip or deflection as a *compound shot*. The skater who tipped or deflected the puck is credited as the shooter, so the initial play action is not officially recorded as a shot attempt. However, we are interested in the initial play action as the tip or deflection is usually the result of traffic, and the initial play action reflects the traffic leading to the tip or deflection. Thus, we replace tipped and deflected shot attempts with their corresponding initial play actions. First, we identify shot attempts with type "tip" or "deflection" in the NHL API data. For each of these shot attempts, we trace the event back to the initial play action by finding the last recorded puck touch by a teammate of the skater who tipped or deflected the shot. We then calculate traffic based on the initial play action while preserving the final shot outcome (e.g., goal, save, block, or miss).

4.2.3 Results of the Shot Identification and Augmentation Process

We analyze data from 891 NHL games played during the 2024-25 season, up to February 9, 2024 (the 4 Nations Face-Off break). We define a shot attempt as *matched* if it is either detected by the NHL's shot classification algorithm or inferred from puck touch data. After excluding games where fewer than 50% of official shot attempts were matched, 870 games remain in our dataset. Across these games, there were 103,948 official shot attempts, with a total of 80,473 detected shot attempts, 12,887 inferred shot attempts, and 10,588 shot attempts for which we were unable to determine a timestamp. The remaining 870 games in our dataset had at least 75% of the official shot attempts matched, and more than half of the games have 90% or more matched. Table 4.1 summarizes the results of the shot-matching process.

4.2.4 Shot Duration Considerations

We now turn to the question of whether or not a window of time should be used when determining traffic and if so, how to determine an appropriate window. Capturing traffic only at the exact moment of the shot release may miss skaters who obstruct the shot

Matching Stage	Shot Count	Percentage
Total official shot attempts in 870-game dataset	103,948	100.0%
Matched via NHL shot event data (“detected shot attempts”)	80,473	77.4%
Matched via puck touch data (“inferred shot attempts”)	12,887	12.4%
Total matched shot attempts (with PPT data)	**93,360**	**89.8%**
Unmatched shot attempts (no PPT timestamp available)	10,588	10.2%

Table 4.1: Summary of shot identification and matching process for 870 games in dataset.

attempt later in its trajectory or skaters seen by the shooter or goaltender just prior to the shot attempt. Consider Figure 4.3 which presents three screenshots from the broadcast of a shot attempt during the Utah Hockey Club versus Chicago Blackhawks game on October 8, 2024. In the top picture, Utah player #22 (Jack McBain) is positioned in front of the goaltender. As Utah player #11 (Dylan Guenther) prepares to shoot, McBain likely obstructs the goaltender's view of the puck and Guenther's shooting lane. By the time the shot attempt is released in the middle picture, McBain has moved out of the shooting lane. There appears to be little traffic, aside from the two skaters near the left post. In the bottom picture, Chicago player #8 Ryan Donato, whose body was not in the shooting lane at the time of the shot release, blocks the shot attempt. Although Donato's stick may appear to be in the shooting lane in the middle picture, we are unable to capture it as each player's location is tracked using a single LED embedded in the sweater, as detailed in the next section. This example, along with many others like it, highlights the need to consider measuring traffic over a window of time rather than just at the moment of release.

Given these observations, we define two time windows around each shot attempt: a *pre-release window* which captures the shot wind-up and decision-making phase and a *post-release window* which approximates the shot duration. The NHL shot data set does not have end times for all attempts, thus, to ensure uniform treatment across all shot attempts, we do not rely on these end times and instead approximate the duration of each shot attempt using a predefined post-release window for all shot attempts. While we could attempt to determine the shot duration ourselves, estimating this proved challenging and ultimately falls within the NHL's shot detection process. Our examination of video for a sample of shot attempts suggests that 0.5 seconds is a reasonable approximation for the time required to release a puck, as well as for the duration of a shot attempt. However, even a short post-release window (e.g., 0.1 seconds) often included post-shot events such as rebounds which are unrelated to the initial attempt. For this reason, for the analysis in this thesis, we use only a 0.5-second pre-release window and omit the post-release window. This approach also best reflects the information available to the shooter at the time of the shot attempt which is ideal for shooter analysis and shot prediction models.

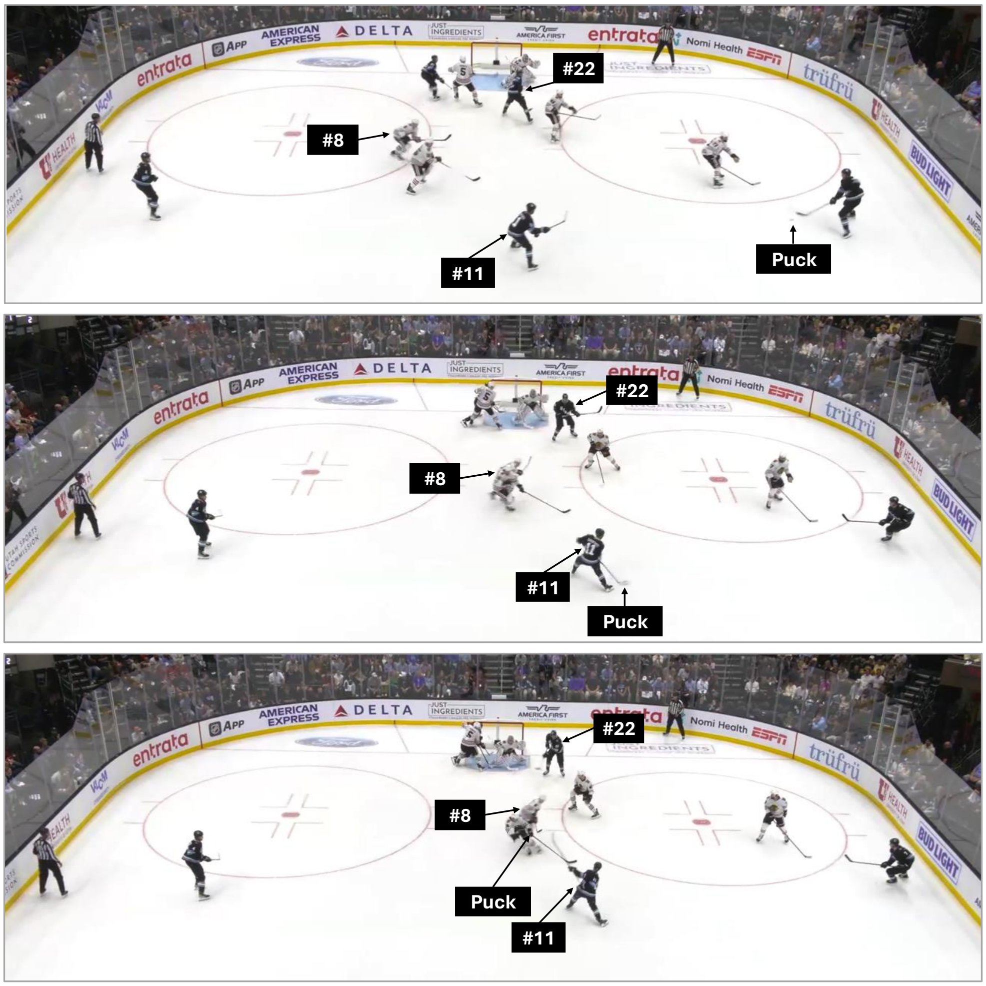

Figure 4.3: Sequence of screenshots showing traffic before, during, and after a shot attempt. The top and middle frames are 29 broadcast frames apart ($\approx 0.48\,\text{s}$) and the middle and bottom are 14 frames apart ($\approx 0.23\,\text{s}$). Utah vs. Chicago, October 8, 2024.

Selecting a suitable approach for a shot duration window is challenging as different choices carry trade-offs that may impact the results. We explore the impact of these considerations in Section 4.6. Although we find that different choices may yield slight variations in specific results, the overall trends remain consistent.

4.3 Defining and Calculating Traffic

Once we have determined the time window in which to measure traffic for each shot attempt, the next challenge is identifying whether skaters are in or near the shooting lane during the shot window using their location data. Each player wears a single light-emitting infrared diode (LED) positioned in their sweater between the top of their back and right shoulder, as illustrated in Figure 2.1. Because this is the only tracked point on the player, we are unable to determine the position of their limbs or stick. To account for this, we introduce a buffer around the shooting lane to define the broader *traffic lane*, which represents skaters that may be obstructing the shot attempt even if their LED is not directly within the shooting lane.

Figure 4.4 illustrates a power-play goal by Ottawa player #18 (Tim Stützle) at 5:52 of the 1$^{\text{st}}$ period in Ottawa's game against Florida on October 10, 2024. The top image shows the game broadcast at the time of the shot, while the bottom image shows a frame from an animation we produced. The shooting lane is highlighted in blue, and the traffic lane is defined as the combined area of the blue triangle and adjacent green rectangles. Skaters detected within the traffic lane are shown in red. The goaltender is not shown as we ignore them in traffic calculations.

After examining a sample of shot attempts, we determined that placing rectangles two feet outside the shooting lane reasonably matched our visual assessments of traffic. Typically, if a skater's LED is within two feet of the shooting lane, it indicates that they are likely obstructing the shot attempt with their body. Conversely, if a skater's LED is not within this range, they are unlikely to significantly impact the shot attempt, though exceptions do occur. Similar to the shot duration window, we recognize the potential variability in results based on the chosen buffer size (green rectangle). To address the variability, we perform a sensitivity analysis in Section 4.6 where we repeat our computations with different buffer sizes.

4.4 Dividing the Offensive Zone into Regions

As shown in our initial analysis (Section 4.1), traffic levels are strongly correlated with shot location. To accurately evaluate the impact of traffic on shot outcomes independently from shot distance and angle, we group similar shot attempts together. To do this, we start by dividing the offensive zone into ten regions based on shot distance and angle. We chose ten regions because we wanted a large enough number of regions to have a limited

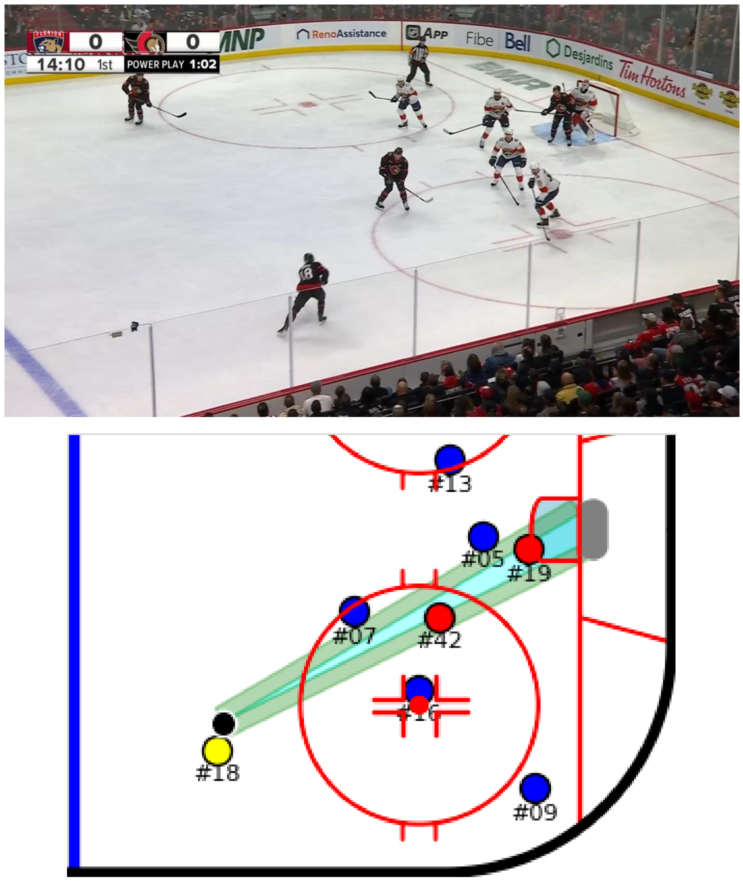

Figure 4.4: Traffic lane analysis for a goal scored during the Florida Panthers vs. Ottawa Senators game on October 10, 2024. The yellow skater is the shooter, red skaters are in the traffic lane, and all blue skaters are outside of the traffic lane

range of distance and angle within each region while keeping the number of regions small enough so that they could be represented graphically in a meaningful way. We analyzed the impact of the number of regions and found that while the quantitative results may change, the qualitative results still stand.

To define the regions, we implemented an algorithm that minimizes within-region variation in shot distance and angle. The goal is for shot attempts within each region to have similar distances and angles but vary in traffic. Shot attempts taken from below the goal line or outside the offensive zone are excluded from this analysis. Distance and angle variation within each region are defined below. 90 degrees is used as the denominator for angle as it represents the maximum possible shot angle deviation, while the maximum distance within each region normalizes variation relative to the range within each region.

$$\text{Distance Variation} = \frac{\max(\text{avg distance}) - \min(\text{avg distance})}{\max(\text{avg distance})} \tag{4.1}$$

$$\text{Angle Variation} = \frac{\max(\text{avg angle}) - \min(\text{avg angle})}{90} \tag{4.2}$$

To minimize variation in shooting location within each region, the algorithm evaluates all combinations formed using four distance regions and either two or three angle regions per distance region for a total of 10 regions. Specifically, the two closest distance regions are divided into two angle segments, while the two farthest are divided into three, resulting in a total of 4,088,304 combinations. For each combination, we compute the variation in distance and angle within each region using the formulas above. We then sum the distance and angle variation for each region and evaluate each combination based on the highest such sum among its regions. The ideal set of regions is the one that minimizes this maximum within-region variation. This approach ensures that no single region has disproportionately high spatial variance which could otherwise confound the effects of traffic with those of shot location.

Resulting Regions: Unless otherwise specified, all numeric ranges in the rest of the thesis are expressed with exclusive lower bounds and inclusive upper bounds (e.g., 9–23 refers to (9-23]). We omit brackets and parentheses for readability. The resulting regions shown in Figure 4.1 separate the shot attempts into four distance ranges: 0-9, 9-23, 23-45, and 45-90 feet. The two smaller-distance ranges are further divided into two angle ranges: center (0-37° of the meridian line) or wide (37-90°). The two larger-distance ranges are divided into three angle ranges: center (0-29° of meridian), off-center (29-45°), and wide (45-90°). Across all regions, the maximum distance variation of any region was 12% of the maximum distance in each region, and the maximum angle variation was 6% of 90°. We note that this approach assumes that the traffic lane is symmetric for the left and right sides of the net. Factors such as goaltender and player handedness may introduce asymmetries in how traffic forms. While we do not explicitly account for these possibilities, they represent an area for future exploration.

4.5 Results: Traffic Versus Shot Outcomes

In Figure 4.5, the x-axis represents each of the ten distance-angle regions, ordered by distance from smallest to largest, and then by angle, from center to wide. Within each region, the shot attempts are grouped by traffic. We refer to each of these bars as a traffic-region group. The y-axis shows the percentage of all shot attempts that result in one of four outcomes: goals (Goal%), shots on goal that are saved by the goaltender (Saved%), blocked shot attempts (Blocked%), and missed shot attempts (Missed%). The combination of Goal% (blue) and Saved% (green) together forms the shots on goal (SOG) percentage. The rightmost group aggregates all shot attempts for comparison. To ensure meaningful statistical comparisons, we exclude any traffic-region group that has less than 100 shots on goal.

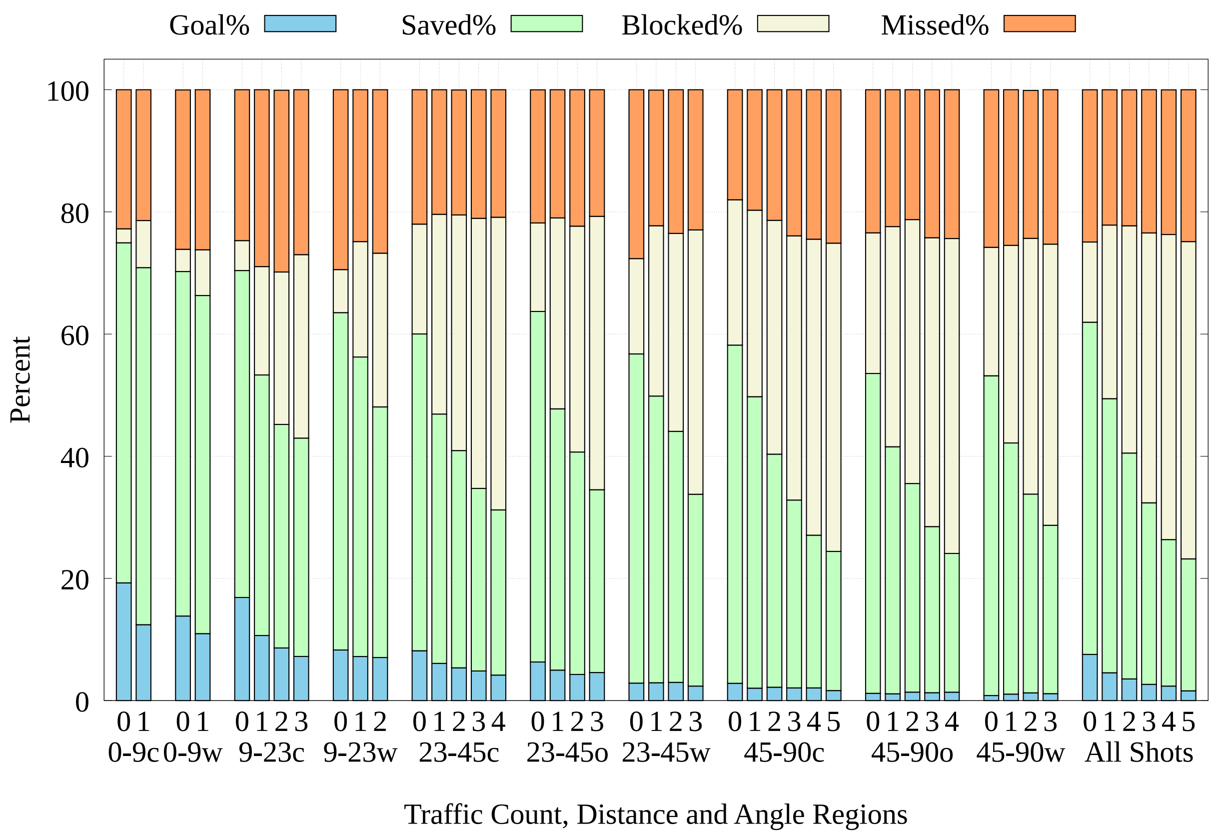

Figure 4.5: Shot outcome distributions across traffic levels and location regions

To assess whether the patterns observed in our results are statistically significant, we apply two types of tests. When analyzing ordered traffic levels (e.g., 0, 1, 2, 3 players) against a binary outcome such as Goal vs. No Goal, we apply the Cochran-Armitage trend test [12] [5]. This test evaluates linear trends across the ordered groups. For comparisons between two traffic groups (e.g., attacking versus defending), we use the chi-squared test [46], which compares two independent proportions. Throughout our analysis, we report only results that are statistically significant ($p < 0.05$) or moderately significant ($p < 0.1$). In tables, statistically significant results are marked with an asterisk (*) and moderately significant results with a dagger (†).

Traffic versus the percentage of shot attempts that miss the net (Missed%): To our surprise, the percentage of shot attempts that miss the net (Missed%) remains fairly constant across regions and traffic levels, showing that players' frequency of missing the net is not strongly affected by location or traffic. The overall Missed% is 22.4%. Across traffic levels it varies by only 1.5% and across shooting locations it varies by 2.5% (values are the coefficients of variation). The only traffic-region groups where Missed% deviates noticeably are zero-traffic shot attempts in the 9-23w and 23-45w regions. We believe that this may be due to these being common one-timer locations where players take quick shot attempts from sharp angles, slightly increasing the possibility of missing the net.

Traffic versus the percentage of shot attempts that are blocked (Blocked%): The percentage of shot attempts that are blocked (Blocked%) increases with traffic, particularly for mid-to-long range shot attempts ($p < 0.05$ for all traffic-region groups). However, what stood out is that 29% of all shot attempts in our dataset are blocked and 91% of those blocked shot attempts are by skaters on the defending team. 29% is a remarkably high proportion, especially considering the modern trend toward fewer total shot attempts as teams increasingly prioritize puck possession and higher-quality scoring chances [6]. This finding has important implications for shot prediction or expected goal (xG) models, which are now widely used in ice hockey analytics [60] [58] [66] [59] [23]. Most public xG models either exclude blocked shot attempts entirely or include them in limited ways because the NHL does not release shot location data for blocked attempts. As a result, *nearly one-third of all shot attempts are often ignored in these models*, which may limit their completeness and accuracy. In future work it would be interesting to study whether including blocked shot attempts in xG models would increase their accuracy.

Interestingly, a notable number of shot attempts with zero recorded traffic are blocked. This might seem counterintuitive but can be explained by the limitations of traffic detection. Many shot attempts are blocked by sticks rather than bodies, or by skaters stepping into the traffic lane during the shot attempt's flight. Capturing all such instances as traffic would require larger buffer and window sizes which could result in including players who do not meaningfully affect the shot attempt. We investigate the impact of different choices for buffer size and time window duration in Section 4.6.

Traffic versus the percentage of shot attempts that result in a shot on goal

(SOG%) and goal (Goal%): As shown in Figure 4.5, the percentage of shot attempts that are on goal (SOG%) declines as traffic increases ($p < 0.05$ for all traffic-region groups), indicating that traffic reliably reduces the likelihood of a shot attempt reaching the goaltender. Notably, the percentage of shot attempts that result in a goal (Goal%) is highest for shot attempts from the 0-9c and 9-23c regions when there is no traffic. For NHL front office staff, coaches and players, this highlights the value of prioritizing shot attempts in these traffic-region groups. In general, while Goal% also tends to decrease as traffic increases, this trend is only statistically significant in regions 0-9c, 0-9w, and 9-23c. Because traffic makes it hard to get a shot on goal, very few goals are scored in high-traffic situations, making it difficult to assess how traffic affects scoring itself. To better understand this, we focus next on Goal% of SOG, which captures whether traffic helps or hurts once a shot attempt reaches the goaltender, more directly reflecting the possible effects of tips, deflections, and screens. Note that Goal% of SOG corresponds to what the NHL has historically referred to as "shooting percentage", though we use the term Goal% of SOG as it more clearly describes the metric.

Traffic versus the percentage of shots on goal that result in a goal (Goal% of SOG): To better understand the quality of shot attempts that are on goal, we examine the proportion of shots on goal that result in a goal (Goal% of SOG). Figure 4.6 presents these results with 95% confidence intervals. In Table 4.2, we summarize the direction of the trend in Goal% of SOG as traffic increases for each region. To provide context for interpreting these values, Figure 4.7 shows the total number of shot attempts in each traffic-region group, which helps explain the width of the confidence intervals.

0-9c	**9-23c**	**45-90c**	**45-90o**	**45-90w**
(↓) 0.01^*	(↓) 0.04^*	(↑) 0.01^*	(↑) 0.01^*	(↑) $0.05^\dagger$

Table 4.2: Trend and p-value of Goal% of SOG across traffic levels for each region. Only regions with significant ($p < 0.05$) or moderately significant ($p < 0.1$) trends are shown. * indicates statistical significance, † indicates moderate significance. ↑ indicates an increasing trend and ↓ indicates a decreasing trend.

We focus our analysis on regions with significant or moderately significant trends. For shot attempts in the 0-9c and 9-23c regions, Goal% of SOG decreases as traffic increases, suggesting a decline in shot quality. This means that even when shot attempts from these regions reach the goaltender, they are less likely to result in goals if there is traffic. Teams may benefit from avoiding taking heavily contested shot attempts in these regions, instead seeking ways to create or move into space to shoot with minimal traffic at a similar location. In contrast, for long-range shot attempts (45+ feet), Goal% of SOG increases with

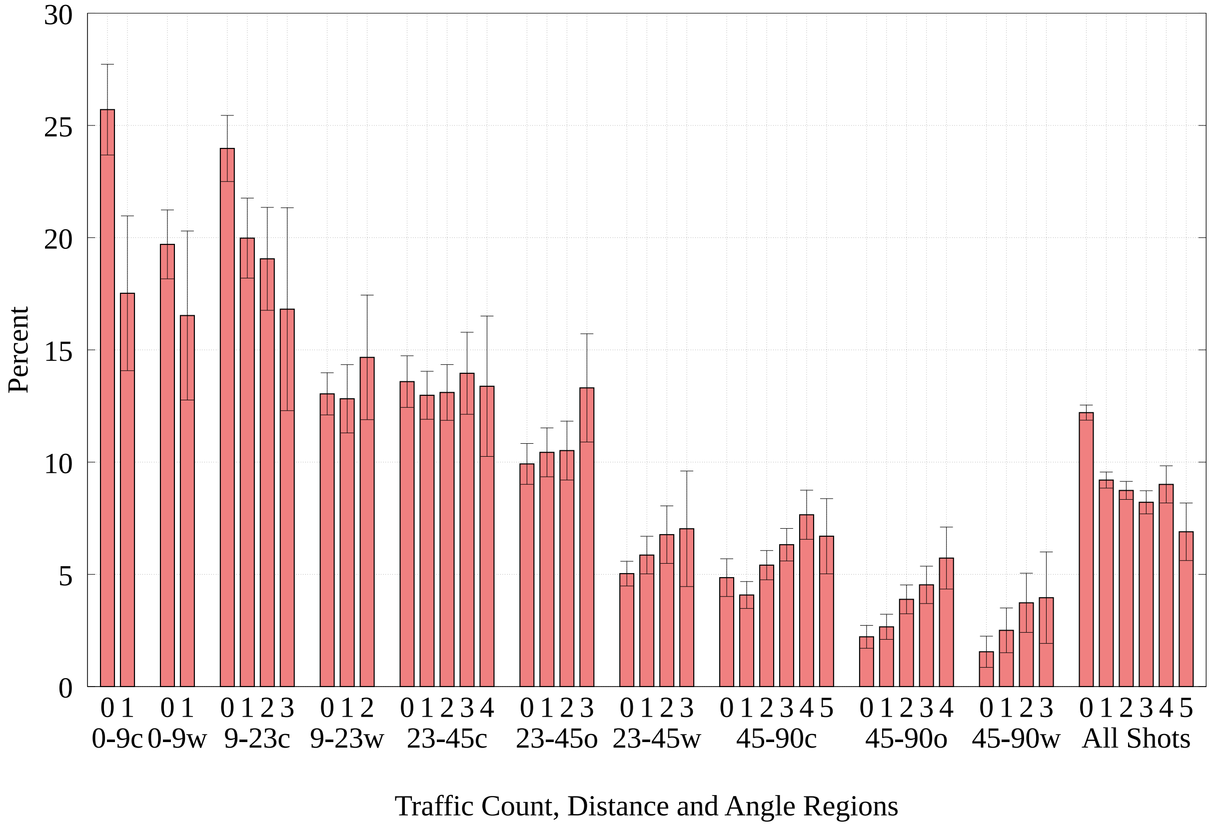

Figure 4.6: Goal% of SOG across traffic levels and location regions with 95% confidence intervals.

traffic. This indicates that while traffic reduces the chance of the shot attempt reaching the goaltender, it improves scoring success when it does, possibly due to tips, deflections or screens that impair the goaltender's view of the puck. These findings highlight a strategic tradeoff. For long-range shot attempts, traffic reduces total shot attempt success but increases the probability of scoring for the shot attempts that do reach the goaltender. As a result, teams may benefit from deliberately placing traffic in front of the net for long-range attempts, accepting fewer total shots on goal in exchange for more dangerous ones.

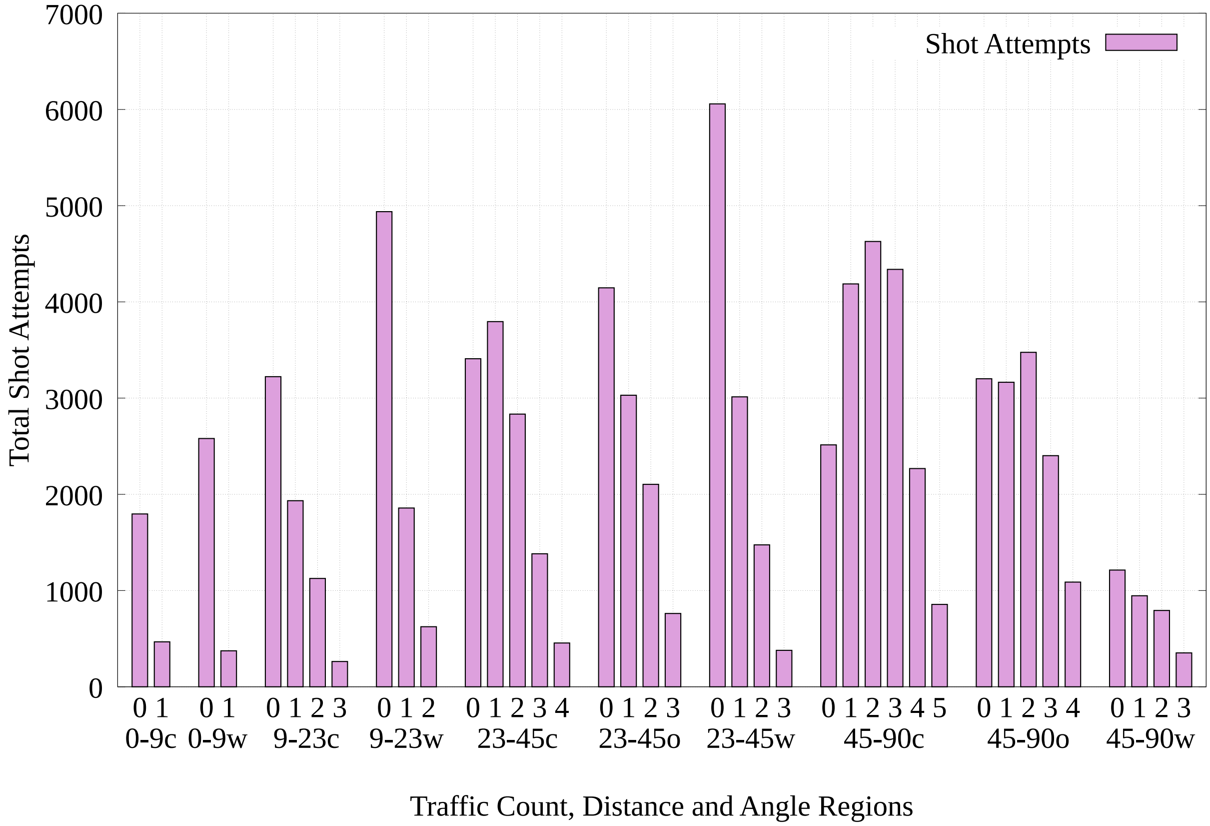

Figure 4.7: Total shot attempts across traffic levels and location regions.

4.5.1 Attacking Versus Defending Traffic

We next analyze how the balance of attacking versus defending skaters affects shot outcomes. A heavy defensive presence could lead to more blocked shot attempts and fewer goals, while a strong offensive presence might create tips, deflections, screens, and more goals. However, the opposite could also be true as defensive skaters may screen their own goaltender and offensive skaters might inadvertently block shot attempts. One possible approach would be to analyze each unique combination of attacking and defending players in traffic. However, this quickly results in a large number of categories (e.g., 2 attackers and 1 defender, 3 defenders and 0 attackers, etc.), many of which are rare and thus yield unreliable comparisons. Instead, in Figure 4.8, we categorize shot attempts in each region into three groups: (1) more attacking than defending traffic (A), (2) more defending than attacking traffic (D), and (3) equal attacking and defending traffic (E). This figure and subsection aim to highlight differences between attacking and defending traffic rather than the overall magnitude of traffic. Consequently, shot attempts with one attacking skaters are treated the same as those with two or three attacking skaters, and shot attempts with zero traffic are excluded from this analysis.

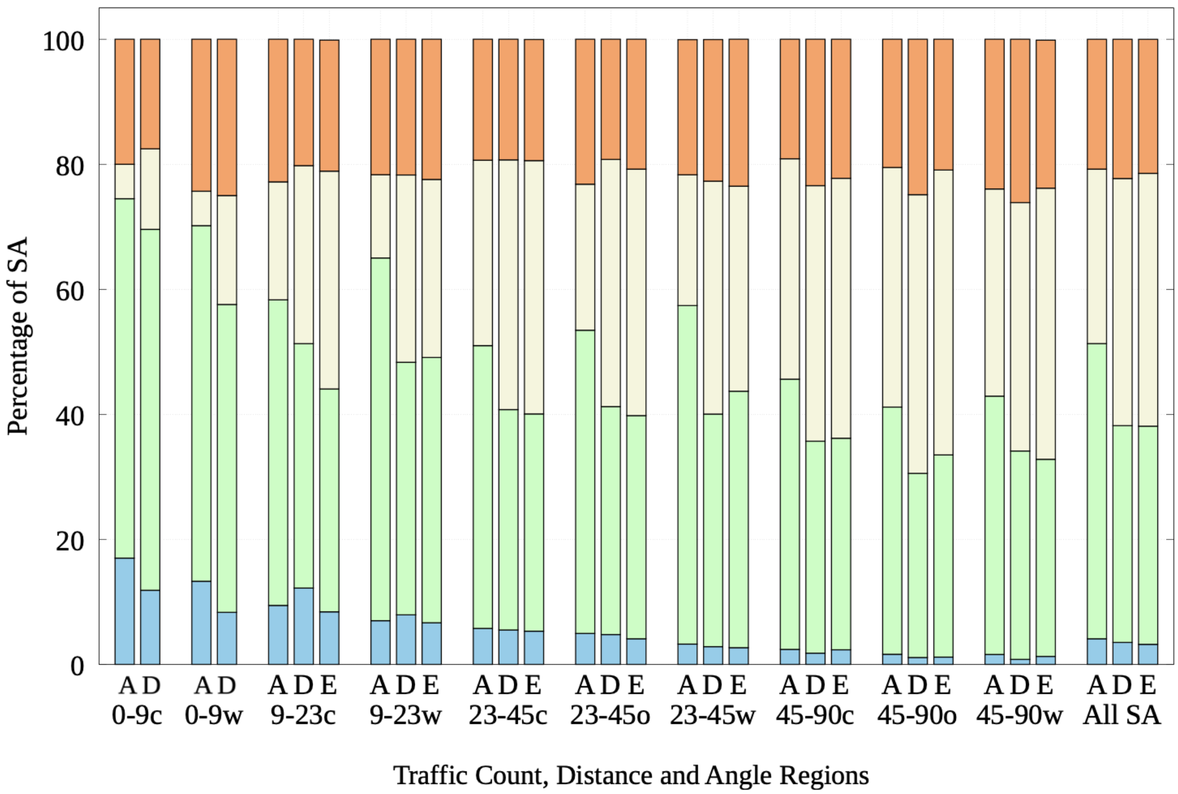

Figure 4.8: Shot outcome distributions for attacking versus defending traffic.

For each region, when defending traffic exceeds attacking traffic (D), SOG% is lower and Blocked% is higher compared to when the reverse is true (A) ($p < 0.05$ for all traffic-region groups). This is expected as defenders are typically positioned to block shot attempts whereas attackers aim to tip, deflect, or screen shot attempts. However, the effect on goal scoring is less straightforward. Similar to the previous section, to investigate this, we focus on the proportion of shots on goal that result in goals (Goal% of SOG). This allows us to evaluate whether offensive or defensive traffic is more beneficial (or harmful) to scoring in each region. Results are shown in Table 4.3.

In all mid-range distance regions (9–45 ft), shot attempts taken with more defending traffic (D) have higher Goal% of SOG (or shooting percentage) than those with more attacking traffic (A) ($p < 0.1$ for all four traffic-region groups). This may be because defenders unintentionally screen their own goaltender or redirect the puck in unpredictable ways. For coaches, this finding highlights an important risk in collapsing defenders to attempt to block mid range shot attempts as defensive traffic may sometimes impair the goalitender more than it helps. The optimal defensive approach may vary by shot location: limiting mid-range screens while emphasizing blocks on both short-range and long-range attempts.

9-23c	**9-23w**	**23-45c**	**23-45o**	**23-45w**
0.00^{*}	0.01^{*}	$0.06^{\dagger}$	$0.07^{\dagger}$	$0.09^{\dagger}$

Table 4.3: Effect of traffic balance on Goal% of SOG for each region. Each cell reports the p-value from a chi-squared test comparing shot attempts with more defending traffic (D) to those with more attacking traffic (A), testing whether D has a significantly higher Goal% of SOG. Only regions with statistically ($p < 0.05$) or moderately significant ($p < 0.1$) differences are shown. * denotes statistical significance, † denotes moderate significance.

4.6 Sensitivity

To examine the impact that our choice of buffer size and pre-release and post-release windows have on the results in this thesis, we repeat the analysis using alternate buffer sizes and shot attempt duration windows. For buffer size, we compared our 2-foot default to both a 0-foot and 4-foot buffer. For the shot attempt duration window, we evaluate three alternative approaches: specifically using only the shot release timestamp, a window of 0.5 seconds before and after the shot release, and a window of 0.5 seconds extending only after the release. In each case, we recalculate traffic and compare shot outcomes across the same distance-angle regions used in the main analysis. Figure 4.9 shows the resulting distributions for the 9-23c region across parameter settings. We focus on this region because it is the source of the most goals (18% of all goals), making trends across configurations easier to observe. These alternative configurations sometimes show slightly steeper declines in shot success, but the overall trends remained consistent, helping to demonstrate the robustness of our findings with respect to the choice of buffer size and shot window.

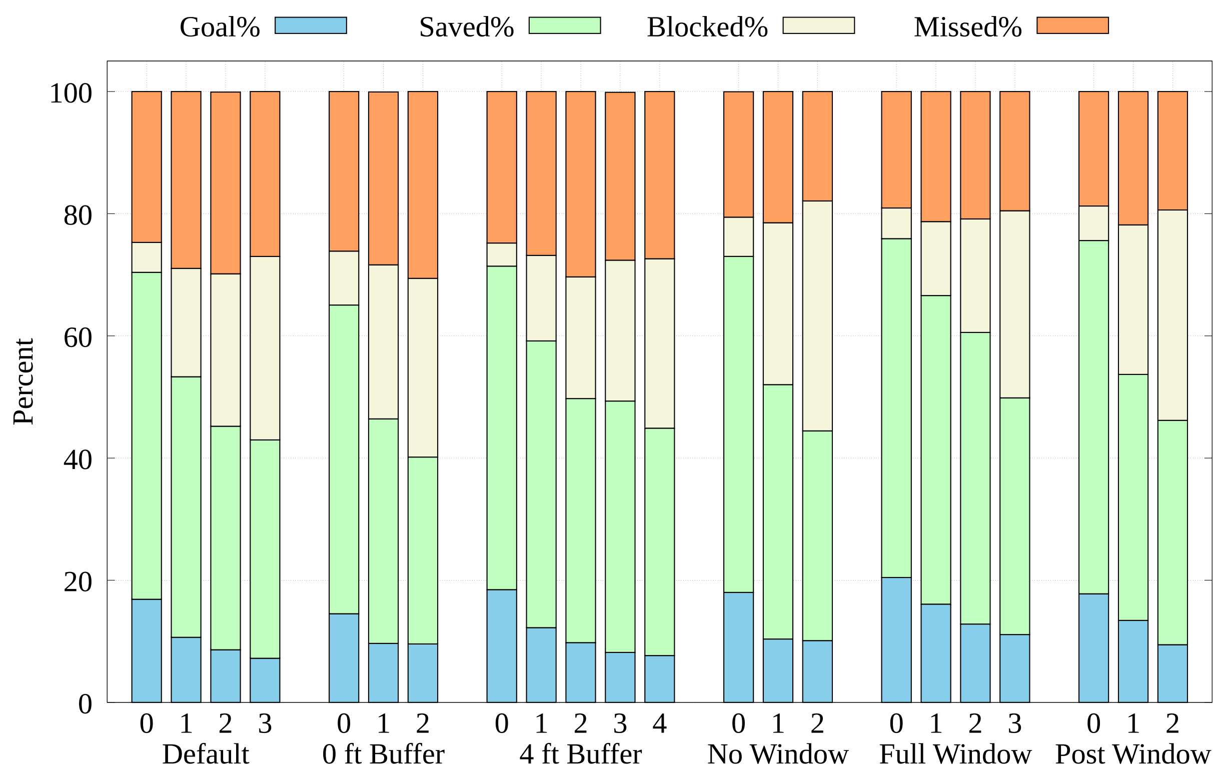

Figure 4.9: Comparison of shot outcome distributions by traffic level in the 9-23c region across buffer and window settings.

Chapter 5

Conclusions and Future Work

5.1 Thesis Summary

This thesis leverages the NHL's puck and player tracking data to explore two core but historically under-analyzed elements of hockey: team puck possession and traffic during shot attempts. Both are widely emphasized by analysts, coaches, and players as critical to success, yet their effects have rarely been quantified.

In the first part of this thesis, we investigate the relationship between team possession metrics and success. We begin by developing a robust data processing pipeline to clean and filter the NHL's player possession data. This includes removing duplicate and overlapping possessions, improving detection of possession end points, incorporating puck location to determine zone information, integrating game context such as strength and score state, and aggregating individual player possessions into continuous team-level sequences. With this augmented dataset, we analyze average team possession percentage and find that it has only a moderate correlation with team success metrics such as average goals for and goal differential (r=0.56 for both) . To gain deeper insight, we introduce a novel metric, Average Offensive Zone Possession Time Differential (Avg. OZPTD), which quantifies the difference between how long a team maintains possession in its offensive zone versus how long its opponents maintain possession in their offensive zone. This metric correlates strongly with average goal differential (r=0.77), outperforming traditional possession proxies like SAT% (also known as Corsi). We show that Avg. OZPTD is not only more predictive but also stable across games, capable of distinguishing between teams, and independent from existing metrics. These qualities support its use as a metric for understanding team performance.

In the second part of the thesis, we shift focus to the analysis of traffic during shot attempts. We develop a methodology for detecting and augmenting shot attempts using the NHL's puck and player tracking data. To analyze traffic, we define a spatial-temporal window around each shot and incorporate skater roles to distinguish between offensive and defensive presence. We find that traffic is strongly correlated with shot distance (longer shot attempts are typically subject to more traffic) and moderately, negatively correlated with shot angle (shot attempts from sharper angles typically have less traffic). To isolate the impact of traffic from the confounding factor of shot location, we develop an algorithm that divides the offensive zone into regions that minimize within-region variation in distance and angle, allowing us to study the impact of traffic separately from distance and angle. We find that for all regions, increased levels of traffic significantly increase the percentage of shot attempts that are blocked and reduce the chance of a shot attempt resulting in a shot on goal. At short ranges (0-23 feet), goals with no traffic exceed the combined total of all other traffic levels. Across all regions, the highest number of goals occurs from the 9–23c region with zero traffic, highlighting the value of passing or moving into space before shooting. For long-range shot attempts (45-90 feet), 38% of shot attempts are blocked, compared to 29% for all shot attempts in our dataset. However, long-range shot attempts through higher levels of traffic are associated with an increased likelihood of a goal.

Our findings also suggest that defensive traffic is generally effective at suppressing shot attempts as 91% of blocked shot attempts are by skaters on the defending team. However, for mid-range shot attempts (23-45 feet), when there are more defenders than attackers creating traffic, the percentage of shot attempts on goal that result in a goal is significantly higher than when there are more attackers than defenders creating traffic. This shows that defensive traffic can unintentionally aid scorers possibly due to screening the goaltender or deflecting shot attempts unpredictably.

Together, the contributions in this thesis introduce new techniques and empirical findings that enhance our understanding of puck control and skater positioning during shot attempts in the NHL. The methods developed offer a foundation for analysts interested in possession and traffic metrics, while the insights gained can inform NHL front office staff, coaches, players, and fans seeking a deeper understanding of how team possession relates to performance and how traffic affects shot outcomes in ice hockey.

5.2 Discussion and Future Work

The availability of puck and player tracking data represents a turning point for ice hockey analytics, enabling new research directions and a deeper evaluation of what makes a player or team successful. This thesis lays the groundwork for multiple future investigations.

One important avenue is the continued development of possession-based metrics. Future work could explore the predictive value of offensive zone possession count differentials, which emphasize frequency over duration, as well as metrics that capture how possessions begin, evolve, and end (e.g., starting location of possessions, sustained cycles, and rebound recoveries). We would also like to study whether the time spent in the offensive zone correlates with team success or if puck possession is a key component. In addition, our methodology for preparing and filtering NHL possession data can support individual-level player analysis. For example, one of our subsequent studies applied this methodology to evaluate individual player possessions and their relationship to player success [20].

For traffic analysis, while our method offers a new way to quantify traffic, it has limitations. First, the tracking data does not always contain explicit shot start and end times. To approximate shot duration, we construct a timing window based on empirical estimates. While our sensitivity analysis suggests the findings are robust to reasonable variations, having true shot start and end times would improve precision. Additionally, the data is not official and detecting the point of release of a shot is difficult and may not be 100% accurate (despite our improvements). Lastly, our definition of traffic considers only the number of skaters in the traffic lane without considering skater orientation, posture, or stick position. This is a necessary simplification due to the use of a single LED per skater.

In another subsequent study, we build on our shot detection methodology augmented with computer vision techniques to estimate player poses and define a new metric, net visibility, which we define as the fraction of the goalmouth visible from the puck's perspective [21]. This metric enables deeper insights into shot-blocking effectiveness, shooting decisions, and goaltender positioning. While current limitations make net visibility difficult to fully automate, it reinforces the value of the traffic framework developed in this thesis and highlights the need for increasingly detailed models of obstruction.

To our knowledge, this thesis is the first study in hockey that systematically quantifies the impact of traffic using puck and player tracking data. Our methods will enable analysts to explore not only league-wide patterns, but also team and individual player tendencies, such as the ability to create, avoid, or shoot through traffic. For NHL front office staff, coaches, players, and fans, our findings provide new insights into when traffic helps or hinders scoring, depending on shot location and context.

5.3 Concluding Remarks

This thesis demonstrates how tracking data can help to examine aspects of hockey that have long been discussed anecdotally but rarely quantified. Through the introduction of new metrics and methodologies, we offer a more complete picture of how puck control and skater positioning during shot attempts impact success in ice hockey. We believe that these contributions not only advance academic understanding but also offer practical techniques for analysts and insights for coaches and players seeking to gain a competitive edge. In short, possession and traffic matter, but their impact is nuanced. Not all possessions are equally valuable, and not all traffic is equally effective. This thesis provides a framework to evaluate those nuances and opens new possibilities for both research and practice in hockey analytics.

References

[1] Part 1.1: The relationship between possession and winning. https://www.fifatrainingcentre.com/en/community/fifa-research/possession-analysis/part-1-1-the-relationship-between-possession-and-winning.php, 2023. Posted on *FIFA Training Centre*.

[2] Amazon Web Services. NHL and AWS Unveil Opportunity Analysis: A New Hockey Statistic. https://aws.amazon.com/sports/nhl-opportunity-analysis/, 2023.

[3] Pedro Andrez, Manuel Silva, and Christopher Carling. Analysis of team success based on match technical and running performance in a professional soccer league. *BMC Sports Science*, 14:82, 2022.

[4] Jeremy Arkes. Is controlling the rushing or passing game the key to NFL victories? *The Sport Journal*, 14:1, 2011.

[5] Peter Armitage. Tests for linear trends in proportions and frequencies. *Biometrics*, 11(3):375–386, 1955.

[6] Hockey's Arsenal. Goodbye to the Mid-Range Shot: Why NHL Teams are Shooting Differently. https://hockeysarsenal.substack.com/p/goodbye-to-the-mid-range-shot, 2023. Posted on *Hockey's Arsenal*.

[7] C. J. Bellis. Reaction Time and Chronological Age. *In Proceedings of the Society for Experimental Biology & Medicine*, 30:801–803, 1933.

[8] Songlin Bi, Yonggang Gu, Jiaqi Zou, Lianpo Wang, Chao Zhai, and Ming Gong. High Precision Optical Tracking System Based on near Infrared Trinocular Stereo Vision. *Sensors*, 21(7):2528, 2021.

[9] Henry Bushnell. Why the NFL still uses refs, not technology, to spot the ball. https://sports.yahoo.com/why-the-nfl-still-uses-refs-not-technology-to-spot-the-ball-153102648.html, 2021. Posted on *Yahoo Sports*.

[10] Claudio A Casal, Rubén Maneiro, Toni Ardá, Francisco Marí, and José Losada. Possession zone as a performance indicator in football: The game of the best teams. *Frontiers in Psychology*, 8:1176, 2017.

[11] Julen Castellano, David Casamichana, and Carlos Lago Peñas. The use of match statistics that discriminate between successful and unsuccessful soccer teams. *Journal of Human Kinetics*, 31:139–147, 2012.

[12] William Cochran. Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4):417–451, 1954.

[13] Andrew Cohen. How NHL's sensor-embedded puck allows for better broadcast statistics and graphics, new types of wagers, more. https://www.sportsbusinessjournal.com/daily/issues/2022/10/28/technology/nhl-smt-puck-upgraded-sensors-leds-player-tracking-statistics, 2022. Posted on *Sports Business Journal*.

[14] Christian Collet. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *Journal of Sports Sciences*, 31(2):123–136, 2013.

[15] FJ Dynamics. GPS System Accuracy: What Factors Affect It? https://www.fjdynamics.com/blog/industry-insights-65/gps-system-accuracy-361, 2023.

[16] George Foussekis. Hockey's Next Generation Technologies: Puck and Player Tracking Systems. https://novacapsfans.com/2015/07/15/hockeys-next-generation-technologies-puck-and-player-tracking-systems/, 2015. Posted on *Nova Caps Fan Blog*.

[17] Alexander Franks, Alexander D'Amour, Daniel Cervone, and Luke Bornn. Meta-analytics: Tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, 12:151–165, 2016.

[18] José Miguel Gama, Gonçalo Dias, Micael S. Couceiro, and Tiago Sousa. Networks metrics and ball possession in professional football. *Complexity*, 21(S2):342–354, 2016.

[19] Aristotelis Gioldasis and Stergios Komsis. Winners and losers in top level soccer. How do they differ? *Journal of Physical Education and Sport*, 14(3):398–405, 2014.

[20] Evan Iaboni, Sebastian Negulescu, Miles Pitassi, Fauzan Lodhi, and Tim Brecht. Individual puck possessions Part I: Frequency, duration, and distance travelled. In *Proceedings of the Linköping Hockey Analytics Conference, Research Track (LINHACK)*, 2025. Under submission.

[21] Evan Iaboni, Sebastian Negulescu, Miles Pitassi, Amirt Nazemi, Jerrin Bright, Vasyl Chomko, David Clausi, Sven Dickison, and Tim Brecht. New views of shots: Towards measures of net visibility and reachability. In *Proceedings of the Linköping Hockey Analytics Conference, Research Track (LINHACK)*, 2025. Under submission.

[22] Dale Johnson. Premier League semi-automated VAR offside: All you need to know. https://www.espn.com/soccer/story/id/44038727/premier-league-fa-cup-semi-automated-var-offside-all-need-know, 2025. Posted on *ESPN*.

[23] Ken Krzywicki. Shot Quality Model: A Logistic Regression Approach to Assessing NHL Shots on Goal. https://hockeyanalytics.com/2005/01/shot-quality/, 2005. Posted on *hockey::analytics()*.

[24] Ken Krzywicki. NHL Shot Quality 2009-10: A Look at Shot Angles and Rebounds. https://hockeyanalytics.com/2010/10/nhl-shot-quality-2010/, 2010. Posted on *hockey::analytics()*.

[25] Joe Lemire. How Sports Technology Titan SMT Acquiring Rival Sportvision Will Change the Graphics Game. https://smt.com/how-sports-technology-titan-smt-acquiring-rival-sportvision-will-change-the-graphics-game/, 2016. Posted on *SMT.com*.

[26] Daniel Link, Steffen Lang, and Philipp Seidenschwarz. Real-Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLOS ONE*, 11(12), 2016.

[27] Hongyou Liu, Miguel Ángel Gomez, Carlos Lago-Peñas, and Jaime Sampaio. Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, 33(12):1205–1213, 2015.

[28] Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. Quality vs Quantity: Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data. In *MIT Sloan Sports Analytics Conference*, 2015.

[29] Dom Luszczyszyn. Inside 3-on-3 OT: Strategies and stats for the NHL's most chaotic game state. https://theathletic.com/1672268/2021/02/08/inside-3-on-3-ot-strategies-and-stats-for-the-nhls-most-chaotic-game-state, 2021. Posted on *The Athletic*.

[30] Alejandro López-Valenciano, Adam McRobert, Humberto Sarmento, Manuel Reina, Richard Page, and Paul Read. A goalkeeper's performance in stopping free kicks reduces when the defensive wall blocks their initial view of the ball. *European Journal of Sport Science*, 2021.

[31] Brian Macdonald. An Expected Goals Model for Evaluating NHL Teams and Players. In *MIT Sloan Sports Analytics Conference*, 2012.

[32] Sean McIndoe. How the NHL's new stats and analytics hub will take fans inside the game. https://www.nytimes.com/athletic/4983970/2023/10/23/nhl-edge-stat-tracking/, 2023. Posted on *The Athletic*.

[33] Bob McKenzie. The Real Story of How Corsi Got Its Name. https://www.tsn.ca/mckenzie-the-real-story-of-how-corsi-got-its-name-1.100011, 2014. Posted on *TSN*.

[34] MoneyPuck. https://moneypuck.com/data.htm, 2025. Includes expected goals (xG) and expected shots on goal (xShotWasOnGoal) metrics. Accessed March 2025.

[35] Felipe Moura, Luiz Eduardo Barreto Martins, and Sergio Augusto Cunha. Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, 32(20):1881–1887, 2014.

[36] National Football League. NFL Next Gen Stats. https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/, 2025. Posted on *NFL.com*.

[37] National Hockey League. How NHL Edge Data Helps Teams Analyze Scoring Opportunities. https://www.nhl.com/news/how-nhl-edge-data-helps-teams-analyze-scoring-opportunities/c-343703756.

[38] National Hockey League. NHL API. https://api-web.nhle.com/v1/, 2025.

[39] National Hockey League. NHL Edge Goal Visualizer. https://www.nhl.com/ppt-replay/goal/2024020013/182, 2025. Posted on *NHL.com*. Accessed March 4, 2025.

[40] Natural Stat Trick. https://naturalstattrick.com/, 2025. Includes expected goals (xG%). Accessed March 2025.

[41] The Hockey News. Smart Chip on Finns' Shoulders: Inside Liiga's Player Tracking Technology. https://thehockeynews.com/all-access/smart-chip-on-finns-shoulders-inside-liigas-player-tracking-technology, 2023. Posted on *The Hockey News*.

[42] NHL. NHL Hockeyverse Animated Programming to Engage Fans. https://www.nhl.com/news/nhl-hockeyverse-animated-programming-to-engage-fans, 2024.

[43] NHL EDGE. https://edge.nhl.com, 2025. Displays player and goalie statistics segmented by offensive zone location. Accessed March 2025.

[44] Edward Norton and Bryan Dowd. Log odds and the interpretation of logit models. *Health Services Research*, 53:859–878, 2018.

[45] Arik Parnass. How Can We Quantify Power Play Performance In Formation? https://hockey-graphs.com/2016/04/25/how-can-we-quantify-power-play-performance-in-formation, 2016. Posted on *Hockey-Graphs*.

[46] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175, 1900.

[47] Miles Pitassi, Tim Brecht, and Mingyue Xie. Puck Possessions and Team Success in the NHL. In *Proceedings of the Linköping Hockey Analytics Conference, Research Track (LINHACK)*, pages 51–66. Linköping University Electronic Press, 2024.

[48] Quuppa. Quuppa Intelligent Locating System. https://www.quuppa.com/offering/quuppa-intelligent-locating-system/, 2025.

[49] David Radke, Tim Brecht, and Daniel Radke. Identifying Completed Pass Types and Improving Passing Lane Models. In *Proceedings of the Linköping Hockey Analytics Conference*, 2022.

[50] David Radke, Jaxin Lu, Jackson Woloschuk, Tin Le, Daniel Radke, Charlie Liu, and Tim Brecht. Analyzing passing metrics in ice hockey using puck and player tracking data. In *Proceedings of the Linköping Hockey Analytics Conference*, pages 25–39, 2023.

[51] David Radke, Daniel Radke, Tim Brecht, and Alex Pawelczyki. Passing and Pressure Metrics in Ice Hockey. In *Artificial Intelligence for Sports Analytics (AISA) Workshop at IJCAI21*, 2021.

[52] NHL Research and Development. Fuzzy Matching Algorithm for Shot Detection.

[53] Dan Rosen. OT rule changes, face-off clarity discussed at GM meetings. https://www.nhl.com/news/overtime-rule-changes-face-off-clarity-discussed-at-gm-meetings, 2023. Posted on *NHL.com*.

[54] Alan Ryder. Shot Quality: A Methodology for the Study of the Quality of a Hockey Team's Shots Allowed. https://hockeyanalytics.com/2004/01/isolating-shot-quality/, 2004. Posted on *hockey::analytics()*.

[55] Frank Seravalli. NHLPA reminds players of their right to control or destroy wearable tech data. https://www.dailyfaceoff.com/news/nhlpa-reminds-players-of-their-right-to-control-or-destroy-wearable-tech-data, 2022. Posted on *Daily Faceoff*.

[56] Stephen Shea and Christopher Baker. *Hockey Analytics: A Game-Changing Perspective*. CreateSpace Independent Publishing Platform, 2017.

[57] Nikolay Shmakov. Application and Development of the Expected Goals for Hockey Player Evaluation. In *International Symposium on Artificial Intelligence and Computational Engineering (ISACE)*, pages 235–248. Springer, 2024.

[58] Harry Shomer. Analyzing the Game of Hockey: Evaluating my Shooter xG model. https://fooledbygrittiness.blogspot.com/2018/03/evaluating-my-shooter-xg-model.html, 2018. Posted on *Fooled by Grittiness*.

[59] Lars Skytte. Hockey-Statistics: Building an xG Model – v. 1.0. https://hockey-statistics.com/2022/08/14/building-an-xg-model-v-1-0, 2022. Posted on *Hockey-Statistics*.

[60] Sprigings, D (also known as DTMAboutHeart). Expected Goals are a Better Predictor of Future Scoring than Corsi and Goals. https://hockey-graphs.com/2015/10/01/expected-goals-are-a-better-predictor-of-future-scoring-than-corsi-goals, 2015. Posted on *Hockey-Graphs*.

[61] Stathletes Inc. Big Data Cup 2025. https://www.stathletes.com/bigdatacup/, 2025.

[62] Stats Perform. STATS Acquires Prozone. https://www.statsperform.com/press/stats-acquires-prozone/, 2015. Press release. Accessed March 4, 2025.

[63] Tore Purdy (also known as JLikens). Loose Ends - Part I: Predicting Future Success. https://objectivenhl.blogspot.com/2011/03/loose-ends-part-i-predictive-validity.html, 2011. Posted on *Objective NHL*.

[64] Eric Tulsky, Geoffrey Detweiler, Robert Spencer, and Corey Sznajder. Using Zone Entry Data to Separate Offensive, Neutral, and Defensive Zone Performance. In *MIT Sloan Sports Analytics Conference*, 2013.

[65] K. Vats, P. Walters, M. Fani, D.A. Clausi, and J.S. Zelek. Player Tracking and Identification in Ice Hockey. *Expert Systems with Applications*, 127, 2023. Canada.

[66] Josh Younggren and Luke Younggren. A New Expected Goals Model for Predicting Goals in the NHL. https://evolving-hockey.com/blog/a-new-expected-goals-model-for-predicting-goals-in-the-nhl, 2021. Posted on *Evolving-Hockey*.

[67] Miguel Ángel Gómez, Evangelos Tsamourtzis, and Alberto Lorenzo. Defensive systems in basketball ball possessions. *International Journal of Performance Analysis in Sport*, 6:98–107, 2017.