PROJECTING THE FUTURE INDIVIDUAL CONTRIBUTIONS OF NHL PLAYERS

A Thesis

by

AARON THOMAS KNODELL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	James Caverlee
Committee Members,	Zhangyang Wang
	Irina Gaynanova
Head of Department,	Dilma Da Silva

May 2019

Major Subject: Computer Engineering

ABSTRACT

Professional sports are a multibillion-dollar industry with millions of people invested in the outcomes of games and seasons. Owners, management, and fans sit on the edges of their seats wondering what will happen next. Lots of work has been done forecasting success at the team level across a variety of sports, but player level predictions are less common. Predictive work related to the NHL is even rarer.

This thesis explores the ability to predict NHL player performance in a given season using publicly available information via statistical learning methods. Data featured in the analysis includes play-by-play and shift information, box score statistics, a variety of composite and catch-all statistics, injury information, and player biographical information. Data was compiled and analyzed to find meaningful relationships between past and future performance. The results of the analysis found the most predictive values in the given data set and leveraged them to build a predictive model. The findings suggest that a previous season's raw numbers can be supplemented with more information to improve predictive power.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor James Caverlee and Zhangyang Wang of the Department of Computer Science and Engineering and Professor Irina Gaynanova of the Department of Statistics. All work for the thesis was completed independently by the student.

Funding Sources

Graduate study was supported by a teaching assistantship from the Department of Computer Science and Engineering.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

A common challenge in all team sports is separating and identifying the contributions of an individual player from those of his or her teammates. This discussion is ultimately rooted in trying to determine which players are the most responsible for driving the success of their teams. Is a player good because of the team, or is the team good because of the player? These questions are fun for fans to debate, but crucially important for the managers in charge of building professional teams.

Isolating and quantifying individual player contributions goes back to the beginning of keeping stat sheets for games and tracking individual player statistics. These box score statistic have evolved over time, with new ones being added in every sport and old ones being tweaked throughout their history. For example, in the National Hockey League (NHL), the maximum number of assists per goal varied from a single assist to three, before the modern maximum of two was settled on [1]. Total time on ice was not tracked for skaters in the NHL until the 1998-99 season [2].

These box score statistics are useful for counting the major events that happen in a game, but they are sorely lacking in the context that allows for a more accurate reflection of a player's true talent level, such as the role they may be asked to play by a coach. They also each measure just a single part of a player's ability, meaning that a comparison between two players involves investigating multiple statistics. However, this is an inherently arbitrary process, since each person evaluating a player will value each contribution differently. Single-number or catch-all statistics are an attempt to get around the bias by combining several individual box score numbers into a single number. Passer rating is an early example, commissioned by the National Football League (NFL) to determine the top quarterback for the season [3]. Most early catch-all statistics arbitrarily weighted their components based on what their creator's felt was fair. Modern methods have used statistical methods and machine learning techniques to try to separate the play of individuals from their teammates and to remove the bias of the creator towards each component.

Quantifying past performance is a good first step to building a team, but managers should be less interested in what a player has done in the past than what they can do in the future. This is where the real value of isolating individual talent becomes apparent. The ability to identify a player who is on the verge of breaking out or falling off a cliff, performance-wise, and making the appropriate roster moves is the difference between teams that can sustain success and those that fail. The New England Patriots appear to be especially good at this, cutting ties with major contributors before they become drags on the team and consistently finding new players to take their place without missing a beat [4].

The NHL presents several challenges that make player evaluation and projection more difficult than the NFL, Major League Baseball (MLB), or the NBA (National Basketball Association). The sport itself makes player evaluation difficult. Baseball has a finite number of clearly defined states for each play with regards to a team being on offense or defense, the number of players on base and number of outs within an inning. Football also has clear game states for each defined by number of downs, distance to a first down, and position on the field. Even basketball, which is more continuous than baseball or football can be clearly broken down into possessions when a team is on offense or defense. All three sports also maintain consistent personnel for each play made within a game, with substitutions only occurring during stoppages.

Hockey, by contrast is an extremely fluid game. There are no clearly defined offensive or defensive plays since teams transition back and forth multiple times between stoppages. The presence of the neutral zone also complicates analysis since teams can be in possession of the puck without being considered on offense or defense. Players also make substitutions "on the fly", changing the interactions between players within a play. It is also possible for man power to change during a shift when a penalty expires allowing the penalized player to return to the ice, or a goaltender is pulled and replaced with an extra skater. Hockey is also lower scoring than the other three sports, with most teams averaging a total of between five and six scoring events in a game. This scarcity introduces a large amount of variance at the player level, making shooting percentages and goal totals difficult to predict from year-to-year [5].

Additionally, the NHL trails the other leagues in terms of the sophistication of its statistics. Unlike MLB, which has detailed records for over 100 years of play, detailed NHL statistics cover a relatively short period of time, with the modern era of tracked statistics only covering the past eleven full seasons. The other three sports have also introduced play tracking software to track the motion of each player and the ball [6, 7, 8]. The NHL, however, still tracks player substitutions and event locations by hand [9]. This leads to obvious opportunities for mistakes in the data recorded. This also means that player movement away from the puck and puck movement that is not part of the tracked events, such as passes, is completely excluded from official NHL numbers. NHL data is also relatively spread out with many different platforms compiling their own composite metrics from box scores, play-by-play data, and press releases.

However, despite these limitations, player projection is still a worthwhile exercise. Although it has a smaller market share than the big three professional sports leagues in North America, the NHL is still a multibillion-dollar industry with contracts for top players exceeding $100 million [10]. With this much money at stake, every gain matters and teams should seek as much certainty as possible when making decisions about which players to sign to new contracts or acquire in trades. Furthermore, hockey has been shown to be a sport where having the best player on the ice is more important than not having the worst player [11] and knowing whether a player will improve or not is important in building a winning team. Players can also use this information when determining how to maximize their career earnings, either as a tool in contract negotiations or to analyze their own abilities and modify their routines to prolong their careers. Finally, this information is valuable to fans. More information creates more avenues for fan engagement and allows fans to appraise the moves made by their teams with more accuracy.

1.1 Contributions

In this thesis, the following contributions are made:

- **Centralized Dataset of Regular Season NHL Statistics:** NHL data was collected from multiple sources, cleaned and consolidated into a single centralized dataset containing regular season statistics since the 2007-08 season.
- **Comparative Analysis of Context Adjustments:** The effectiveness of different context adjustments for box score stats on predicting future performance was examined.
- **Analysis of Additional Factors:** The impact of past playing style, coach usage, experience, and prior injuries on future performance was examined.
- **Predictive Model:** A model was developed for predicting a player's contributions in the next season.

The major findings suggest that predictions can be improved by supplementing raw statistics with additional information. However, there is a limit to predictive accuracy that can be gained using current information. Certain context adjustments and additional factors weakened the predictive power of other components. Additionally, there were a surprising number of errors in the official NHL records. However, despite this, the results are encouraging for future predictive work as new information, such as player tracking data, are implemented and made available for further research.

2. BACKGROUND AND RELATED WORK

2.1 A Brief History of NHL Statistics

The earliest statistics recorded by the NHL were simply the goal, point, and penalty totals beginning in the inaugural season. In the 1933-34 season, goals could be broken down by strength states [1]: even strength (EV), power play (PP), and shorthanded (SH). In the 1959-60 season, Plus/Minus was introduced, which attempted to evaluate the net impact of a player by subtracting the goals scored against his team while he was on the ice from the goals scored for his team when he was on the ice [2]. Individual shots also began to be recorded in the same season, allowing shooting percentage to be calculated. The 1997-98 saw the introduction of faceoffs, hits, missed shots, shots, giveaways, and takeaways, known as Real-Time Scoring Stats (RTSS) [2]. Shift charts were also introduced in the same season and time on ice was introduced. In the 2002-03, blocked shots were added, completing the current group of compiled statistics [2]. In the 2007-08 a new system was introduced for the play-by-play records which included all RTSS events as well as rink coordinates for each event [9]. These statistics have formed the basis for almost all subsequent developments by other individuals and organizations. Table 2.1 contains a description of each event recorded by the RTSS system. These events have formed the basis of most modern public analysis of hockey, including the development of shot attempt-based statistics such as Corsi (the sum of all blocked, missed, and saved shots and goals) and Fenwick (the sum of all missed and saved shots and goals), which are central to most current advanced statistics.

Event	Description
SHOT	The puck is directed towards the opposition net and is stopped by the goaltender. Location of the shooter and shot type included.
GOAL	The puck is directed towards the opposition net and enters it, crossing the goal line completely. Location of the shooter and shot type included.
MISS	The puck is directed towards the opposition net and misses without being contacted by the goaltender or a defending skater. Location of the shooter and shot type included.
BLOCK	The puck is directed towards the opposition net and is stopped by a defending skater. Location of the blocker and shot type included.
FAC	The puck is dropped between two opposing players to start play. Location of the faceoff included.
HIT	A body check against the puck carrier. Location of the hit included.
PENL	A violation of the NHL rulebook. The time of the penalty records the stoppage of play.
GIVE	A change in possession of the puck caused by an unforced error from the player in possession of the puck.
TAKE	A change in possession of the puck caused by the puck being actively taken from an opposition player, not as the result of a hit.

Table 2.1: NHL RTSS events

2.2 Catch-All Statistics

2.2.1 Baseball

Win Shares are an early attempt at creating a catch-all statistic in baseball [12]. A team's marginal runs are calculated using their total runs scored and allowed, league averages for runs scored an allowed, margin calculations for hitters and pitchers. Marginal runs are then used to calculate an expected wins total, which is divided by the team's actual win total to derive a scaling factor for each player's contributions. Expected wins are then proportioned to each player by the number of runs they created and scaled down based on the team's actual performance. Although Win Shares are not widely used anymore, they have been influential in designing subsequent catch-all statistics for baseball and other sports. The attempt to account for the quality of a player's team and the season or era he played in has been incorporated into many of its followers.

Arguably the most famous catch-all statistic in sports in baseball's Wins Above Replacement (WAR). WAR attempts to value a player's contributions in terms of wins the created above what a replacement level player would contribute. The precise definition of a replacement player varies depending on the implementation of WAR, several of which exist, but the general idea is that a replacement player is one who can be easily replaced without impacting a team's quality. Two of the most popular models were developed by the websites FanGraphs and Baseball-Reference. These models use the same method of determining replacement level, which is that a team of replacement players would be expected to be worth 48 wins in a full season [13, 14]. However, they use different methods for estimating the value a player contributes. The website Baseball Prospectus has a version of WAR called Wins Above Replacement Player, which directly states that it is only interested in describing what a player has done [15]. In their terms, a player's performance in a season is the population, not a sample. All three of these methods stress that they are estimates, not exact calculations of a player's value, but do not include uncertainty. Although these publications are open about their methods, the details of their calculations are still not publicly available. A newer implementation called openWAR has also been developed, which is notable for making the code used in calculating its output available and for explicitly including uncertainty estimates in the results [16].

2.2.2 Basketball

Basketball has several versions of catch-all statistics that have gained popular acceptance. NBA Win Shares is an attempt developed for the website Basketball-Reference using a similar methodology to baseball's Win Shares [17]. NBA Win Shares are calculated by determining a team's marginal points scored for and against and determining how much a player contributed to each based on his individual production. Offensive and defensive win shares are derived separately and added together to get the total win shares for a season.

Box Plus/Minus (BPM) is an attempt at determining relative contributions for a player in terms of points that uses box score statistics to compare the performance of a team with a player on the court to the average performance in the league [18]. BPM is calculated relative to the league average rate of points per 100 possessions. Value Over Replacement Player is a WAR analogue and is calculated directly from BPM using the percent of available minutes a player plays in the games they play and the percentage of possible games a player plays in a season.

Adjusted Plus/Minus (APM) is a technique of determining a player's impact on a play-by-play level using linear regression. The first technique used ordinary least squares regressions and looked at each play or possession [19]. A player's presence on the court was denoted by a variable and the coefficients for each variable were used to determine their relative impact on the results. A later iteration of APM separated each possession into an offensive possession and a defensive possession, calculating offensive and defensive APM for each player, rather than a single number for overall impact [20]. Finally, a newer method was introduced using ridge regression to reduce variance in the ratings and handle the issue of collinearity for players that play together frequently [21].

2.2.3 Hockey

In hockey, the objective of each game is to score more goals than the other side. However, like basketball and unlike football or baseball, hockey is a continuous sport, meaning each player must transition from offense to defense and back without a pause in the flow of play. Each player's value on offense and defense can be summed together to provide a single number for their contribution. Offensive contributions are usually measured in goals, either for the individual or the team, while defensive contributions are usually measured in goals against or some calculation for goals prevented.

Plus/Minus can be considered the first attempt to attach a single number to a player's contribution. Players are credited with a "plus" when they are on the ice for a goal scored for their team, excluding power play goals and a "minus" when they are on the ice for a goal against, excluding goals scored while their team is shorthanded. If their final Plus/Minus is positive, it indicates a positive contribution to the team and vice versa. Plus/Minus has recently received a substantial amount of criticism for several reasons [22]. First is that it arbitrarily includes or excludes goals scored in its calculation. Power play goals are not counted for either team since it is determined that one team has an advantage, but goals scored with a goalie pulled or into an empty net are counted, even though the teams are not evenly matched in this situation. This has the effect of not crediting players who contribute goals on the power play and not penalizing players that allow a high number of shorthanded goals. Plus/Minus has also been criticized for not being comparable between players. Players that play on good teams tend to have a higher Plus/Minus. Additionally, as a counting statistic, players that receive more playing time also tend to have more extreme Plus/Minus ratings than players with little playing time. Finally, Plus/Minus has been criticized for being based entirely on goals. Goals scoring for and against is also impacted by the other players on the ice, particularly goalies, meaning Plus/Minus does a poor job of isolating individual ability. Goals are also a relatively rare event in hockey, which makes Plus/Minus highly variable from year-to-year and not predictive of future Plus/Minus.

Later attempts at single number statistics attempted to represent a player's value relative to league numbers for a season, making it easier to compare players across seasons and eras.

Two of the earliest were Player Contributions [23] and Point Shares [24], which were both based on baseball's Win Shares. Both were derived by calculating a team's marginal goals for and against and crediting them to the players on their rosters based on position and playing time and individual production. Of the two, Player Contributions was more detailed, notable for breaking down contributions by even strength, power play, and shorthanded strength states. However, Player Contributions and Point Shares both allocate each player's contributions somewhat arbitrarily. Point Shares, for instance, treated a goalie as contributing as much as to the team a forward or defenseman. They were also both heavily based on goals against to determine defensive contributions, which led to some criticism. Player Contributions has not been maintained, while Point Shares is still available at the website Hockey-Reference but is not widely used in the hockey community.

Goals Versus Threshold (GVT) was a metric used by the website Hockey Prospectus with a similar methodology to Player Contributions and Point Shares [25]. However, it was distinct in two important ways. The output was expressed in terms of goals instead of standings points, which is an important part of many subsequent statistics. It also used shot statistics instead of goals against to represent defensive impact, separating a skater's defensive impact from the goalie. GVT made some arbitrary decisions about each player's contribution to his on-ice results, such as defensemen contributing twice as much to defense as forwards. GVT has not been maintained in recent years and updated values are not available.

Game Score was developed as an attempt to highlight outstanding individual game production [26]. It is weighted average for various box score statistics weighted by the frequency of each event and scaled to be comparable to point totals for a season. However, despite its relative simplicity compared to other catch-all statistics, it was discovered to strongly predictive of future Game Score when calculated at a season level. Game Score is currently available at the website Corsica Hockey and its formula is publicly available as well.

Regression-Adjusted Plus/Minus (RAPM) is an adaptation of basketball's APM for hockey [27]. RAPM follows a similar methodology to the final version of APM, utilizing ridge regression and including each sequence of play twice, once for a team's offense and once for defense. Unlike APM, RAPM performs its regression on shifts instead of possessions. Hockey faces a difficult challenge in separating offensive and defensive possessions since these are not tracked by the NHL and transitions can occur multiple times between pauses of play. Shifts are defined as a period of play in which no players are substituted. RAPM is calculated for even strength, power play, and shorthanded play separately for each shot-based metric (goals, shots, Fenwick, and Corsi). Macdonald's implementation of RAPM has not been updated to the current season but a separate implementation is up-to-date at the website Evolving Hockey [28].

Total Hockey Rating (THoR) was the first catch-all statistic for hockey to utilize advanced statistical techniques [29]. Unlike RAPM, THoR considers all the events recorded by the RTSS system. THoR attempted to derive player ratings by utilizing ridge regression and attempts to account for home ice advantage and the effects of starting play in the offensive zone versus the defensive zone. Outputs were expressed as Wins Created but were not compared to an average or replacement level player. THoR has not been updated for recent seasons.

The first WAR model in hockey was developed for the website War-On-Ice [30]. The goal of this model was to maximize its year-to-year predictive power, meaning there was an emphasis on filtering out the randomness in hockey and focusing on the true ability of the player. The next major WAR model was developed by Dawson Sprigings at the website Hockey-Graphs [31]. This model placed even more of an emphasis on evaluating the true value of a player, rather than accurately representing what he had done. Both these models were notable for their breakdown of component parts and for being the first metrics in hockey to represent player value relative to a baseline. However, neither is currently available.

There are two WAR models still in use in hockey analysis. The first was developed for Corsica Hockey (cWAR) [32] and the second for Evolving Hockey (eWAR) [31]. Both models follow a similar high-level methodology, breaking down value into component parts, representing a player's contributions in terms of Goals Above Average (GAA) for that component, converting GAA to Goals Above Replacement (GAR) based on the average performance of a replacement level player, and finally converting GAR to WAR and summing the component parts together to get an overall value. However, the models differ in philosophy, methods, how component parts are broken down, and how replacement level is defined, leading to differences in the final outputs.

The goal of cWAR is to be a mix of predictive and descriptive [31]. To this end, player value is calculated holistically without breaking down value by strength state like other models have done [23, 31, 33]. Instead, value is broken down by independent aspects of hockey that exist at all strength states: shot rate creation and prevention, shot quality creation and prevention, shooter quality, penalty impacts, and transition ability [32]. Value for each component is derived using a variety of regression techniques. Replacement level for cWAR is defined as the average performance of all players on a league minimum contract.

By contrast, eWAR aims to be as descriptive as possible, with no emphasis being placed on predictive power [31]. Player contributions are broken down by general strength state, even strength, power play, and shorthanded, but not into specific strength states (i.e., 5-on-5 and 4-on-4 are both treated equally as even strength). Player impacts are derived from Evolving Hockey's implementation of RAPM and converted into offensive and defensive GAA [34]. Offensive and defensive values are used for even strength, while only offense is used for power play value and only defense is used for shorthanded value. Penalty impacts are calculated separately as their own component. Replacement level for eWAR is defined each season as the average performance of each team's thirteenth most used forward and seventh most used defenseman for forwards and defensemen respectively since NHL teams usually play twelve forwards and six defensemen in a given game.

2.3 Context Adjustments in Hockey

Due to the random nature of hockey and the difficulties in recording events accurately, several attempts have been made to adjust play-by-play records to correct for human error and bias. Location information is difficult to measure because of the speed of the game, but some rinks have been observed to record shot locations that differ substantially from league norms and a location adjustment method has been proposed by Schuckers and Curro to correct for this [29]. Rink bias has also been observed for other play-by-play events with some rinks consistently recording event totals significantly above or below totals in other rinks. Some rinks also include a home team bias or "homer" effect in event totals that are beyond what would be expected from the home team. One method of accounting for these effects is to only use away game totals when calculating a player's output, since this averages out opposing rink biases. Another method has been proposed by Schuckers and Macdonald that calculates a rink and homer bias for each rink [35]. Adjusted totals for this method are not up to date, but their methodology is described clearly and can be recreated.

Additionally, some intrinsic aspects of the sport complicate analysis. Home ice advantage, also called venue effect, is a known feature of hockey and is included as an independent factor in other adjustment methods [29, 35]. Score effects are another well-documented phenomenon in which the trailing team tends to outshoot the leading team . One proposed method of adjusting for score effects proposed by Tulsky is essentially a weighted average of a team's shot attempt share while in various score states and their performance relative to league average in those states [36].

Another method has been proposed by McCurdy that accounts for score and venue effects simultaneously [37]. This method has been shown to produce more reliable values in terms of year-to-year predictability than Tulsky's method. Score and venue adjustments have not been performed for non-shot metrics, but the methodology can be extended to them as well.

2.4 Playing Styles in Hockey

Playing styles can be difficult to define but are considered an integral part of hockey player analysis. In the *NHL* video game series, for example, every player is given a playing style in addition to their overall skill ratings. However, establishing the cutoffs between different styles is tricky, with some styles containing significant overlap. Some playing styles have also been linked to different aging patterns [38], making playing style a potentially important aspect to consider when making projections for future contributions.

There have been several attempts to identify playing styles statistically. Neutral zone playing styles have been examined in terms of how often a player is tasked with entering the offensive zone and the way they do so (carrying the puck in versus shooting it in) [39]. Stimson proposed classifying overall playing styles for forwards and defensemen using *k*-means clustering and shot and passing statistics [40]. Goal-scoring style has also been analyzed using *k*-means clustering based on a wide variety of statistics [41].

2.5 Projections and Predictions

2.5.1 Baseball

As with catch-all statistics, baseball is at the forefront of play projections. The simplest method of forecasting in the Marcel the Monkey Forecasting System (Marcels), designed to be "the most basic forecasting system you can have, that uses as little intelligence as possible" [42]. The idea being that any decent projection system should be able to provide better forecasts. Marcels take a weighted average of a player's previous three seasons, with the most recent season receiving the most weight, apply a regression to mean production based on playing time, and apply an age adjustment.

A more sophisticated model for MLB is PECOTA, designed for Baseball Prospectus [43]. PECOTA's details are black-boxed, but in general it works by computing similarity score between a given player and other players with similar careers. PECOTA makes use of minor league statistics and combines that with similar careers to make its projections. Adjustments are also made based on calculations to determine a player's true talent, and control for randomness or luck that may have been involved in their previous production.

2.5.2 Basketball

A projection system for the NBA called CARMELO has been developed for the website FiveThirtyEight [44]. CARMELO uses a method comparable to PECOTA based on similarity scores for players and the career arcs of similar players. CARMELO projects a player's offensive and defensive impact separately in terms of Box Plus/Minus. A notable innovation for CARMELO is putting a player's projected value in terms of dollars, which is a useful idea for a league with a salary cap.

2.5.3 Hockey

Hockey has had a small number of projections systems for individual contributions with more work being concentrated on team projections [45, 46, 47]. Aging curves have been researched generally [38, 48], but this work has not been directly incorporated into any current projection systems. Marcels have also been adapted for skaters [49] and goalies [50], but up-to-date forecasts are not available using either of these methods. A system called VUKOTA that was similar to PECOTA was developed, but it has not been maintained. Projections have been done for the website The Athletic using Game Score, which is then converted into a metric called Game Score Value Added, but this is done mainly in service of team-level projections [51].

3. METHODS OF ANALYSIS

This project consisted of several distinct phases. The first step was searching for and collecting relevant data for the project. After that, consolidating it and analyzing it for useful information took the bulk of the time and effort. Once that was completed, preliminary predictive models were designed as a baseline to judge subsequent models against. Lastly, a final model was developed for predicting the contribution of a player in the following year.

The following statistical learning methods were utilized over the course of the project.

- Elastic-Net Regression as part of the exploration and analysis of effects and biases on play-by-play data.
- *k*-means clustering to define playing styles and coaching usage.
- Linear models using least means to model future individual contributions.
- Generalized additive models to compare with linear models and examine the possibility of non-linear relationships between the predictor and target variables.

3.1 Data Collection

Data was collected from seven different sources. Each source provided unique information not available from other sources. A breakdown of the volume of data collected from each source is presented in Table 3.1.

3.1.1 Fenwicka Play-By-Play

The largest data set was a compilation of all shift and play-by-play events for each regular season and playoff game from the 2007-08 season to the 2017-18 season. The data set was originally scraped from the NHL's official play-by-play records, ESPN's play-by-play records, and the NHL's shift charts and made available by Perry through his server Fenwicka. Events recorded included all nine RTSS events described previously, ON and OFF events

Source	Number of Files	Total Data Points
Fenwicka	11	13,255,986
Money Puck	1	1,173,844
Evolving Hockey	3	30,583
Corsica	2	18,229
eSA Spreadsheet	1	8,480
NHL Injury Viz	1	6,853
NHL.com	2,244	2,244

Table 3.1: Summary of data sources

indicating on-ice personnel changes, beginnings and endings of games and periods, stoppages caused by icings, goalies, or pucks leaving the ice, and coaches challenges. Each season contained over 1,000,000 events, except for the lockout shortened 2012-13 season, which contains over 700,000 events. Each event has up to 57 fields, which are detailed in Table 3.2. Notable fields include the adjustments, which provide the adjustment coefficients for McCurdy's score and venue adjustment method for each shot attempt event, and the probability of a goal for each Fenwick event, which is the output of Perry's expected goals model.

3.1.2 Money Puck Shot Data

In addition to the play-by-play data from Fenwicka, a second set of shot data was collected from the website Money Puck. This set contains the same set of Fenwick events as the Fenwicka set, however it contains much more detail about each event with over 120 fields for each entry. In total there are over 1,100,000 Fenwick regular season and playoff events from 2007-08 to 2017-18. Notable additional information included in the Money Puck data are indicators for if the shot was a rush or rebound shot, the expected goal value of each shot according to Money Puck's expected goals model, adjusted event coordinates according to the

Game ID	Index	Season	Date	Session	Period
Event Time	Type	Description	Detail	Event Team	Player 1
Player 2	Player 3	Length	X Loc	Y Loc	Substitutions
Home On 1	Home On 2	Home On 3	Home On 4	Home On 5	Home On 6
Away On 1	Away On 2	Away On 3	Away On 4	Away On 5	Away On 6
Home Goalie	Away Goalie	Home Team	Away Team	Home Skaters	Away Skaters
Home Score	Away Score	Score State	Strength State	Highlight Code	Distance
Angle	Rink Side	Circle	Zone	Home Zone	Prob Goal
Prob Save	Home Corsi Adjustment	Home Fenwick Adjustment	Home Shot Adjustment	Home Goal Adjustment	Away Corsi Adjustment
Away Fenwick Adjustment	Away Shot Adjustment	Away Goal Adjustment			

Table 3.2: Fenwicka play-by-play event fields

method laid out by Schuckers and Curro, and attempts to measure pre-shot puck movement by measuring distance travelled and angle to the goaltender since the previous event.

3.1.3 Evolving Hockey RAPM and WAR

Season level data was collected from Evolving Hockey. Data was available for all regular seasons from 2007-08 to 2017-18. Raw outputs for the Evolving Hockey implementation of RAPM were collected for goals, expected goals, and Corsi. These impacts were calculated at the player season level, meaning players who were traded during a season did not have their results split between teams. RAPM outputs relative to a player’s teammates were also collected, which were split by team. Lastly, the outputs for the eWAR model and its available component parts were collected. These outputs were available at a player-season-team level. The component parts of the eWAR model are expressed in terms of GAR.

3.1.4 Corsica WAR and Quality of Teammates and Competition

Additional season level data was also collected from Corsica. Quality of Teammate (QoT) and Quality of Competition (QoC) context data was collected for regular season 5-on-5 play from 2007-08 to 2017-18, adjusted for score and venue according to McCurdy's method. QoT and QoC consist of a weighted average of a player's teammates and opponents shot metric rates without a player, weighted according to the time spent on the ice with a player. The outputs of the cWAR model and its components were also collected. The cWAR model is computed at the season level and includes the playoffs. The cWAR data was collected in GAR form to make interpretation and comparison to eWAR easier.

3.1.5 Estimated Shot Assist Data

Estimated shot assists (eSA) were collected for all regular season 5-on-5 play from 2007-08 to 2017-18. Shot assists have been manually tracked publicly. However, the tracking data only goes back to the 2014-15 season. The shot assist estimates were derived by a linear regression outlined at Hockey Graphs [52]. The full data set was stored in a Google spreadsheet and retrieved from there.

3.1.6 Injury Data

NHL injury data was collected from the NHL Injury Viz Tableau page. Injury data for the site is sourced from TSN's player biographical information, but only extends back to the 2009-10 season. Injury information includes each incident of an injury report, how many games were missed, and the official description of the injury type. NHL teams are notoriously secretive about injuries [53], so the data is likely missing minor injuries that were not announced or announced injuries that were misleading ("flu-like symptoms" is commonly used to cover for other injuries) or vague ("upper body injury"). It also does not include injuries if no games were missed. For instance, Jamie Benn played through an injury and underwent double hip surgery during the 2015 offseason and had abdominal surgery the next summer, but did not miss any games during either season, so both incidents are left out.

3.1.7 Player Biographical Data

Finally, player biographical data was collected from the NHL's official records via their statistics API. This process was by far the most intensive. The NHL API requires the player's NHL ID to access player information. However, player IDs were not included in any of the other data sets that were collected. To retrieve each player ID, I scraped each official game log using game IDs, which follow an obvious formula. Information is returned in JSON format, which includes the full roster for each team. Using this information, a list of the IDs of all players who had appeared in a regular season or playoff game from the 2007-08 season to the 2017-18 season was compiled and their date of birth, first season, and handedness was collected for each player through the API. Data was scraped and processed using Python and written to a CSV file.

3.2 Data Consolidation and Analysis

Once the data was collected from each source, it was examined in greater detail. Unfortunately, many errors were discovered with the accuracy of the play-by-play set, which will be elaborated on in the results section. Major issues included missing players, both on ice and as actors in the event, missing event length, and inconsistencies in player identification. A substantial amount of time was expended to correct these errors manually, however the volume was too great and the gain in accuracy was not likely to justify correcting every possible error. The player identification issues were rectified, but the decision was made to exclude the data that involved missing event lengths or player information.

Statistic	Description
Corsi	GOAL, SHOT, MISS or BLOCK event
Fenwick	GOAL, SHOT, or MISS event
Shot	GOAL or SHOT event
Turnovers	GIVE event for or TAKE event against

Table 3.3: Composite statistics

The next step was extracting meaningful information from the mass of data points. I decided to focus on 5-on-5 play since it forms the bulk of playing time in a given season. Additionally, power play and shorthanded time are highly structured and considered more dependent on coaching than even strength play. Counts for each event for each player were compiled from the play-by-play data at a season and team level. Each event that a player was listed as the primary actor for was credited as an individual event for. Each event that a player was listed as the secondary actor for was credited as an individual event against. Second and third players listed on goal events were credited with primary and secondary assists respectively. The same aggregation for events a player was on the ice for and against. A player's individual counts are included in the total for team counts. Additionally, zone starts were included, which are the number of faceoffs taken in the offensive, defensive, and neutral zones that the player was on the ice for. This resulted in 36 columns of raw counts for each player. Next these counts were combined into the composite metrics described in Table 3.3. Counts were then expressed in terms of rate per 60 minutes of play and shares. Shares for individual statistics are the individual count divided by the for count; shares for on ice statistics are the team for count divided by the total on ice count. In total, this created 180 variables in the data set.

3.3 Adjusting for Effects and Bias

3.3.1 Score and Venue Effects

After compiling the raw counts for each player, the focus turned to adjusting the values to improve their predictive power. Two adjustment methods were tested, one that produces a single coefficient for score and venue (home or away) effects and one that produces separate coefficients for rink and homer effects. These adjustments cover discrete external effects on the flow of play, and theoretically can be "stacked" on top of each other. The score and venue adjustments were calculated following the method described by McCurdy [37]. The total count for each event type for each team was summed and binned according to the goal differential for the home team with goal differentials of more than three being binned together "3 or more" and "-3 or less". The coefficient for each event, score state, and venue was then calculated according to the following formula:

$$C_{esv} = \frac{2T_{esv}}{T_{es}}$$

where *C* is the coefficient value; *T* is the count for an event; *e* is the event type; *s* is the score differential, relative to the home team; and *v* is the venue; either home or away. Coefficients were calculated for all RTSS events, excluding faceoffs, as well as Corsi, Fenwick, and shots. Regular season counts for all available seasons were used when calculating the adjustment coefficients.

3.3.2 Rink and Homer Bias

Rink and homer adjustments were also calculated according to the method described by Shuckers and Macdonald. This method includes score, venue, and team effects as independent variables to isolate them from the rink and homer effects. Adjustments are calculated for each event using an elastic net Poisson regression according to the following formula:

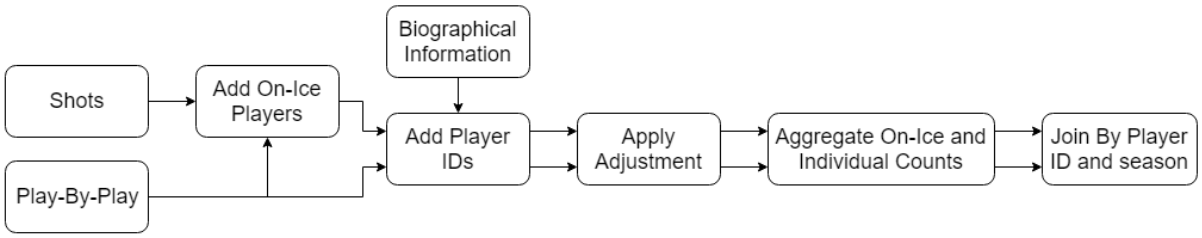

Figure 3.1: Pipeline for aggregation of adjusted season level statistics

$$ln(Y_i^*) = \mu + \gamma_j + \beta ASD_i + \phi_k + \omega_l + \eta I_{j=k} + \eta\gamma I_{j=k} + \epsilon$$

where Y_i^* is the per 60-minute rate of the event for game i; μ is the intercept; γ_j is the rink effect for rink j; βASD_i is the average score differential for game i, an average of the differential weighted by the time spent at each differential; ϕ_k is the for team; ω_l is the against team; η is the home effect; $\eta\gamma$ is the homer effect; $I_{j=k}$ is an indicator function indicating whether the for team is the for team and ϵ is the error term. Each game was listed twice, once with the home team as the for team and once with the away team as the for team. The regression was performed using the glmnet library in R. Coefficients were calculated for each RTSS event plus Corsi, Fenwick, and shots by rink, for each season.

Once all the coefficients were calculated, regular season 5-on-5 totals were then aggregated for each type of adjustment, as illustrated in Figure 3.1, as well as all possible combinations of adjustments. This resulted in seven additional tables of season level statistics for each player. Next, the adjusted statistics were then tested for repeatability, using year-to-year R^2, a standard method in sports analytics [37, 54, 5]. The adjustment that had the most predictive power was selected as the basis for future analysis.

3.4 Style and Usage Analysis

Next, playing style and coach usage were estimated using *k*-means clustering to act as a proxy for similarity scores between players, since the sample of available seasons in relatively small. Playing style, particularly physicality of a player, has been linked to increased or decreased speed of aging [38]. Similarly, coach deployment seems to have a possible effect on aging since some situations (shorthanded time, defensive deployments) are considered more physically difficult that others (power play time) [55]. Clusters were estimated separately for forwards and defensemen, since the playing opportunities are fundamentally different enough that there should be meaningful differences between the two groups. Defensemen play farther away from the opposing net when on offense and closer to their own net on defense, affecting their involvement in aspects such as shot generation (less involved) and shot suppression (more involved). Teams also typically play only six defensemen on three pairs, compared to twelve forwards on four lines, meaning even a bottom pair defenseman will play more substantial minutes in a given game than a bottom line forward.

Playing styles were separated into offense and defense, as these two processes have been shown to be discrete [56]. Adjusted values were used for clustering. Features selected for offensive playing style clustering were individual Corsi per 60 minutes and individual Corsi share to estimate shooting volume, the percentage of estimated shot contributions that were shot assists and the estimated shot assists per 60 minutes to estimate playmaking ability, and the expected goal value from the Money Puck data per individual Fenwick attempt to estimate shot quality. The percentage of rush shots and rebound shots of a player’s total shots were also included for forwards, but not for defensemen since defensemen rarely are part of an offensive rush or near enough to the opposing net to shoot off a rebound. For defensive playing style, hits per 60 minutes, hit share, and penalties per 60 minutes were included to estimate physicality in defending. Blocks per 60 minutes and block share were included to estimate the ability to prevent offense by positioning. Takeaways per 60 minutes were included to estimate the ability to transition from defense to offense. For coaching usage, zone start ratio and time on ice share for even strength, power play, and shorthanded situations were included. Players were assigned clusters at the season level. Each classification style was tested on two to nine clusters and the number of clusters were selected by examining the within cluster sum of squares.

3.5 Modelling

Once the processes of data aggregation, transformation, and adjustment and clustering were complete, the outputs were joined with QoT and QoC values from Corsica, the raw and relative RAPM outputs from Evolving Hockey, shot assist data, injury data, component parts for eWAR and cWAR, and player age and the number of years since their debut, resulting in over 250 available features to be used in constructing a model. A target variable also had to be selected and eWAR was chosen for several reasons. First, it is a single number metric and although projecting constituent parts is worthwhile, but ultimately not the goal, since a team cannot acquire just a player's goal scoring ability, for instance. Projecting multiple components also increases the uncertainty for final projections. Therefore, a single number seemed an appropriate target for this thesis. Second, it is readily available. Other models such as Point Contributions, THoR, GVT, and the War-On-Ice and Sprigings WAR models are not currently available for a variety of reasons. Third, it mainly avoids arbitrary decisions in weighting different components by utilizing regression techniques, unlike Point Shares. Fourth, it attempts to be as descriptive a possible, valuing what a player has done in a season, versus what he was expected to do, unlike cWAR. Professional sports are results oriented by nature, and teams and fans are more concerned with what a player tangibly produces than what he could have produced. Finally, it is easily broken down by strength state, unlike cWAR, which calculates its components at all strength states. This is different than being split by constituent parts since a team does have control over which strength state to use a player at. A team can choose not to play a player shorthanded if they provide negative value. Even strength eWAR ($eWAR_{EV}$) was chosen as the focus as the model since most of a game is played at even strength. Additionally, power plays are highly structured, and a player's contributions are dependent on coaching strategy and the role a player is asked to play on the power play. Shorthanded value is also highly variable and may not be a skill except for a very small number of players [34, 57]. Lastly, GAR value was used instead of WAR value because the values are larger, making it a bit easier to interpret the magnitude of results.

Player seasons to use for building a model were selected according to the following criteria: the top 690 players by ice time each season from 2009-10 to 2016-17 and the top 713 for the 2017-18 season. An NHL roster is 23 players, so this serves as a rough estimate of who was a fulltime player in each season. Next, players were removed if they did not appear at least twice since this meant they did not have both a target season and predictor season . This removed 375 player seasons, bringing the total to 7,238.

The first round of modelling was inspired by the Marcel forecasting system. The rate of $eGAR_{EV}$ ($eGAR_{EV}/60$) was chosen as a target variable for projection, using just previous years' rates and an age adjustment. Rates were chosen over counts to allow direct comparisons between player. This was to establish a baseline for subsequent models. Independent variables were also added for years a player had been in the league to measure experience, a player's position, and whether to included missed games due to injuries in the previous season. A linear model using ordinary least squares regression and a generalized additive model were both used, with age and experience being treated as second-degree polynomials and a spline functions respectively, due to the “curve” shape associated with aging. The models were run for every combination of previous years included from one to five, and inclusion of age, years in the league, position, and previous games missed with injuries. Previous season rate variables were also treated as either a single dependent variable or as an interaction between the previous count and time played. In total, this resulted 320 variations on the baseline model. The 2007-09 seasons were excluded since they did not include injury data, reducing the number of seasons to 5,900. Models were trained and evaluated using 5-fold cross-validation.

With this baseline, using different target variables and predictors were examined. Total $eGAR_{EV}$ rather than rate was tested as a target, as well as the change (Δ) in total $eGAR_{EV}$ and $eGAR_{EV}/60$ between seasons. Features were selected for experimentation based on how well they correlated with the predictor variable and how distinct they were from each other to try to capture the broadest possible scope. Over 100 additional variation on models were tested using this method using the same 5-fold cross-validation technique. A linear model was the primary method used for most of the testing, though a robust linear model from the MASS package in R was also tested. Transformations for target variables were explored using the bestNormalize package and an inverse hyperbolic sine transformation of $\Delta eGAR_{EV}/60$ was ultimately selected as the target variable.

4. RESULTS AND DISCUSSION

4.1 Adjustments

4.1.1 Score and Venue Effects

The results of the score and venue adjustments for shot-based metrics followed the expected pattern. Teams tended to take more shots the further they were behind. There is also a persistent bias in favor of the home team compared to the complimentary score state. These effects are illustrated in Figure 4.1 and the coefficients for the Corsi adjustments are detailed in Table 4.1. Goals buck this trend slightly, with a much smaller range of coefficients and with the home team having a persistent advantage. This could be the result of random variance given the relatively small sample, compared to other shot event types. For non-shot events, hits are the only event that is consistently affected by the score, with the relationship being positive. That is, the greater a team's lead, the more hits were recorded for them. This is consistent with the shot event data, since a team must possess the puck to shoot it and must not possess the puck in order to throw a hit. Giveaways had a persistent bias towards the home team with away coefficients ranging from 1.15 to 1.24, regardless of the score differential. Takeaways, on the other hand,

Home Lead	Home Total	Away Total	Home Coefficient	Away Coefficient
3 or more	31,552	40,582	1.14	0.889
2	46,498	56,286	1.11	0.913
1	108,461	119,730	1.05	0.953
0	233,721	21,9347	0.969	1.03
-1	112,386	89,363	0.898	1.13
-2	47,663	34,906	0.866	1.18
-3 or less	28,810	19,628	0.841	1.23

Table 4.1: Score and venue Corsi adjustment coefficients

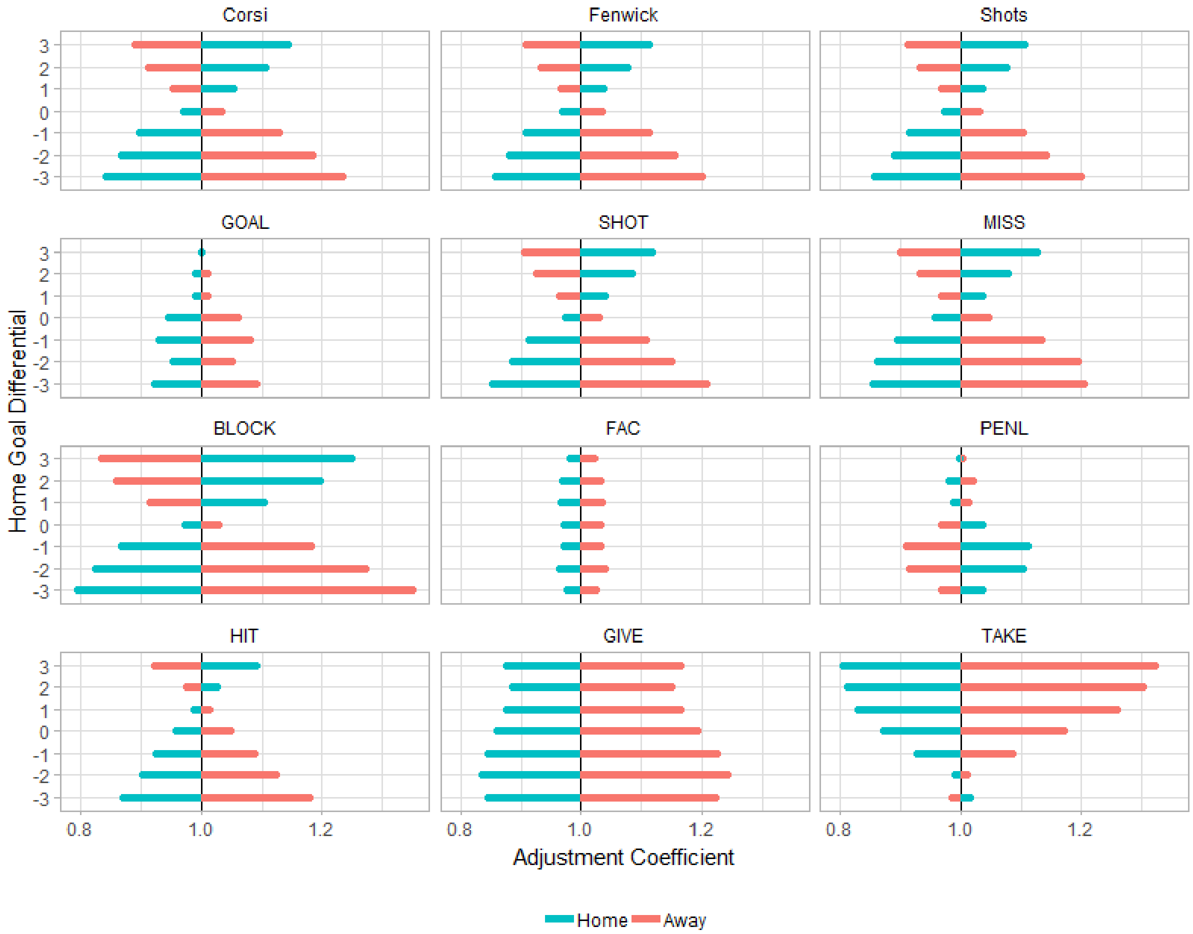

Figure 4.1: Score and venue adjustment coefficients by home goal differential

also displayed a strong bias in favor of the home team, but also showed a positive relationship with the score state. Finally, penalties were distributed evenly between the home and away teams regardless of the score. There was a slightly stronger bias toward calling penalties in favor of the team when they trailed by one or two goals. This could be explained by crowd influence on the referee, or perhaps some other unconsidered factor.

4.1.2 Rink Effects

The rink adjustment effects varied much more widely than the score and venue adjustments. Giveaways and takeaways especially had massive ranges compared to other events

Rink	Season	Event	Rink Effect	Coefficient
ARI	2007-08	GIVE	0.319	3.13
ARI	2008-09	GIVE	0.346	2.89
CBJ	2009-10	GIVE	0.362	2.76
ARI	2009-10	GIVE	0.372	2.69
CBJ	2007-08	GIVE	0.399	2.51
TOR	2012-13	GIVE	1.92	0.522
EDM	2009-10	GIVE	1.94	0.516
CAR	2017-18	TAKE	2.06	0.486
NYI	2009-10	TAKE	2.09	0.479
EDM	2010-11	GIVE	2.10	0.475

Table 4.2: Five largest and smallest rink adjustment coefficients

and were the most affected overall, as shown in Table 4.2. The largest coefficient values were 3.13 for giveaways and 2.13 for takeaways, while the smallest values were 0.475 for giveaways and 0.479 for takeaways. The next largest and smallest coefficients for a different event were 1.78 for missed shots and 0.728 for blocked shots. This makes some sense, given the ambiguity in how these events are defined. Giveaways and takeaways require the scorer to determine whether the turnover was forced or unforced and assign credit for the event. Missed and blocked shots also require judgement from the scorer in determining whether to award credit for shots that intentionally miss the net and inadvertent blocks by defensive players. Hits had the third largest standard deviation of all events at 0.0255 and displayed a fairly large range of effect coefficients ranging from 0.732 to 1.59. The composite statistics shots, Fenwick and Corsi had the second to fourth smallest standard deviations, behind only the non-composite SHOT event, suggesting that missed shot and blocked shot rink effects did not often coincide.

4.1.3 Homer Effects

The homer effects were much more tightly distributed than the rink effects, though Giveaways and takeaways were still the most variable, as shown in Table 4.3. Shuckers and

Rink	Season	Event	Homer Effect	Coefficient
BOS	2008-09	GIVE	0.663	1.51
BOS	2010-11	TAKE	0.694	1.44
BOS	2007-08	TAKE	0.775	1.29
ANA	2014-15	TAKE	0.776	1.29
PHI	2017-18	TAKE	0.791	1.26
VGK	2017-18	TAKE	1.46	0.686
SJ	2015-16	TAKE	1.51	0.661
SJ	2007-08	GIVE	1.52	0.658
SJ	2008-09	GIVE	1.52	0.657
WSH	2008-09	GIVE	1.52	0.657

Table 4.3: Five largest and smallest homer adjustment coefficients

Macdonald proposed that the terms themselves carry negative and positive connotations, which may influence the scorekeeper when deciding whether to award a giveaway or takeaway [35]. Blocks and misses were also subject to higher variance in effect. Shots, Fenwick, and Corsi were all in the bottom four again for variance of the coefficients. The distribution of positive and negative biases was also more skewed than rink effects. Where rink effects benefitted the home and away teams about equally, the number of rinks with positive homer effects outweighed the negative effects by about a 2:1 ratio.

When the adjusted season level counts were tested for repeatability, the results shown in Table 4.4 were somewhat surprising. The score and venue adjustment performed well, improving the year-to-year predictability for 94 of its adjusted statistics, while worsening only 75. However, the rink and homer adjustments did very poorly. The rink adjustments were much worse than the unadjusted raw season numbers with over 120 metrics losing predictive power and only 48 gaining. The negative value of rink adjustments was also much larger for several metrics, illustrated in Figure 4.2. Additionally, stacking the effects did not improve the predictability. One possible reason for this is that the rink and homer effects were done at a

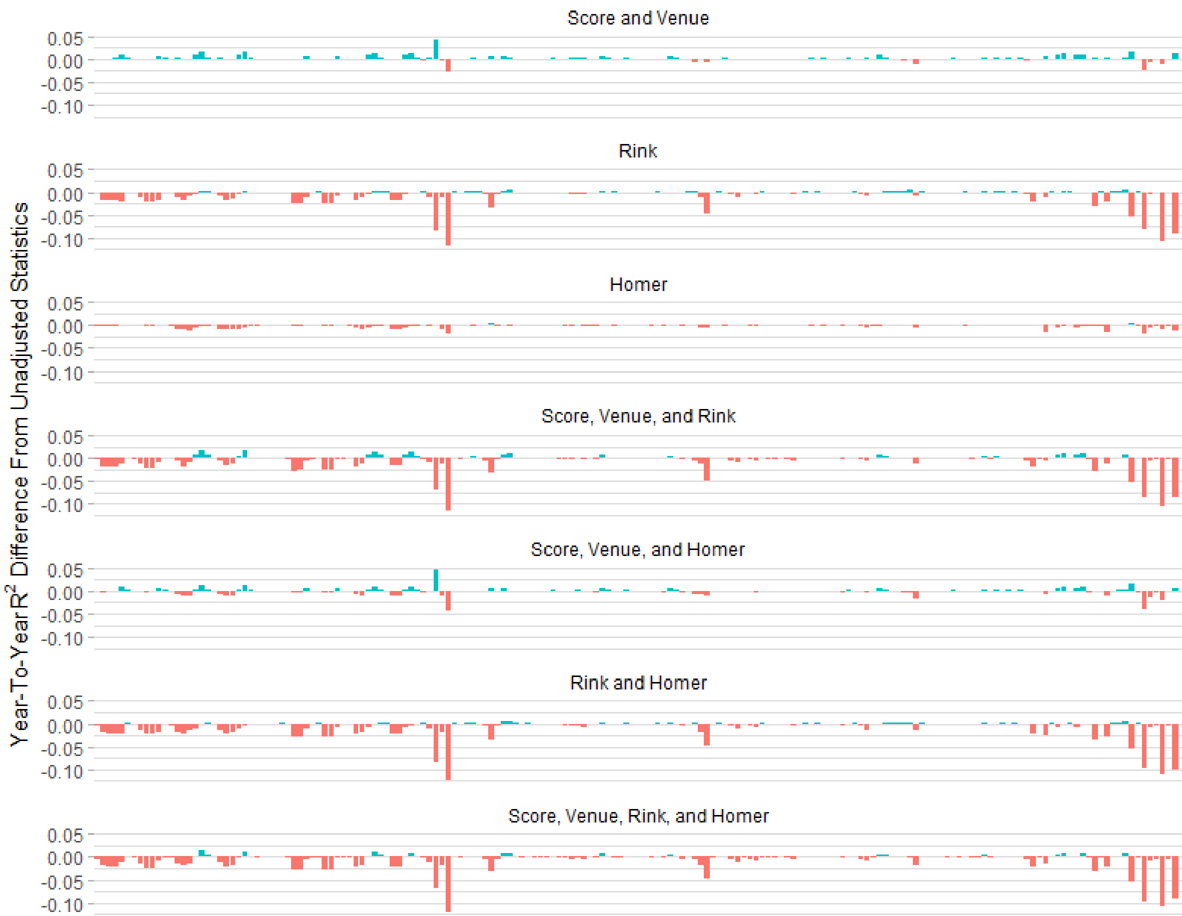

Figure 4.2: Magnitude of change in year-to-year R^2 for each adjustment method compared to unadjusted statistics

season level, which may have introduced more noise and overfitting to the coefficients. Each team plays a maximum of 57 home games in a season, and most play fewer than 50, and it is possible that isn't enough for the effects to stabilize. It is interesting that the simplest method produced the most accurate results.

4.2 Style Clustering

Players were classified according to their offensive and defensive playing styles and their deployment by the coaching staff. This was an attempt to incorporate something akin to the

Score & Venue	Rink	Homer	Improved R^2	Worsened R^2
T	F	F	**94**	**75**
F	T	F	48	121
F	F	T	65	104
T	T	F	62	107
T	F	T	83	86
F	T	T	49	120
T	T	T	65	104

Table 4.4: Breakdown of year-to-year R^2 improvements by adjustment method

similarity scores used by PECOTA and CARMELO, despite the limitations of the sample size. The reasoning was that players that play a similar style or face similar usage would develop in similar ways, making comparisons between those players possible, even if only part of their career was covered by the available data.

4.2.1 Offensive Style

For forwards, four clusters were found to be the best fit for the features selected, detailed in Table 4.5. The first group featured a lower individual shot attempt rate and share, but a high percentage of their estimated shot contributions come from shot assists, and their shot assists are the highest of the four groupings. Their expected goals per attempt were the highest of the group, suggesting more careful shot selection. This fit the traditional description of a playmaker or passer. The second group featured the lower shot attempt rate and a low expected goals per attempt, but a relatively balanced ratio of individual shot attempts and estimated shot assists. This profiled as a middle of the lineup forward with some offensive ability, also referred to as balanced or two-way players. The third group had an extremely low estimated shot assist rate and a relatively high individual attempt rate. The high percentage of rush shots suggests an inability to sustain time in the offensive zone. Due to their high individual output, but low

Position	iCorsi/60	iCorsi Share	eSA%	eSA/60	Rush Shot%	Rebound Shot%	iXGoals/ iFenwick
Forward	11.1	0.195	0.610	17.2	0.00207	0.0539	0.0784
Forward	9.74	0.209	0.583	13.4	0.00217	0.0489	0.0717
Forward	12.4	0.221	0.572	6.30	0.00231	0.0499	0.0743
Forward	15.7	0.273	0.498	15.4	0.00221	0.0461	0.0704
Defense	9.94	0.174	0.516	3.96	-	-	0.0287
Defense	10.6	0.187	0.496	10.4	-	-	0.0284
Defense	6.83	0.134	0.594	10.1	-	-	0.0281

Table 4.5: Even strength offensive playing style cluster centers

overall output, they were classified as low-offense shooters. The last group had the highest shot attempt rate and shot attempt share in addition to the highest estimated shot assist rate. Due to their overall high rate of offense and their high individual shot rates, they were be classified as high-offense shooters.

For defensemen, three clusters were the best fit and produced three easily separable styles, shown in Table 4.5. The first group was noteworthy for its high attempt rate and low estimated shot assist rate and was classified as shoot-only defensemen. The second group had both the highest individual shot attempt rate and the highest estimate shot assist rate, which easily fit the traditional offensive defenseman style. The third group had the lowest shot rate and a high shot assist rate, suggesting they prefer to let other players shoot. They were classified as having a deferential offensive style.

4.2.2 Defensive Style

The defensive style clusters were similar for forwards and defensemen and are detailed in Table 4.6. Both groups performed best with three clusters, though the means for each metric differed by position. One cluster for each position was a highly physical style with high hit rates and shares, penalty rates, and blocked shot rates and shares and the lowest takeaway rates. A

Position	Hits/60	Hit Share	Penalties/60	Blocks/60	Block Share	Takeaways/60
Forward	8.20	0.275	1.08	1.94	0.135	1.53
Forward	15.5	0.383	2.23	2.11	0.143	1.21
Forward	3.09	0.135	0.652	1.67	0.119	1.87
Defense	9.72	0.301	1.22	4.86	0.324	0.651
Defense	2.53	0.105	0.568	4.23	0.301	0.928
Defense	5.53	0.201	0.823	4.39	0.309	0.808

Table 4.6: Even strength defensive playing style cluster centers

second group for each position had the highest takeaway rates and low hit, penalty and block rates and shares. This style was considered transitional defense, meaning the emphasis was on taking the puck away from the opposing team and transitioning back to offense. The third group for each position had the middle value for all metrics. A potentially an insightful finding from the analysis of these groupings is that takeaway rates seem to be inversely related to physical play, with physically demanding and potentially injurious actions such as hitting and blocking shots decreasing as takeaways increase.

4.2.3 Coach Usage

The coach deployment clusters were clearest at four clusters for both forwards and defensemen, shown in Table 4.7. The forwards categories are easily interpretable. The first group is the depth line players. They get the smallest share of even strength ice time and almost no time on the power play or shorthanded. They are also deployed with about an even zone start ratio. The second group is the top offensive specialists. They receive a large amount of even strength ice time with offensively skewed zone starts, heavy power play time and almost no shorthanded time. The third group is top two-way players. They get about even zone starts and play large minutes in all situations. Finally, the last group is defensive specialists. They get heavy defensive zone starts, play fewer minutes at even strength, and get the most shorthanded time and almost no power play time. The defensemen follow a similar distribution. The first

Position	Zone Start Ratio	EV TOI %	SH TOI %	PP TOI %
Forward	0.501	0.211	0.0575	0.101
Forward	0.542	0.276	0.0380	0.487
Forward	0.497	0.281	0.299	0.469
Forward	0.429	0.223	0.347	0.0565
Defense	0.498	0.363	0.432	0.509
Defense	0.476	0.332	0.449	0.0732
Defense	0.546	0.321	0.090	0.463
Defense	0.522	0.284	0.153	0.0720

Table 4.7: Coach deployment cluster centers

group is the top pair defensemen getting the most even strength ice time, the second most shorthanded time, and the most power play time. The second group is the shutdown defensemen. They get the second most time at even strength along with the highest share of defensive zone starts and the most time shorthanded, but hardly any time on the power play. The third group is in reverse, offensive defensemen who get heavy offensive zone starts, and lots of power play time, but very little time of the penalty kill. Finally, the fourth group of defensemen is the bottom pairing defensemen. They have the lowest ice time at even strength and are not strong enough offensively or defensively to get time on the power play or penalty kill. They are also “protected” at even strength by starting in the offensive zone and away from their own net as often as possible. These groupings follow conventional hockey coaching strategies.

4.3 Predictive Models

4.3.1 Baseline Models

Over 300 combinations of the baseline model were run using one to five seasons and options for age, years in the league, position, games missed with injuries, and either the previous seasons’ $eGAR_{EV}/60$ or an interaction between the previous seasons’ total $eGAR_{EV}$ and time on ice. Each combination of independent variables was tested using both a linear model and a GAM. Adding additional data was able to improve the predictive power of the models, but there

Model	Seasons	Age	Years	Position	Injury	Interaction	R^2	AIC	RSME
LM	1	No	No	No	No	No	0.0984	-	**0.329**
LM	1	No	No	No	No	Yes	0.111	-	0.324
LM	3	Yes	No	Yes	Yes	No	0.175	-	0.312
LM	5	Yes	No	No	No	Yes	0.202	-	**0.310**
GAM	1	N	No	No	No	No	-	391	0.329
GAM	1	No	No	No	No	Yes	-	381	0.324
GAM	3	Yes	No	Yes	Yes	No	-	310	0.312
GAM	5	Yes	No	No	No	Yes	-	299	0.311

Table 4.8: Sample baseline model comparison

appears to be a hard ceiling in terms RMSE on test data and the goodness of fit measures. The worst of these models used only a single year of previous data and none of the options and had an RMSE of 0.329. The best model was a linear model using five previous seasons, age, and the interaction between count and time instead of rate and had an RMSE of 0.310. Table 4.8 contains a sample of results including the best and worst performing linear models and the GAM counterpart.

It was interesting to see that the linear model outperformed the generalized additive model on all the tests that included the age variable. The age variable was modeled using a spline function in the GAM and a second-degree polynomial in the linear model. Like the adjustment coefficients, the simpler method proved more effective. This could be the result of overfitting the spline to the training data. Another interesting finding was that as more years were added to the model, the significance of the other factors decreased, except for age. More information about a player is gained through their past performance than by external factors, except in outlier cases.

Another interesting finding was that including previous seasons' information for up to five seasons continued to improve the predictive power of the models. In general, three years is considered a "good enough" limit for how much previous information to include in a model,

Variable	Estimate	Std. Err.	T-Value	P-Value
$eGAR_{EV}/60$	-8.43e-01	2.11e-02	-39.8	<2e-16
Age^2	-1.38e-04	2.43e-05	-5.69	1.46e-08
PP TOI %	1.85e-01	4.21e-02	4.39	1.20e-05
Primary Points Share	1.63e-01	4.58e-2	3.57	3.69e-04
$Impact_{CF\text{-}xGF}/60$	1.12e-02	3.46e-03	3.56	3.72e-04
Defensive $Style_{F3}$	4.17e-02	1.32e-02	3.17	1.55e-03
$Impact_{CA\text{-}xGA}/60$	-9.48e-03	3.46e-03	-2.74	6.19e-03
iBlocked Shot Share	2.74e-01	1.10e-01	2.49	1.28e-02
EV TOI%	4.57e-01	1.87e-01	2.45	1.45e-02
$cGAR_{Rates}$	5.61e-03	2.81e-03	2.00	4.60e-02

Table 4.9: Coefficients for the variables included in the final model

such as the Marcel system [58]. However, older information does still have some value, though there are diminishing returns to reaching further into past performance.

4.3.2 Final Model

These models, like the Marcels that inspired them, were only intended to serve as baselines and minimum standards for a more complex models to surpass. Requiring five years of previous experience to make more accurate projections is not realistic for teams looking for a competitive edge since most players are considered known quantities once they have played five years. A five-year cutoff also significantly limits the number a model would be able to provide projections for. With this motivation, a model was developed for projecting a player's $\Delta eGAR_{EV}/60$ using just a single previous season. The features of the model out outlined in Table 4.9. The most impactful variable, by far, in predicting the next season's performance was $eGAR_{EV}/60$. The negative relationship fits with the idea that single season $eGAR_{EV}/60$ is unstable and will regress to a player's true ability in the coming season. Age was the second most impactful variable, which matches prior expectations about its effect on performance. Power play and even strength ice time share also a have positive relationship to $\Delta eGAR_{EV}/60$, implying that coaches have some sense of who their best player are and distribute ice time proportionally, even if results are not present in a given season. The impact rates and $cGAR_{Rate}$ are attempts to measure a player's underlying performance, without focusing on the results and act as a proxy for true talent. This is another relationship that matches the assumption of a player regressing to the mean, in this case the mean on ice shooting and save percentages.

The only playing style with a significant impact was the third forward defensive style. This was the high-takeaway, low hit and block rate style. There is an implication, due to the positive relationship, that a less physical style is conducive to a higher positive impact. What's interesting though is that the same does not hold true for the corresponding defensive style. This could be because takeaways do not translate to offensive output as directly for defensemen. It's also noteworthy that the individual block shot share, a component of the defensive style, for a player has a slightly positive impact. This could be because blocked shots are not included in expected goals models since the recorded location is the blocker, not the shooter. Therefore, a higher blocked shot share could result in a lower impact for expected goals against, increasing the defensive value a player contributes.

Surprisingly unimportant factors included the number of games missed in the previous season due to injury. Even when games injured were binned to reduce noise, there was very little information gained. This suggests that in the aggregate, injuries do not affect a player's playing ability the following season. Of course, some injuries are more serious than others and games missed is a somewhat crude metric for evaluating the severity and impact of an injury on performance. Injuries should be taken on a case-by-case basis when evaluating their impact on a player's future. Relatedly, it was surprising to discover than physicality, as measured by hits thrown and received, penalties taken and drawn, and shots blocked was not predictive of decline, though it could be because the effect of this wear and tear is cumulative over multiple seasons

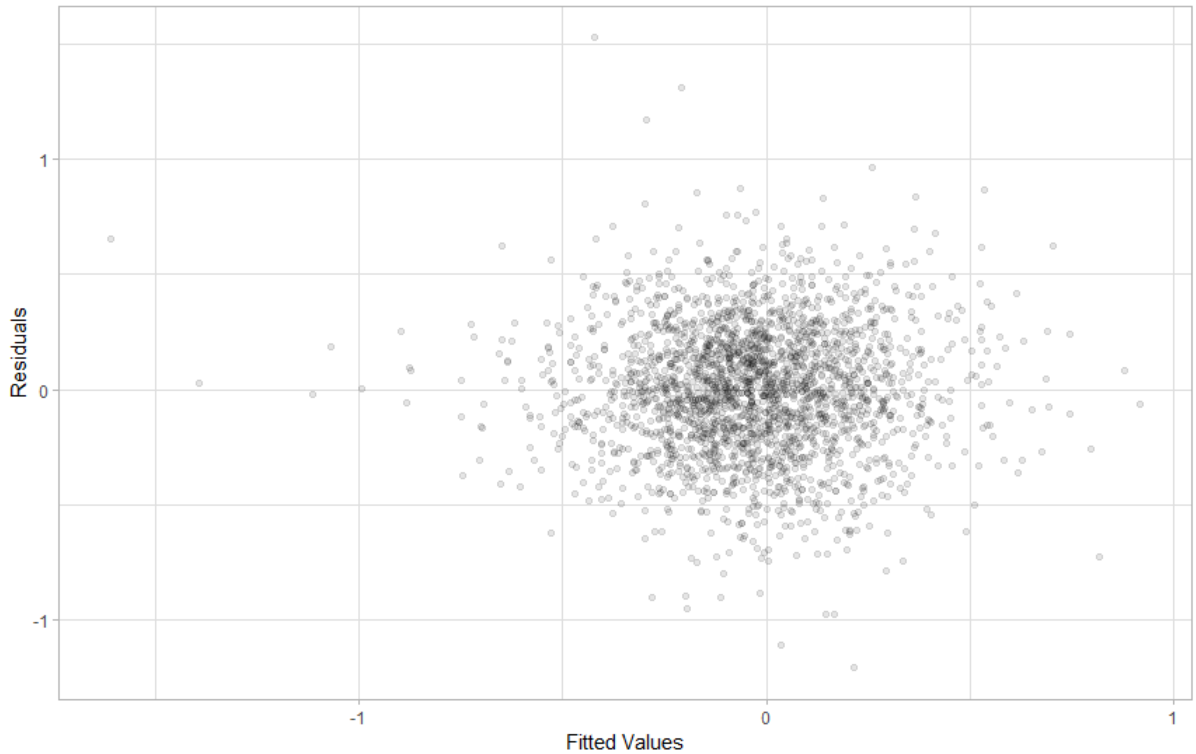

Figure 4.3: Plot of fitted values versus residuals for the final model

and a single season does not provide much information. It was also interesting that for the most part the playing and usage styles did not have significant impacts, although some component parts of the clusters did, such as power play and even strength ice time shares.

The R^2 for the final model was 0.417. RMSE on the test set was 0.299. Examining the relationship between the fitted values and the residuals, shown in Figure 4.3, does not suggest any linear patterns that have been missed by the model. The ten most and least accurate predictions are listed in Table 4.10 and Table 4.11 respectively. In the best case, the model predicted the change in $eGAR_{EV}/60$ almost exactly. It also handled some large changes well, including one swing of over 0.5, and some very small ones. In the worst predictions, the residual error was nearly 2.0, which was a huge miss. Most of the misses came from bottom of the lineup

Player	Season	$GAR_{EV}/60_n$	$GAR_{EV}/60_{n+1}$	Δ_{Actual}	$\Delta_{Predicted}$	Residual
Craig Adams	2009-10	-0.0720	-0.126	-0.0538	-0.0537	-8.77e-05
Nikolaj Ehlers	2015-16	0.383	0.292	-0.0911	-0.0910	-9.84e-05
Joffrey Lupul	2009-10	0.698	0.124	-0.574	-0.574	1.97e-04
Artemi Panarin	2015-16	0.357	0.269	-0.0880	-0.0878	-2.51e-04
Ryan Wilson	2009-10	0.235	0.0875	-0.148	-0.148	3.36e-04
Niklas Kronwall	2012-13	0.0496	0.0828	0.0331	0.0328	3.83e-04
Shea Weber	2011-12	0.357	0.214	-0.144	-0.144	4.26e-04
P.K. Subban	2015-16	0.485	0.199	-0.286	-0.285	-4.60e-04
Ryan Suter	2013-14	0.197	0.185	-0.0126	-0.0121	-4.65e-04
Connor Murphy	2016-17	0.249	0.0431	-0.206	-0.207	5.30e-04

Table 4.10: Ten most accurate predictions for the final model

Player	Season	$EVGAR/60_n$	$EVGAR/60_{n+1}$	Δ_{Actual}	$\Delta_{Predicted}$	Residual
Matt Gilroy	2011-12	0.274	-1.85	-2.12	-0.169	-1.95
Bret Carson	2009-10	0.220	1.86	1.64	-0.170	1.81
Kimmo Timonen	2013-14	0.167	-1.64	-1.81	-0.0736	-1.73
Sidney Crosby	2010-11	0.953	2.30	1.35	-0.346	1.69
Joffrey Lupul	2011-12	0.431	1.77	1.34	-0.233	1.58
Peter Holland	2016-17	-0.198	-1.36	-1.16	0.273	-1.44
Jack Hillen	2012-13	0.603	-1.24	-1.84	-0.412	-1.43
Andrej Meszaros	2011-12	0.320	-1.28	-1.60	-0.170	-1.43
Paul Bissonette	2011-12	-0.609	1.31	1.92	0.527	1.39
Alex Tanguay	2012-13	-0.123	1.43	1.56	0.173	1.38

Table 4.11: Ten least accurate predictions for the final model

players either having unexpectedly productive years or regressing after an unexpected productive year. Perhaps including career averages or multiple years of data could improve predictions in those cases.

4.4 Erroneous Data

The biggest surprise in compiling this project was the sheer volume of errors in the data set. The data errors came in two forms, missing information and false or incorrect entries. These issues mostly arose in the Fenwicka play-by-play data and existed in the NHL sources, making them more difficult to discover and evaluate. In the case of missing data, many events from the

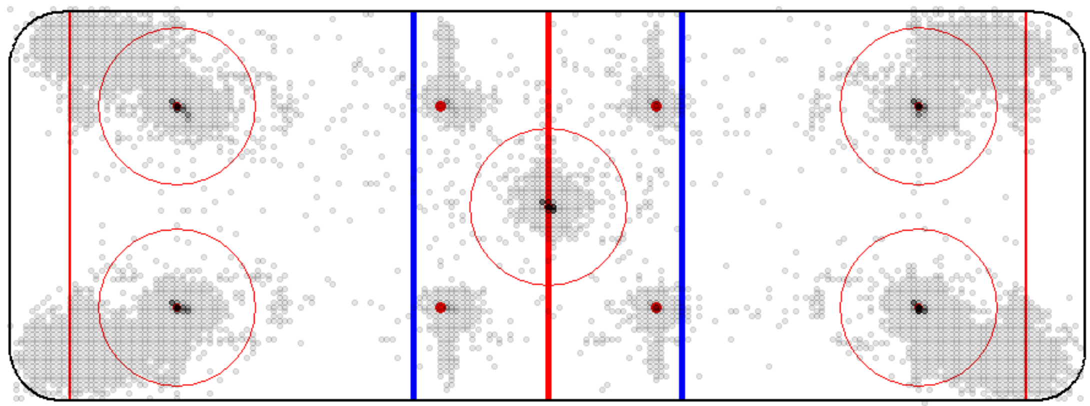

Figure 4.4: Plot of all faceoff locations recorded from the 2007-08 season to the 2017-18 season

2007-08 and 2008-09 seasons lack location data, which is a known issue. A more surprising issue were errors in the on-ice personnel data. Nearly 3,000 events, not including ON and OFF were listed with an invalid combination of players on the ice, meaning no players were given credit for the on-ice events. Additionally, these events contained about 500 minutes of playing time. These numbers are not a large set of the data set, but in at least two instances, entire periods or games have been affected, impacting a small number of players' numbers more directly than if the errors had been spread through the data set. There are also many events with missing actor information. Most egregiously, there are over 14,000 giveaway events with no credited player. These number are also the basis for many of the publicly available stat sites, and it is not clear how much effort was put into auditing the data before publishing it.

Location data is also known to be suspect, and work has been done to try to correct for this [29]. This makes sense; hockey is a very fast sport and the scorekeepers are asked to mark down where events occurred while also trying to keep track of many other things. However, it was surprising to discover over 4,000 different locations for faceoffs, as illustrated in Figure 4.4. Faceoffs can only be performed at one of the nine faceoff dots, which are fixed and constant from rink to rink. Furthermore, since faceoffs are the event which initiates play, the players are stationary at the start of a faceoff. It is difficult to have faith in the accuracy of the data for other events, beyond the zone they occurred in.

Finally, the most difficult data issue encountered on this project was player identification. The play-by-play events use a player's name, while the NHL.com API uses an ID number. Within the play-by-play events, 111 players appeared with two separate names, including some well-known players, such as Hall of Famers Mike Modano and Rob Blake appearing a few times as Michael Modano and Robert Blake. Matching the aliases to each other had to be done manually and was a fairly time intensive process. An additional three players appeared under three names in the data set. Player naming conventions were standard across sources either. The NHL Injury Viz database used TSN's names, for example, 21 of which did not have a match in the play-by-play, while Money Puck used its own naming convention. Finally, there were five cases of players having the exact same name. Fortunately, they never played for the same team at the same time, so it was possible to sort them out, but this was still time consuming and introduced errors later in the process that were disruptive and required repeating a lot of previous work.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

This project developed a model for predicting an individual player's contribution to their team. The first major lesson learned is the difficulty in predicting a single player's performance. Hockey is a highly random sport and there are lots of factors that contribute to a player's season totals that do not show up when aggregated. Adjusting for score and venue effects was effective in improving predictive power, but adjustments for rink and homer bias in subjective statistics was not.

There seemed to be a hard limit on the predictive ability for current metrics, which is consistent with predictive work done at the team level [45, 46, 47]. External factors only explain so much of the variance in the outcomes between different years. Playing style and usage were also largely irrelevant in projecting future performance at a high level. Including information from multiple years only significantly improved the overall accuracy of projections, as measured by RMSE, when additional information from the immediately prior season was not included. Age was the only significant factor in predicting future performance that was not derivable from on-ice play.

Perhaps the biggest takeaway from this project are the limitations of the data available. Some limitations were known beforehand, such as the lack of puck and player movement data and the secrecy surrounding injury data. Errors were also somewhat expected in fan-based data, since this data is largely compiled as a hobby. However, the extent of issues in the official league records was surprising for such a large industry. A large amount of time was spent examining mistakes in the data and trying to determine how to cope with them. Better auditing and validation from the NHL could improve the quality of future work.

5.2 Future Work

There are many possible avenues for future work in this area. One potentially major effect that was difficult to model, given the available information, was coaching. Coaches have impact beyond ice time distribution in determining the strategy and tactics of a team. Work involving rookies could also be improved. Rookies were excluded from this project due to not having a previous NHL season, but research on minor and foreign leagues could have a major impact on draft strategies. Conversely, more work can be done on age-related decline specifically. That is, projecting when a player's ability will drop off.

More work can also be done with data management, especially with cataloging errors. Although I noted many erroneous data points, there are over 13 million entries and many of them do not have easily identifiable errors. Further data consolidation is another possibility for future work. With the hockey analytics community growing, more data is being produced all the time and it is largely decentralized. Gathering it into a single source would improve future research by reducing the amount of time spent collecting, cleaning, and matching from different places.

Finally, the NHL has announced that it will be introducing tracking data in the 2019-20 season for players and the puck. This information can potentially provide much more insight into what a player is doing and what they are capable of that could massively improve both the measurement of individual contributions and the predictive ability of models for future performance. With even more information available, the possibilities are exciting.

REFERENCES

[1] NHL, "Historical Rule Changes," NHL, [Online]. Available: https://records.nhl.com/history/historical-rule-changes. [Accessed 2019].

[2] NHL, "Glossary," NHL, [Online]. Available: http://www.nhl.com/stats/glossary. [Accessed 2019].

[3] Pro Football Hall Of Fame, "NFL's Passer Rating," Pro Football Hall Of Fame, 2 January 2005. [Online]. Available: https://www.profootballhof.com/news/nfl-s-passer-rating/. [Accessed 2019].

[4] G. Easterbrook, "So What's the Patriots' Secret to Success?," New York Times, 27 October 2015. [Online]. Available: https://www.nytimes.com/2015/10/28/upshot/so-whats-the-patriots-secret-to-success.html. [Accessed 2019].

[5] dtmaboutheart [D. Sprigings], "Expected Goals are a better predictor of future scoring than Corsi, Goals," Hockey-Graphs, 1 October 2015. [Online]. Available: https://hockey-graphs.com/2015/10/01/expected-goals-are-a-better-predictor-of-future-scoring-than-corsi-goals/. [Accessed 16 September 2018].

[6] S. Highkin, "NBA's new tracking stats give fans a window to the future," USA Today, 1 November 2013. [Online]. Available: https://www.usatoday.com/story/sports/nba/2013/11/01/nba-analytics-sportvu-cameras-stats-chris-paul/3358057/. [Accessed 2019].

[7] MLB, "Statcast," MLB, [Online]. Available: http://m.mlb.com/glossary/statcast. [Accessed 2019].

[8] S. Walder, "Player tracking data is next step in NFL's analytics revolution," ESPN, 23 August 2018. [Online]. Available: http://www.espn.com/nfl/story/_/id/24445965/player-tracking-data-next-step-nfl-analytics-revolution. [Accessed 2019].

[9] S. Kasan, "Off-ice officials are a fourth team at every game," NHL, 24 October 2008. [Online]. Available: https://www.nhl.com/news/off-ice-officials-are-a-fourth-team-at-every-game/c-388400. [Accessed 2019].

[10] B. Adgate, "Hockey's Big Pay Day Is Coming," Forbes, 15 January 2018. [Online]. Available: https://www.forbes.com/sites/bradadgate/2019/01/15/hockeys-big-pay-day-is-coming/#620716014e44. [Accessed 2019].

[11] A. Novet, "Strong and Weak Links: Talent Distribution within Teams," Hockey-Graphs, 14 March 2017. [Online]. Available: https://hockey-graphs.com/2017/03/14/strong-and-weak-links-talent-distribution-within-teams/. [Accessed 21 August 2018].

[12] Baseball-Reference, "Win Shares," Sports Reference, 29 April 2007. [Online]. Available: https://www.baseball-reference.com/bullpen/Win_Shares. [Accessed 6 January 2019].

[13] Baseball-Reference, "Baseball-Reference.com WAR Explained," Sports Reference, March 2013. [Online]. Available: https://baseball-reference.com/about/war_explainer.shtml. [Accessed 6 January 2019].

[14] S. Slowinski, "What Is WAR?," FanGraphs, 15 February 2010. [Online]. Available: https://library.fangraphs.com/misc/war/. [Accessed 6 January 2019].

[15] C. Wyers, "Reworking WARP: The Series Ahead," Baseball Prospectus, 21 August 2013. [Online]. Available: https://www.baseballprospectus.com/news/article/21586/reworking-warp-the-series-ahead/. [Accessed 6 January 2019].

[16] B. S. Baumer, S. T. Jensen and G. J. Matthews, "openWAR: An Open Source System for Evaluating Overall Player Performance in Major League Baseball," *Journal of Quantitative Analysis in Sports*, vol. 11, no. 2, pp. 69-84, 2015.

[17] Basketball-Reference, "NBA Win Shares," Sports Reference, [Online]. Available: https://www.basketball-reference.com/about/ws.html.. [Accessed 7 January 2019].

[18] D. Myers, "About Box Plus/Minus (BPM)," Sports Reference, 7 January 2015. [Online]. Available: https://www.basketball-reference.com/about/bpm.html. [Accessed 7 January 2019].

[19] D. T. Rosenbaum, "Measuring How NBA Players Help Their Teams Win," 82Games, 30 April 2004. [Online]. Available: http://www.82games.com/comm30.htm. [Accessed 7 January 2019].

[20] S. Ilardi and A. Barzalai, "Adjusted Plus-Minus Ratings: New and Improved for 2007-2008," 82Games, 2008. [Online]. Available: http://www.82games.com/ilardi2.htm. [Accessed 7 January 2019].

[21] J. Sill, "Improved NBA Adjusted +/- Using Regularization and Out-Of-Sample Testing," in *2010 MIT Sloan Sports Analytics Conference*, Boston, 2010.

[22] G. Hohl, "Behind the Numbers: Why Plus/Minus Is the Worst Statistic in Hockey and Should Be Abolished," Hockey-Graphs, 1 November 2016. [Online]. Available: https://hockey-graphs.com/2016/11/01/behind-the-numbers-why-plusminus-is-the-worst-statistic-in-hockey-and-should-be-abolished/. [Accessed 23 August 2018].

[23] A. Ryder, "Player Contribution a Method for Allocating Credit for a Team’s Performance to the Individual Contributors on a Hockey Team," 2003. [Online]. Available:

http://www.hockeyanalytics.com/Research_files/Player_Contribution_System.pdf. [Accessed 27 August 2018].

[24] J. Kubatko, "Calculating Point Shares," Sports Reference, [Online]. Available: https://www.hockey-reference.com/about/point_shares.html. [Accessed 27 August 2018].

[25] C. Turtoro, "Catch-All Statistics Part I: GVT versus Point Shares," All About the Jersey, 8 August 2014. [Online]. Available: https://www.allaboutthejersey.com/2014/8/8/5978425/catch-all-statistics-part-i-gvt-versus-point-shares. [Accessed 27 August 2018].

[26] ogmitsdomi [D. Luszczyszyn], "Measuring Single Game Productivity: An Introduction To Game Score," Hockey-Graphs, 13 July 2016. [Online]. Available: https://hockey-graphs.com/2016/07/13/measuring-single-game-productivity-an-introduction-to-game-score/. [Accessed 29 August 2018].

[27] B. Macdonald, "Adjusted Plus-Minus for NHL Players Using Ridge Regression with Goals, Shots, Fenwick, and Corsi," *Journal of Quantitative Analysis in Sports,* vol. 8, no. 3, 2012.

[28] EvolvingWild [J. Younggren and L. Younggren], "Reviving Regularized Adjusted Plus-Minus for Hockey," Hockey-Graphs, 14 January 2019. [Online]. Available: https://hockey-graphs.com/2019/01/14/reviving-regularized-adjusted-plus-minus-for-hockey/. [Accessed 14 January 2019].

[29] M. Schuckers and J. Curro, "Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events," in *7th Annual MIT Sloan Sports Analytics Conference*, Boston, 2013.

[30] @acthomasca [A. C. Thomas], "The Road to WAR (for Hockey), Part 1: The Single-Number Dream," War-On-Ice, 13 August 2014. [Online]. Available: http://blog.war-on-ice.com/index.html%3Fp=37.html. [Accessed 10 November 2018].

[31] EvolvingWild [J. Younggren and L. Younggren], "Wins Above Replacement: History, Philosophy, and Objectives (Part 1)," Hockey-Graphs, 16 January 2019. [Online]. Available: https://hockey-graphs.com/2019/01/16/wins-above-replacement-history-philosophy-and-objectives-part-1/. [Accessed 16 January 2019].

[32] E. Perry, "Wins Above Replacement," Corsica Hockey, 2018. [Online]. Available: http://corsica.hockey/misc/war_notebook.html. [Accessed 24 August 2018].

[33] M. B. McCurdy, "The Magnus Prediction Models," Hockey Viz, 24 September 2018. [Online]. Available: https://hockeyviz.com/txt/magnusEV. [Accessed 24 September 2018].

[34] EvolvingWild [J. Younggren and L. Younggren], "Wins Above Replacement: The Process (Part 2)," Hockey-Graphs, 17 January 2019. [Online]. [Accessed 17 January 2019].

[35] M. Schuckers and B. Macdonal, "Accounting for Rink Effects in the National Hockey League's Real Time Scoring System," *arXiv preprint arXiv:1412.1035,* 2012.

[36] Eric T. [E. Tulsky], "Adjusting for Score Effects to Improve Our Predictions," Broad Street Hockey, 23 January 2012. [Online]. Available: https://www.broadstreethockey.com/2012/1/23/2722089/score-adjusted-fenwick. [Accessed 16 October 2018].

[37] M. B. McCurdy, "Better Way to Compute Score-Adjusted Fenwick," Hockey Viz, October 2014. [Online]. Available: https://hockeyviz.com/txt/senstats. [Accessed 17 November 2018].

[38] money puck [pseud.], "Big Deals for Big Hitters: How Physical Players Age," Hockey-Graphs, 25 November 2015. [Online]. Available: https://hockey-graphs.com/2015/11/25/big-deals-for-big-hitters-how-physical-players-age/. [Accessed 3 September 2018].

[39] A. Novet, "Neutral Zone Playing Styles," Hockey-Graphs, 6 July 2016. [Online]. Available: https://hockey-graphs.com/2016/07/06/neutral-zone-playing-styles/. [Accessed 9 October 2018].

[40] R. Stimson, "Identifying Playing Styles With Clustering," Hockey-Graphs, 4 April 2017. [Online]. Available: https://hockey-graphs.com/2017/04/04/identifying-player-types-with-clustering/. [Accessed 29 August 2018].

[41] A. Novet, "Goal Scorer Cluster Analysis," Hockey-Graphs, 4 January 2018. [Online]. Available: https://hockey-graphs.com/2018/01/04/goal-scorer-cluster-analysis/. [Accessed 29 August 2018].

[42] Tangotiger [T. Tango], "Marcel 2012," 2012. [Online]. Available: http://tangotiger.net/marcel/. [Accessed 28 August 2018].

[43] Baseball Prospectus, "Glossary: PECOTA," Baseball Prospectus, [Online]. Available: https://legacy.baseballprospectus.com/glossary/index.php?context=6&category=true. [Accessed 28 August 2018].

[44] N. Silver, "We’re Predicting The Career Of Every NBA Player. Here’s How," FiveThirtyEight, 9 October 2015. [Online]. Available: https://fivethirtyeight.com/features/how-were-predicting-nba-player-career/. [Accessed 24 August 2018].

[45] M. J. Lopez, G. J. Matthews and B. S. Baumer, "How often does the best team win? A unified approach to understanding randomness in North American sport," *The Annals of Applied Statistics,* vol. 12, no. 4, pp. 2483-2516, 2018.

[46] Josh W. [Joshua Weissbock], "Machine Learning and Hockey: Is there a theoretical limit on predictions?," Arctic Ice Hockey, 1 August 2013. [Online]. Available: https://www.nhlnumbers.com/2013/8/1/machine-learning-and-hockey-is-there-a-theoretical-limit-on-predictions. [Accessed 20 August 2018].

[47] E. Perry, "On Salad And Predicting Hockey Games," Corsica Hockey, 6 October 2017. [Online]. Available: http://www.corsica.hockey/blog/2017/10/06/on-salad-and-predicting-hockey-games/. [Accessed 20 August 2018].

[48] EvolvingWild [J. Younggren and L. Younggren], "A New Look at Aging Curves for NHL Skaters (part 1)," Hockey-Graphs, 23 March 2017. [Online]. Available: https://hockey-graphs.com/2017/03/23/a-new-look-at-aging-curves-for-nhl-skaters-part-1/. [Accessed 30 August 2018].

[49] Mimicohero [D. Galamini Jr.], "Using Marcels to Forecast Player Performance in Hockey," Own the Puck, 5 September 2015. [Online]. Available: https://ownthepuck.wordpress.com/2015/09/05/using-marcels-to-forecast-player-performance-in-hockey/. [Accessed 28 August 2018].

[50] garik16 [pseud.], "2015 Midseason Goalie Projections Using Hockey Marcels," Hockey-Graphs, 2 February 2015. [Online]. Available: https://hockey-graphs.com/2015/02/02/2015-midseason-goalie-projections-using-hockey-marcels/. [Accessed 28 August 2018].

[51] D. Luszczyszyn, "NHL Season Preview: Projections FAQ," The Athletic, 31 August 2017. [Online]. Available: https://theathletic.com/90596/2017/08/31/nhl-season-previews-preview-projections-faq/. [Accessed 20 August 2018].

[52] loserpoints [pseud.], "Estimating Shot Assist Quantities for Skaters," Hockey-Graphs, 12 December 2017. [Online]. Available: https://hockey-graphs.com/2017/12/12/estimating-shot-assists/. [Accessed 30 August 2018].

[53] C. Rush and I. Khurshudyan, The Washington Post, 17 January 2018. [Online]. Available: https://www.washingtonpost.com/sports/some-nhl-teams-hide-injury-information-but-it-might-be-hurting-players/2018/01/17/373a98f0-fb07-11e7-ad8c-ecbb62019393_story.html?noredirect=on&utm_term=.f0e4447e4cc1. [Accessed 17 January 2018].

[54] S. Slowinski, "Sample Size," FanGraphs, 18 February 2010. [Online]. Available: https://library.fangraphs.com/principles/sample-size/. [Accessed 25 August 2018].

[55] plusandminus, "How tiring was/is it to play ES, PP, PK?," HFBoards [forum], 12 September 2011. [Online]. Available: https://hfboards.mandatory.com/threads/how-tiring-was-is-it-to-play-es-pp-pk.982883/. [Accessed 2019].

[56] M. B. McCurdy, "Shot Generation and Suppression are Mostly Independent," Hockey Viz, 15 September 2015. [Online]. Available: https://hockeyviz.com/txt/indepGenSupp. [Accessed 13 January 2019].

[57] piyer97 [pseud.], JJfromkansas [pseud.], P. Iyer, Asmean [pseud.], "Assessing the Impact of Deployment on Individual Shorthanded Shot Suppression," Winging It In Motown, 16 June 2016. [Online]. Available:

https://www.wingingitinmotown.com/2016/6/16/11923116/assessing-the-impact-of-deployment-on-individual-shorthanded-shot. [Accessed 12 Decemer 2018].

[58] E. Tulsky, "Quantifying the Added Importance of Recent Data," SBNation, 21 January 2014. [Online]. Available: https://www.sbnation.com/nhl/2014/1/21/5329992/nhl-stats-projections-data. [Accessed 24 August 2018].