Patrick Lambrix
Niklas Carlsson
Mikael Vernblom (Eds.)

Linköping Hockey Analytics Conference - LINHAC 2022

June 6-8, 2022, Linköping, Sweden

Proceedings of the Linköping Hockey Analytics Conference LINHAC 2022 Research Track

LINKÖPING UNIVERSITY

LINKÖPING

Organizing Committee

Patrick Lambrix, Linköping University, Sweden (chair)
Niclas Carlsson, Linköping University, Sweden (co-chair)
Mikael Vernblom, Linköping Hockey Club, Sweden (co-chair)

Nanna Holm Glaas, Linköping Hockey Club, Sweden (conceptualization)
Kim Martin Hasson, Linköping Hockey Club, Sweden (conceptualization)

Program Committee for the Research Track

Niklas Carlsson, Linköping University, Sweden (chair)
Martin Arlitt, University of Calgary, Canada
Micah Blake McCurdy, Hockeyviz.com, Canada
Tim Brecht, University of Waterloo, Canada
Meghan Chayka, Stathletes, Canada
David A Clausi, University of Waterloo, Canada
Lars Magnus Hvattum, University College Molde, Norway
Ulf Johansson, Jönköping University, Sweden
Patrick Lambrix, Linköping University, Sweden
Erik Lignell, Frölunda HC, Sweden
David Thomas Radke, University of Waterloo, Canada
Michael Schuckers, Statistical Sports Consulting and St. Lawrence University, USA
Oliver Schulte, Simon Fraser University, Canada
Mikael Svarén, Dalarna University, Sweden
Sean Tierney, Sportlogiq, Canada
Sam Ventura, Buffalo Sabres and Carnegie Mellon University, USA
Erik Wilderoth, Färjestad BK, Sweden

Editors: Patrick Lambrix, Niklas Carlsson, and Mikael Vernblom
Linköping Electronic Conference Proceedings 191
eISSN: 1650-3740
ISSN: 1650-3686
ISBN: 978-91-7929-492-2
DOI: https://doi.org/10.3384/ecp191

Gold Sponsors

Stretch on Sense AB

The Swedish Research Council for Sport Science

In cooperation with

The City of Linköping

Alliance of European Hockey Clubs

Preface

LINHAC 2022 took place from June 6-8, 2022, and was organized by Linköping University (Patrick Lambrix and Niklas Carlsson) and Linköping Hockey Club (Mikael Vernblom). LINHAC brought together professionals and academics with an interest in hockey analytics. It featured the latest research in hockey analytics in academia and companies, discussions with analysts and coaches, industry sessions with the latest hockey analytics products, and an analytics competition for students.

The program included invited research talks by Oliver Schulte, Tim Brecht, Carleen Markey, and Patrick Lambrix and Niklas Carlsson. Further, seven papers were selected for presentation at LINHAC. These research track proceedings contain these research papers as well as abstracts and papers related to the invited research talks.

In addition to the research track, Dan Tagnes, head coach of EV Zug, who won his second consecutive title in Switzerland, talked about the use of data and analytics from a coach perspective. Sean Tierney from Sportlogiq talked about the state of the art in hockey analytics from an industry perspective.

Further, there were four panel discussions moderated by Mike Helber. The first panel was made up of analysts from different SHL teams (Petter Carnbro from Leksands IF, Patrik Hall from Växjö Lakers, Erik Lignell from Frölunda Hockey Club, and Erik Wilderoth from Färjestad BK) as well as a representative from Sportloqiq (Sean Tierney). The second panel with Adam Albelin (Swedish Ice Hockey Association), Meghan Chayka (Stathletes), Erika Holst (Frölunda Hockey Club) and Carleen Markey (Carnegie Mellon University), discussed the state of the art and future of analytics for women's hockey. Thorsten Apel (Sportcontract), Martin Rumo (OYM), Sean Tierney (Sportlogiq), and Morgan Zeba (Spiideo) discussed hockey analytics from the industry perspective. In the final panel representatives from the entertainment industry discussed the use of analytics in TV broadcasts. The panel members were Mike Kelly (NHL Network), Alison Lukan (ROOT Sports NW and Seattle Kraken), Björn Oldeen (CMORE) and Håkan Södergren (Viaplay).

There were several discussion sessions. Mikael Verblom led a discussion on analytics for goaltenders with Justin Goldman (The Goalie Guild), Thomas Magnusson (Swedish Ice Hockey Association) and Jonas Gustavsson (former NHL and SHL goaltender). Mike Helber discussed with Karl Schwarzenbrunner from the German Ice Hockey Association about knowledge transfer and coaching the coaches. Adam Albelin, Adam Almqvist Andersson, Mikael Vernblom and Matheus Vieweg, coaches on different levels of the Swedish national teams, discussed the use of hockey analytics in their jobs.

Several companies presented their products: PwC Hungary - Sports Advisory, Spiideo, Sportcontract, Sportradar, Stathletes, Stretch On Sense AB, and Wisehockey.

Finally, there was a student competition where the task was to provide insights based on sequences of events in a hockey game. Data was provided by the SHL and Sportlogiq.

This conference was the first in its kind in Europe and as far as we know the first hockey analytics conference that dealt with all aspects related to hockey analytics.

We thank our collaborators the Alliance of European Hockey Clubs and the City of Linköping, as well as sponsors the Swedish Research Council for Sport Science and Stretch on Sense AB.

June 2022

Patrick Lambrix,
Niklas Carlsson,
Mikael Vernblom

Contents

Invited talks

Valuing Actions and Ranking Hockey Players With Machine Learning (Extended Abstract)

Oliver Schulte
School of Computing Science
Simon Fraser University
Burnaby, Canada

Linkökping Hockey Analytics Conference
May 2022

Abstract

A fundamental goal of sports analytics is to rank player performance. A common approach is to assign a value to each player action and rank a player by their aggregate action value. A recent AI-based approach is to measure the value of a player's action by how much it increases their team's chance of success, that is, their team's chance of scoring the next goal. This requires a model that outputs a success probability estimate, given a match context and an action. This talk describes machine learning techniques for building success probability models from data. The techniques range from easy-to-implement probabilistic classifiers to advanced reinforcement learning methods. The results of success probability models are illustrated with action values and player rankings for the National Hockey League.

1 Introduction: Success Probabilities in Sports Analytics

During a match, each action by a sports team is directed towards maximizing the chance of future success. Therefore the *probability of (future) success* is a key statistical quantity for evaluating the strength of a team, the impact of an action, and the contributions of a player. This note accompanies a talk that shows how success probabilities can be used to value actions and ranking players, and describes techniques for estimating them from data.

2 Defining Success Probabilities

For the purposes of this note, success is flexibly defined as a binary event that a team seeks to bring about. An analyst is free to define success in different ways depending on the question they are investigating. Examples of success concepts that have appeared in the literature include the following (using hockey examples where possible).

- Winning the Match [Pettigrew, 2015].
- Scoring a goal within a short time interval (e.g. 1 minutes) [Schuckers and Curro, 2013].
- Scoring a goal within the next 5 actions [Decroos et al., 2019].
- Scoring the next goal in the match [Routley and Schulte, 2015].
- Drawing the next penalty [Routley and Schulte, 2015]. This is a failure event that would be interesting to a coach who is concerned to minimize the number of penalties incurred by their team.

The notation

$$P(S_i|\mathbf{X}_t)$$

denotes the probability that i achieves (future) success given the current match context $\mathbf{X}_t$. We discuss in Section 5.1 below how a context vector can be computed from play-by-play data.

A success probability is a dynamic quantity; a success probability ticker shows an estimated probability for each time in a match [Liu and Schulte, 2018]; see Figure 1.

3 From Success Probabilities to Action Values

From success probabilities we can assign a value to actions called the **impact** of an action occurring at time $t+1$ [Liu and Schulte, 2018].

$$impact_i(t+1) \equiv P(S_i|\mathbf{X}_{t+1}) - P(S_i|\mathbf{X}_t)$$

Thus the impact of an action is the difference in success probabilities before and after the action occurred. Figure 3 shows boxplots for impact values. Note that impact values can vary widely for the same action, depending on context.

4 From Action Values to Player Ranking

We can compute a player performance metric from impact values in a straightforward way: for each player, and for each of their actions, we can compute the impact of the action. The **goal impact metric** (GIM) is simply the total impact of the player's actions.

Our metric can be used to *identify undervalued players*. For instance, Johnny Gaudreau and Mark Scheifele drew salaries below what their GIM rank would suggest. Later they received a $\$5M+$ contract for the 2016-17 season.

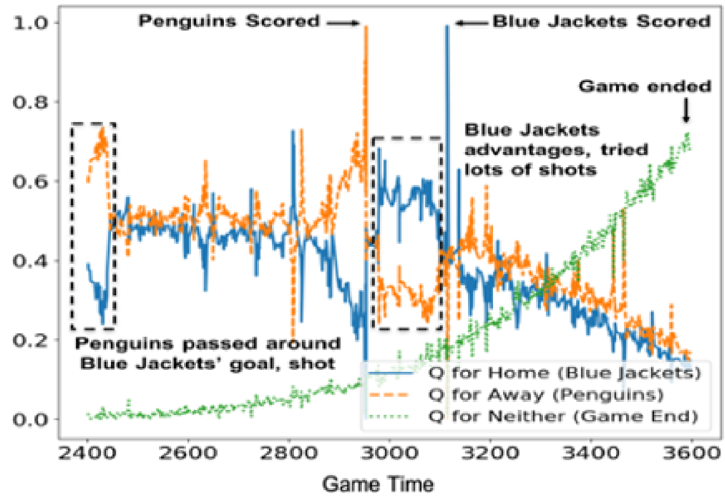


Figure 1: A success probability ticker for a match between the Penguins and the Blue Jacket. The y-axis shows the estimated probability of scoring the next goal.

While we do not have ground truth for evaluating player rankings, the goal impact player rankings have been validated indirectly in several ways.

1. GIM correlates well with standard success metrics (e.g., Points) [Liu and Schulte, 2018].
2. GIM converges close to half-way through the season. This means that the beginning of the season can be used to evaluate player strength (predict the player's final ranking).
3. The GIM values per player correlate well across different seasons [Routley, 2015, Pettigrew, 2015], which is evidence that they measure a stable quality of players.

The impact metric approach has also been validated in other sports, such as soccer [Decroos et al., 2019, Liu et al., 2020a, Fernández et al., 2021] and basketball [Cervone et al., 2014].

5 Learning Success Probability Models

Given the usefulness of success probability models, a major concern of machine learning for sports analytics is to develop machines for building such models from data. In the following I will discuss methods for building success probability models from event data, also known as play-by-play data. Estimating success probabilities from tracking data is less studied because tracking data is less

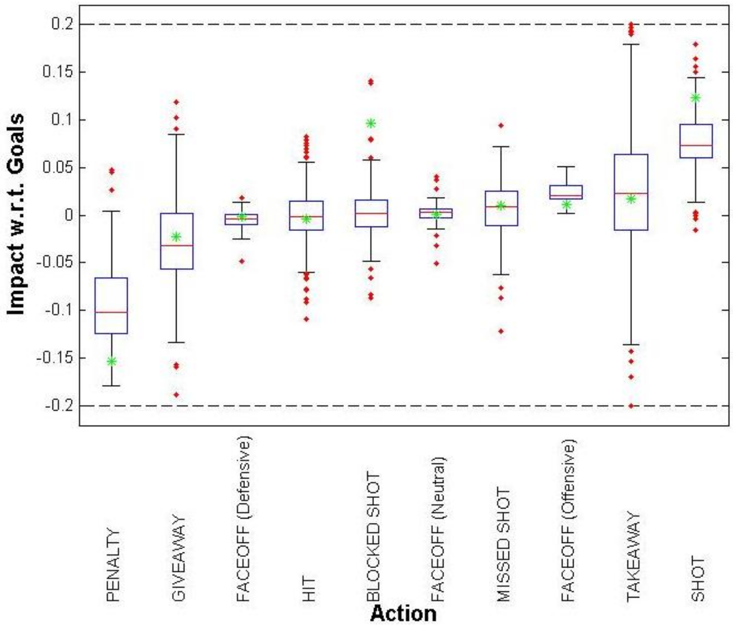


Figure 2: Impact on the probability of Scoring the Next Goal. Higher numbers are *better* for the team that performs the action. Action Impact Values vary with context. The central mark is the median, the edges of the box are the 25th and 75th percentiles. The whiskers are at the default value, approximately 2.7 s.d. Based on the model of [Routley and Schulte, 2015].

commonly available (but see [Dick and Brefeld, 2019, Fernández et al., 2021]). Table 2 illustrates play-by-play data.

5.1 Classifier Approach

A straightforward approach is to annotate each event at time t with a binary target $Y_{i,t} \in \{0,1\}$ that denotes whether team i acting at time t achieved future success after time t. For example in the game of Figure 1, the Penguins scored around time $t = 3,900$ sec. So for all previous times $2,400 < t' < 3,900$, we would have $Y_{Penguins,t'} = 1$ and $Y_{Flyers,t'} = 0$. Then estimating success probabilities can be modelled as predicting a binary label given the information $\mathbf{X}_t$ available at time t.

It is straightforward to include in the context vector $\mathbf{X}_t$ values for time indexed features score differential, manpower differential, time remaining, location etc. [Routley and Schulte, 2015, Liu et al., 2018]. The main difficulty is how to include the match history prior to time t. A simple approach is to fix a window size k and then to append to $\mathbf{X}_t$ the time-indexed features for the previous times $t-1, \ldots, t-k$. See [Decroos et al., 2019] for a model of this approach applied

Table 1: 2015-2016 Top-20 Player Impact Scores. Based on the model of [Liu and Schulte, 2018].

Name	Impact	Assists	Goals	Points	+/-	Salary
Taylor Hall	96.40	39	26	65	-4	$6,000,000
Joe Pavelski	94.56	40	38	78	25	$6,000,000
Johnny Gaudreau	94.51	48	30	78	4	$925,000
Anze Kopitar	94.10	49	25	74	34	$7,700,000
Erik Karlsson	92.41	66	16	82	-2	$7,000,000
Patrice Bergeron	92.06	36	32	68	12	$8,750,000
Mark Scheifele	90.67	32	29	61	16	$832,500
Sidney Crosby	90.21	49	36	85	19	$12,000,000
Claude Giroux	89.64	45	22	67	-8	$9,000,000
Dustin Byfuglien	89.46	34	19	53	4	$6,000,000
Jamie Benn	88.38	48	41	89	7	$5,750,000
Patrick Kane	87.81	60	46	106	17	$13,800,000
Mark Stone	86.42	38	23	61	-4	$2,250,000
Blake Wheeler	85.83	52	26	78	8	$5,800,000
Tyler Toffoli	83.25	27	31	58	35	$2,600,000
Charlie Coyle	81.50	21	21	42	1	$1,900,000
Tyson Barrie	81.46	36	13	49	-16	$3,200,000
Jonathan Toews	80.92	30	28	58	16	$13,800,000
Sean Monahan	80.92	36	27	63	-6	$925,000
Vladimir Tarasenko	80.68	34	40	74	7	$8,000,000

to soccer (they used $k = 3$ as the window size). After extracting the window information as context, the data will be a list of $\langle \mathbf{X}_t, Y_{i,t} \rangle$ pairs, which is the standard format for any classifier package available in systems like R, Weka, scikit-learn.

An alternative to using a fixed window size is to apply a recurrent neural network, which can take as input a sequence without the need for preprocessing.

5.2 Reinforcement Learning

Reinforcement learning (RL) is the branch of machine learning that studies learning to act [Sutton and McCallum, 2007]. Estimating success probabilities from sequential data is one of the basic well-studied problems in RL. In RL, a mapping from match states to success probabilities is known as a **value function** and estimating a value function is called the **prediction problem**. The classifier approach described in the previous subsection (implicitly) treats all match states as independent, and hence ignores the correlations between success probabilities due to the temporal dynamics of ice hockey. In contrast, reinforcement learning seeks to exploit the temporal dynamics to efficiently learn success probabilities.

If we discretize the spatial rink coordinates, we can model hockey dynamics

Table 2: Sample Play-By-Play Data in Tabular Format.

gameId	playerId	period	teamId	xCoord	yCoord	Manpower	Action Type
849	402	1	15	-9.5	1.5	even	lpr
849	402	1	15	-24.5	-17	even	carry
849	417	1	16	-75.5	-21.5	even	check
849	402	1	15	-79	-19.5	even	puckprot.
849	413	1	16	-92	-32.5	even	lpr
849	413	1	16	-92	-32.5	even	pass
849	389	1	15	-70	42	even	block
849	389	1	15	-70	42	even	lpr
849	389	1	15	-70	42	even	pass
849	425	1	16	-91	34	even	block
849	395	1	15	-97	23.5	even	reception

in a framework known as a **discrete Markov decision process** [Routley and Schulte, 2015, Schulte et al., 2017a,b]. The key parameters in a Markov decision process are *state transition probabilities* that describe what is likely to happen next in a hockey game. Given an estimate of state transition problems, the dynamic programming algorithm can be used to compute success probabilities for any match state.

While discretization can simplify learning and in many cases increases the interpretability of success probabilities, it also loses information and introduces unnatural discontinuities in a success probability model. Reinforcement learning provides so-called model-free methods for learning success probabilities that do not require discrete state transition probabilities. Combining model-free methods with neural networks provides a method for learning success probabilities that can take as input continuous spatio-temporal data “as is” without the need for discretization or fixing a window size. Model-free deep RL has been developed in several recent approaches for sports dynamics [Liu et al., 2018, 2020b,a]. Figure 3 summarizes the options for learning success probabilities discussed.

6 Conclusion

Estimating success probabilities is a basic statistical problem in hockey analytics. A good success probability model can be leveraged to solve important analytics problems such as quantifying the value of an action and the contributions of a player. Machine learning models can include rich match contexts to provide useful success probabilities. Probabilistic classifiers based on a sliding window are relatively straightforward to implement and can serve as a strong baseline for evaluating the usefulness of success probabilities in an application. Reinforcement learning is especially suitable for handing complex dynamic domains like ice hockey and provides a powerful set of tools for increasing the complexity and accuracy of a hockey model.

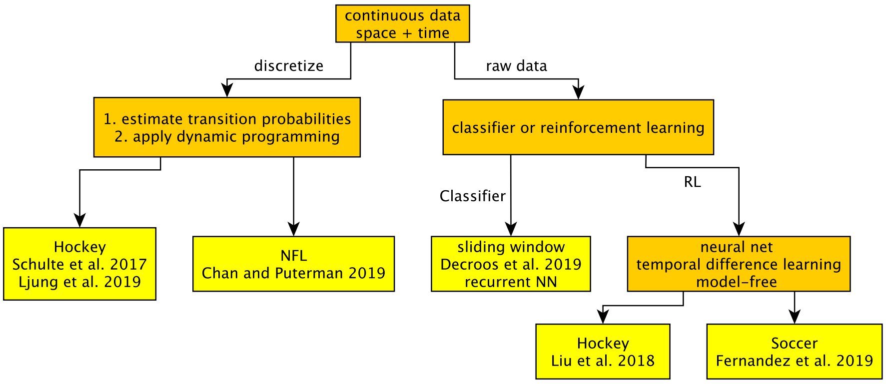


Figure 3: Approaches for Learning Success Probabilities

References

Stephen Pettigrew. Assessing the offensive productivity of nhl players using in-game win probabilities. In *9th Annual MIT Sloan Sports Analytics Conference*, 2015.

Michael Schuckers and James Curro. Total hockey rating (thor): A comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events. In *MIT sloan Sports Analytics Conference*, 2013.

Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th International Conference on Knowledge Discovery & Data Mining (KDD-19)*, pages 1851–1861, 2019.

Kurt Routley and Oliver Schulte. A markov game model for valuing player actions in ice hockey. In *Uncertainty in Artificial Intelligence (UAI)*, pages 782–791, 2015.

Guiliang Liu and Oliver Schulte. Deep reinforcement learning in ice hockey for context-aware player evaluation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 3442–3448. International Joint Conferences on Artificial Intelligence Organization, 7 2018. doi: 10.24963/ijcai.2018/478. URL https://doi.org/10.24963/ijcai.2018/478.

Kurt Routley. A markov game model for valuing player actions in ice hockey. Master's thesis, Simon Fraser University, April 2015.

Guiliang Liu, Yudong Luo, Oliver Schulte, and Tarak Kharrat. Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Min. Knowl. Discov.*, 34(5):1531–1559, 2020a.

doi: 10.1007/s10618-020-00705-9. URL https://doi.org/10.1007/s10618-020-00705-9.

Javier Fernández, Luke Bornn, and Daniel Cervone. A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, 110(6):1389–1427, 2021.

Dan Cervone, Alexander D'Amour, Luke Bornn, and Kirk Goldsberry. Point-wise: Predicting points and valuing decisions in real time with NBA optical tracking data. In *8th Annual MIT Sloan Sports Analytics Conference, February*, volume 28, 2014.

Uwe Dick and Ulf Brefeld. Learning to rate player positioning in soccer. *Big data*, 7(1):71–82, 2019.

Guiliang Liu, Oliver Schulte, Wang Zhu, and Qingcan Li. Toward interpretable deep reinforcement learning with linear model u-trees. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML)*, volume abs/1807.05887, pages 1–16, 2018. URL http://arxiv.org/abs/1807.05887.

Charles Sutton and Andrew McCallum. An introduction to conditional random fields for relational learning. In *Introduction to Statistical Relational Learning*, chapter 4, pages 93–127. MIT Press, 2007.

Oliver Schulte, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. Apples-to-apples: Clustering and ranking nhl players using location information and scoring impact. In *MIT Sloan Sports Analytics Conference*, 2017a.

Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. A Markov game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, pages 1–23, 2017b. ISSN 1573-756X. doi: 10.1007/s10618-017-0496-z. URL http://dx.doi.org/10.1007/s10618-017-0496-z.

Guiliang Liu, Oliver Schulte, Pascal Poupart, Mike Rudd, and Mehrsan Javan. Learning agent representations for ice hockey. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2020, December 6-12 2020*, 2020b.

Puck and Player Tracking: Challenges and Opportunities

Tim Brecht

Cheriton School of Computer Science
University of Waterloo, Waterloo, ON, Canada

Abstract. The National Hockey League (NHL) is using a puck and player tracking (PPT) system that records the location of the puck and players during games. Data is recorded 12 times per second for each player on the ice and 60 times per second for the puck. There are tremendous opportunities for the use of this data, including the development of new metrics that can be used for a variety of purposes. However, there are also significant challenges that need to be overcome. In this talk I first describe several such challenges and opportunities. I then focus on one of the opportunities we have been exploring, which is to develop several passing metrics. I briefly describe some of these metrics, the intuition behind them and outline some possible uses for a few metrics.
This is talk is based on joint work with David Radke, Daniel Radke and Alex Pawelczyk.

Performance Metrics for Ice Hockey Accounting for Goal Importance

Patrick Lambrix   Niklas Carlsson

Linköping University, Sweden

Abstract. The evaluation of player performance is an important topic in sports analytics and is used by coaches for team management, in scouting and in sports broadcasts. When evaluating the performance of ice hockey players, many metrics are used, including traditional metrics such as goals, assists, points and modern metrics such as Corsi. One weakness of such metrics is that they do not consider the context in which the value for the metric was assigned. Other advanced metrics have been introduced, but as they are not easily explainable to practitioners, they may not make it into the hockey discourse.

In this paper we introduce new goal-based metrics that (i) are based on traditional and well-known metrics, and thus easily understandable, (ii) take context into account in the form of time, manpower differential and goal differential and (iii) add a new aspect by taking into account the importance of the goals regarding their contribution to team wins and ties. We describe the intuitions behind the metrics, give formal definitions, evaluate the metrics using the eye test and show correlations to the traditional metrics. We have used data from the NHL seasons 2007-2008 to 2013-2014 and show which players stand out with respect to the number of goals and the importance of goals.[1]

1 Introduction

When evaluating the performance of ice hockey players, it is most common to use metrics that attribute a value to the actions the player performs (e.g., scoring a goal for the goals metric or giving a pass that leads to a goal for the assists metric) and then compute a sum over all those actions. Some extensions to these traditional metrics have been proposed, e.g., for the +/- metric [7, 1]. There is also work on combining metrics such as in [2]. Some of the approaches for player performance metrics take game context into account such as event impacts [11]. Other works model the dynamics of an ice hockey game using Markov games where two opposing sides (e.g., the home team and the away team) try to reach states in which they are rewarded (e.g., scoring a goal) [14, 3, 9, 12, 13, 5, 10, 6]. One critique of these more advanced metrics is that they are not easily understandable by or explainable to practitioners such as coaches, players and GMs. An approach to predict the tier (e.g., top 10%, 25% or 50%) to which a player belongs is presented in [4].

[1] This paper is a revised and extended version of [15].

Although some metrics take context into account for goals, e.g., the location of the shot, few take into account the importance of goals. For instance, a goal scored when the team is in the lead with 5–0 at the end of the game is most likely not crucial for winning. In contrast, scoring a goal when the score is tied at 1–1 with some seconds left of the game is of more importance for winning. Furthermore, some players have a reputation to often make important goals, while others may have the reputation to mainly score when the team is playing 'easier' games. For instance, during the 2013-2014 season the Washington Capitals' Alexander Ovechkin ranked the highest regarding game-tying and lead-taking goals while he only ranked 29^{th} regarding goals scored when the team is already in the lead. The importance of goals was taking into account in the added goal value metric in [8].

In this paper, our aim is to introduce new goal-based metrics for evaluating the performance of players. The metrics should take into account the importance of the goals in the sense of having important contributions to winning or tying games. Further, the metrics should be easily understandable and based on well-known traditional metrics. To achieve these goals, we introduce variants of the traditional goals, points[2], assists and +/- metrics that take into account the importance of the goals. By accounting for the importance of each goal, compared to these traditional metrics, our metrics better capture how much each player's goals, assists, or on-ice presence may have contributed to a positive game outcome (e.g., by scoring game deciding goals) and give less weight to players that score most of their goals when the outcome of a game may already be decided.

2 Defining a metric

When defining a metric, several questions must be addressed. First, there are some questions regarding the purpose of the metric and its definition.

- *What are the intuitions behind the metric?* It is important to know why a new metric is introduced. Usually, interesting observations regarding the game, that are not addressed by existing metrics, lie at the base of introducing new metrics. Therefore, a new metric should measure something that is not already measured by other metrics.
- *How is the metric defined?* Once the intuitions and purpose of the new metric are clear, a formal definition of the metric is needed that allows us to compute the values for the metric.

Further, we need to evaluate the metric. This is not a simple task as we usually do not have a gold standard against which to evaluate. Therefore, the metric's behavior is usually considered from different points of view, including

[2] Defined as the number of goals plus the number of assists for the player and often denoted by P or TP. In this paper, we also use the points that a team receives for a win or a tie, which are used to produce a ranking of the teams, often denoted by PTS. To avoid confusion, we call this latter kind of points 'game points'.

passing the eye test, finding correlations with existing metrics and looking at a metric over different seasons.

– *Does the metric pass the eye test?* Although there is no gold standard, based on the intuitions behind the metric, experts may expect a certain ranking of the players based on the new metric. The eye test checks whether the actual ranking according to the new metric makes sense according to the expectations of the experts.
– *Are there correlations with existing metrics?* A perfect correlation to existing metrics would mean that these metrics essentially measure the same thing. This could be interesting as an insight or in the case that it is easier to measure the new metric than existing metrics. However, as the intuitions behind the new metric usually deal with aspects that were not taken into account by existing metrics, there will not be a perfect correlation and this is what we would want. However, it is still interesting to check the correlation between the new metric and well-established metrics. A high correlation would show that the metric behaves in a similar way to a well-established metric, but still brings something new.
– *Is the metric stable?* The values for metrics will differ from each other over different seasons. However, unless good reasons, they should not change too drastically.

Finally, it is interesting to look at whether the value of the metric can be predicted.

– *Can one predict the value of the metric at the end of a season based on data for part of the season?* For some traditional metrics the value of a metric after half of the season gives a good indication of the value at the end of the season. Therefore, it is interesting to check whether data for part of the season would allow to predict the value of the metric at the end of the season.

3 Data

We have used play-by-play data from the NHL, seasons 2007-2008 to 2013-2014. The data was generated by Sportlogiq and used for the work in [9]. It is available at https://www2.cs.sfu.ca/~oschulte/sports/.

4 Intuitions - Game points importance value

The observations on which our new metrics are based, are the following. First, we investigated when goals are scored. We did this for different time intervals from seconds to minutes. Fig. 1 shows the results of goals per minute for the 2013-2014 season and this is representative for all seasons and most time intervals. We note that few goals are scored in the first minute of the game. Further, during the last minute of the game, at least three times as many goals are scored than for

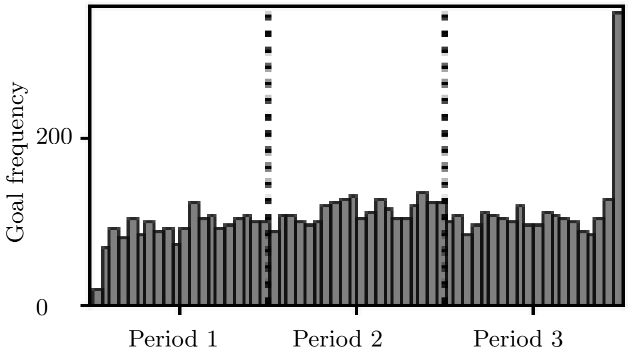


Fig. 1. Goal frequency for each minute of the first three periods in the NHL during the 2013-2014 season.

any other minute in the game. A possible explanation is the higher frequency of 6 on 5 situations at this time of game, in which a team's gamble to pull their goaltender often results in either them scoring a goal (in part helped by their extra attacker) or the other team scoring an empty-net goal. We also note that power-plays more often result in goals and that shorthanded goals are not that common. A team's strategy may also shift depending on the current score. Our metrics therefore take time, goal differential and manpower differential into account.

Another observation is that not all goals are equally important for producing game points, i.e., 2 PTS for a win, 1 PTS for an overtime loss and 0 PTS for a regular time loss in the NHL. For instance, scoring the 6th goal for the team when already leading with 5-0, will most likely not be contributing much for obtaining 2 PTS. The team would most likely win anyhow. However, a goal that ties the game in the last second of the game normally secures 1 PTS (while just before the goal the team would have 0 PTS) and therefore is an important goal. Our new metrics take the importance of a goal for producing PTS into account.

5 Metrics definition - GPIV-weighted performance metrics

5.1 Game Points Importance Value

As a basis for our new metrics we need to formally define the importance of a goal. Our intuition is that the importance of the goal represents the change in probability of the team taking points for the game (PTS) before and after the goal has been scored.[3] Further, as discussed earlier, we take into account time (t) for which we choose one second intervals, goal differential (GD) and manpower differential (MD). This we call a context.

[3] In [8] only the change in win probability is considered.

We note that in this paper we focus on regular time and leave overtime for future work. That means that the outcome of a game is one of win, tie, or loss.

We next define the probability of an outcome of a game given a context, as the ratio of the number of occurrences of the context that have resulted in the outcome and the total number of occurrences of the context in our dataset:

$$\mathrm{P}(\text{outcome} \mid \text{context}) = \frac{\mathrm{Occ}(\text{context with outcome})}{\mathrm{Occ}(\text{context})}.$$

We then attribute a game points importance value (GPIV) to a context. Intuitively, the GPIV represents how much a goal in a particular context increases or decreases the expected game points taking into account that a win gives 2 PTS, a tie gives 1 PTS and a loss 0 PTS. When a goal is scored, the context after the goal (context AG) has the same time as the context before the goal (context BG), but the GD is changed by one and the MD may (minor penalty power-play goal) or may not change (even strength, short-handed, or major penalty power-play goal). Based on this intuition, we define the GVIP (for regulation time in the NHL) as follows:

$$\begin{aligned}\mathrm{GPIV}^{\mathrm{reg}}_{\mathrm{NHL}}&(\text{context BG})\\ &= 2 \cdot [\mathrm{P}(\text{win} \mid \text{context AG}) \text{ - } \mathrm{P}(\text{win} \mid \text{context BG})]\\ &\quad + 1 \cdot [\mathrm{P}(\text{tie} \mid \text{context AG}) \text{ - } \mathrm{P}(\text{tie} \mid \text{context BG})].\end{aligned}$$

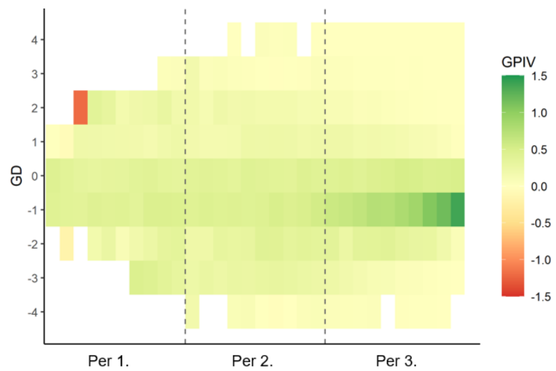


Fig. 2. GPIV versus GD for the 2013-2014 season. Each bin is two minutes. Less than two observations for each bin are left out.

In Figs. 2 and 3 we show representative visualizations of the characteristics of GPIV. From Fig. 2 we note that the value of GPIV is high when the GD is -1 or 0 at the end of the third period, as scoring then will tie the game (going from 0 to 1 PTS) or result in a 1 goal lead (going from 1 to 2 PTS). However, as the scoring frequency in the last minute is three times higher than at any other arbitrary minute in the game (see Fig. 1), this increase in GPIV may not be as high as expected.

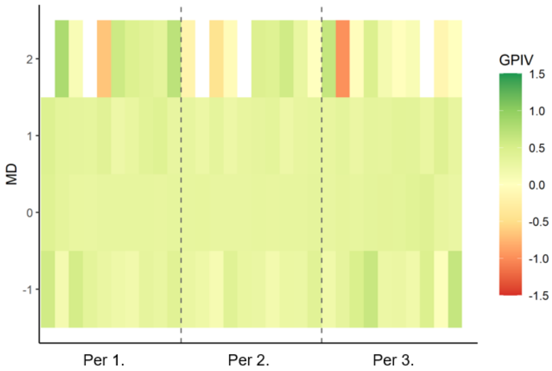


Fig. 3. GPIV versus MD for the 2013-2014 season. Each bin is two minutes. Less than two observations for each bin are left out.

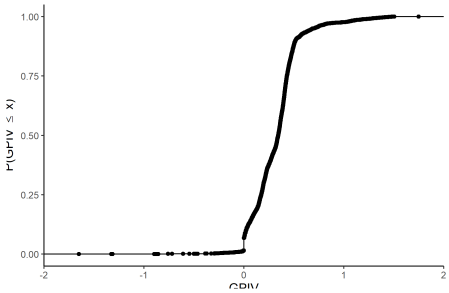


Fig. 4. Cumulative distribution function of GPIV.

Scoring goals is not always positive for the probability of taking game points. We noted that, although this situation rarely appears, taking a 3-goal lead early in the game may have negative consequences. This could be explained by the possibility of the leading team becoming too complacent with a comfortable lead. In general, negative consequences were limited to the first period or special MD cases.

In Fig. 4 we see that the probability of a negative GPIV is 1.57%. Approximately 86% of the GPIV range is between 0 and 0.5. Furthermore, 12% of the GPIV range is from 0.5 to 1.64. What is interesting with this last group is that they have the same or greater GPIV (0.5) as typical game deciding goals scored in overtime (which results in the team directly being awarded an extra point instead of - on average - getting the extra point with probability 0.5).

5.2 New metrics

We define new variants of the traditional metrics goals (G), assists (A), points (P) and +/- which we call GPIV-G, GPIV-A, GPIV-P and GPIV-+/-, respectively. In the traditional metrics the value is raised by 1 when a player scores a goal

Table 1: Top 10 players for GPIV-P for the 2013-2014 season.

P-rank	GPIV-P-rank	Rank change	Player	Position	P	GPIV-P
1	1	0	Sidney Crosby	C	102	34.698
8-9	2	7	Nicklas Bäckström	C	78	29.038
12	3	9	Alex Ovechkin	R	75	28.810
27-28	4	23	Blake Wheeler	R	65	27.735
4	5	-1	Tyler Seguin	C	83	27.264
2-3	6	-3	Claude Giroux	C	85	26.524
10	7	3	Joe Pavelski	C	77	26.404
23-24	8	14	Anze Kopitar	R	67	25.901
6-7	9	-3	Phil Kessel	C	77	25.871
29	10	19	Bryan Little	R	64	25.170

(for G and P), a player gives an assist to a goal (for A and P) or the player is on the ice when a goal is scored by the player's team (for +/-). For the latter when a goal is scored by the opposing team the value is decreased by 1. For the variants of the metrics, instead of raising or decreasing by 1, we raise or decrease the value by the GPIV of the goal. The new metrics value the amount of goals as well as the importance of these goals. Some of the highest ranked players are involved in many goals, while others may be involved in fewer goals, but with higher importance.

6 Eye test for GPIV metrics

Table 1 shows the top ranked players for GPIV-P during the 2013-2014 season.[4] Looking closer at the results, several players stand out. First, Alex Ovechkin went from a rank 9 (P) to being ranked 3rd (GPIV-P) when using the new metric. This is a considerable difference in rank, but can be explained by the many important goals he scored that season. For example, as mentioned already in the introduction, Alexander Ovechkin had the most game-tying and lead-taking goals while he only ranked 29th regarding goals scored when the team is already in the lead. Other players on the top-10 list that saw significant increases in their relative point-based rankings were Blake Wheeler (Winnipeg Jets) and Anze Kopitar (LA Kings). Similar to Alexander Ovechkin, the latter of these has proven to take the game to the next level during the play-offs (when goals are tougher to get by and each goal is typically considered of greater value).

Results for the other metrics and seasons are available at https://www.ida.liu.se/research/sportsanalytics/projects/conferences/LINHAC-22

[4] Note that we only take into account regular time, so the numbers for the traditional metrics do not conform with the numbers at nhl.com that also include overtime data.

7 Correlations of GPIV metrics with traditional metrics

Figs. 5-8 show for the top-30 players in the GPIV-based rankings for goals, assists, points and +/-, respectively, what their change in rank is with respect to the traditional metrics. Players on the black line have the same ranking. Players in red have lower ranking in the new metric than in the traditional metric and players in green have raised their ranking. Here, the points shows the actual rank assigned with the different metrics and the length of the lines indicates the absolute differences in rank (shown away from the black line so to make the points close to the line easier to identify). The figures show that the new metrics differ from the old metric and do lead to changes in rankings.

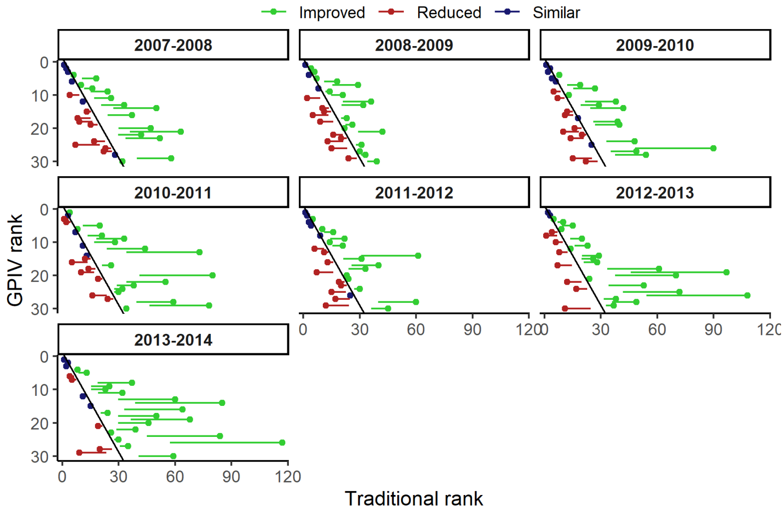


Fig. 5. Rank comparisons for traditional goals and GPIV-goals.

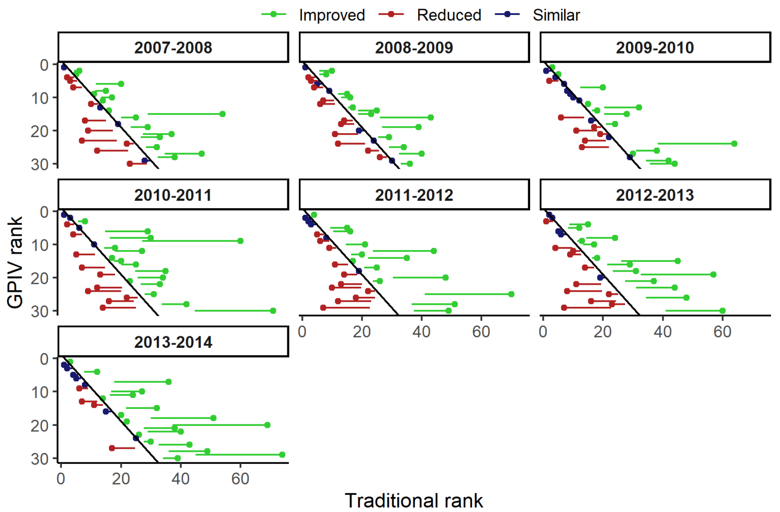


Fig. 6. Rank comparisons for traditional assists and GPIV-assists.

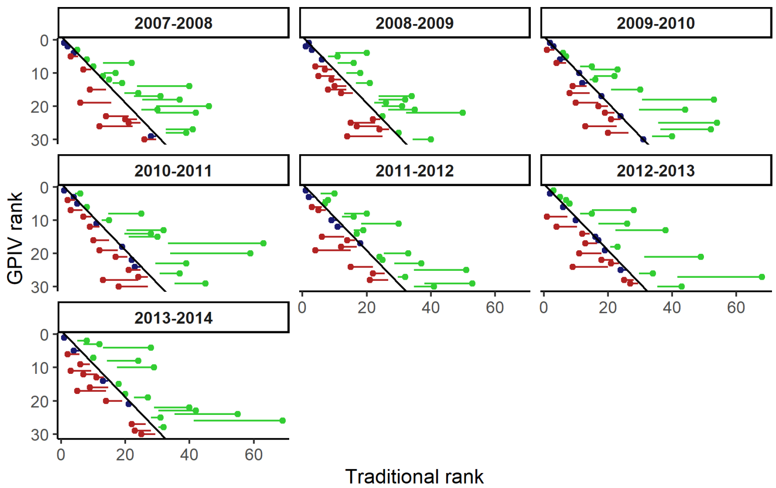


Fig. 7. Rank comparisons for traditional points and GPIV-points.

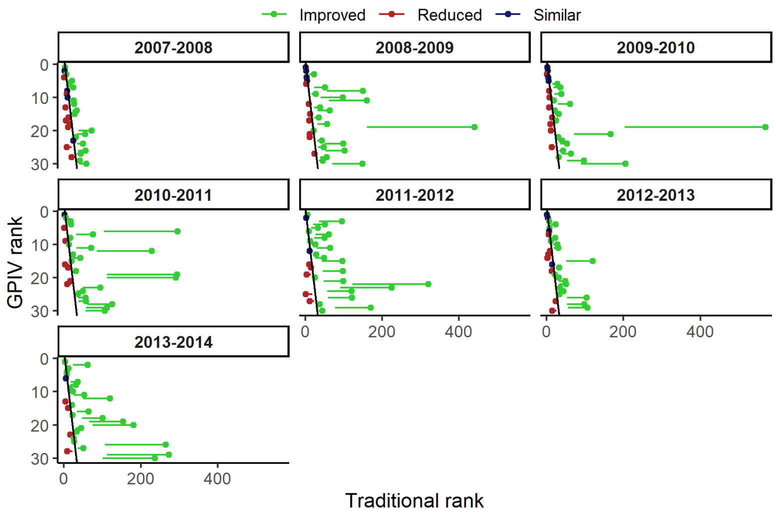


Fig. 8. Rank comparisons for traditional +/- and GPIV-+/-.

In Figs. 9-12 we show the Spearman correlation of the traditional metrics and their respective new GPIV-based metrics. For goals the correlation is between 0.915 and 0.968, for assists between 0.960 and 0.979, and for points between 0.972 and 0.987. These are high correlations, indicating that the new metrics have a similar behavior as well-accepted metrics, but they do introduce new insights. For +/- the correlation is lower being between 0.715 and 0.821.

8 GPIV metrics over different seasons

We check now how the metrics behave over different seasons. In Table 2 we show the maximal values for the traditional goals, assists, points and their GPIV-

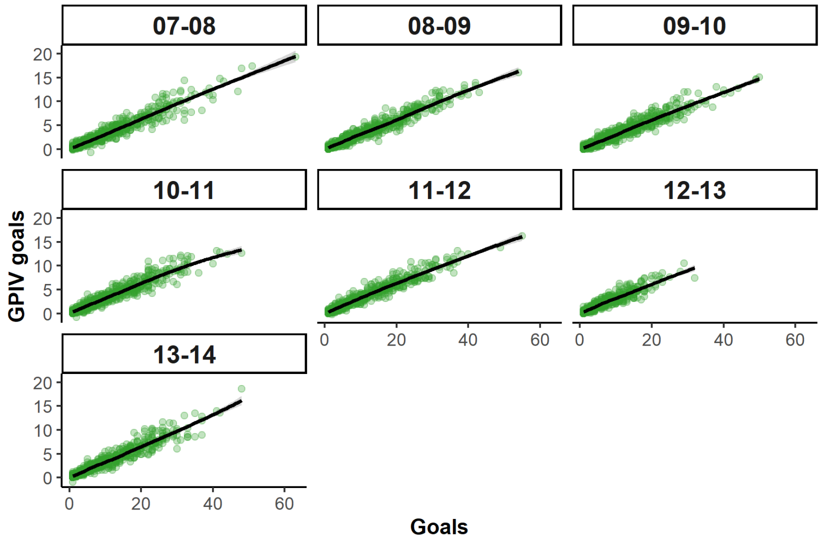


Fig. 9. Correlation traditional goals and GPIV-goals.

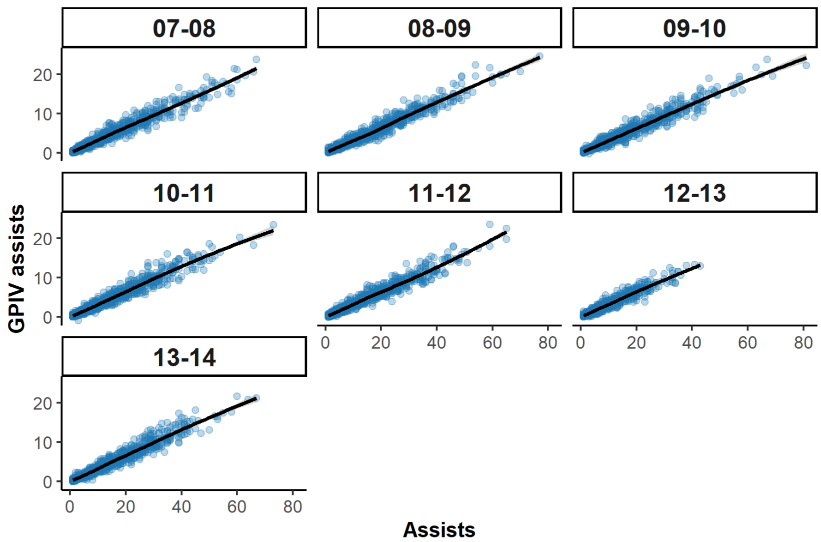


Fig. 10. Correlation traditional assists and GPIV-assists.

based counterparts. The minimum values for the traditional metrics is 0, while for the GPIV-based metrics there are a few players per season that receive a negative value for the GPIV-based metrics.

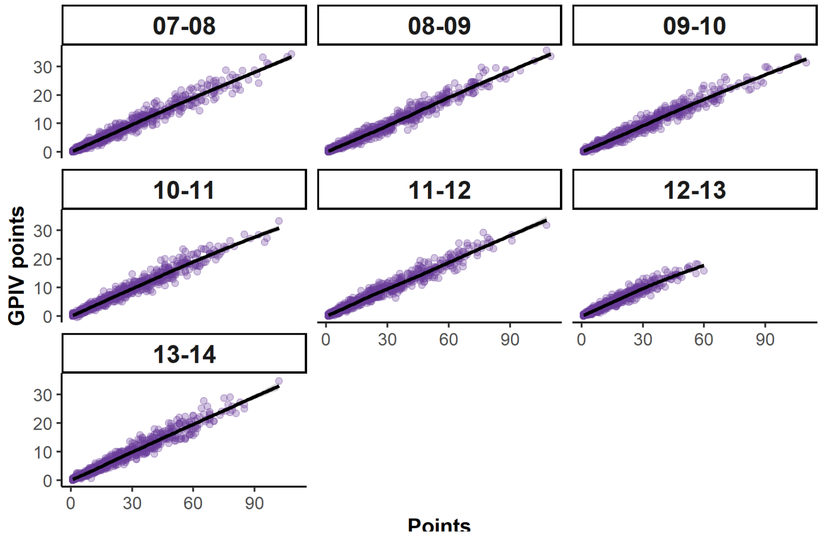


Fig. 11. Correlation traditional points and GPIV-points.

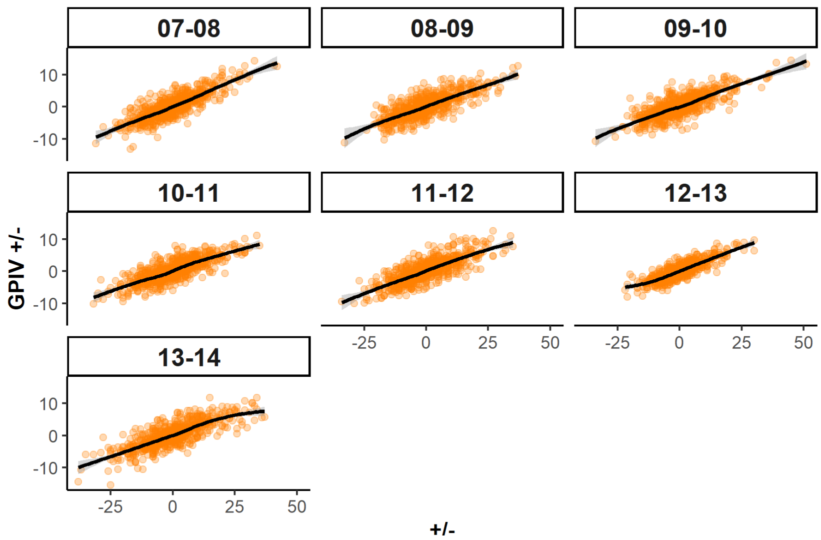


Fig. 12. Correlation traditional +/- and GPIV-+/-.

For the maximal values we note that there is a variation in values for the traditional metrics for different seasons which is followed by the GPIV-based metrics.[5]

[5] The values for the 2012-2013 season are lower as it was a shortened season.

Table 2: Maximum values for the metrics. Notes below table.

	Goals	GPIV-G	Assists	GPIV-A	Points	GPIV-P
2007-2008	63	19.311	67	23.689	108	34.322
2008-2009	54	16.091	77	24.493	110	(6) 35.601
2009-2010	50	15.056	81	(2) 23.740	110	(7) 33.164
2010-2011	48	(1) 13.115	73	23.411	102	33.355
2011-2012	55	16.255	65	(3) 23.530	108	31.799
2012-2013	29	10.482	43	(4) 13.300	57	(8) 18.305
2013-2014	48	18.580	67	(5) 21.657	102	34.698

Table notes:
(1) Corey Perry 48/12.621 vs Daniel Sedin 41/13.115
(2) Henrik Sedin 81/22.123 vs Brad Richards 67/23.740
(3) Henrik Sedin 65/22.447 and Claude Giroux 65/19.739 vs Joe Thornton 59/23.530
(4) Martin St. Louis 43/12.987 vs Sidney Crosby 13.300
(5) Sidney Crosby 67/21.222 vs Nicklas Bäckström 60/21.657
(6) Evgeni Malkin 110/33.443 vs Alex Ovechkin 108/35.601
(7) Henrik Sedin 110/31.210 vs Alex Ovechkin 106/33.164
(8) Steven Stamkos 57/18.150 vs Sidney Crosby 56/18.305

Table 2 (with accompanying table notes) also shows that the players with the highest value for the traditional metric were not always the players with the highest value for the GPIV-based counterpart and vice-versa. For instance, Henrik Sedin topped the assists ranking in 2009-2010 and in 2011-2012, but did not have the highest rank according to the GPIV-based assists. On the other hand Ovechkin topped the GPIV-based points in 2008-2009 and 2009-2010, but not the traditional points.

9 Prediction of GPIV metrics

In this section we investigate whether data from part of the season can be used to predict the value of the metric at the end of the season. We do this by dividing the data in partitions. For n partitions, we use the value of the metric after $\frac{1}{n}$-th part of the season, multiply with n and compare with the actual result of the metric at the end of the season. We do this for the traditional metrics as well as for the new metrics.

Fig. 13 shows for different seasons and different numbers of partitions, the Pearson correlation between a metric (final result after the season) and a value obtained by using the partitions (called 'generalized' in the figure) for all players.

We note that for all metrics, the more partitions, the lower the correlation. This is as expected. For instance, after half of the season (n=2) we have more data to base our prediction on than after one tenth of a season (n=10).

Further, for traditional metrics (in red color) as well as the new metrics (in orange) there is a high correlation between the final value and 2 times the value after half of the season. When we have less data, i.e., the number of partitions becomes higher, there is a slightly higher correlation for the traditional metrics than for the new metrics.

The other colors show predictability between traditional metrics and new metrics, which relates back to the correlation between the metrics.

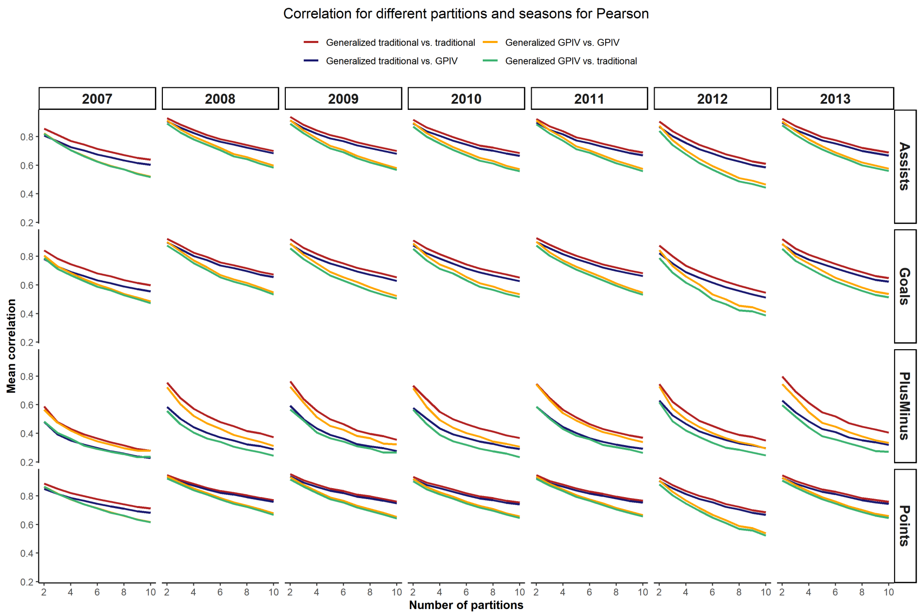


Fig. 13. Correlations for partitions for different metrics.

10 Conclusions

In this paper we have introduced new metrics that are variants of the well-known traditional metrics G, A, P, +/-. In addition to the number of goals scored, these new metrics also take into account the importance of goals with respect to earning PTS. This ensures that the metrics favor players that have greater impact on the outcome of the game (e.g., by scoring game deciding goals) over players that score most of their goals when the outcome of a game may already be decided. As the metrics are based on well-known metrics, they are easily understandable for the practitioners. The new metrics also pass the eye test. For G, A and P there is a high correlation between the traditional metrics and the GPIV-based counterparts.

Acknowledgements

Thanks to Rasmus Säfvenberg for implementing the current versions of the metrics, and to Rabnawaz Jansher, Sofie Jörgensen, Min-Chun Shih and Jon Vik for implementations in early stages that were used to decide on the current versions of the metrics.

References

1. Robert B. Gramacy, Shane T. Jensen, and Matt Taddy. Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis in Sports*, 9:97–111, 2013.
2. Wei Gu, Krista Foster, Jennifer Shang, and Lirong Wei. A game-predicting expert system using big data and machine learning. *Expert Systems with Applications*, 130:293–305, 2019.
3. Edward H Kaplan, Kevin Mongeon, and John T. Ryan. A Markov Model for Hockey: Manpower Differential and Win Probability Added. *INFOR Information Systems and Operational Research*, 52(2):39–50, 2014.
4. Timmy Lehmus Persson, Haris Kozlica, Niklas Carlsson, and Patrick Lambrix. Prediction of tiers in the ranking of ice hockey players. In Ulf Brefeld, Jesse Davis, Jan Van Haaren, and Albrecht Zimmermann, editors, *Machine Learning and Data Mining for Sports Analytics. MLSA 2020*, volume 1324 of *Communications in Computer and Information Science*, pages 89–100, 2020.
5. Guiliang Liu and Oliver Schulte. Deep reinforcement learning in ice hockey for context-aware player evaluation. In Jérôme Lang, editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 3442–3448, 2018.
6. Dennis Ljung, Niklas Carlsson, and Patrick Lambrix. Player pairs valuation in ice hockey. In Ulf Brefeld, Jesse Davis, Jan Van Haaren, and Albrecht Zimmermann, editors, *Machine Learning and Data Mining for Sports Analytics. MLSA 2018*, volume 11330 of *Lecture Notes in Computer Science*, pages 82–92, 2019.
7. Brian Macdonald. A Regression-Based Adjusted Plus-Minus Statistic for NHL Players. *Journal of Quantitative Analysis in Sports*, 7(3), 2011.
8. Stephen Pettigrew. Assessing the offensive productivity of NHL players using in-game win probabilities. In *MIT Sloan Sports Analytics Conference*, 2015.
9. Kurt Routley and Oliver Schulte. A Markov Game Model for Valuing Player Actions in Ice Hockey. In Marina Meila and Tom Heskes, editors, *Uncertainty in Artificial Intelligence*, pages 782–791, 2015.
10. Carles Sans Fuentes, Niklas Carlsson, and Patrick Lambrix. Player impact measures for scoring in ice hockey. In Dimitris Karlis, Ioannis Ntzoufras, and Sotiris Drikos, editors, *MathSport International 2019 Conference*, pages 307–317, 2019.
11. Michael Schuckers and James Curro. Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events. In *MIT Sloan Sports Analytics Conference*, 2013.
12. Oliver Schulte, Mahmoud Khademi, Sajjad Gholami, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. A Markov Game model for valuing actions, locations, and team performance in ice hockey. *Data Mining and Knowledge Discovery*, 31(6):1735–1757, 2017.

13. Oliver Schulte, Zeyu Zhao, Mehrsan Javan, and Philippe Desaulniers. Apples-to-apples: Clustering and Ranking NHL Players Using Location Information and Scoring Impact. In *MIT Sloan Sports Analytics Conference*, 2017.
14. A.C. Thomas, Samuel L. Ventura, Shane Jensen, and Stephen Ma. Competing Process Hazard Function Models for Player Ratings in Ice Hockey. *The Annals of Applied Statistics*, 7(3):1497–1524, 2013.
15. Jon Vik, Min-Chun Shih, Rabnawaz Jansher, Niklas Carlsson, and Patrick Lambrix. Not all goals are equally important - a study for the NHL. In J. James Reade, editor, *MathSport International 2021*, pages 26–31, 2021.

Keeping Count: Archiving Women's Hockey Analytics for Accessibility

Carleen Markey[1[0000−0003−0629−8840]], Nix Dube, Alyssa Longmuir, Mike Murphy, An Nguyen, Jessika Ross, and Anna Severn

Carnegie Mellon University, Pittsburgh, PA 15213, USA carleenmarkey@gmail.com

Abstract. Women's hockey analytics has historically lacked a centralized repository for research, data, and other projects, despite other areas of hockey analytics having such central resources. In this paper, we attempt to fill in this missing piece of women's hockey analytics by holding an archiving event in which volunteers methodologically gathered as many details on past women's hockey research and data as possible. Each piece of research and data was then turned into an entry on MetaHockey according to standardized instructions. This event resulted in almost one hundred new women's hockey focused entries on MetaHockey, whose characteristics largely align with trends in men's hockey analytics. Examining these entries also empirically reveals an exponentially increase trend in women's hockey analytics entries year-over-year, demonstrating that both continuing to archive works and taking advantage of this new research in the private and public spheres is conducive to the growth of the field.

Keywords: Women's Hockey · Hockey Analytics · Archiving.

1 Introduction

Women's hockey has gained considerable momentum in recent years internationally, and the men's hockey analytics movement has followed a similar trajectory in the same timeframe. The women's hockey analytics research and data sources have grown exponentially thanks to these explosions in popularity of the two central components. However, this community has historically lacked easily accessibly centralized sources of data and projects, raising the barrier of entry of an already niche subject. Additionally, events focusing on women's hockey analytics have begun to occur, such as the Big Data Cup [15] and WHKYHAC [8]. Building off of previous works is a crucial aspect to both events and to the progress of this field of research, growing the need for an easily accessed archive of projects.

There have been several websites that have attempted to create such a centralized resource. Even-Strength [1], TheirHockeyCounts [14], CWHL Tracker [2], and pick224 [9] have centralized summary statistics, respectively for advanced PHF statistics, SDHL & NCAA D1/D3 counting statistics, and various leagues' & international competitions' counting statistics. While being quite thorough in their compilation of current women's hockey statistics, they do not document the research and projects being developed in this space.

MetaHockey [4], a site for archiving men's hockey analytics projects, was founded to fill this void in hockey analytics in general. Until October 2021 however, the website contained only seven entries and data sources related to women's hockey [5]. As documented in the WHKYHAC presentation "Contextualizing Historical Data and Current Projects in Women's Hockey" in July 2021, well over ten times that many public women's hockey analytics projects have been created since 2015 alone [10]. Additionally, the authors of this paper, as major participants in the women's hockey analytics research, determined that MetaHockey's organizational system in it's current form, does not adequately serve the women's hockey analytics researchers' current needs of an archival website. Among other things, the website is hard to navigate when trying to find code repositories in data, the tagging system does not differentiate between various women's leagues as it does with men's leagues, and documented advances in women's hockey analytics often do not take the form of formal books or articles, which are the two categories available for publications submitted to MetaHockey. Published advances in women's hockey often take the form of twitter threads or Tableau-based tools, which to not fall under either of these categories.

In this project, we take the first step towards fully satisfying the need for easy access to historical women's hockey projects and data sources, as well as continuing MetaHockey's original purpose of serving all sides of hockey analytics. We do this by compiling a detailed list of as many women's hockey analytics projects as possible that were publicly accessible as of October 2021 and adding them to MetaHockey's article repository, with permission and help of MetaHockey site owners and editors. Modifying the MetaHockey website itself to serve users better is left to a future project.

2 Methodology

To proceed with adding women's hockey analytics works to MetaHockey, we followed the methods below in designing the archiving process, designing the archiving materials, putting on the archiving event, and uploading everything to MetaHockey.

2.1 Designing the Archiving Process

Since the authors observed that there is no common publication spot for women's hockey analytics works except for Twitter threads, the most effective way of obtaining the maximum amount of publications, events, and resources was first creating a collaborative list of the people who have created women's hockey analytics projects and compiled data sources, a list of known websites of compiled data sources, a list of direct data sources, and a list of events featuring women's hockey, such as conferences. This is the "To Archive" document [3]. Then, specific publications, events, and data sources that would become MetaHockey entries would be gathered by searching Twitter and Google for each person/data source/event on the four lists, and creating entries in a Google Sheet ("Whockey MetaHockey Entries") for the publications, events, and data sources found to be related to them [16]. The "Whockey MetaHockey Entries" would then be copied into the Google Sheets-based MetaHockey back-end to get all the entries onto MetaHockey.

This is a time- and labor-intensive process, and the authors recognized the expedited need for the completion of this project by the beginning of the Big Data Cup in spring 2022. As a result, a call was put out for volunteers to help with the searching for and creating MetaHockey entries, and a date was set for an event in which some of the authors would be available over Zoom to help with both [13].

An additional note on this method: simply searching something like "women's hockey analytics" or "women's hockey data" in Twitter's or another website's search engine would have not returned the maximal results for archival entries, as creators tend to title their projects and datasets with the relevant league and area of study/statistics, as seen in entries 714-812 of now-archived women's hockey analytics projects [4].

2.2 Designing the Archiving Materials

Once the three lists were compiled and the overall methodology distilled, an instructional guide, "How to Archive", was designed for volunteer archivists to use for each entry type [11]. The first three pages outline exactly how to go about gathering entry details and adding them into the central archival spreadsheet for each list [16]. The first page of the guide is shown in Fig. 1 and was designed to be used to search for entries using the "people" list from the "To Archive" document.

The second and third pages of this guide are similar to the first in general flow, but with specific modifications for collecting entry data using the "events" list and the "websites"/"direct data sources" lists respectively.

The fourth page, "Creating Entries", continues the workflow from pages 1-3, and outlines how to format the details for each possible entry into the MetaHockey specific format and enter it to the "Whockey MetaHockey Entries" sheet. This fourth page can be viewed in Fig. 2. It is important to note here that volunteer archivists chose the keywords for each entry, as they were women's hockey analytics researchers and therefore familiar with the source material or had help from members like this.

The fifth and final page contains an appendix of common terms used in the "How to Archive" document and instructions on how to select keywords from a suggested list. Keywords not on this list were also able to be added manually for an entry.

Fig. 1. The first page of the "How to Archive" instructional guide designed to guide a layperson through how to start with a name from the people section of the "To Archive" document, search for the publications, data sources, and events they have been involved with and create detailed MetaHockey entries from those search results.

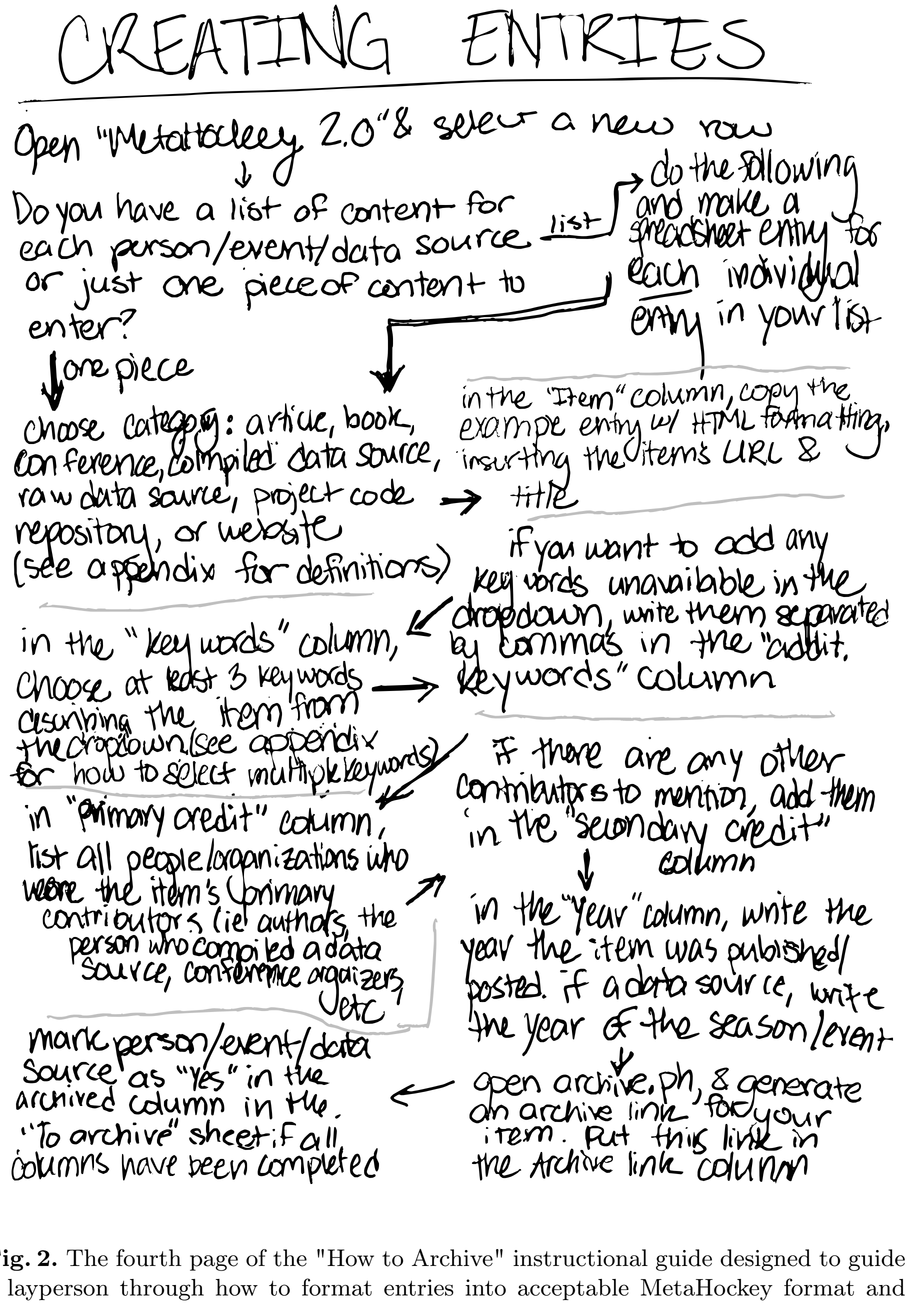


Fig. 2. The fourth page of the "How to Archive" instructional guide designed to guide a layperson through how to format entries into acceptable MetaHockey format and enter it into the "Whockey MetaHockey Entries".

2.3 The Archiving Event and MetaHockey Upload

The archiving event that used this process and these materials occurred on Oct. 23rd, 2021 from 5-8pm EST, with an option for volunteer archivists to keep adding to the "Whockey MetaHockey Entries". Once the event was over, duplicate entries were removed and chosen keywords were checked to be accurate in the "Whockey MetaHockey Entries" Sheet.

Unfortunately at this point, the authors of this paper lost contact with the main editor of the MetaHockey site, and spent the next several months reaching out to various editors of MetaHockey to see if they had site access. The last step of uploading entries onto MetaHockey was finally completed on Feb. 11, 2022 when an editor with site editing access was finally contacted and they agreed to do the current Google Sheet upload, as well as future uploads of entries.

Inevitably with the method used in this paper, publications, data, and events will be missed, since the three lists relies on the authors' memories of such things and ability of volunteers to get accurate search results. Nonetheless, we proceed with this method because the goal is progress, not perfection.

3 Results

As a result of this archiving effort, there are 98 new women's hockey analytics projects, data sets, and research tools on MetaHockey, for a total of 105 entries of the 812 existing MetaHockey entries in the Articles section being works pertaining to women's hockey. Given that this is the first quantitative survey of the field of women's hockey analytics, after qualitative examinations at OTTHAC 2022 [12] and WHKYHAC 2021 [10], it's important to briefly examine these entries statistically. Starting with Table 1, the counts of women's hockey analytics entries are broken down by MetaHockey category label.

Table 1. Popularity of MetaHockey categories among women's hockey entries.

MetaHockey Category Label	# of Entries
Article	33
Files - Raw Data	18
Files - Compiled Data	16
Website (Blog / Tableau Profile / Stats / Etc.)	14
Book	7
Conference	3
Project Code Repository	3

The object of note in this list is the prevalence of files of data, raw and compiled. Several of the book entries are also compiled records of data. The focus on data is likely caused by something the authors are familiar with: the ever-looming possibility of data loss. The authors have heard anecdotes of years

IIHF data being lost to a basement flood, experienced the loss of NWHL/PHF play by play and location data from the league website, and lost access to CWHL statistics when the league ceased operations. It has become a priority of women's hockey analytics researchers to preserve data whenever possible, as seen with the websites mentioned in the introduction.

The other part of this table that may be surprising to some is the lack of project code repositories. The proposed explanation for this a matter of common practices in the community: project code does not often stand on its own and are often linked within articles to support those projects. Therefore, there is a fundamentally low amount of entries in this category.

Moving past the entry type and onto entry focus, Table 2 is a list of the top 25 keywords associated with entries, excluding the obviously highest use of the women's hockey tag.

Table 2. Top 25 most popular keywords for women's hockey entries on MetaHockey.

Keyword	# of Entries
Counting Stats	39
NWHL / PHF	32
CWHL	29
Goalies / Goaltending	23
Big Data Cup	16
NCAA	13
xG	13
Shots / Shooting	8
Passing	8
Central Ontario Women's Hockey League	7
Western Women's Hockey League	7
Pre-Shot Movement	7
National Women's Hockey League (old)	6
PWHPA	5
Olympics	5
Prediction	4
Shot Quality	4
IIHF	4
Play By Play	4
Tracking	4
Advanced Stats	4
SDHL	3
Team	3
World Championships	3
Model(s)	3

The majority of these keywords are linked to either data sources or books, which preserve leagues both defunct and active, as well. The non-data source focused keywords are in line with general trends of hockey analytics study since

2015, namely the focus on xG, shooting, passing, and pre-shot movement. Curiously, goaltending makes a highly ranked appearance on this list. 20 of the 23 entries referencing goalies and/or goaltending can be attributed to one women's hockey analytics researcher who has been preserving goaltending data for the CWHL, NWHL/PHF, and the SDHL since at least 2016 [6].

Lastly, Fig. 3 looks at the number of women's hockey analytics projects (including all projects under all MetaHockey Categories) published in each year since 2014, which is the year of publishing of the oldest project found.

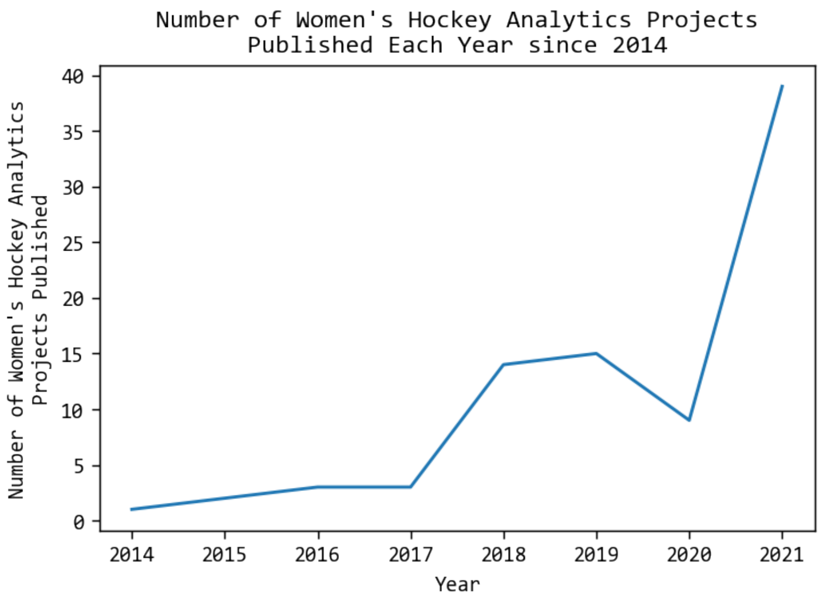


Fig. 3. A chart displaying an increasing trend of women's hockey analytics projects each year, with the exception of 2020. 2020 was when most women's hockey leagues and tournaments were inactive due to the COVID-19 pandemic [7], and therefore there was no new data to work with.

As described in the introduction, women's hockey analytics is on a trajectory of exponential growth. Fig. 3 shows that this is not just conjecture or wishful thinking. Teams, leagues, and researchers would be wise to turn attention to this field of research as women's hockey analytics continues on on the rise in the public and private spheres.

4 Conclusion

Women's hockey analytics has a history of projects and data that needed to be centrally archived in order for the community and research to continuing to grow. Thanks to a volunteer-based effort, an integral first step has been made towards fulfilling this need. By investigating all available avenues in which projects and data might be found in a procedural manner, details and archived copies of nearly one hundred women's hockey projects, data, and events have made it onto MetaHockey. This set of now-archived research and data reflect the recent priorities of the women's hockey analytics community of data preservation and bringing the field up to speed with men's hockey analytics. It also shows a concrete trend of women's hockey analytics research exponentially growing in the last few years. In the future, we hope to continue archiving women's hockey analytics research in a more periodic manner and hope that as the community gains more momentum, referencing previous works will become more prevalent and will be used to accelerate the public and private development of the field.

Acknowledgements Thanks to Mike Gallimore, Prashanth Iyer, and Mike Pfeil for their permission and help with uploading articles to MetaHockey. We would also like to acknowledge and thank all the volunteers who helped search for and create MetaHockey entries, for all of this would not have been possible without them.

References

1. Matt Barlowe, Jake Flancer, Alyssa Longmuir, and CJ Tuturo. Even-Strength. https://web.archive.org/web/20210415171201/https://even-strength.com/
2. Jeff Craig. CWHL Tracker. https://cwhl-tracker.herokuapp.com/
3. Nix Dube, Alyssa Longmuir, Mike Murphy, Carleen Markey, An Nguyen, Jessika Ross, and Anna Severn. "To Archive". https://bit.ly/3KI2kGr
4. Mike Gallimore and Prashanth Iyer. Metahockey. https://bit.ly/3LM4ar8
5. Mike Gallimore and Prashanth Iyer. Metahockey, Dec. 11, 2020. https://bit.ly/3saOxBE
6. Geremy/Giants in the Crease. https://creasegiants.wixsite.com/index/single-post/2016/08/20/top-goalies-in-the-cwhl-draft
7. Emily Kaplan. "How the COVID-19 pandemic widened the gender gap in hockey". https://www.espn.com/olympics/hockey/story/_/id/31394538/how-covid-19-pandemic-widened-gender-gap-hockey
8. Alyssa Longmuir, Carleen Markey, and Mike Murphy. WHKYHAC. https://www.whkyhac.com/
9. Dave MacPherson. Pick224. pick224.com/women
10. Carleen Markey. "Contextualizing Historical Data and Current Projects in Women's Hockey". https://youtu.be/WyhSZpKxRFM
11. Carleen Markey. "How To Archive". https://bit.ly/3KIgCqc
12. Carleen Markey. "On the Breakaway: Accelerating the evolution of women's hockey analytics". https://youtu.be/r-iVM1WACHo
13. MetaHockey Twitter Account. https://bit.ly/3OTc185

14. Mike Murphy. TheirHockeyCounts. https://theirhockeycounts.com/
15. Stathletes. Big Data Cup. https://www.stathletes.com/big-data-cup/
16. "WHockey MetaHockey Entries". https://docs.google.com/spreadsheets/d/1J-Lg18bUdwVC71O5g47UTYch2ZxUR-2zUeR226GpDL4/edit?usp=sharing

Research papers

Scouting Automated Ratings Analyzing Habits (SARAH): A Statistical Methodology for Scouting and Player Development

Noah Chaikof[1], Katia Clément-Heydra[2], Carleen Markey[3], Mikael Nahabedian[2], Adam Pilotte[5], and Mairead Shaw[2]

[1] Ottawa 67's, Ottawa, Ontario, K1S 3W7, Canada
[2] McGill Marlets Hockey, Montreal, Quebec, H2W 1S4, Canada
[3] Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[4] Saint Laurent Patriotes, Saint-Laurent, Quebec, H4L 3X7, Canada
[5] McGill Redbirds Hockey, Montreal, Quebec, H2W 1S4, Canada

Abstract. The project serves a two-fold purpose: to reduce the time that scouts and coaches spend trying to identify what players have foundational on-ice habits, and to streamline the process of evaluating the developmental progress of a players' habits. Essentially what we did was first look at the various national women's hockey teams and identify the set of "habits" a player regularly executes (i.e., edgework, catching the puck in the hip pocket, pass placement, etc). Combining the dataset of players' habits with a set of players' microstats (entries via pass/stickhandling, exits via stickhandling/pass, accurate/inaccurate passes, etc.), we developed a random forest classification model to accurately predict if a player possesses a certain habit based on their set of microstats. We also used random forest regression on our data to see how habits impacted each specific microstat. Combining this with an estimate of how frequently players used each habit, we created a Player Development Matrix for a player's habits based entirely on their microstats. To help coaches, scouts, and anyone else access & use these tools, we've also created an interactive visualization for these models using our training dataset of national women's hockey teams in the last Worlds and Olympics.

Keywords: Player Development, Analytics, Coaching, Scouting.

1 Introduction

This paper provides a comprehensive overview of a newly designed player-evaluation framework for women skaters at the 2021 IIHF tournament and 2022 Olympic games using a 'habit-tracking' system. Building on the work of Bryce Chevallier[1], Jack Han[2], and Darryl Belfry[3], the goal of this study is to explore the validity of using micro habit-tracking as a supportive scouting technique (player-ranking system) and utilize habit-tracking as a foundation to uncover the highest priority areas for player development staff to hone in on meaningful skill improvements in their players or clients.

Our study demonstrates a statistically significant ability to accurately link a player's "habit-score" to statistical events for scouting purposes (micro stats such as zone exits, zone entries, type of pass,...), and uncovers 'habits-of-focus' for player development staff based on a player's advanced stats. Lastly, the study explores a habit-improvement framework using a Player Development Matrix[4] to analyze the habits of highest importance for development staff relative to the rest of that player's skill set.

The core technique of this study is the novel development of a complete list of habits and categorization of those habits into 7 different skill set areas. A comprehensive tracking model was used to obtain a baseline habit-score of all players, this data (combined with enriched data from InStat[5]) was used as the basis for the two models and the development matrix outlined below.

1.1 Motivation

The aim of this study is to offer a quantitative tool to both player evaluators (coaches and scouts) and player development staff as they are challenged with examining/improving the skill sets of large groups of players.

Scouting. The motivation for the study is to attempt to add a complimentary quantitative approach to traditional scouting and player evaluation analysis. Under the current model of scouting across hockey leagues, scouts are faced with a tremendous challenge of ranking players across broad skill categories, as evidenced by the sheer number of NHL draft rankings alone [6,7,8]. It is a significant challenge to rank a player's skill set (i.e. passing) on a scale from 1-10 and subsequently justify why a player's rating in that category will vary so significantly across scouts watching the same player.

The goal of the study with the creation of a binary habit-tracking system (said habit positively impacts a player's game or not) will enable certain player evaluators to bring a more quantitative approach to their rankings and give teams an edge in their scouting process.

Player Development. Similarly, player development staff are facing a tremendous challenge in trying to prioritize their limited time with each player and design a personalized skill development plan to drive improvement in their game[9]. The binary tracking system will allow player development staff to hone in on more exact skill gaps and work directly on improving those habits. Additionally, as a larger dataset of player habit-tracking is built over time, player development coaches can uncover which groups of habits are most critical to player success at different points in their careers, and how player habits may evolve over time.

2 Methodology

2.1 Statistical Methodology

The core statistical methodology/tracking technique used in this study is a novel binary-habit evaluation model developed below. In lay terms, the contributors of this study developed a list of habits (edgework, neutral zone angling etc.) and categorized those habits into different skillset areas (skating, puck reception, stickhandling, physicality, play away from the puck, passing & shooting) in an attempt to break down a player's game into micro attributes.

The selection of these various habits cover a broad spectrum of skills that may be displayed over the course of a hockey game (both offensive and defensive) but are highly specific in nature. Each habit was selected only if it can be measured clearly in the tracking process and the presence of that habit in a player's game is associated with driving impactful results during their time on ice. The following table summarizes the different skill sets and habits identified as part of the project. Refer to Appendix for a brief description of each habit identified as part of this project.

Table 1. Skill Sets & Habits

Skating	Puck Reception	Stickhandling	Physicality	Play Away from the Puck	Passing	Shooting
Edgework Outside	Catching Puck in Hip Pocket	Loading Puck to Hip Pocket	Initiating Contact	Shoulder Checks	Slip Passes	Coordination
Backwards Skating	Dynamic Catch	Underhandling of Puck	Puck Protection with Body	NZ Angling	Leveraging & Creating Seams	Weightransfer
Stride Recovery	Getting Off the Boards	Handedness Versatility	Fitness Level	Unassisted Stops	Pass Placement	Tip
Skating Mechanics		Deception with Puck		Jumping in Shot Lanes	Vision	
Crossovers				Awareness Without Puck		
Shouldering Speed				Net Front Presence		
Feet in Motion						

2.2 Tracking Technique

In order to build a sample with over 7500 observations to train the models on a period per period basis, the tracking technique used for the study relied on observing a minimum of three periods of a player's ice-time and assigning a binary score for each of the habits underscored above. The sample time-on-ice from the three periods were each tracked from three different games to adjust for strength of opponent and variances in a player's effort and effectiveness from game to game. In total, the data set included habits for 262 players from 12 different teams.

Based on whether a player demonstrated that habit more often than not when given the opportunity to do so during their observed ice-time, they were given a score of '1' (habit positively impacting a player's game) or '0' (habit not positively impacting a player's game). This resulted in a total unweighted score out of 30 for each roster player based on the number of habits they possessed during the sample period.

Table 2. Skill Sets & Habits

Rank	Name	Team	Position	Score (on 30)
1	Marie-Philip Poulin	Canada	F	30
2	Jenni Hiirikoski	Finland	D	29
T-3	Kendall Coyne Schofield	USA	F	28
T-3	Jocelyne Larocque	Canada	D	28
T-3	Ronja Savolainen	Finland	D	28
T-6	Mélodie Daoust	Canada	F	27
T-6	Brianna Decker	USA	F	27
T-6	Sarah Fillier	Canada	F	27
T-6	Rebecca Johnston	Canada	F	27
T-10	Michelle Karvinen	Finland	F	26
T-10	Claire Thompson	Canada	D	26

2.3 Modelling

SARAH 1 - Identifying events or advanced metrics expected based on player habits. The first model used in this project (random forest regression model[10]) was created to identify the different events or advanced statistics that one would expect to see a player possess based on whether they have a given habit. The random forest used in SARAH 1 and 2 consists of generating a number of decision trees, each of which are only given a random part of the dataset. Each decision tree then decides how each independent variable affects the dependent variable based on the random subset of the data it sees and makes predictions for each player in the entire dataset based on their independent variable data. The predictions from all the trees are then averaged to create one prediction for each player.

This model utilizes the event specific data from InStat (i.e. controlled entries and inner slot shots etc.)[5] for each player, with the intended goal of **finding which habits yield results in specific advance statistics or event categories**. *Subconsciously, scouts complete this same exercise when evaluating a player's effectiveness and instincts. For example, one would expect a player who exhibits linear crossovers and keeps their feet in motion following a puck catch, to complete successful controlled entries at a higher rate than a player without these habits.* In this model, the independent variables are the habits (variables X), with event data being treated as the dependent variable (variable y).

SARAH 1 included 17 separate sub-models, with each of the sub-models representing one of the 17 different event types adjusted per 60 minutes that were observed in the study. This is also referred to as "event-based advanced stats" later in the paper. The events included in the model are the following:

Table 3. Event Types (Microstats)

Accurate passes	Puck battles lost
Breakouts via pass	Puck battles won
Breakouts via stickhandling	Puck losses
Dump ins	Puck recoveries
Dump outs	Shots blocking
Entries via pass	CF
Entries via stickhandling	CA
Inaccurate passes	Shots
Passes to slot	

Significance Threshold for Linking Habit to Event and Selection Process. A critical component of this event-to-habit linking methodology is to identify habits that meaningfully impact the event/advanced statistical metrics. In this study - any habit with an importance above the 0.0325 threshold is considered having a strong influence on the likelihood of a player-habit meaningfully impacting that statistic or advanced stat category. Below is an example of the 10 main habits that meet the threshold for the event pertaining to "puck battles won".

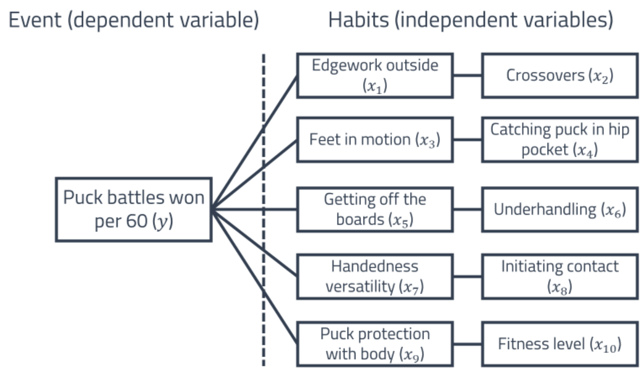


Fig. 1. An example of the 10 main habits that meet the threshold for the event pertaining to "puck battles won"

The considerations used in selecting the threshold of 0.0325 include;

1. Finding a threshold figure that aligns with the knowledge of the contributors to the report and removes results that do not align with hockey-logic (i.e. neutral zone angling should have no relation to shots statistic).

2. We wanted to select a threshold that ensured each habit would be meaningfully connected to a minimum of five events. If this was not the case - the habit was removed for lack of importance to the model).

Weighted Average Consideration for Event Statistics. Lastly, a weighted-average accounting for both the number of events completed and time-on-ice in the period was relied on in the SARAH 1 analysis.

This was done to adjust for problematic tracking outcomes when a player may have a high volume of events on a low base of ice-time (i.e. 4 successful completed passes in 3 minutes of ice-time in a given period) that would result in non-representative per/60 minute data. Therefore, greater weight was assigned to events that occurred over a larger period of ice time than in smaller sample sizes.

SARAH 2 - Predicting the probability of a habit meaningfully impacting a player's game. After establishing the impactful event-habit relationships in the first set of models, SARAH 2 reverses the variables and attempts to make a prediction about the probability of a habit successfully being completed by a given player.

This second set of models serves a dual purpose. First, it provides scouts with a baseline to precisely quantify habit evaluation. In other words, if the event-based advanced stats are available, this model can be seen as an automated habit-evaluation tool. However, SARAH 2 can also be used in conjunction with video scouting, allowing player evaluators to compare the statistical results versus their personal assessment of habits for different skaters.

Secondly, by precisely evaluating the success probability of various habits for skaters through the steps described below, this set of models enables skills coaches to uncover development opportunities for players and measure their progress over time in a systematic way.

The starting point of SARAH 2 is the meaningful event-habit relationships identified as part of SARAH 1, based on the 0.0325 threshold discussed in the previous section. However, flipping the variables in the case of SARAH 2 allows us to statistically estimate the probability of successful habit completion based upon a set of event-based advanced statistics for a given player.

For instance, when attempting to predict the success probability of the "outside edgework" habit, the first step is to highlight that this habit is strongly impacting the following 8 event-based advanced stats in SARAH 1. After identifying these strong event-habit relationships, the idea of SARAH 2 is to use these events to predict the successful completion of the "outside edgework" habit, as exemplified below:

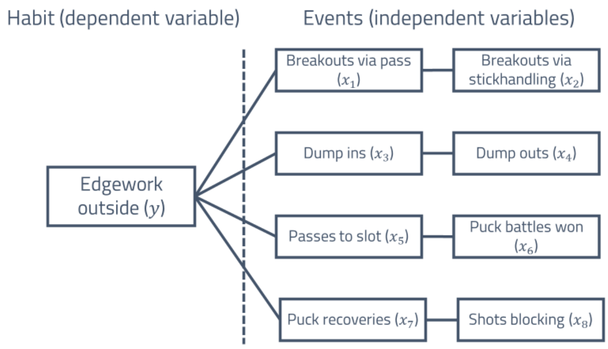


Fig. 2. An example of the 8 main events that meet the threshold for the "outside edgework" habit

In this example, as part of SARAH 1, we had identified that the "puck battles won-outside edgework" event-habit relationship was meaningful. For this reason, as part of SARAH 2, the "puck battles won" statistic is incorporated, among other events, as one of the predictors of the "outside edgework" habit.

While not visualized in the previous section, similarly for the 7 other events listed above (e.g., breakouts via pass, puck recoveries,...), it was established in SARAH 1 that the "outside edgework" habit is also meaningfully driving part the results for these other events. As such, in addition to "puck battles won'", these 7 other event-based advanced stats are also incorporated as predictors in SARAH 2 for this specific habit.

In short, SARAH 2 is built as a *random forest classification model [11] in which the event-based advanced statistics are the independent variables (X variables) and the habits are the dependent variable (y variable).*

SARAH 2 included 30 separate sub-models, with each of the sub-models representing one of the 30 different habits that were tracked in the study.

The outcome of SARAH 2 is that for each player, all of the habits measured will be assigned a value between 0 and 1 (considered a percentage probability) that a respective habit yields positive results while on ice.

For instance, in the case of Laura Stacey, a Canadian forward who initiates a high volume of controlled exits, dump entries and puck recoveries, the probability that she successfully completes the "outside edgework" habit is around 80%.

It is important to note that the outcome of this model is only identifying the success probability of a habit completion (i.e. a 0.8 score is not necessarily better than a 0.65), it is only significant in that it creates a probability based prediction on which habits are likely strengths and weaknesses for a given player.

For this random forest model, any habit with a score above 0.5 implies that when a player has the opportunity to exhibit this habit, they are more likely to complete this micro-ability well. As we had established that the event-habit relationships are meaningful, the successful completion of said habit is inherently related to driving impactful results on the ice.

SARAH 2 Testing - Hyper-Parameter Tuning. SARAH 2 went through hyper-parameter tuning in order to optimize the number of trees to use for probabilistic prediction of habits. The process described below yielded an accuracy score 82%.

For this hyper-parameter tuning, part of the data was used as the test set and was separated from the training data. The test set was utilized to compare predictions to tracked habits.

The resulting closeness of the predicted outcomes made by the training data set compared to the actual test-data enables us to be confident in the prediction made by our model.

3 Outcome - Player Matrices of Success Probability and Frequency

The outcome of this study is that each player will have their habits mapped out in a 2x2 matrix based on the amount of times that habit is exhibited (driven by event-data) and the success probability expected when that habit is completed (probabilistic figure uncovered in SARAH 2).

3.1 Frequency and Success Probability – Measurement Techniques

Frequency. This number is driven by the number of times a player exhibited that habit - which is uncovered through their time adjusted event data.

Example - A player with a significant volume controlled entry via pass or stickhandling (after establishing the connection between those events and the efficient use of crossovers as a habit) allows us to conclude that crossovers are frequently utilized by this player.

We can predict that a player will utilize crossovers habit more often because of this higher volume of event data.

Success Probability. The probabilistic figure between 0-1 discussed in SARAH 2 that provides a percentage probability that a player will complete that habit successfully when the opportunity presents itself, which is inherently related to driving impactful results on the ice.

3.2 Matrix Deep Dive - Quadrant Breakdown (Player Development Matrix)

As introduced in the public sphere by Jack Han in his newsletter[4], the matrix presented below has four quadrants, which is designed to enable player-development staff and scouts to identify the habits of strength and weakness for players. In its current form, skills on the the Player Development Matrix are estimated qualitatively and plotted on the chart. To instead quantitatively determine where skills should go on this matrix, we plot the calculated frequency against the success probability for each player. An example of this novel quantitative iteration of matrix is included below.

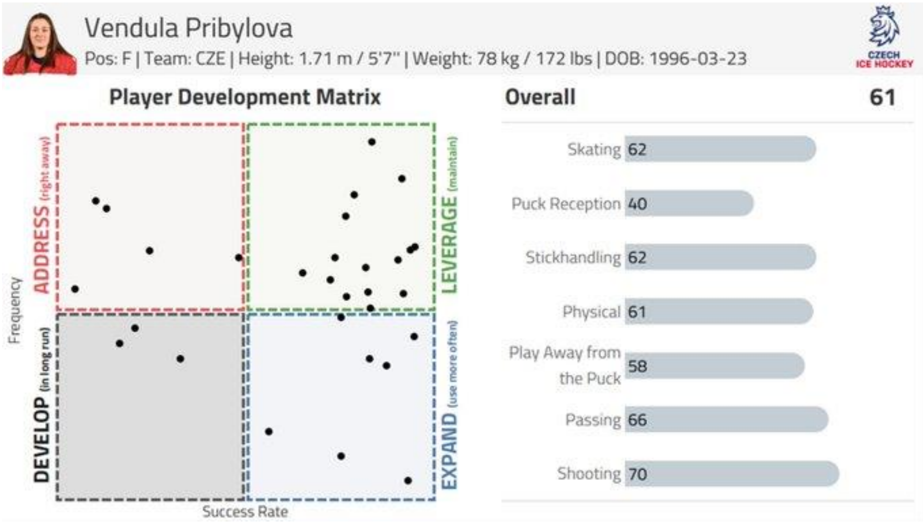


Fig. 3. The development matrix of Vendula Pribylova. Her data and development matrix has been included in this publication with her permission. The development matrix is used with permission from its creator, Jack Han.

A breakdown of the interpretation of the four quadrants is provided below:

Green Quadrant (LEVERAGE) - High success probability and high frequency; a player is expected to use this habit quite frequently and when completed it is done well (these are the skills that enable them to drive strong play).

Blue Quadrant (EXPAND) - High success probability and low frequency; these are habits completed well when attempted, but player development staff should encourage these habits to occur more often because they are being underutilized.

Red Quadrant (ADDRESS) - High frequency and low success rate; highest priority items to fix for player development given it occurs often but is done very poorly (high failure rates and likely holding the player back).

Black Quadrant (DEVELOP) - Low frequency and low success probability; staff should target long run improvement for these habits, the player does not have the opportunity to complete these habits often, but they are not executed well when the situation presents itself. This should be the lowest priority items for player development staff and may be unimportant to a player's archetype (i.e., grinder does not need to exhibit x skill).

Each matrix is relative to only that player's broader skill set. For example, Marie-Philip Poulin's red quadrant habits may still be elite in comparison to 95%+ of her opponents but it is weak relative to the rest of her habit score. The reason this matrix was created on a relative basis was to allow player development staff to focus on a personalized plan for each player, rather than the most elite players having almost no areas of improvement.

3.3 Skill Set Scores Methodology

In order to estimate the score on different skill sets, a weighted average calculation was used to incorporate both the effects of success probability and frequency of habits. As initially outlined in the tracking methodology, 7 different skill sets were determined with the goal of linking statistical techniques to more traditional scouting techniques (video analysis) containing the following habits.

As such, weighting was applied to the frequency of different habits in each skill set to calculate the average success probability for the skill set.

4 Conclusion and Future Works

In short, this paper introduces a new approach to linking traditional scouting methods to advanced and micro stats in hockey through an automated scouting tool that can be used to improve the quantitative evaluation and player development processes of organizations. In terms of future work, three possible model expansions that could be explored are the following:

- Developing a multi classification model combined with a non-binary habit tracking system would allow the incorporation positive impact (or lack thereof) of a habit to different degrees. For instance, a player that is developing a habit, while not fully mastering it could receive a score of 0.5 for said habit instead of simply limiting the choices to binary options (0 or 1).

- The current model could also be extended to identify player archetypes at the micro-habit level in order to characterize the strengths and the weaknesses of different groups of players more precisely.

- Finally, the idea of skill stacking could be incorporated into the modeling process in the form of interactions between the different habits and multilevel targets in SARAH 1 and 2 respectively.

5 Appendix

The code for this project can be found at https://github.com/mnahabedian1/WHKY-Player-Development-Project. The interactive player development matrix tool can be found at https://public.tableau.com/app/profile/mikael.nahabedian1483/viz/PlayerDevProject-PublicVersion/Dashboard32.

Below are the definitions for the habits included in Table 1.

5.1 Skating

Edgework Outside – Ability to access outside edges with ease (usually with a bow-legged basic posture).
Backwards Skating – Focus on pivot (without crossing feet) + stride mechanics yielding grip & smoothness.
Stride Recovery – Back leg just under full extension and recovers underneath the body to allow for recovery in the next stride.
Skating Mechanics – Knee flexion to generate power on each stride. Joints are stacked (shoulders, knees and toe caps form a line).
Crossovers – Use of crossovers when carrying the puck to change direction or build speed (every 4 to 5 strides).
Shouldering Speed – Movement patterns allowing smooth transition during changes of direction or to move from one play to the next.
Feet in motion – Following cutbacks or puck receptions, ability to create separation with the opponent.

5.2 Puck Reception

Catching puck in Hip Pocket – Ability to receive the puck on the side of the body (let it through body).
Dynamic Catch – Feet position (open) + catch in a weight shift or crossover.
Getting off the boards – Ability to catch the puck along the boards in a favourable posture to get away.

5.3 Stickhandling

Loading Puck to Hip Pocket – Ability to load the puck on the side of the body (good attack position).
Underhandling of Puck – Handling the puck efficiently without unnecessary stick motions.
Handedness Versatility – Being able to play the puck both on the forehand and backhand.
Deception w/ puck – Able to pull in players with the puck or give the illusion of making a specific play.

5.4 Physical

Initiating Contact – In board battles, willingness to initiate contact with the opponent to win the puck.
Puck Protection with Body – Ability to use body as a shield between puck and opponent.
Fitness Level – Overall ability to keep up with the pace of the game (& have reasonable shift lengths).

5.5 Play Away from Puck

Shoulder Checks – Making meaningful checks behind the play before retrieving the puck/in the DZ.

NZ Angling – Close space to ensure that threats are angled and neutralized in the NZ.
Unassisted Stops – Getting out of structure and swiftly killing plays early without opening seams in DZ.
Jumping in Shot Lanes – Purposefully & voluntarily jumping in front of shots in DZ.
Awareness without puck – Reading plays correctly yet understanding the purpose of playing inside structure.
Net Front Presence – Box out + goalie presence in DZ and OZ respectively.

5.6 Passing

Slip Passes – Ability to identify seams under or above the stick of opponents.
Leveraging & creating seams – Ability to create seams through movement and accurately leverage them.
Pass Placement – Ability to provide good pucks to teammates.
Vision – Ability to identify the best passing option.

5.7 Shooting

Coordination – Feet placement (front towards net) + application of downward force for accuracy/power.
Weight transfer – Transfer of weight to generate velocity on the shot.
Tip – Ability to tip shots/generate shots that are tip-able (usually low and through the defense).

References

1. Chevallier, B, Architecte Hockey, https://architecte-hockey.com/methods/
2. Han, J., The Hockey Tactics Newsletter, https://jhanhky.substack.com/
3. Belfry, D., Belfry Hockey, https://belfryhockey.com/
4. Han, J., https://jhanhky.substack.com/p/how-to-ruin-a-player?s=r
5. InStat Hockey, https://instatsport.com/hockey
6. NHL Central Scouting, http://www.nhl.com/ice/draftprospectbrowse.htm
7. Pronman, C.https://theathletic.com/3297887/2022/05/10/nhl-draft-lottery-ranking/
8. Mckeen's Hockey, https://www.mckeenshockey.com/draft/
9. Han, J. https://jhanhky.substack.com/p/upgrading-the-player-development
10. Scikit-learn, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
11. Scikit-learn, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

How Analytics is Changing Ice Hockey

Ulf Johansson, Erik Wilderoth, and Arsalan Sattari

Dept. of Computing, Jönköping University, Sweden
ulf.johansson@ju.se, erik.wilderoth@gmail.com, arsalan.sattari@ju.se

Abstract. While ice hockey is often considered to lag behind the other major sports in advanced analytics, the relatively straightforward metric Corsi has now been used for more than a decade. In this paper, we investigate how the introduction of Corsi and later xG has affected ice hockey. As seen from an extensive quantitative study, two different eras can be identified; the Corsi era, where the number of shots is the most important criterion, and the xG era where shot quality is prioritized. Looking at how the teams later performed in the playoffs, the analysis show that until approximately five years ago, regular season Corsi was the best indicator, but now it is instead xG. In the study, we specifically identify and reason about differences and similarities between NHL and SHL.

1 Introduction

Originating in Major League baseball, the utilization of advanced data-driven analytics has during the last decades become the norm in all major sports. While the derived information is valuable in itself, in particular for evaluating players, it is also obvious that the use of analytics has changed the approach of both players and teams. In this paper, we investigate how ice hockey has changed the last decade, arguing that the usage of analytics has played a big part in this.

More specifically, we look at two standard metrics often employed in ice hockey, Corsi and expected goals (xG), and see how the awareness of their importance has increased, ultimately affecting how ice hockey is played. Using Corsi first, we demonstrate that relatively high values generally indicate a successful season, in both the NHL and the SHL. After that, we see how the two leagues have actually evolved quite differently when trying to maximise Corsi. Finally, we show the rather significant effects of xG, over time, becoming recognized as more important than Corsi. In summary, the overall purpose of this paper is to look into how the number of shots taken in ice hockey has changed over time based on the current understanding of what makes teams successful.

2 Background and related work

In baseball, Sabermetrics, as made popular with the Moneyball [4] phenomenon, has lead to a number of drastic changes in team strategies. Two striking examples are the reduction in attempted sacrifice bunts and stolen bases. Looking at only the American League, where a designated hitter bats for the pitcher, the number of attempted stolen bases per team and game was in 2020 0.59, while the corresponding numbers for 1990, 2000 and 2010 were 1.01, 0.89 and 0.9. Similarly, the number of sacrifice bunts per team and game was 0.07 in 2020, which should be compared to 0.26, 0.20 and 0.24 in the years 1990, 2000 and 2010 respectively. The reason for this is that when analysing the effect of attempting to steal a base, it became obvious that the chance of success would need to be extremely high to make the decision to send the runner correct [2]. Regarding the sacrifice bunt, analytics discovered that in a large majority of all situations, even a successful sacrifice bunt will actually *reduce* the number of expected runs in that inning [7]. Another example is the frequent use of the so-called *defensive shift*, where the infield is positioned in an unorthodox way. Specifically, against a left-handed batter prone to pull the ball, three infielders are positioned to the right of second base, often with the second baseman playing very deep. While the success of the shift is somewhat questionable, see e.g., [5], it was in 2010 used in total 1707 times in the American League. In 2015, the number of at bats with a shift on was 14147 and in 2019 27592. In fact, traditionalist are now arguing for a ban of the shift.

In addition to these fundamental changes in strategy, it could be argued that players now approach the game in a different way. Specifically, pitchers are looking for more strike outs, and batters for more home runs. As a consequence, the proportion of at bats ending with a ball put in play has gone down significantly. The $K\%$, i.e., the number of strike outs divided by the number of at bats, has gone up from 17.5% in 2010 to 23.0% in 2019. At the same time, the proportion of at bats ending with a home run has increased from 2.85% to 4.16%.

In basketball, analytics, very simply put, showed that taking relatively hard shots inside the three-point line should generally be avoided. Instead, the shots should either be for three points, or taken from very close to the basket. As a consequence, the shot locations have changed dramatically during the last decade. Fig. 1 below shows the 25 most common shot locations for the NBA teams in the season 2006-2007 (Left) compared to the 2019-2020 season (Right).

Fig. 1. NBA shot locations from NBA.com as posted on Instagram by Owen Phillips

Ice hockey has traditionally been a conservative sport regarding analytics. Since Plus/Minus was introduced back in the 1959/60 season in NHL, it took more than 40 years for another metric to evaluate a player's contribution to the team except for scoring. When Alan Ryder came up with the *Player Contribution* in 2003 [6] and Tom Awads the *GVT* (Goals versus threshold) [1], that were two groundbreaking metrics. Both these metrics try to give one single value describing how good players are. Technically, the two metrics were based on goals, assists and Plus/Minus, i.e., still very rudimentary.

Since the 2010/11 season, NHL has published event data from all games. This enabled data-driven approaches producing metrics like *Corsi* and *Fenwick*, see [3]. According to Vollman [9] Corsi negates some of the major flaws of Plus/minus including, e.g., sample size, team effects, zone starts and goalkeeping.

Following Bill James in baseball, Vollman, who is since 18/19 hired by LA Kings as an senior analyst, started to write yearly editions of *Hockey Abstracts* to highlight the advances of hockey analytics [8]. As in baseball, this made the interest for quantitative approaches rise with both fans and teams. Consequently, the NHL organisations have the last couple of years expended their analytics departments a lot and by the season of 2021 there are 75 analysts hired by the 31 teams

3 Corsi in NHL and SHL - an historical view

The Corsi metric is very straightforward, simply calculating the attempted shots. Often, it is broken down into CF (Corsi for) and CA (Corsi against) with the obvious meaning. Sometimes it is aggregated into one number, CF%, which is $CF/(CF + CA)$, meaning that a team with a CF% over 0.5 has more shot attempts than their opponents.

In NHL, the goal of course is to win the playoffs, becoming the Stanley Cup champions. Table 1 below gives an overview of the importance of Corsi in the NHL. Interestingly enough, we see that having a good CF% rank is often more important than the regular season finish. This is true in particular for the earlier years, i.e., up to 2016, where the Stanley Cup champion often had one of the best CF% ranks in the regular season, and corollary, the best team according to CF% in the regular season very often made it to the final four.

Table 1. Corsi history NHL. Ranks are for the regular season.

Season	Champions (Reg. seas., CF%)	Regular Season Winner (CF%, end of the road)	Best CF% Team (end of the road)
07/08	DET(1,1)	DET (1, champions)	DET (champions)
08/09	PIT (8,19)	SJ (5, 1st)	DET (runner-up)
09/10	CHI(3,1)	WSH(3, 1st)	CHI (champions)
10/11	BOS(7,14)	VAN(6, runner-up)	SJ (conf final)
11/12	LA(13, 2)	VAN(7, 1st)	DET (1st)
12/13	CHI (1,4)	CHI(4, champions)	LA (conf final)
13/14	LA (9,1)	BOS (4, 2nd)	LA (champions)
14/15	CHI (7,2)	NYR (20)	LA (no playoffs)
15/16	PIT (4,2)	WSH (14, 2nd)	LA (1st)
16/17	PIT (2,16)	WSH (4, 2nd)	LA (no playoffs)
17/18	WAS (7,24)	NSH (8, 2nd)	CAR (no playoffs)
18/19	STL (12,10)	TBL (9, 1st)	SJ (2:nd)
19/20	TBL (3,5)	BOS (13, 2nd)	VGK (conf final)

We now, in Table 2 below, take a similar look at SHL (Swedish Hockey League), often considered the third strongest ice hockey league in the world after NHL and the Russian KHL. Here, Corsi data are only available for the 15/16 season and later, and it should be noted that for the 19/20 season, the playoffs were cancelled due to Covid-19. While the sample size thus is very small, it is interesting to see that the champions actually had the best regular season CF% in three of the four years.

Table 2. Corsi history SHL. Ranks are for the regular season

Season	Champions Regular season rank	Regular Season Winner (CF% rank, end of the road)	Best CF% Team (end of the road)
15/16	Frölunda (2)	Skellefteå (2, runner-up)	Frölunda (champions)
16/17	HV71 (2)	Växjö (3, quarter-final)	HV71 (champions)
17/18	Växjö (1)	Växjö (1, champions)	Växjö (champions)
18/19	Frölunda (3)	Färjestad (5, semi-final)	HV71 (quarter-final)

Based on this, the overall picture is that teams with high Corsi-values in the regular season have generally been successful in the playoffs. Specifically, CF% has been a much better indicator of how far the team will make it in the playoffs than the regular season finish, despite the fact that a high finish in the regular season by design leads to lower ranked opponents, and a home-field advantage.

4 Corsi development in NHL and SHL

We now address the question of whether the importance of high Corsi values, in particular CF%, has changed the way ice hockey is played. To answer this, we first look into how the number of shots, i.e., CF has changed over the years. To get unbiased results, we divide the number of shot attempts with the total time played with both teams at full strength. The values in Table 3 represent CF per 60 minutes. From these numbers, in particular when looking at the moving averages over the last three years (MA-3), the trend in NHL is quite clear; teams attempt more and more shots. In SHL, though, we see only small changes during the five years.

Table 3. CF development in NHL and SHL

Season	NHL		SHL	
	CF/60	MA-3	CF/60	MA-3
07/08	50.5	50.5		
08/09	53.1	51.8		
09/10	53.9	52.5		
10/11	55	54		
11/12	54.1	54.3		
12/13	53.8	54.3		
13/14	54.4	54.1		
14/15	54.4	54.2		
15/16	54.1	54.3	50.68	50.68
16/17	55	54.5	51.65	51.17
17/18	57.4	55.5	50.86	51.06
18/19	56.9	56.4	50.26	50.92
19/20	55.6	56.6	48.55	49.89

To further analyze this, we divide the teams into four categories based on their CF/60 and CA/60. In the NHL, we set the threshold to 55, i.e., Low represents values smaller than 55, and High values over 55. We use the following names for the categories:

- DULL: Low CF and Low CA
- BAD: Low CF and High CA
- GOOD: High CF and Low CA
- FUN: High CF and High CA

Fig. 2 below shows how the teams in NHL have developed over thirteen seasons. Starting with the earlier seasons, most teams are actually DULL. Specifically, in 07/08, no team is categorized as FUN. After that, and until the 17/18 season, there is a clear movement from the top-left quadrant (DULL) towards the lower right (FUN), i.e., most teams shoot more, but also receive more shots. In the 18/19 and 19/20 seasons, however, the trend is reversed, with teams leaving the FUN quadrant. Actually, in 19/20, a number of teams are again categorized as DULL.

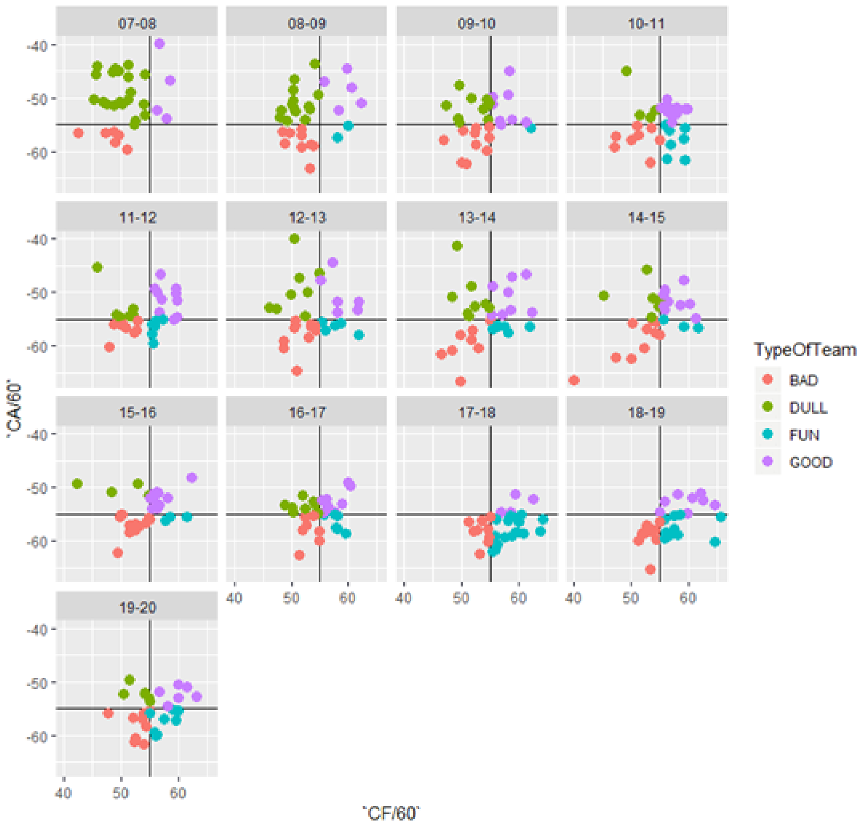


Fig. 2. NHL team development

Fig. 3 below presents the corresponding development in SHL. Here, however, since the number of shots is generally lower, due to the larger rinks, the threshold was set to 50 instead of 55. In SHL, the trend is actually quite different, with more and more teams appearing to minimizing the number of shots from the opponent, rather than taking more shots of their own. So, the two leagues take different approaches to maximizing CF%, in SHL the approach is more defensive, and in NHL more attacking.

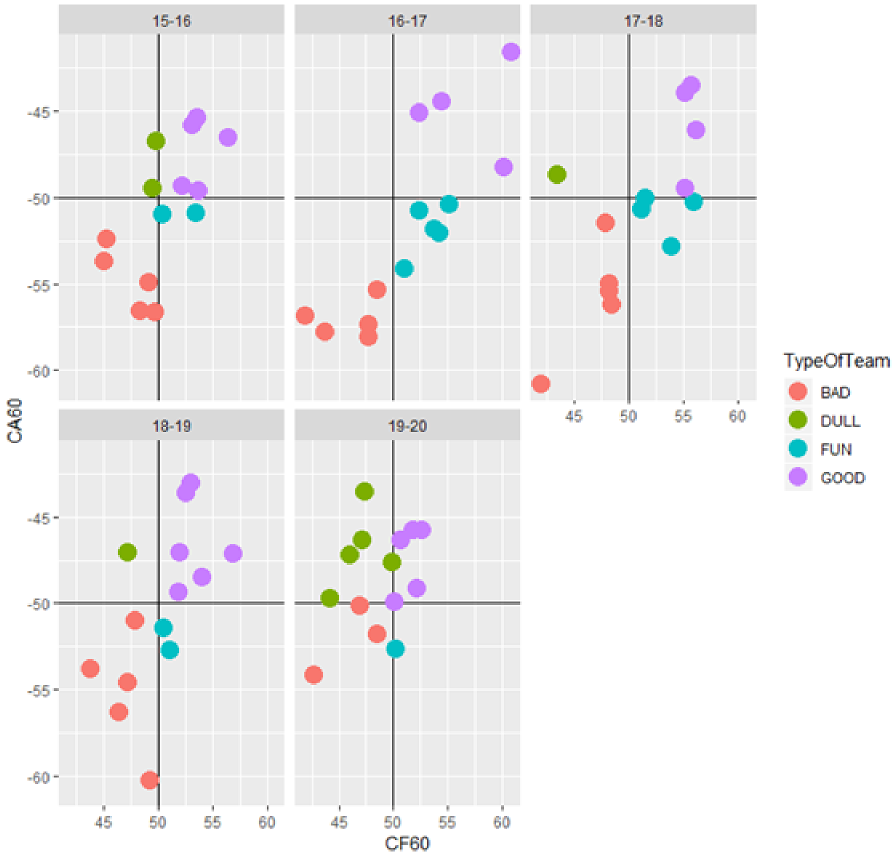


Fig. 3. SHL team development

5 The quality of shots – incorporating xG

The Corsi metric is blind to the quality of the shots. All attempts, regardless of the actual probability that it will score a goal, are taken into account. To incorporate shot quality, we add expected goals (xG) to the analysis. The xG of a shot is, loosely put, the likelihood of that shot scoring a goal, so the higher xG per shot, the higher the quality. Just by inspecting the relationship between CF/60 and xG/60 between the seasons 07/08 and 19/20 in Fig. 4, the change in quality per shot is obvious. We argue that this graph shows the rise and fall of the "Corsi Game". Between the seasons 10/11 and 15/16 the two lines are separated with CF/60 on top, i.e., while more shots were taken, the quality was low. From the season 15/16, however, we can see the rise of xG, and in the last seasons the xG line is for the first time actually higher than the Corsi line, showing that the quality of the shots has increased.

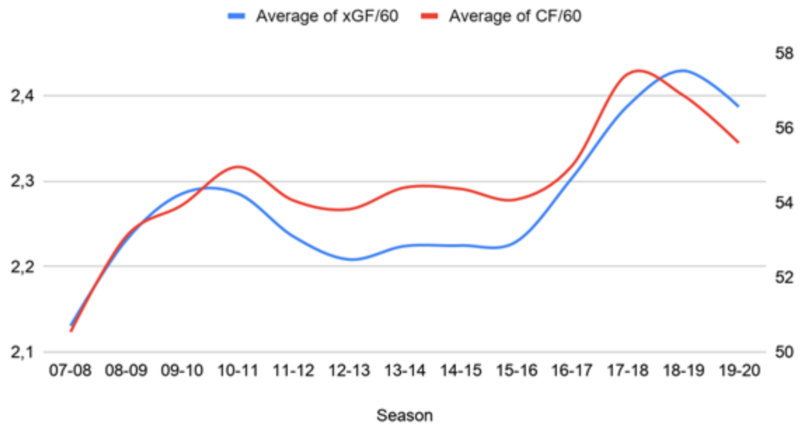


Fig. 4. Corsi vs. xG in NHL

Adding to this, we compare the success in the playoffs of the regular season winner, the best CF% team and the best xGF%. For this, we use a linear scale:

- 5 points: Stanley Cup Champions
- 4 points: Runner-up
- 3 points: Conference final
- 2 points: Second round
- 1 point: First round

Using this scale, Table 4 below shows the average points for the regular season winners, the best Corsi team and the best xG team, for the two different periods before and after the 14/15 season. While it should be noted that we only look at how individual teams fare in the playoffs, the differences between the two eras are striking. Specifically, in the Corsi era, the best CF% team averaged the conference final as the end of the road. In the xG era, it is barely a playoff team. On the other hand, in the xG era, the best xG team reaches almost one round further into the playoffs, on average. Actually, before the season 15/16 no Stanley Cup champion had ever had a higher rank in xGF% than CF%. After that season, no champion has had a higher CF% rank than xGF% rank.

Table 4. Corsi and xG eras

	Regular season winner	Best CF%	Best xGF%
Corsi era (until14/15)	2.75	3.25	2.00
xG era (after 14/15)	1.80	1.20	2.80

While a full analysis of how the quality of shots has increased is left for future work, we give two important explanations. First of all, as seen in Figs. 5 and 6 below, where the most common shot positions for the 30 teams in 2010 and 2018 are shown, shots are now generally taken from closer to the goal. Second, the number of one-timers has increased rapidly the last few years. Specifically, in NHL the increase is 30.9% during the last three years, and in SHL it is 11.6% for the last two seasons.

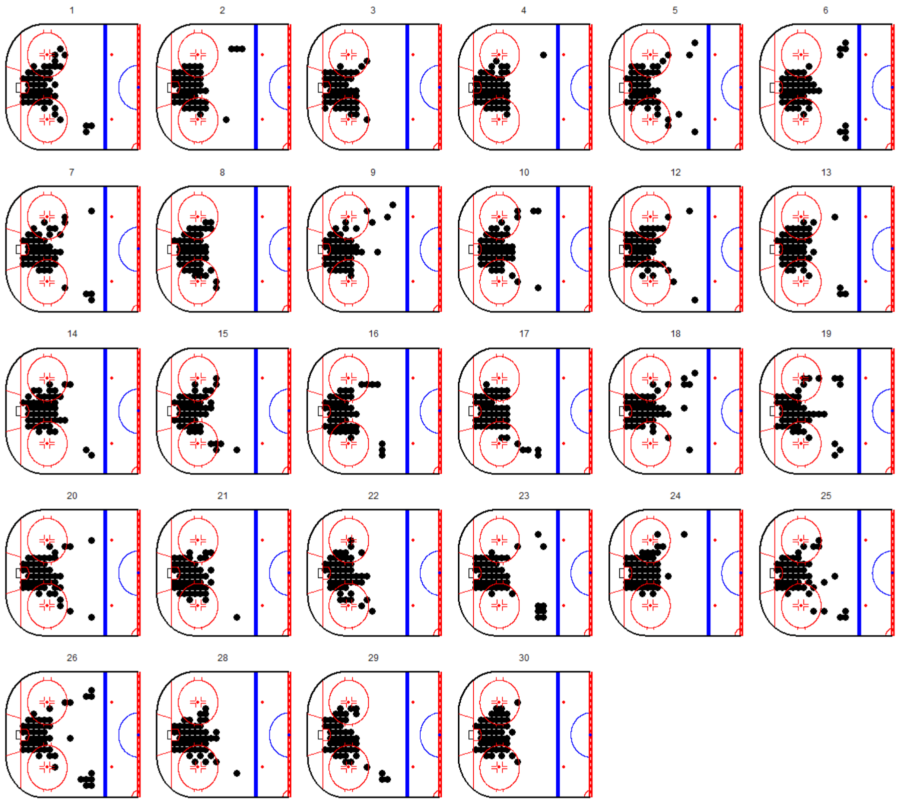


Fig. 5. NHL shot positions 2010

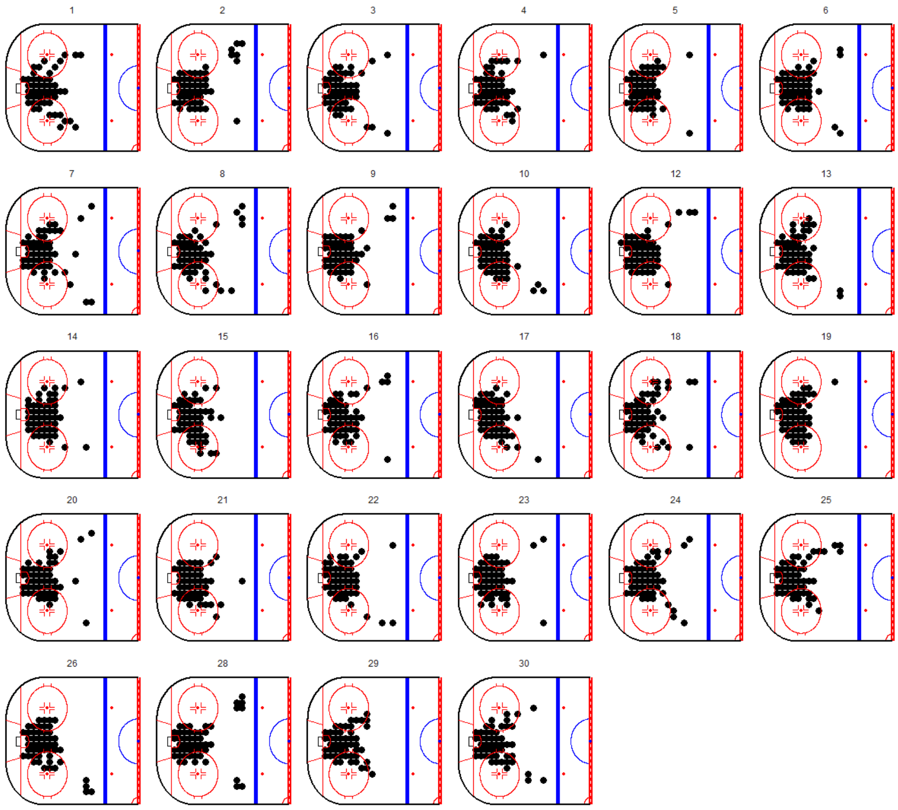


Fig. 6. NHL shot positions 2018

6 Concluding remarks

We have in this paper described how advanced analytics has influenced ice hockey. From the analysis, we identified two very different eras; the Corsi era and the xG era. In the Corsi era, the teams strived to take many shots, resulting in that the overall number of shots increased, especially in the NHL. In the last five years, however, the quality of the shots, as measured by xG, has become more important. The logic behind this is confirmed by comparing the playoff success of the best Corsi, xG and regular season teams. For many years, the best Corsi teams did in fact also have the most success in the playoffs, but now this position is taken over by the best xG team. Another strong indication is that shots in the NHL are now taken from closer to the net.

References

1. Awad, T.: Understanding GVT, part 3. Hockey Prospectus, Numbers on Ice, http://www.hockeyprospectus.com (2003)
2. Demmink, H.: Value of stealing bases in major league baseball. Public Choice 142(3), 497–505 (2010)
3. Hammond, R.: Understanding advanced stats, part one: Corsi and fenwick. https://www.matchsticksandgasoline.com (2011)
4. Lewis, M.: Moneyball : the art of winning an unfair game. W. W. Norton (2003)
5. Model, M.W.: Hitting around the shift: Evaluating batted-ball trends across major league baseball. Department of Statistical Science Honors Thesis, Duke University (2020)
6. Ryder, A.: Player contribution a method for allocating credit for a team's performance to the individual contributors on a hockey team. http://www.hockeyanalytics.com (2003)
7. Tango, T.M., Lichtman, M.G., Dolphin, A.E.: The Book: Playing the Percentages in Baseball. Potomac Books, Inc., New York (2007)
8. Vollman, R.: Rob vollman's hockey abstract, `http://www.hockeyabstract.com`
9. Vollman, R.: Hockey Abstract Presents... Stat Shot: The Ultimate Guide to Hockey Analytics. ECW Press (2016)

Predicting the NHL Draft with Rank-Ordered Logit Models

Brendan Kumagai, Ryker Moreau, and Kimberly Kroetch

Simon Fraser University, Burnaby BC, Canada
brendan_kumagai@sfu.ca, ryker_moreau@sfu.ca, kimberly_kroetch@sfu.ca

Abstract. The National Hockey League Entry Draft has been an active area of research in hockey analytics over the past decade. Prior research has explored predictive modelling for draft results using player information and statistics as well as ranking data from draft experts. In this paper, we develop a new modelling framework for this problem using a Bayesian rank-ordered logit model based on draft ranking data obtained from scouting sites and media outlets. Rank-ordered logit models are designed to model multicompetitor contests such as triathlons, sprints, or golf through a sequence of conditionally dependent multinomial logit models. We apply this model to a set of draft ranking data from the 2021 NHL draft and use it to provide a consolidated ranking for the draft and estimate the probability that any given player will be selected at any given pick.

1 Background and Motivation

Over the past two decades, the National Hockey League (NHL) has imposed a hard salary cap to limit player salaries and control a team's ability to retain and add talented players in an effort to enforce competitive balance throughout the league. This has forced teams to become increasingly savvy in how they allocate resources. The NHL has three main outlets where a team can add, lose or maintain talent: free agency, trades, and the entry draft. Acquiring players through free agency or trades can often be an expensive endeavour costing valuable cap dollars or assets. On the other hand, the draft is a low-risk, high-reward way to find and develop NHL-level talent.

Every NHL team employs a department of scouts to identify and evaluate the top draft-eligible players throughout the season and inform the team's draft selections each year. To strategize and obtain the players they desire, teams make assumptions on how long a player will last before being selected in the draft. Previous research has explored predictive modelling approaches for the outcome of the entry draft in both hockey [1] and other sports [2,3].

In this paper, we take a new approach to this problem by building a rank-ordered logit (ROL) model to estimate the probability that any given draft-eligible player will be selected at any given pick in the NHL draft. ROL models are typically used in sports that involve multicompetitor contests such as sprinting, triathlons, or golf. Primarily, our work was inspired by a discussion with Tyrel Stokes on this topic and his work with ROL models in the 100m dash [4].

In multicompetitor sports, there are generally dozens of major events per year that can be used to fit the ROL model and predict the outcome of future events. However, the NHL draft only occurs once a year and has a completely different crop of players each year. To address this issue we scrape draft rankings from various draft experts that provide ranking lists on scouting sites (i.e., Elite Prospects, Dobber Prospects, etc.) and media outlets (i.e., TSN, Sportsnet). We will refer to these media outlets and scouting sites hereinafter as 'agencies'. Additionally, we will refer to each ranking list from an agency hereinafter as a 'ranking set'. These ranking sets from various agencies are used as input into our model.

2 Methods

2.1 Multinomial Logit Models

We begin with a brief review of multinomial logit models. A multinomial logit (MNL) model is a method used in statistics to classify observations into one of two or more discrete outcome categories.

In particular, we are concerned with a special case of the MNL where we consider one trial (draft pick) being taken from c categories (available draft-eligible players). The goal of this model is to predict probabilities that each player is selected with a particular draft pick. In other words, we wish to estimate probabilities, $[\pi_1, \pi_2, \ldots, \pi_c]$
such that π_k is the probability of player k being selected with the draft pick of interest out of the c available draft-eligible players.

In the MNL model, these probabilities are derived as

$$\pi_k = \frac{\exp(\theta_k)}{\sum_{j=1}^{c} \exp(\theta_j)} \tag{1}$$

where θ_k is an 'ability' parameter for player k that we wish to estimate by fitting this model [5].

As an example, suppose we wish to model the outcome of the 1st overall pick in the 2021 NHL draft given draft rankings from various agencies. By using the 1st overall ranked player from these ranking sets, we can estimate the values of θ_k for all available players $k = 1, \ldots, c$, and consequently, obtain estimates for the probability that player k is selected 1st overall, π_k from (1), for all available players.

2.2 Rank-Ordered Logit Model

The MNL provides us with a simple framework for estimating the probability that a player is selected with the first pick in the draft, but there are still questions that this model cannot answer alone such as: What is the probability of a player being drafted 2nd, 3rd or beyond? How would these probabilities differ depending on which player(s) were selected prior? If a player is consistently ranked top 5 but is never ranked 1st, would his probability, π_k, of being selected 1st be the same as a player rarely ranked in the top 200?

These questions can be addressed using a rank-ordered logit model. A ROL model can be thought of as a series of conditional multinomial logit models where the 1st overall pick is modelled as a MNL model with a single pick from the pool of all draft-eligible players, then the 2nd pick is modelled as a MNL model with a single pick from all draft-eligible players excluding the player selected 1st, and so on until the nth player, who is modelled using the MNL model with a single pick from all draft-eligible players excluding the $n-1$ players that have already been selected.

To define this model, let θ_i be the underlying ability parameter for player i and let Y_i be the latent evaluation of player i's ability by the agency that developed the ranking set.

A key assumption in this model is that the latent evaluation by the agency is a realization from a Gumbel distribution with a location parameter of θ_i and a scale parameter of 1. That is, $Y_i|\theta_i \sim \text{Gumbel}(\theta_i, 1)$ [6]. If we let the true performance Y_i equal $\theta_i + \epsilon_i$, where ϵ_i is an error term, then this assumption is equivalent to assuming that the distribution of the error is Gumbel with $\mu = 0, \beta = 1$ where μ and β are the location and scale parameters of the Gumbel distribution, respectively. The convenience of this assumption is made clear by Luce and Suppes [7], who show that a Gumbel assumption of the errors implies a logit formula for the choice probabilities; furthermore a logit formula for the choice probabilities implies a Gumbel distribution for the errors [8]. In practice, this assumption is almost identical to an assumption of independent, normal errors, although extreme value distributions have fatter tails [9]. This assumption allows us to define the likelihood for a single draft ranking set in this model as

$$P(Y_1 > Y_2 > \cdots > Y_n \mid \theta_1, \ldots, \theta_n) = \prod_{i=1}^{n-1} \frac{\exp(\theta_i)}{\sum_{j=i}^{n} \exp(\theta_j)} \tag{2}$$

For example, consider a ranking set by TSN. Suppose TSN ranks Shane Wright 1st, Logan Cooley 2nd, and Juraj Slafkovsky 3rd, and θ_1, θ_2, and θ_3 correspond to Wright, Cooley and Slafkovsky's underlying abilities, respectively. This implies that Y_1, Y_2, and Y_3 correspond to the TSN evaluation of Wright, Cooley and Slafkovsky's abilities, respectively, where $Y_1 > Y_2 > Y_3$.

We do not observe these scores directly from any ranking sets. However, we operate under the assumption that some sort of rating scale exists for each ranking set. To add some intuition behind the latent Y_i's, imagine that the scouting team at Elite Prospects gets together and collaboratively comes up with a player grading scheme with scores ranging from 0-100. They may have scored Wright as 93/100, Cooley as 89/100, Slafkovsky as 88/100, and everyone else as 86/100 or below.

2.3 Accounting for Unranked Players

We can improve on the basic rank-ordered logit model specified in Section 2.2 by accounting for unranked players in our model likelihood.

Consider two ranking sets. In ranking set A there are 32 players ranked; Aatu Räty is ranked 8th while Fyodor Svechkov is ranked 20th. In ranking set B there are also 32 players ranked; Aatu Räty is not ranked in the top 32 while Fyodor Svechkov is ranked 22nd.

When we attempt to fit this model and estimate the θ_i's, the likelihood from the base ROL model as defined in Section 2.2 will take into account that Räty ranked 8th in set A but will not penalize Räty for being unranked all together in set B. On the other hand, the likelihood will take into account that Svechkov was ranked 20th and 22nd in sets A and B, respectively.

This example highlights an issue with the basic ROL model in the NHL draft setting. Players with more volatile rankings (i.e., players that are ranked highly by some agencies and are left unranked entirely by others) will have overestimated ability parameters because the cases where they are left entirely unranked do not factor into the likelihood at all.

To address this, we leverage the extension to the rank-ordered logit model for ranking the top m competitors out of a pool of M total competitors as outlined by Fok et al. [10]. The likelihood for a single draft ranking set in this case is expressed as follows:

$$P(Y_1 > Y_2 > \cdots > Y_m > \max(Y_{m+1}, \ldots, Y_M) \mid \theta_1, \theta_2, \ldots, \theta_M) = \prod_{i=1}^{m} \frac{\exp(\theta_i)}{\sum_{j=i}^{M} \exp(\theta_j)} \tag{3}$$

Here we assume that a ranking set ranks m players out of a pool of M total players available. Referring back to the above example, this would now account for the fact that Aatu Räty was unranked in ranking set B and adjust his θ_i estimate accordingly.

2.4 Considering Changes in Rankings Over Time

At the beginning of the 2020-21 season, Aatu Räty was ranked as a likely candidate for the 1st overall pick. However, Räty struggled to perform well in his draft year and as the season wore on, he rapidly fell down every agency's draft rankings until he was eventually selected 52nd overall in the 2021 NHL draft.

Suppose we were in the days leading up to the draft in June 2021, and ranking set A from September 2020 had Räty ranked 1st overall, while ranking set B from May 2021 had Räty ranked 45th overall. Using the ROL model as we have defined it so far would allow both ranking sets to influence the θ_i estimates equally. However, ranking set B is likely more relevant to how the draft will play out in reality since it was built with an entire season of information that ranking set A did not observe.

This can be addressed by allowing player abilities to vary over time by assuming that the θ_i's follow an autoregressive process through the season as done in Glickman and Hennessey [11]. To do so, we divide the season into time periods. Typically, this could be done according to key dates throughout the season, but the 2020-21 season had inconsistent scheduling across leagues due to COVID-19. We thus split the season into four three-month time periods as follows:

$$t = \begin{cases} 1, & \text{if between 2021-11-01 and 2021-02-01} \\ 2, & \text{if between 2021-02-01 and 2021-05-01} \\ 3, & \text{if between 2021-05-01 and 2021-07-23.} \end{cases}$$

We define $\boldsymbol{\theta}_t$ as the ability parameters for all players in time period t. Recall that the autoregressive process assumes that

$$\boldsymbol{\theta}_{t+1} = \nu\boldsymbol{\theta}_t + \boldsymbol{\delta}_{t+1}$$

$$\boldsymbol{\delta}_{t+1} \sim \mathcal{N}(\mathbf{0}, \tau^2\boldsymbol{I}).$$

Essentially, the ability parameter from the previous time period, θ_{it}, is regressed towards zero by the autoregressive parameter $\nu \in [0, 1]$ while varying by the random $\delta_{t+1} \sim N(0, \tau^2)$ component to obtain the updated $\theta_{i(t+1)}$.

3 Model Setup

Now that we have laid out a ROL model for the NHL draft, we can move on to implementing the model in R [12] and Stan [13]. We opted to use Bayesian inference to fit this model as it involves a complex autoregressive hierarchical structure that is beyond the scope of any current ROL model packages available in R. The computation time for this model took approximately 55 minutes to run using the 'sampling' function from the 'rstan' package in R [14].

The likelihood used in our ROL model is simply the product of (3) from Section 2.3 over all draft ranking sets in all time periods as defined below. Here, K_t represents the number of draft ranking sets from time period t with m_{kt} and M_{kt} representing the total number of players ranked and the total number of draft-eligible players available to be ranked from our database, respectively, in the kth ranking set of the tth time period.

$$L(\boldsymbol{\theta}, \nu, \tau) = \prod_{t=1}^{3} \prod_{k=1}^{K_t} P(Y_1 > \cdots > Y_{m_{kt}} > \max(Y_{m_{kt}+1}, \ldots, Y_{M_{kt}}) \mid \boldsymbol{\theta_t}) \tag{4}$$

We assume a simple multivariate normal prior on the ability parameters in the first time period, $\boldsymbol{\theta}_1$. Each subsequent time period leverages the autoregressive process described in Section 2.4 to set a prior on $\boldsymbol{\theta}_t$, $t = 2, 3$. Additionally, we assume hyperpriors on ν and τ of Unif(0,1) and Inv-Gamma(2,1), respectively.

Since the variance of $Y_i|\theta_{it} \sim \text{Gumbel}(\theta_{it}, 1)$ will remain constant at $\frac{\pi^2}{6}$ for any value of θ_{it} [15], $\boldsymbol{\theta}_t$ is only identifiable up to an additive constant. To address this, we impose a constraint on the model that all player ability parameters in a given time period must sum to zero. As a result, the ability parameters should be interpreted as ability relative to the other players being considered.

4 Results

4.1 Parameter Estimates

We obtain estimates for the player ability parameters, θ_{it}, in each time period via posterior distributions from our Bayesian ROL model. Figure 1 displays the top 32 players based on their posterior means of θ_{i3}. These ability estimates allow us to get a consolidated draft ranking based on our input data and determine the most likely draft outcome (by ordering abilities from greatest to least).

4.2 Draft Simulations

These player ability parameter estimates are much more powerful than a tool for basic comparison between players. We can also use these abilities to estimate the probability that player i will be selected with the next pick given the remaining pool of players $i+1, \ldots, M$ available at that pick. This probability can be expressed as the following equation:

$$P(Y_i > \max(Y_{i+1}, \ldots, Y_M) \mid \boldsymbol{\theta}_t, Y_1 > Y_2 > \cdots > Y_{i-1}) = \frac{\exp(\theta_{it})}{\sum_{j=i}^{M} \exp(\theta_{jt})} \tag{5}$$

With the player ability parameters estimated, we can now use (5) to simulate entire drafts. At each pick we use (5) to calculate the probability of each remaining player being selected at the pick of interest, then use these probabilities to

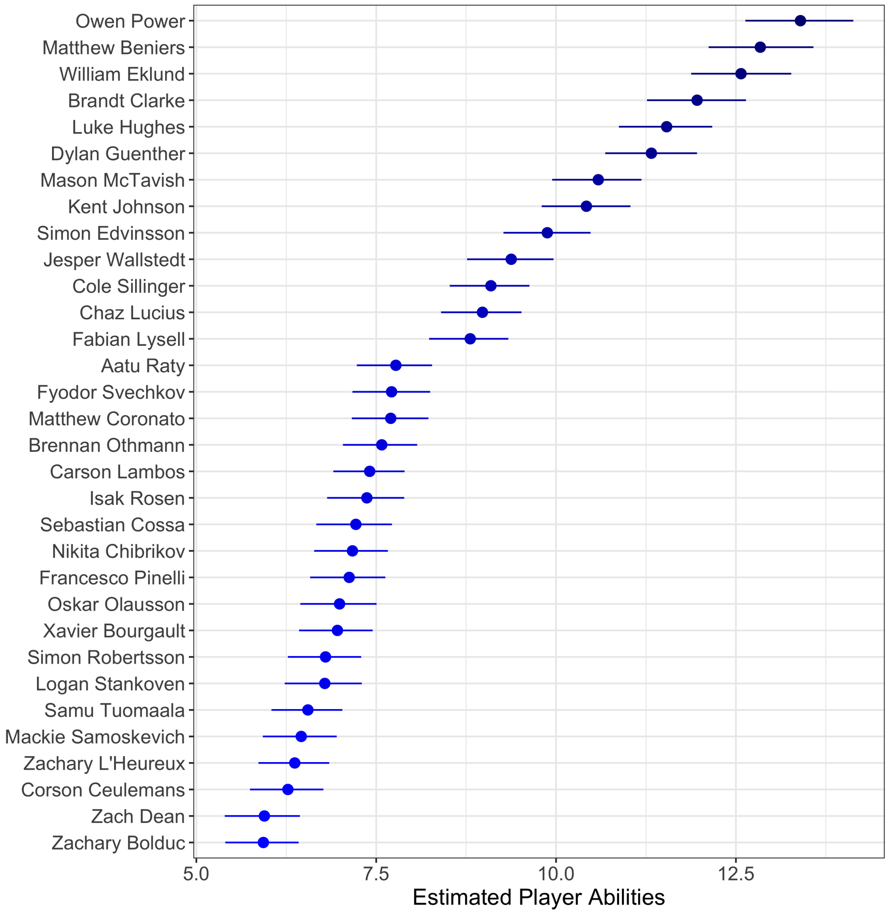


Fig. 1. Top 32 players in the NHL draft based on ability parameter posterior estimates from our rank-ordered logit model in time period 3 (2021-05-01 to 2021-07-23). Points represent the posterior means of θ_{i3} for player i; lines represent the corresponding 95% credible intervals of the posterior.

take a multinomial draw of size 1 from the remaining players to simulate the next player selected.

Figure 2 provides an illustration of the probability distribution of pick/player combinations in the 2021 NHL draft as determined by these draft simulations. The probabilities are determined by taking the total number of cases where a player was drafted in a certain position and dividing it by the total number of simulations. We ran 10,000 simulations of the NHL draft based on posterior draws to produce this visual.

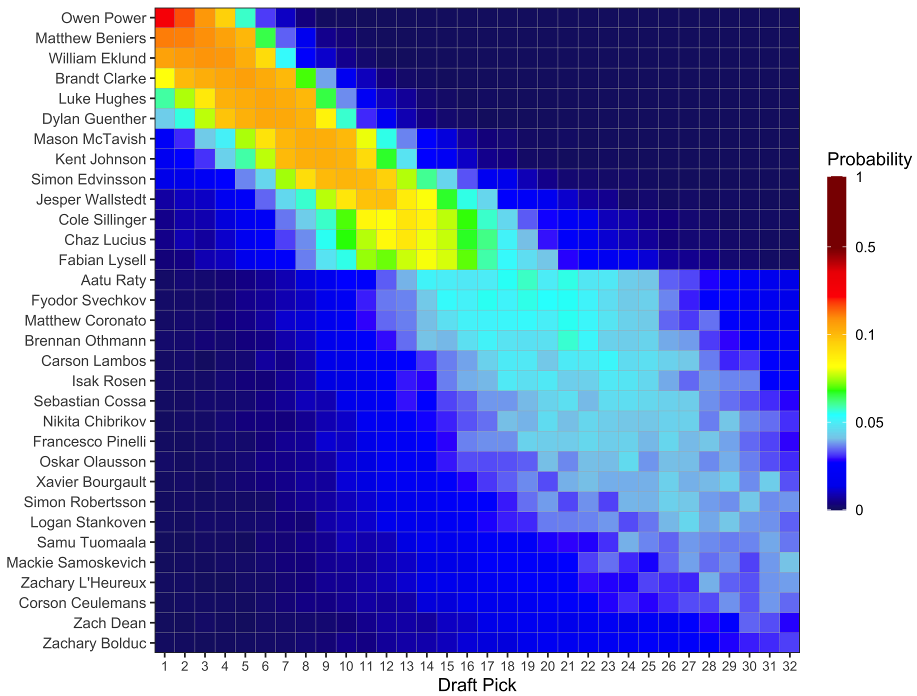


Fig. 2. A visualization of the probability that any of the top 32 players are selected with any of the top 32 picks; the colour of a square indicates the probability that the corresponding player will be picked with the corresponding pick.

The purpose of these draft simulations is to estimate the probability that a player is selected at any given draft pick. Ideally, we would compute this directly by calculating the probability for every possible permutation of the draft then summing up the total probability that player i is selected at pick j for all pairs of i, j; however, this is computationally infeasible. Assuming we consider 400 draft-eligible players and select 224 (7 picks for each of 32 teams), there are ${}_{400}P_{224} = 3.23565 \times 10^{548}$ possible draft outcomes. By simulating the NHL draft 10,000 times we can gain estimates of these probabilities without as much of a computational burden.

4.3 Player Ranking Distributions

Upon simulating the NHL draft using the posterior estimates of the player ability parameters, we can obtain discrete probability distributions for the pick number at which a player will be selected, which we call a 'player ranking distribution'. For example, Figure 3 displays the player ranking distributions for Owen Power and Matthew Beniers.

To provide an example of how this model can be used by a team, consider a team with the 7th pick in the draft. Lets suppose they believe Matthew Beniers is going to be a superstar. From the cumulative distribution function (blue) provided in Figure 3, we can see that the probability that he is selected prior to the 7th pick is roughly 90%. Thus, to have a better shot at selecting Beniers, the team would have to consider trading their 7th overall pick plus additional assets in order to acquire a higher pick in the draft where Beniers will have a higher probability of being available.

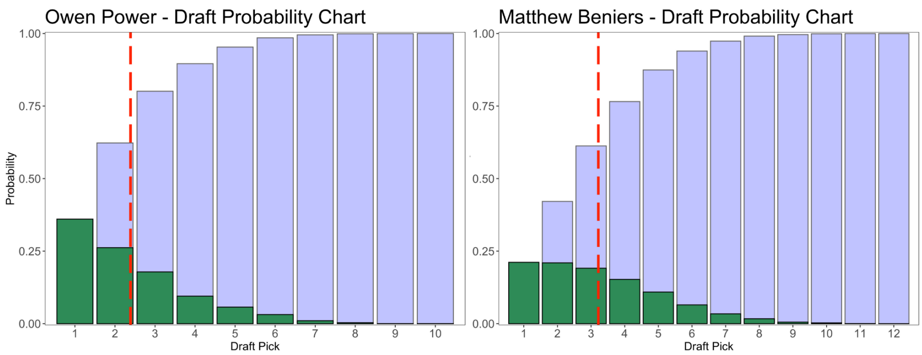


Fig. 3. Player ranking distributions for Owen Power (left) and Matthew Beniers (right) based on 10,000 NHL draft simulations. Red, dashed lines represent expected draft pick based on the distribution.

5 Concluding Remarks

In summary, we built a rank-ordered logit model based on NHL draft ranking data. This model allows us to estimate the ability of draft-eligible players relative to their peers, simulate draft outcomes, and estimate a probability distribution for the pick at which each player will be selected.

This model is still a work in progress and we feel there are many different routes that we can take to improve its performance and accuracy. Primarily, we intend to model the ability parameter θ by a linear predictor of player covariates with coefficients that assume a hierarchical structure to allow the model to adjust for team and agency tendencies. We expect that both agencies and teams will value particular traits (such as skating, shooting, passing, grit, etc.) differently and teams may draft players to address certain team needs (e.g., draft a defenceman when their roster and prospect pipeline are lacking talent on defence).

Additionally, we do not directly address the between-ranking correlation due to communication/collaboration between agencies. Two agencies may share thoughts amongst each other and, as a consequence, bias each other's evaluation of certain players. This has not been acknowledged directly in our paper and is an area that we hope to address with future work.

Acknowledgements

We would like to thank Tyrel Stokes and Dr. Rachel Altman for their help with this project.

Tyrel was instrumental in helping us develop this idea, find resources to improve our model, and implement the model in Stan. We appreciate him taking the time to help us debug various issues with the multivariate priors in Stan, among other things.

Dr. Altman provided us with thought-provoking insights and questions regarding our model and was very generous with her time when discussing our ideas.

References

1. Sprigings, D.: NHL Draft Probability Tool. (2016). Retrieved from https://hockey-graphs.com/2016/06/08/nhl-draft-probability-tool/#more-9010
2. Robinson, B.: Grinding the Bayes: A Hierarchical Modeling Approach to Predicting the NFL Draft. (2020). In proceedings of the 2020 Carnegie Mellon Sports Analytics Conference. https://www.stat.cmu.edu/cmsac/conference/2020/assets/pdf/Robinson.pdf.
3. ESPN. *Draft Day Predictor Tool.* ESPN Analytics. (n.d.). http://espnsportsanalytics.com/about.html.
4. Stokes, T. Monaco_ranking [github repository]. (2022). Retrieved from https://github.com/tyrelstokes/Monaco_ranking.
5. Agresti, A.: An Introduction to Categorical data analysis. 3rd edn. New York: John Wiley (2019).
6. Train, K.: Discrete Choice Methods with Simulation. 2nd edn. Cambridge University Press (2009).
7. Luce, D., Suppes, P.: Preferences, utility and subjective probability. In: Luce, R. Bush, R., Galanter, E. (eds.) Handbook of Mathematical Psychology, pp. 249–410. John Wiley and Sons, New York (1965).
8. McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: Zarembka, P (ed.) Frontiers in Econometrics, pp. 105– 142. Academic Press, New York (1974).
9. Train, K.: Discrete Choice Methods with Simulation. 2nd edn. Cambridge University Press (2009).
10. Fok, D., Paap, R., Van Dijk, B.: A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. Journal of Applied Econometrics, **27** (5), 831–846. (2010). https://doi.org/10.1002/jae.1223.

11. Glickman, M., Hennessy, J.: A stochastic rank ordered logit model for rating multi-competitor games and sports. Journal of Quantitative Analysis in Sports, **11**(3), 131-144. (2015). https://doi.org/10.1515/jqas-2015-0012.
12. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2022). URL https://www.R-project.org/.
13. Stan Development Team: Stan Modeling Language Users Guide and Reference Manual, Version 2.21.0. (2022). https://mc-stan.org
14. Stan Development Team: RStan: the R interface to Stan. R package version 2.21.5. (2022). https://mc-stan.org/
15. Gumbel.: E.J.: Statistics of Extremes. Columbia University Press, New York (1958).

Identifying Completed Pass Types and Improving Passing Lane Models

David Radke, Tim Brecht and Daniel Radke

David R. Cheriton School of Computer Science, University of Waterloo

Abstract. The implementation of a puck and player tracking (PPT) system in the National Hockey League (NHL) provides significant opportunities to utilize high-resolution spatial and temporal data for advanced hockey analytics. In this paper, we develop a technique to classify pass types in the tracking data as either Direct, 1-bank, or Rim passes. We also address two fundamental limitations of our previous model for passing lanes by modeling 1-bank indirect passes and the expected movement of players. We implement our pass classification and extended passing lane models and analyze 198 games of NHL tracking data from the 2021-2022 regular season. We study the types of completed passes and introduce a new passing metric that shows about 59% of completed 1-bank passes have an equal or more open indirect passing lane than the direct lane. Furthermore, we show that our expected movement addition reduces receiver location error in over 94% of completed passes.

Keywords: Hockey · Passing · Metrics · Passing Lanes · Tracking Data

1 Introduction

Ball and player tracking systems have revolutionized soccer and basketball analytics with extensive implications for scouting, coaching, player development, and fan engagement. Recently, the National Hockey League (NHL) deployed a puck and player tracking (PPT) system that records the location of the puck and every player with high resolution and frequency (60 and 12 times per-second for the puck and players respectively). Traditional methods of performance evaluation in hockey have relied mostly on offensive events like goals and shots despite these representing only a small fraction of the actual game play. Hockey has lagged behind other sports in advanced analytics due to technical challenges caused by the fast pace, small puck, white-colored ice, and other hardware challenges [14, 13, 3]. However, the new tracking system broadens the scope of potential metrics, analysis, and performance evaluations in hockey.

Most of the game play in hockey involves puck possession and passing between teammates. Previously, we developed a model to quantify the availability of passing lanes for completed passes, which associated smaller values with more difficult (or less open) passes [8]. While that model effectively calculates the available space between a passer and receiver, it assumes passes can only be direct (i.e., they are not banked off of or around the boards) and it treats player locations as static with respect to the time the pass was initiated. In reality, players often use the boards to complete passes when direct passing lanes are small or unavailable and pass to where their intended receiver is expected to be instead of where they are at the time the pass is initiated. In this paper, we make the following contributions:

- We develop a model to classify completed passes in PPT data to be either Direct, 1-bank, Rim, or Other passes. This is required to apply passing lanes models that are appropriate for different types of passes.
- We extend our passing lane algorithm [8] to 1) model the available passing lanes for 1-bank indirect passes and 2) include the expected movement of all players while the pass is made.
- We analyze passes using PPT data from 198 games from the 2021-2022 NHL regular season and devise a new metric for comparing completed 1-bank indirect passes with the alternative direct passing lane to the same receiver. We also examine the improvement in receiver location accuracy of our expected player movement model.

2 Related Work

Numerous passing models have been developed for football (soccer) and basketball using tracking data. These are typically used to analyze aspects of the game, such as pass disruptions to defensive formations [5], the expected value of passes [2, 4], and the number of outplayed opponents by passes [11]. The focus of that work is on the impact of passes instead of the actual difficulty or risk associated with a pass, which could provide insight into decision making, skills, and player risk profiles. Expected pass completion models (xPass) have gained popularity in soccer and are used to estimate the probability of passes being completed, or the difficulty of a pass, using physics [9], logistic regression [7], Graph Neural Networks [12], and supervised machine learning [1]. While these models can give significant insight into a player's decision making and passing ability, they rely on data for incomplete passes or ball control which may be difficult to determine in hockey.

To model the availability of passing lanes without relying on data for incomplete passes, Steiner et al. [11] calculate the angle from the direct pass line to the nearest opponent, where smaller angles correspond to less available passes. This model is limited by not including opponents behind the passer or receiver and not scaling for pass length. In response to these limitations, in previous work [8] we defined four key requirements for a passing lane model: **1)** always assign a real numbered value, **2)** incorporate the area surrounding the passer and receiver, **3)** be asymmetric with respect to pass direction, and **4)** scale with respect to the pass length. Our passing lane model presented in [8] assigns a value to each pass (in $\mathbb{R}^+$) that defines how *open* a passing lane is and simultaneously satisfies all four requirements without requiring data for incomplete passes. In this paper, we extend this passing lane model in three ways. We classifying different types of passes from the PPT data, calculate passing lanes for 1-bank indirect passes, and model the expected movement of players.

2.1 Background

Puck and Player Tracking Dataset Location data is collected through tracking technology that is inserted into the sweater of each player (back of the right shoulder) and embedded into pucks. Location information contains x, y, and z-coordinates to record locations in 3-dimensional space. The x and y locations are relative to center ice (which is 0,0) and the z locations are relative to the surface of the ice. The PPT data is recorded at 60 locations per second for the puck and 12 locations per second for each player on the ice, resulting in a total of about 734,400 location readings of main interest in a 60 minute game. Additional location data is obtained once a second for players that are deemed to be off of the ice. The tracking data is accompanied by event data including shots, goals, faceoffs, hits, and completed passes among others. These event labels contain information about the time of the event and the identities of the players involved.

Passing Lane Model To the best of our knowledge, the passing lane model in [8] is the only attempt to quantify the availability of passing lanes in hockey. The model uses the spatial locations of players in PPT data to estimate the available space between a passer p and any receiver r. The model utilizes event labels in the tracking data which have been identified by the data collection company, SportsMEDIA Technology (SMT)[1].

For each passing event, the passing lane model constructs a teardrop-like passing lane shape (shown in Figure 1) between the x, y locations of a passer p and receiver r that simultaneously satisfies all four requirements listed in Section 2. The size of this lane is determined by the locations of opposing players, representing the space between p and r *without* opponents (i.e., the *open* space). The passing lane size and shape is described by a positive real-numbered value γ, where larger γ values represent a wider lane and more open pass.

To determine the value of γ for a pass, we initialize $\gamma = 0$ (the direct line from p to r, pr in Figure 1). In this paper, we relabel pr to be $\vec{pr}$ to use vector notation. Increasing γ expands the passing lane shape until the edge of the lane contacts the location of an opponent. For example, increasing γ in Figure 1 grows the passing lane from the blue, to the green, to the yellow shaded regions. Since opponent 1 (o_1) is contacted first by the growing shape, the passing lane from p to r is represented as the blue shaded region. The resulting γ value is determined to be the passing lane value (for efficiency, we implement binary search instead of unidirectional growth). In Figure 1, $\gamma = 0.6$ since it is the smallest γ value with respect to each opponent (i.e., o_1 was contacted by the growing passing lane first). While γ has no direct correspondence with completion percentage, values of γ can be compared across time, locations, or players. We refer the reader to [8] for a more detailed description about the original passing lane model.

[1] www.smt.com/hockey

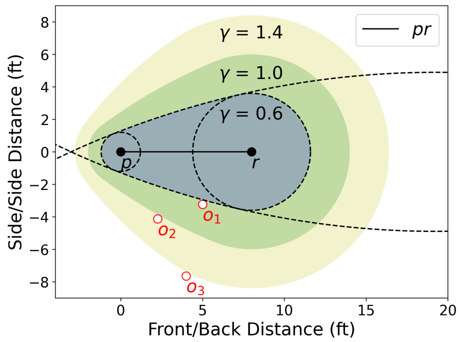


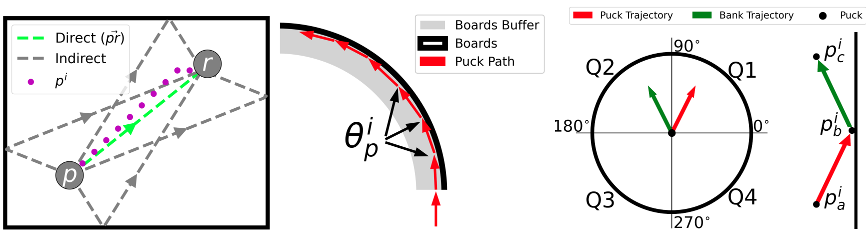
Fig. 1: Passing lane diagram from [8]. This example shows three passing lane shapes regulated by a parameter γ. The passing lane grows until the edge contacts the nearest opponent. The passing lane in this example has value $\gamma = 0.6$ and is the blue shaded region (the others are included as examples if o_1 or o_2 did not exist). In this work, we relabel the direct passing line pr as $\vec{pr}$.

3 Completed Pass Classification

The PPT data includes event labels to identify instances of a completed *pass*; however, there are multiple ways to pass the puck in hockey that should be modeled differently. A passer p can pass directly to r (a *Direct* pass), bank off a flat section of boards (a *1-bank* pass), or rim the puck around a curved corner of the surrounding boards (a *Rim* pass). We construct a model to identify these types of completed passes from the PPT data and overcome several challenges in the process. For example, there may exist some noise in the exact location of the puck (potentially from puck fluttering or position accuracy) or the time labels associated with passes. We found that the trajectory of passes cannot be assumed to compose perfectly straight lines, even for Direct passes. The puck may also contact the boards between consecutive readings of its location (i.e., the puck is traveling towards the boards at time t but traveling away from the boards at time $t+1$). Thus, the puck location never truly *contacts* the boards in the data. Our model uses a sequential filtering approach to differentiate between completed passes that are Direct, 1-bank, and Rim passes, and leave more fine-grained classification for future work.

Let $\mathcal{P}$ be the set of all passes in a game, $P^i \in \mathcal{P}$ be a single pass, and p^i be the set of x, y puck locations for P^i (origin at center ice). Our classification algorithm identifies: 1) Direct ($\mathcal{P}^d$), 2) Rim ($\mathcal{P}^r$), and 3) 1-bank ($\mathcal{P}^1$) passes in that order, each time reducing the set of possible passes to consider (starting at $\mathcal{P}$). The remaining unclassified passes compose a fourth class, *Other*, which we discuss in detail later. Since two consecutive readings close to the boards may represent actual contact with the boards (a challenge described above), all three phases use a value d_b, a distance from the boards, to construct a buffer that is used to determine puck readings that are sufficiently close to the boards.

1) Direct Passes: We identify completed Direct passes using two characteristics: 1) they may never be close to the boards and 2) have relative straight trajectories when compared with the possible indirect passes to the receiver. Our algorithm has two phases. First, if no points in p^i are within distance d_b of the boards, P^i is classified as to be Direct. Since Direct passes *may* also happen close to the boards, if any points in p^i are within distance d_b of the boards, we proceed to the second phase. If not identified as a Direct pass in the first phase, we determine the five possible paths for p to pass to r, ignoring corners (i.e., the direct path $\vec{pr}$, and off of both side-boards and both end-boards). Figure 2a shows this procedure in an example box (not-to-scale). The purple dots represent p^i, which has some change in direction near the receiver (i.e., likely contacting the receiver's stick before being considered *received*). To mimic actual puck behavior, we remove any of the five passing paths that contact the net or a rounded corner since the puck would not follow the projected trajectory following contact. We estimate the error from P^i to each projected path using the total Euclidean distance from each of the points p^i to each of the five possible paths. If $\vec{pr}$ has the least error, the pass is considered Direct.



(a) Direct (b) Rim (c) 1-bank

Fig. 2: (a) Project 5 ways for p to pass to r (excluding corners). "Direct" pass if the path with least error (Euclidean distance from p_i) is $\vec{pr}$ and/or puck is never within distance d_b of the boards. (b) Calculate puck direction changes in the rink corners. "Rim" passes have more than three direction changes in a corner that are greater than threshold t^θ. (c) Identify where the puck trajectory direction changes quadrants of the Unit Circle. "1-bank" passes have at most 3 of these points within distance d_b of the boards for specific changes of direction.

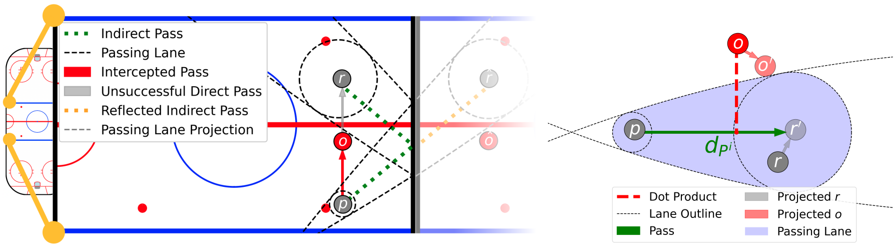
2) Rim Passes: The set of remaining completed passes are those not classified as Direct (i.e., indirect). Some indirect passes may be rims, where p directs the puck around a curved corner of the boards so the puck contacts the boards multiple times (Figure 2b). Our intuition to classify "Rim" passes is that the puck 1) changes direction multiple times and 2) these changes in direction are close to the corner boards. To calculate general puck direction vectors (and reduce change in direction noise), we average every 10 readings for the puck locations for passes $|p_i| > 10$ (red arrows in Figure 2b; shorter passes are not averaged). We calculate the difference in direction between adjacent vectors θ_p^i (in degrees), and define a threshold t^{θ} to determine if a direction change is sufficiently large. Since direction changes can have several causes (e.g., deflection from a stick, player, or referee) we only consider those direction changes that occur within d_b of the corner boards. In our implementation, a Rim pass is determined to have *greater than* three direction changes greater than $t^{\theta} = 4°$ within distance d_b of the corner boards. We choose three points since 1-bank passes should contain at-most three points with specific direction changes (explained next).

3) 1-bank Passes: From the remaining completed passes, we determine the set of 1-bank passes, where the puck only contacts a straight segment of the boards once. Since we are only detecting a single change in direction, the model for Rim passes is unable to be adapted since a change in the puck's direction of travel could happen for any number of reasons (i.e., deflections from sticks, inaccuracies in device readings, or inaccuracies with time labels associated with the pass). Therefore, we build on the intuition of detecting significant types of direction changes since a puck contacting the boards once will completely change its direction of travel. Our model draws on concepts from the quadrants of the Unit Circle in Trigonometry (Figure 2c left). To reduce the noise in the puck's trajectory we use an average of 10 consecutive readings (we do not use averages for short passes). For example, a sequence of 30 points could result in the three points p_a^i, p_b^i, and p_c^i shown on the right side of Figure 2c. We then calculate vectors between these points to determine the general direction of the puck (red and green vectors in Figure 2c). In the right of Figure 2c, the red vector (from point p_a^i to p_b^i) represents the puck traveling towards the boards (at 60°), and the green vector (from point p_b^i to p_c^i) represents the puck traveling away from the boards after the contact (now at 120°). Note that angles are relative to 0° which is the line perpendicular to the boards in this example.

We plot these vectors for the puck traveling to and from the boards on the Unit Circle shown on the left of Figure 2c. For a 1-bank pass, our model identifies the three points (p_a^i, p_b^i, and p_c^i) that comprise two consecutive vectors (red and green) where their directions appear in different quadrants of the Unit Circle ($0°, 90°, 180°, 270°$). It is not possible for a puck to contact a straight segment of boards and continue in the same quadrant of the Unit Circle. Therefore, 1-bank passes are classified if *three or fewer* points associated with puck direction vectors that are within distance d_b of the boards where the direction changes due to the boards (in the example in Figure 2c the angle of the vectors changes quadrants from Q1 to Q2).

4 Passing Lanes for 1-bank Passes

The original passing lane model only considers the direct line from p to r, which only represents Direct passes. For example, the model would consider the passing lane from p to r in Figure 3a extremely small (red arrow) because there is an opponent o directly on the path from p to r. However, a 1-bank pass can avoid the opponent o and is more open than the Direct pass. Our goal in this section is to model such passing lanes.



(a) 1-bank Passing Lane (b) Expected Movement

Fig. 3: (a) We calculate passing lanes for 1-bank passes by reflecting receiver r and opponent o about the boards. (b) We fit the passing lane to the expected movement of r and o using their locations, velocities, and expected pass distance.

Using the theory of geometric reflections, a 1-bank pass off the boards is geometrically equivalent to a Direct pass *through* the boards to a reflected representation of r ($\hat{r}$). This assumes the angle of incidence is equal to the angle of reflection which we acknowledge may not be completely accurate due to puck spin, fluttering, board imperfections, and variables such as drag and energy loss. However, our model is an approximation of the available passing lane instead of modeling the exact trajectory of the pass. We *reflect* all players besides the passer about the boards so that the 1-bank pass can be modeled as a Direct pass (in Figure 3a, green dotted line can be modeled as the orange dotted line extension). In the example in Figure 3a, we keep p at it's location and reflect r and o to locations $\hat{r}$ and $\hat{o}$ respectively. Using $\hat{r}$ as the location of r, we calculate the passing lane with respect to the nearest opponent (also considering their reflections). We acknowledge that 1-bank passes should be considered more difficult than Direct passes. This is accounted for in our model, as γ scales with the pass length and a 1-bank pass would be longer than the Direct pass.

To consider an in-game example when both teams are at even strength (no penalties), consider a passer p. Given any potential receiver r, p has the option to make a Direct pass, or bounce the puck off either side-boards or end-boards (e.g., shown in Figure 2a). Some of these lanes will make more sense than others, since a player is unlikely to pass the puck off their defensive end-wall when in the offensive zone. Since γ decreases as the length increases, excessively long 1-bank passes will receive very low γ values. We calculate γ for all five passing options from p to r with respect to all opponents. The largest γ value is the *most* open passing option for p to pass the puck to r. We expect a similar reflection-based methodology may work for Rim passes, but leave this for future work.

5 Expected Player Movement

To compute γ, the passing lane algorithm in [8] uses the locations of players taken at the time the pass is initiated. The asymmetry of the passing lane shape accounts for opponents closer to r having more time to react to a pass (i.e., skate towards the pass and/or move their stick in an attempt to intercept or disrupt the pass). Furthermore, receiver r will most often not be stationary and receive the pass at a different location than where they were when the pass was initiated. We expand the previous passing lane model to include the expected location of all players when computing the passing lane. In Section 6, we demonstrate how our new model improves the expected location of the actual pass reception. Our method calculates 1) the approximate distance of a pass P^i (d_{P^i}), 2) the expected speed of the pass ($s_{\vec{pr}}$), 3) the duration of the pass (t_{P_i}), and 4) the expected locations of receiver r and opponents o (r' and o'). Visualized in Figure 3b, we fit the passing lane from p to r' with respect to o'.

The approximate pass distance (d_{P^i}) is calculated using the Euclidean distance from p to where the pass is estimated to be received (r'), defined later. Given the approximate pass distance, we train a linear regression model on previous passes π to produce the expected speed of a pass with distance d_{P^i}, so that $\pi(d_{P^i}) \rightarrow s_{\vec{pr}}$ produces a positive real number. We calculate the duration of a pass as $t_{P^i} = \frac{d_{P^i}}{s_{\vec{pr}}}$ and use r's velocity vector (which includes direction) at the time of the pass $\mathbf{v_r}$ to determine their expected location, $r' = r + t_{P^i}\mathbf{v_r}$. This assumes r will continue in the current trajectory for t_{P^i} time.

Since opponents only have until the puck passes their location to disrupt the pass, we only project o's movement for time t_o, the time until the puck passes their location along the pass trajectory $\vec{pr}$. For this computation, consider the example in Figure 3b where o is located between p and r. Taking the dot product of o with respect to p and the direct passing line $\vec{pr}$ determines the perpendicular location of o onto $\vec{pr}$ (the red dashed line in Figure 3b). The expected time for the puck to reach this intersection is calculated to be t_o. If o is behind p, $t_o = 0$, and if o is behind r, $t_o = t_{P^i}$. We solve $o' = o + t_o\mathbf{o_r}$, where $\mathbf{o_r}$ is the velocity vector for o at the time the pass is made. Given the locations of p, r' and o', we calculate the passing lane using the algorithm from [8], shown as the blue shaded region in Figure 3b. In this example we only show one r and o for simplicity; however, we can calculate passing lanes for any r with respect to all opponents. This allows us to determine the receiver with the largest passing lane (i.e., the most open player). The tracking data does not include information on stick location, although this could be included if collected or estimated in future work.

6 Analysis

We implement our pass classification algorithms and passing lane extensions to analyze completed passes using a combination of the raw tracking data and labeled event data. Our dataset is from 198 games played in November of the 2021-2022 NHL regular season. We utilize the pass event labels in the dataset to determine when a pass was made; however, this dataset does not contain labels for passes that were not completed. Additionally, the automated labeling of events is a difficult problem; thus, the dataset may be missing some completed passes and/or include labels for events that are not actually completed passes. This dataset is still considered unofficial by the NHL, and may differ from other datasets that contain complete and/or incomplete passes (e.g., a hand labeled dataset). In this paper, we utilize the event labels provided in the dataset while including techniques to handle some, but not all, inaccuracies. We analyze features of our classification algorithm, 1-bank passing lanes, and expected movement extensions in isolation to identify interesting passing behavior and hypothesize about the potential performance of our models in the absence of ground truth data.

6.1 Pass Classification and Statistics

We first analyze general features of the completed passes in our dataset and how our classification model differentiates them. We observed $d_b = 2.5$ feet (ft) captures multiple adjacent puck readings for Rim and 1-bank passes that are close to the boards in most cases. Table 1 shows the results of our pass classification algorithm on the set of all completed passes. Since our tracking dataset only includes completed passes, the information may be biased towards successful events and not necessarily reflect the game as a whole (i.e., what was attempted and failed). As shown in Table 1, our model identifies 84.4% of the passes labeled complete in this dataset to be Direct, 10.2% to be 1-bank, 2.6% to be Rim, and 2.8% to be Other. Forwards as a whole tend to complete slightly more passes (49.5%) than defence (47.2%); however, when considering there are typically two defence and three forwards on the ice, a defensive player on average completes 43% more passes than a forward. We consider the relatively small percentage of unclassified completed passes (Other) to be acceptable, but is something we plan to examine in future work. After manually inspecting a significant number of these unclassified passes, we believe that most are either mislabeled as passes or consist of edge-cases that are difficult to identify (e.g., inaccurate timestamps resulting in odd changes in trajectory by a player).

Type	Direct	1-Bank	Rim	Other	Total
Forward %	41.9	4.8	1.2	1.5	49.5
Defence %	39.7	5.1	1.2	1.1	47.2
Goalie %	2.8	0.3	0.2	0.1	3.4
Avg/Game %	84.4	10.2	2.6	2.8	100.0

Table 1: Completed pass categorizations. Note that this data is based on events labeled as completed passes. Actual values may differ if labels are incorrect, missing and/or if incomplete passes are included.

Figure 4 compares the paths of completed 1-bank passes made by defence (left) and forwards (right). Darker green represents more 1-bank passes in that region. We see that most of the completed 1-bank passes initiated by defence are behind their own net or off the defensive half-walls. In contrast, the majority of completed 1-bank passes from forwards are made behind the offensive net or off the offensive half-walls (likely passing back to defence).

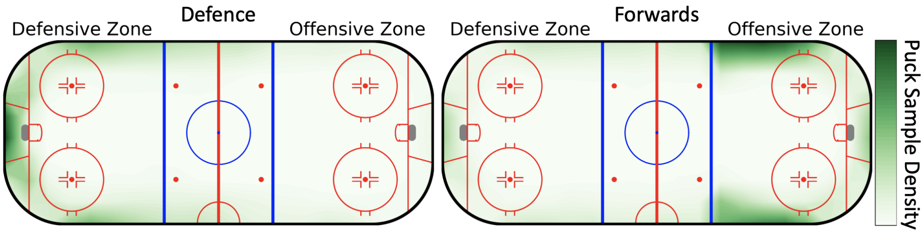
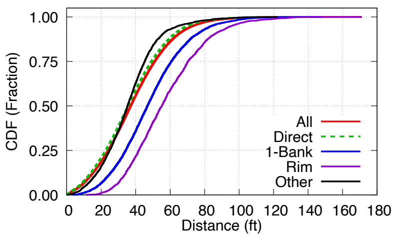
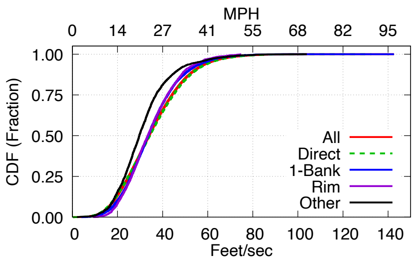


Fig. 4: Heatmap of completed 1-bank passes; defence (left), forwards (right).

Figures 5a and 5b show Cumulative Distribution Functions (CDF) for pass distance (puck travel distance) and pass speed for completed passes, calculated by comparing the total puck travel distance with the duration of the pass. For example, half (0.5) of all completed passes had a distance of about 38 ft or less, shown in Figure 5a. Interestingly, while the distances for completed indirect passes (1-bank and Rim passes) are typically longer than completed Direct passes, we observe almost no distinct difference between these classes for pass speed (Figure 5b). Thus, we hypothesize that players pass the puck harder towards the boards for indirect passes than they would for a Direct pass, to account for the expected energy loss from the boards. The distributions of this data may be different when considering all pass attempts.



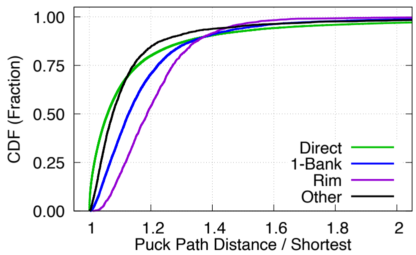
(a) Pass puck travel distances.



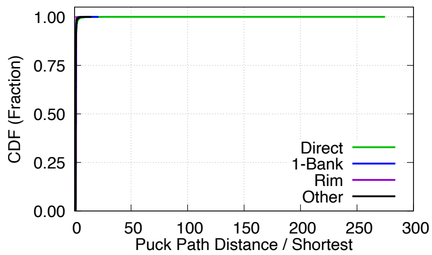
(b) Pass speeds.

Fig. 5: (a) CDF of pass length (distance traveled) for each type of completed pass. (b) Speed of completed passes for each type (distance traveled divided by total duration of the pass).

Figure 6a shows a CDF for the extra distance the puck traveled for each type of pass. We calculate this as $\frac{d_{puck}}{d_{p,r}}$, where d_{puck} is the actual distance the puck traveled and $d_{p,r}$ is the Euclidean distance from p to r (i.e., the shortest path for the puck). In theory, Direct passes should have the least extra distance traveled compared to $d_{p,r}$ and Rim passes must travel corners which accumulates more distance. Figure 6a shows that Direct passes generally do travel the least extra distance, followed by 1-bank, and Rim passes.



(a) CDF Extra Distance Traveled (zoom).



(b) CDF Extra Distance Traveled.

Fig. 6: (a) CDF comparing the distance the puck traveled to the shortest possible distance (Euclidean distance from p to r) for completed passes. (b) Zoomed out CDF to show the long tail, likely due to event labeling or classification errors.

We do note that the actual path of most Direct passes is longer than the shortest possible path ($d_{p,r}$). These passes are those with values on the x-axis greater than 1. Extreme examples of this can be also seen by the long tail in Figure 6b (an un-zoomed version of Figure 6a). We believe this is due to pucks being deflected by sticks or bodies (but the pass should still be considered Direct). Furthermore, inaccuracies in the timestamps of pass events also lead to add additional distances.[2] Motivated by these challenges, our first classification phase considers such passes Direct if the puck is not within distance d_b from the boards during the pass. We hypothesize that the Direct, 1-bank, and Rim ordering for extra distance in Figure 6a provides some insight into the accuracy of our classification algorithm despite these artifacts.

6.2 Passing Lanes for Indirect vs Direct Passes

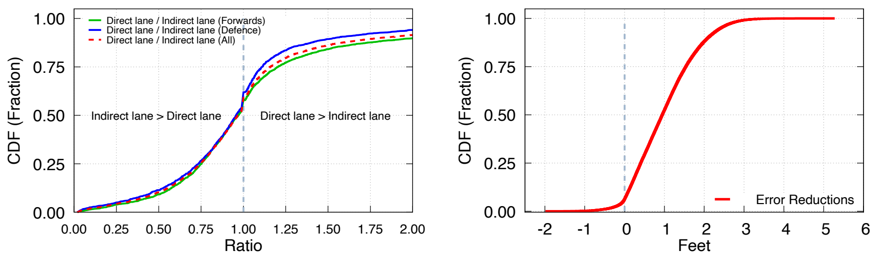
We now analyze our addition to the passing lane model for calculating 1-bank passing lanes. Without any ground truth for how open a passing lane is, our goal is to analyze how our passing lane model captures 1-bank passing behavior by comparing Direct and indirect passing lanes for completed 1-bank passes. For this analysis, we only consider the set of completed 1-bank passes for the reason that a more open indirect passing lane does not always indicate a better play and depends on the context of the game. For example, a player will likely opt

[2] By manually inspecting a significant number of these cases, we observed the timestamp at the end of the pass may occur after the pass was received and the receiver changed directions.

for a Direct pass on a 2-on-1 offensive rush instead of a 1-bank pass, even if the 1-bank is technically more open.

For each completed 1-bank pass, we calculate the value of the indirect 1-bank passing lane γ_i as well as the direct passing lane γ_d for p to pass to receiver r and define a new metric, γ-*ratio* $= \frac{\gamma_d}{\gamma_i}$. If the γ-*ratio* < 1, the indirect passing lane was more open than the direct lane, otherwise the Direct pass was actually more open. Figure 7a shows a CDF of the γ-*ratio* for completed 1-bank passes separated by player position for forwards and defence. We observe that about 59% of 1-bank passes were completed when the 1-bank passing lane was equal to or more open than the direct passing lane size (the γ-*ratio* ≤ 1). There is little difference between the behavior of forwards and defence when the γ-*ratio* < 1; however, when the γ-*ratio* > 1, defence tend to make more 1-bank passes when both lanes are similar (i.e., the γ-*ratio* closer to 1).

Note that in Figure 7a the x-axis is centered around 1 and is limited to a maximum of 2, since if γ_d is much larger than γ_i, the γ-*ratio* grows instead of trending to zero. For an in-game scenario, Figure 8 (left) in Section 7 shows how our model captures the 1-bank passing lane from Player #86 (who has possession of the puck) up to Player #3, whereas our previous model [8] does not. For this pass, the γ-*ratio* $= \frac{0.23}{0.46} = 0.5$ and completing this pass increases the subsequent passing lane to Player #28 from $\gamma = 0.3$ to 0.98 (right side of the figure).



(a) The γ-*ratio* by position. (b) Expected movement improvement.

Fig. 7: (a) CDF of the γ-*ratio* to show the fraction of completed 1-bank passes where the 1-bank passing lane was more open (< 1) or direct lane was actually more open (> 1). (b) Location of r error improvements with expected movement.

6.3 Player Movement

Our motivation for including expected player movement when model passing lanes is to better fit the shape of the passing lane to the location of the receiver when they receive the pass (and opponents to where they would be when the puck passes their location). For the set of all completed passes, we have the labeled location of the receiver at the time when the pass is considered received (r_t^*). Therefore, we calculate the difference between r_t^* and their location when modeled with expected movement (r'_t) and without expected movement (r_t). We calculate the two location errors as the Euclidean distance between 1) r's true location and their projected location with expected movement (r^*_t and r'_t), and 2) r's true location and their location without expected movement (r^*_t and r_t). The difference of these two errors provides insight into whether or not the expected movement model better estimates the location of r when they receive the pass (i.e., there less error). Figure 7b shows a CDF for the difference between these two errors, where positive values correspond with expected movement reducing the location error by the distance along the x-axis (more accurate location of r). We find that expected movement reduces the error of r's location for over 94% of passes, in one case up to 5.3 ft, and increases error in a small fraction of passes by a small amount (at most up to 2.0 ft).

7 Potential Applications

The influx of data in professional sports has given broadcasters and fans the ability to absorb more information about an event, such as shot speed, shift length, or face-off win probabilities; this information is typically presented by overlaying graphics or augmented reality (AR) on the live video broadcast. Our passing lane model can also be used in this context to display the most available passing option for one or more teammates, or the γ of a successful pass. Furthermore, our model could provide more fine-grained metrics that may be useful in fantasy sports or gambling applications. This can increase fan engagement and enjoyment by drawing attention to player formations and passing options.

When reviewing video of games, our passing lane model would give players and coaches quantitative data for the availability of passing lanes to devise new plays or assess performance. For example, the "up-and-over" is a common powerplay sequence to shift the defence to a new side of the ice and open passing lanes to certain players, shown in Figure 8. Using our models, coaches would be able to adjust the location of offensive or defensive players to find positioning to increase passing lane sizes, or to reduce the size of an opponent's passing lanes.

While GMs are tasked with constructing rosters and assessing players, watching every game or shift of a player is often infeasible and current metrics (such as goals and points) provide only a coarse view of player performance skewed towards offense. Our passing lane model could quantify passing behavior in a game or across a season (for assessing consistency). If augmented with incomplete passes, our model could determine how often players force passes when a more open alternative is available and provide insights into the passing skills of players (e.g., whether players manage to complete passes with smaller lanes).

8 Discussion

The high fraction of the γ-*ratio* ≤ 1 (59%) shows that NHL players in our dataset typically complete 1-bank passes when the indirect lane is larger or equal to

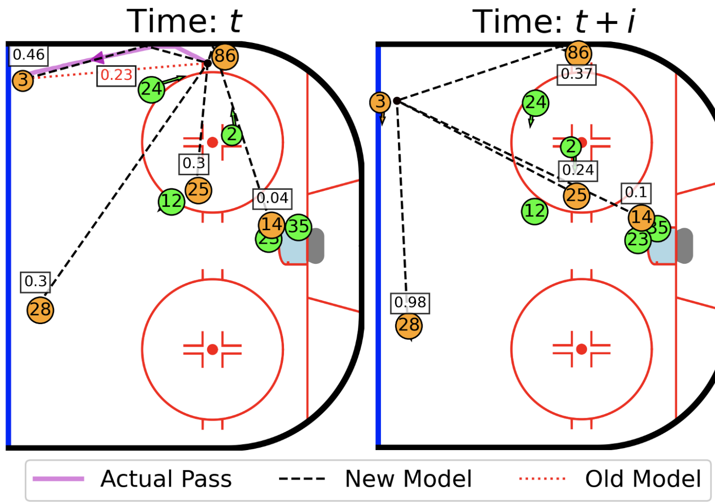


Fig. 8: Powerplay scenario for the Orange team, showing the best passing lanes to each player at times t and $t+i$. At time t (left figure), Player #86 has the puck. Our new passing lane model identifies the 1-bank lane to Player #3 as being the most open (twice as large as the direct lane). Player #86 chooses this lane for their pass (purple line). At time $t+i$ (right figure), after Player #3 receives the pass, the cross-ice lane to #28 increases from 0.3 to 0.98 (a factor of 2.3). Completing this pass is known as an “up-and-over” on the powerplay.

the direct lane defined by our model. Reducing the location error for r in the majority of completed passes (94%) shows that expected movement better aligns with where NHL players pass the puck than when it is not included. However, our analysis has several limitations that are important subjects of future work.

First, since our dataset only contains completed passes, our analysis may not accurately reflect the full behavior of all attempted passes. Another potential application of our model may be to identify incomplete passes based on the movement of the puck; however, this is beyond the scope of this paper.

Second, more accurate time labels for the start and end of passes would improve the precision and scope of future passing models. More accurate time labels would also improve the ability to calculate the speed of passes which has implications on the pass speed model we use for expected movement.

Third, future datasets could allow for more concrete evaluations such as calculating classification accuracy and lead to the development of new models. A ground truth dataset of pass types could be used to evaluate the accuracy of our classification model and allow our system to learn classification thresholds directly from data instead of observing and defining values. Furthermore, a dataset of incomplete passes could help analyze correlations between game context, γ values, and pass completion probabilities.

Fourth, extensions to the current passing lane model could explore a series of different directions. Rim passing lanes are a natural extension of this work. We could further improve the passing lane model to include more advanced methods of expected movement, such as predicting a player's movement with machine learning (i.e, ghosting) [6], physics-based approaches used in soccer [10], or considering handedness, reach, and stick length. When modeling the expected speed of a pass, a future iteration may consider personalized pass profiles by observing previous passes only by a specific player, their location, position, orientation (augmented from a visual dataset since this is not in the PPT data), or type of pass (i.e., Direct, 1-bank, or Rim). Another potential pass classification could be drop passes, which have significantly different dynamics (player movement and puck speed) than most Direct passes. Furthermore, future work can leverage the z coordinate of the puck to analyze who makes *saucer* passes and where, a common pass in hockey that elevates the puck off the ice.

Finally, we would also like to conduct a sensitivity analysis to determine if our classifications are sensitive to d_b, t^{θ}, and other variables.

9 Conclusions

The new PPT system implemented by the NHL has opened the door for a broader scope of hockey analytics to better model higher resolution events of the game. In this paper, we present an algorithm to classify different types of passes from PPT data and extend the passing lane model in [8] to include 1-bank indirect passes and the expected movement of players. Our model estimates that 1-bank passes comprise about 10.2% of all completed passes in our dataset and make up the majority of non-Direct passes completed. We present *gamma-ratio*, a metric to model the relationship between direct and indirect passing lanes available to a passer. Our model calculates the indirect passing lane to be equal or more available than the direct lane for approximately 59% of completed indirect passes. Furthermore, we show that including the expected movement of players reduces the error in modeling the location of the receiver when they receive the puck for over 94% of completed passes. As PPT systems continue to expand and improve, the impact of algorithms to leverage this type of data will only increase.

Acknowledgments

This research is partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), an Ontario Graduate Scholarship, a Cheriton Scholarship, and the University of Waterloo President's Graduate Scholarship. We thank Brant Berglund, Christopher Baker, Keith Horstman, Neil Pierson, and Russell Levine from the National Hockey League Technology, Stats and Information Team for their participation in fruitful discussions and their insights related to this work. In particular we would like to thank Christopher for his timely and insightful comments on drafts of the paper. We thank Jonah

Eisen, Neel Dayal and Oguzhan Cetin from Rogers Communications and Colin Russell and Aaron Pereira from the University of Waterloo, for their help in getting this project off the ground. We especially thank Neel Dayal for his efforts in creating the relationship with the NHL, providing us with access to the dataset and the talented group at the NHL. We also thank Alexi Orchard for her feedback and useful discussions on drafts of this work.

References

1. Anzer, G., Bauer, P.: Expected passes. Data Mining and Knowledge Discovery pp. 1–23 (2022)
2. Bransen, L., Van Haaren, J.: Measuring football players' on-the-ball contributions from passes during games. In: International Workshop on Machine Learning and Data Mining for Sports Analytics. pp. 3–15. Springer (2018)
3. Douglas, A., Kennedy, C.: Tracking in-match movement demands using local positioning system in world-class men's ice hockey. Journal of Strength and Conditioning Research (2019)
4. Fernández, J.: Decomposing the immeasurable sport: a deep learning expected possession value framework for soccer (2019)
5. Goes, F., Kempe, M., Meerhoff, L., Lemmink, K.: Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. Big Data **7 1**, 57–70 (2019)
6. Lindström, P., Jacobsson, L., Carlsson, N., Lambrix, P.: Predicting player trajectories in shot situations in soccer. In: International Workshop on Machine Learning and Data Mining for Sports Analytics. pp. 62–75. Springer (2020)
7. Power, P., Ruiz, H., Wei, X., Lucey, P.: Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1605–1613 (2017)
8. Radke, D.T., Radke, D.L., Brecht, T., Pawelczyk, A.: Passing and pressure metrics in ice hockey. AI for Sports Analytics Workshop (2021)
9. Spearman, W.: Beyond expected goals. In: Proceedings of the 12th MIT Sloan Sports Analytics Conference. pp. 1–17 (2018)
10. Spearman, W., Basye, A., Dick, G., Hotovy, R., Pop, P.: Physics-based modeling of pass probabilities in soccer. In: Proceeding of the 11th MIT Sloan Sports Analytics Conference (2017)
11. Steiner, S., Rauh, S., Rumo, M., Sonderegger, K., Seiler, R.: Outplaying opponents—a differential perspective on passes using position data. German Journal of Exercise and Sport Research **49**, 140–149 (2019)
12. Stöckl, M., Seidl, T., Marley, D., Power, P.: Making offensive play predictable-using a graph convolutional network to understand defensive performance in soccer. In: Proceedings of the 15th MIT Sloan Sports Analytics Conference (2021)
13. Vats, K., Fani, M., Walters, P.B., Clausi, D.A., Zelek, J.: Event detection in coarsely annotated sports videos via parallel multi receptive field 1d convolutions. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 3856–3865 (2020)
14. Walkters, P., Fani, M., Clausi, D., Wong, A.: Bendernet and ringernet: highly efficient line segmentation deep neural network architectures for ice rink localization. Journal of Computational Vision and Imaging Systems (2020)

Score and Venue Adjustment on Transition Data in Hockey

Cédric Ramqaj and Thibaud Châtel

Abstract. Are zone exits and entries influenced by score and venue the same way shots and goals are? Using our proprietary database of over 120,000 transition events, we analyzed how score and venue can impact how much you control your transitions and your success percentage. Playing at home or on the road does not seem to have much impact overall, especially compared to the influence the score of the game has. Trailing teams appear to be able to make more controlled zone exits, with greater success, probably due to a lesser pressure. On the other hand, leading teams tend to dump the puck out of their defensive zone more often. A trailing team would also try more zone entries but the split between controlled and dump attempts surprisingly remains stable, contradicting a common idea that defenses make it harder to enter the offensive zone when protecting a lead. The "play a simple game on the road" mantra, with less controlled transitions, does not seem to hold either, when looking at the data.

1 Introduction

One of the key motivations behind data-driven research in sports has been to confirm or infirm common ideas about the game. In hockey, a sport played on a relatively overcrowded small surface where it is easy to slow down the flow of the game, we know that leading or trailing in the score will push one team to naturally dominate the other in terms of puck possession and shots taken. This also means that the attacking team will face less pressure to exit its defensive zone, as the defending players are more likely to wait in the neutral zone, but will have a tougher time entering the offensive zone in control as they face a tighter wall of defensemen at the blue line.

Earlier in 2022, Micah Blake McCurdy published research [1] on transitions but based on puck movement between the three zones, without any details on the transition events per se. He confirmed some assumptions, namely that trailing teams were exiting their defensive zone faster or that away teams were slower to exit than home teams.

In the last 10 to 15 years [2], multiple studies have analyzed the impact of such a paradigm on shots and goals, pushing data providers, both public and private, to add a "score adjustment" to their data, reflecting the fact that one team is supposed to be attacking and the other is supposed to be defending at some point in the game. That idea was also derived for teams playing at home, or on the road and was called "venue adjustment".

However, such adjustment has never been made on transition data, such as zone exits and entries. Which leads to our main question in this paper: how score and venue (playing at home or away) impact the way a team is transitioning the puck?
To answer it, we investigated how often teams execute zone exits and entries in a given score and venue situation. How they execute such plays (in control or dumping the puck) and with what success rate.

2 Exploring Zone Exits and Entries

2.1 Collecting Transition Events

Hockey games play-by-play data are now largely available around the world and many public initiatives have used them over the last ten years to help us analyze and understand teams and players performances. However, these publicly available datasets are almost entirely shot-related data, and do not include anything regarding how the puck is moving on the ice between two shots.

Transition data, whether zone exits, zone entries or passes, also called “Microstats”, are available through private data providers or public initiatives, such as the All Three Zones project funded by Corey Sznajder [3] for the NHL. Data is collected by individuals, outside any league or private providers, to make it available to the public. NL Ice Data [4] is a project that has been manually collecting data on the Swiss National League since 2019-20, including transition data that will be used in this paper. We acknowledge the dataset includes more games from certain teams based on the work done at the time by NL Ice Data, but every team in the league had enough representation by season so we were not worried about the sample being driven by one or two teams.

2.2 Definitions of Transition Events

The database used for this paper includes events collected in 440 games between 2019-20 and the end of January 2022. It includes 73,778 zone exits and 55,689 zone entries, all made at 5v5.

We defined three types of zone exits. Carry exits happen when a player skates in possession of the puck across his defensive blue line. Pass exits happen when a pass leads the puck to cross the defensive blue line or puts the receiver in an immediate and safe situation to do so. Carry or Pass Exits are successful or failed if the team keeps possession of the puck in the neutral zone. Dump exits happen when a player chips the puck in the air or against the board to send it in the neutral zone or farther away. A successful Dump Exit is retrieved by a teammate in the neutral zone or if the puck reaches the offensive zone. It fails if it becomes an icing though, if an opponent recovers the puck in the neutral zone or if the puck does not even leave the defensive zone.

We defined two types of zone entries. Controlled Entries happen when a player skates in possession of the puck across the offensive blue line or passes it to a teammate in immediate position to do so. It is a success if the attacking player keeps control of the puck for at least two seconds in the offensive zone. Dump Entries happen when the puck is sent in the offensive zone with no passing intent. It is successful if the first or second player to take full possession of the puck is attacking, otherwise it is failed.

Figure 1 shows how many events are included in this research.

Fig. 1.

Number of transition events included in the database for this research

	Home	Away
Entries		
Successful Dump Entries	2,539	2,573
Failed Dump Entries	7,620	7,568
Successful Controlled Entries	13,569	13,116
Failed Controlled Entries	4,286	4,418
Exits		
Successful Dump Exits	4,299	4,387
Failed Dump Exits	3,817	4,065
Successful Carry Exits	8,176	8,271
Failed Carry Exits	1,113	1,186
Successful Pass Exits	13,389	13,182
Failed Pass Exits	5,989	5,904
Total	64,797	64,670

Data: NL Ice Data database, manually collected since 2019-20

3 Calculating Score and Venue Effects on Transitions

We can split the 129,467 transition events from our database between the different score states and venues (Figure 2).

Fig. 2.

Number of transition events per score state and venue

		Carry Exits		Pass Exits		Dump Exits		Contr. Entries		Dump Entries	
Score state	**TOI**[1]	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**
Leading	13,283	3,204	2,350	6,716	4,719	3,636	2,848	6,286	4,256	3,506	2,692
Tied	14,985	3,393	3,364	7,203	7,051	2,801	3,023	6,470	6,265	3,731	3,647
Trailing	13,283	2,692	3,743	5,459	7,316	1,679	2,581	5,099	7,013	2,922	3,802

[1] **TOI**: TOI at 5v5 in minutes

Data: NL Ice Data database, manually collected since 2019-20

3.1 Method

In this paper, we are building on the earlier work by Micah Blake McCurdy back in 2014 [5] on Score-Adjusted Fenwick and with the same rationale behind it. Here, we take controlled entries tried (success or failed) as an example for an event. The adjustment coefficient is the ratio between the rate at which the event happens over all Score & Venue possibilities and the event at a given Score difference (tied for example) and for one of the venues (home team for example).

More formally:

$$Rate\ (per\ 60)\ of\ any\ event = \frac{\sum_{i=trailing}^{leading} \sum_{j=home}^{away} event_{i,j}}{\sum_{i=trailing}^{leading} \sum_{j=home}^{away} TOI_{i,j}}$$

Which leads to the following adjustment coefficient for any event for a home team in a tied game:

$$Adjustment\ coeff. = \frac{Rate\ (per\ 60)\ of\ any\ event}{Rate\ (per\ 60)\ of\ any\ event\ (home\ team, tied\ game)}$$

3.2 Score & Venue Effect

To go back to our example, on average, 51.102 controlled entries are tried per 60 minutes, whatever the score difference and venue context. For a home team in a tied game, on average, 51.811 controlled entries are tried. Using the above formula, the adjustment coefficient for controlled entries tried with a home team in a tied game is then 0.986 (or 51.102/51.811). As home teams in a tied game try more controlled entries on average than in any given context, they should weigh less than 1. Figure 3 shows all our adjustment coefficients, per score difference and venue context.

Fig. 3.

Adjustment coefficient for transtion events

		Carry Exits		Pass Exits		Dump Exits		Contr. Entries		Dump Entries	
Score state	**TOI**[1]	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**	**Home**	**Away**
Leading	13,283	1.084	1.073	1.061	1.096	0.844	0.782	1.043	1.118	1.072	1.014
Tied	14,985	0.996	1.005	0.963	0.984	1.067	0.988	0.986	1.019	0.981	1.004
Trailing	13,283	0.936	0.928	0.947	0.974	1.327	1.189	0.933	0.935	0.934	0.989

[1] **TOI**: TOI at 5v5 in minutes

Data: NL Ice Data database, manually collected since 2019-20

These adjustment coefficients are further discussed in Section 4.4.

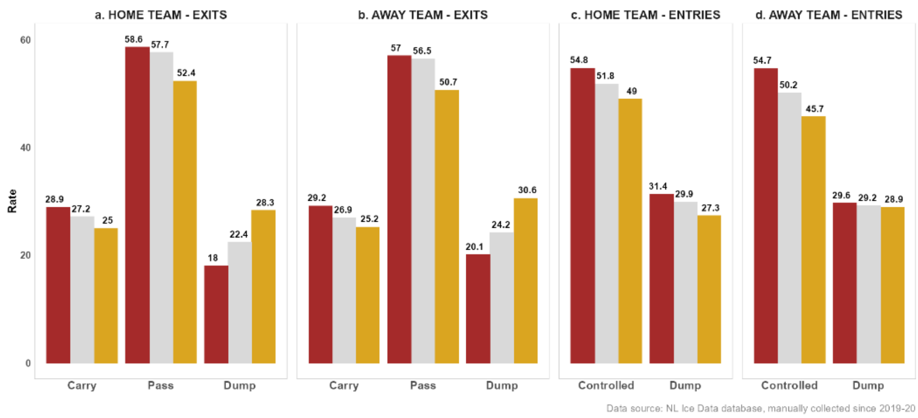
4 Findings

4.1 Rate of Transition Events per Score and Venue

We began our analysis by looking at the rate of transition events during a game. And it immediately appeared that the score was heavily driving how often each team would transition the puck.

Fig. 4: Rate of Event by Score & Venue per 60 minutes

Rate per Scoring state: Trailing, Tied, Leading.

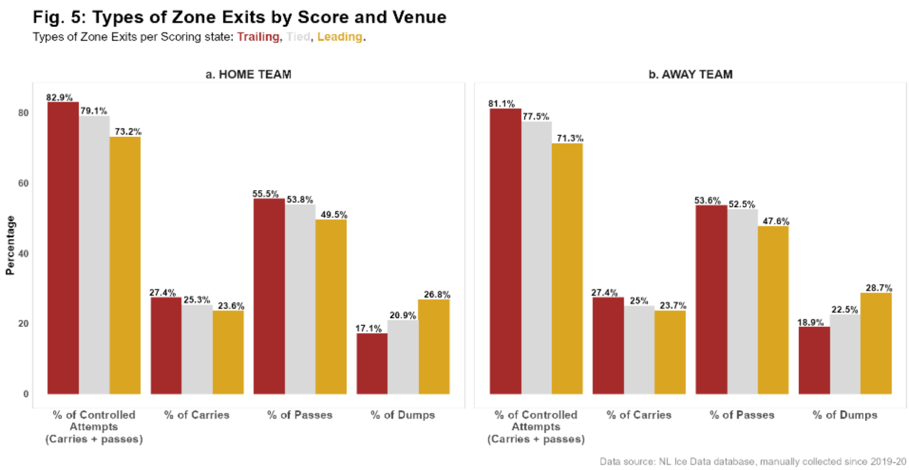


It appears that a trailing team would add about 10 controlled exits (carry or pass) per 60 minutes compared to a leading team (Figure 4, a). And a leading team would perform about 10 more dump exits compared to a trailing team (Figure 4, a), which represents a 57% difference.

A trailing home team would also perform around 5 more controlled entries compared to when leading the game (Figure 4, c). Interestingly, a trailing team on the road would add almost 10 controlled entries compared to a leading away team (Figure 4, d), a 20% difference. We also see that, unlike what we could have thought, the rate of dump entries does not increase much when a team is trailing (Figure 4, c, d). When chasing the score, teams are more likely to add more controlled attempts than dumps-in.

4.2 Zone Exits

Intuitively, the collective knowledge, or also called "eye test", would state that exiting your defensive zone at 5v5 can often become an easy thing if you are trailing, as the leading team is entering shell mode in the neutral zone.



And this historical intuition is supported by numbers. On average, there is a 10 points of percentage drop in the share of exits attempted in control between a team leading or trailing (Figure 5, a, b). A leading home team would attempt 73% of its exits in control, 79% if the score is tied, and 83% if they are trailing (Figure 5, a). A leading away team would attempt 71% of its exits in control, 77% if the score is tied, and 81% if they are trailing (Figure 5, b). The dynamics at play are the same here: the score driving the change of style more than playing at home or on the road. A leading team will use less carry or pass exits and increase the number of pucks dumped out of their zone. On the contrary, a trailing team would use less dumps and equally more carry or pass exits.

We still see a tiny difference created by home ice advantage, especially with a tied score, but it is maybe less than expected. It is to be noted that the difference comes from more pass exits tried by the home team, when carry exits are not impacted. One theory here would be that carry exits are driven by individual talents, players that would execute their play no matter the home ice advantage.

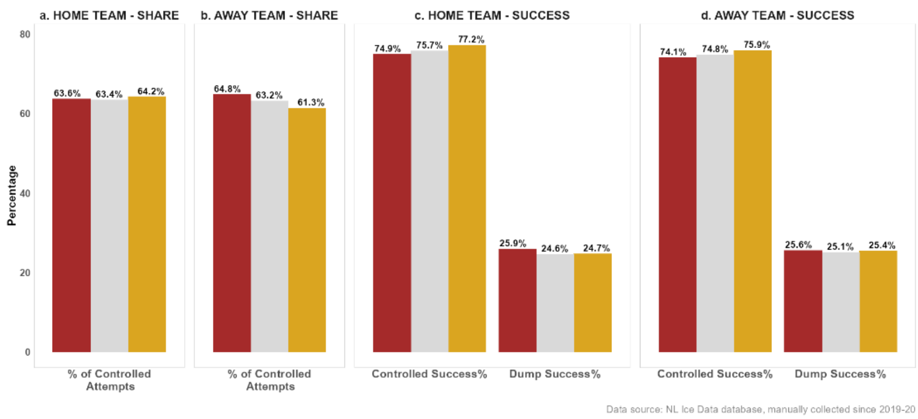
In terms of success rates, they seem to be less impacted by the score or venue than the style chosen to exit. A trailing team would see its success rate on carry and pass exits increase, especially in the third period, as per our data, probably from the lack of pressure. But there is not much difference otherwise before the third period, or overall if you are leading or in a tied game.

4.3 Zone Entries

Do we see a similar dynamic for zone entries? But if zone exits see a change from a sole reduced forecheck, entries might have a double dynamic, with the defending team tightening its play on their defensive blue line, and with the offensive team having a choice between still trying to enter in control, or simply dump the puck in.

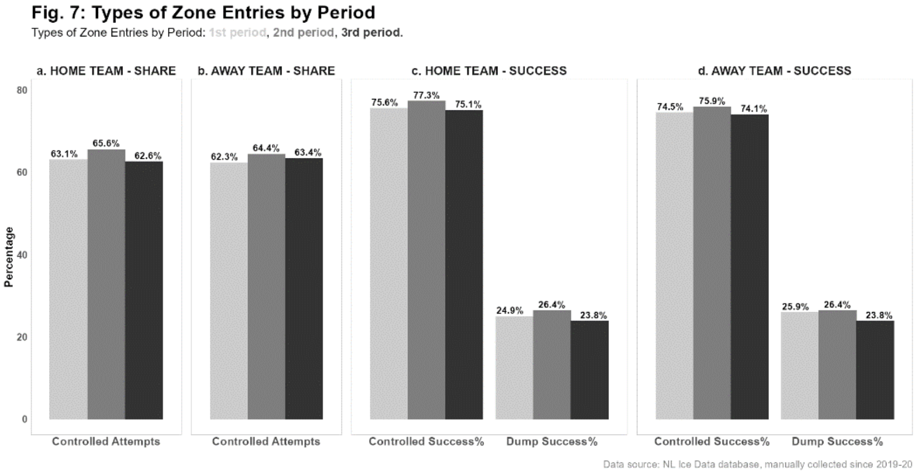
Fig. 6: Types of Zone Entries by Score and Venue

Types of Zone Entries per Scoring state: Trailing, Tied, Leading.



And here, the historical preconception might be a bit off. First, a trailing home team would barely change its style between controlled and dump attempts (Figure 6, a). A trailing away team, however, would increase their share of controlled attempts (Figure 6, b), which goes probably against the “play a simple game on the road” mantra. One common thing is the slightly reduced success rate on controlled attempts when trailing the score (Figure 6, c, d), showcasing that it gets harder to get through the defensemen at the blue line. Dumps success rates barely move, or even from a few decimal points in favor of the trailing team. Does an increased pressure from the trailing forwards compensate the fact defensemen are playing tighter? Defensemen might also let forwards recover the dump in order to pin them along the boards.

And what could drive how zone entries are performed might be how easily the defense can set up and send fresh legs on the ice: namingly the location of the benches.



We clearly see a small but steady increase in the second period, both in the share of controlled attempts and the success rate of those (Figure 7, a, b, c, d). And if a team uses fewer dumps in the second period, their success rate also improves. On the other hand, the first and third periods are almost copycats on all metrics. Based on this, teams willing to build on controlled entries could intentionally push harder for them during the second period of games.

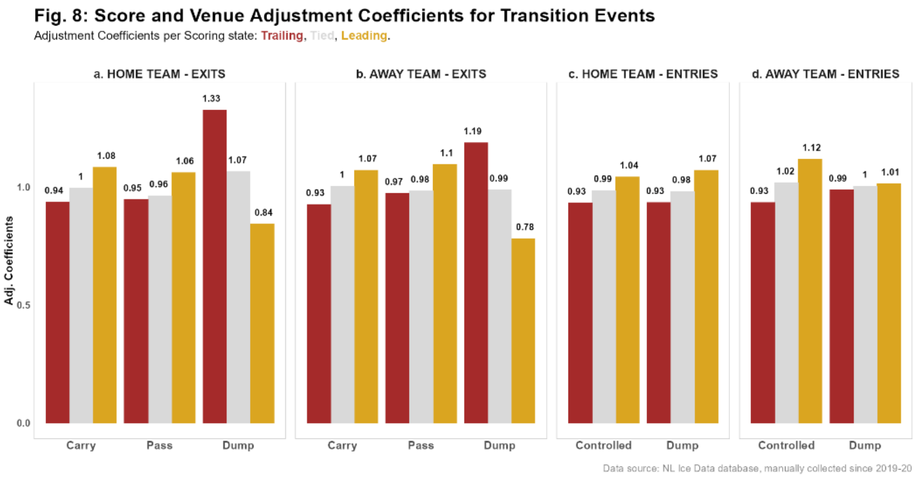
4.4 Score and Venue Adjusted Transition Values

If indeed score and venue impact the way teams are transitioning the puck in a hockey game, it seems possible to now use score and venue adjusted values for exits and entries data when collecting them. More importantly, using adjusted numbers would benefit talented players and teams able to keep on executing controlled plays despite a less favorable context and increased pressure. And, of course, penalize players and teams unwilling to face tougher adversity.

That means, instead of each event having a value of 1, the adjusted value would depend on the score and venue situation, following the formula detailed in Section 3.1.

$$Adjustment\ coeff. = \frac{Rate\ (per\ 60)\ of\ any\ event}{Rate\ (per\ 60)\ of\ any\ event\ (home\ team, tied\ game)}$$

The adjustment for a play made harder by score and venue, for example a controlled zone exit when leading the score, would give that event a value higher than 1, rewarding the play. However, an easy or expected play, for example a controlled exit when trailing the score, would have an adjusted value lower than 1, highlighting the easier context surrounding the event.



Here we chose to group successful and failed events, as we position ourselves ahead of the transition, when the player must choose how he will execute the play. Findings are very similar for home and away teams. A leading home team, facing increased pressure from the trailing forwards would see carry exits (1.08) and pass exits (1.06) (Figure 8, a) bonified to reward the will to keep control of the puck instead of getting rid of it to escape forecheck. On the other hand, dumping the puck as the leading home team is expected and one dump exit would now be worth 0.84 (Figure 8, a), not penalizing the player responsible.

The opposite dynamic is witnessed for the trailing home team. As zone exits get easier, your carry or pass attempts are now worth 0.94 or 0.95 each (Figure 8, a). Dumping out the puck as the trailing home team is not something you are supposed to do and a dump would now be worth 1.34 (Figure 8, a), penalizing the player responsible in his stats, as most agree that dump exits are to be avoided in general because they generate less offense [6].

On zone entries, we discovered that trailing did not mean less controlled entries. Therefore, you would not be rewarded for trying to enter the offensive zone in control when chasing the score. A controlled entry for a trailing team, home or away, would now be worth 0.93 (Figure 8, c, d). Even if controlled entries become harder to complete when trailing, the fact that you are trying many more is driving the adjustment down.

One thing here is also to remember that the side of the ice mattered more than the score on entries, and the third period, where the score would most impact the game, has benches on the easy side for defensemen.

5 Conclusions, Limitations and Future Work

In the end, historical assumptions seem to mostly hold. A leading team would control transitions less and dump the puck more, when a trailing team would face easier zone exits. However, the fact that trailing teams increase their number of zone entry attempts quite a lot, leading to more controlled entries, was a bit surprising.

It also appeared clearly that score dynamics are a much stronger driver than venue dynamics. And that the net difference in style or success between home and away teams is very close, making us wonder if the old saying "play a simple game on the road" is a thing of the past, or even ever existed.

One unexpected finding concerned the impact of playing far from your bench during the second period. It leads to more controlled transitions and better success rates, probably as defenders are more tired and lines get stretched over the ice. Knowing this, teams should really push harder during that second period if they can, also knowing the risk they face defensively.

The next step in our studies would be to look at how trailing teams specifically decided to approach zone entries. Do the way trailing teams approach transitions help them tying the game? What are your probabilities to score based on how much controlled and successful your transitions are at that time? That way, we could possibly highlight the most effective strategies to score goals under the pressure of losing a game.

It would be interesting to see how our work hold for other professional leagues (KHL, NHL, Liiga, SHL, …), envisioning a difference between European hockey, played in big rink, and North-American hockey.

We tracked games during the 2022 Olympics, played in a small rink, and controlled entries percentage tended to be 5 to 10 points of percentage lower than our average numbers in Switzerland. Defending zone entries in a small rink is indeed much easier and running the same analysis with NHL data could bring different conclusions.

Furthermore, if score and venue dynamics probably explain a non-neglectable part of the results, what other variables or aspects of the game could help us understand the observed differences? Does this Score & Venue adjustment offer an improvement in the repeatability of the different transition measures? Would any adjustment of time be justified? In another research [7], Micah Blake McCurdy stated that "*time-adjustment for possession calculations is not justified*".

References

1. Micah Blake McCurdy, Blueline Traversals 2022, https://hockeyviz.com/txt/bluelineTraversals
2. Tyller Dellow on http://www.mc79hockey.com/ Resource no longer available
3. Corey Sznajder, All Three Zone project: https://www.allthreezones.com/
4. https://nlicedata.com/
5. Micah Blake McCurdy, Better Way to Compute Score-Adjusted Fenwick, https://hockeyviz.com/txt/senstats

6. Thibaud Chatel, Introducing Offensive Sequences and The Hockey Decision Tree https://hockey-graphs.com/2020/03/26/introducing-offensive-sequences-and-the-hockey-decision-tree/
7. Micah Blake McCurdy, Adjusted Possession Measures,

Evaluating deep tracking models for player tracking in broadcast ice hockey video

Kanav Vats, Mehrnaz Fani, David A. Clausi, and John S. Zelek

Department of Systems Design Engineering, University of Waterloo
{k2vats, mfani, dclausi, jzelek}@uwaterloo.ca

Abstract. Tracking and identifying players is an important problem in computer vision based ice hockey analytics. Player tracking is a challenging problem since the motion of players in hockey is fast-paced and non-linear. There is also significant player-player and player-board occlusion, camera panning and zooming in hockey broadcast video. Prior published research perform player tracking with the help of handcrafted features for player detection and re-identification. Although commercial solutions for hockey player tracking exist, to the best of our knowledge, no network architectures used, training data or performance metrics are publicly reported. There is currently no published work for hockey player tracking making use of the recent advancements in deep learning while also reporting the current accuracy metrics used in literature. Therefore, in this paper we compare and contrast several state-of-the-art tracking algorithms and analyze their performance and failure modes in ice hockey.

Keywords: ice hockey · deep learning · tracking.

1 Introduction

Ice hockey is played by an estimated 1.8 million people worldwide [10]. As a team sport, the positioning of the players and puck on the ice are critical to team offensive and defensive strategy [22]. The location of players on the ice is essential for hockey analysts for determining the location of play and analyzing game strategy and events. In ice hockey, prior published research [15, 5] perform player tracking with the help of handcrafted features for player detection and re-identification. Okuma *et al.* [15] track hockey players by introducing a particle filter combined with mixture particle filter (MPF) framework [23], along with an Adaboost [24] player detector. The MPF framework [23] allows the particle filter framework to handle multi-modality by modelling the posterior state distributions of M objects as an M component mixture. A disadvantage of the MPF framework is that the particles merge and split in the process and leads to loss of identities. Moreover, the algorithm does not have any mechanism to prevent identity switches and lost identities of players after occlusions. Cai *et al.* [5] improve upon [15] by using a bipartite matching for associating observations with targets instead of using the mixture particle filter framework. However, the algorithm is not trained or tested on broadcast videos, but performs tracking in the rink coordinate system after a manual homography calculation.

Remarking that there is a lack of publicly available research for tracking ice hockey players making use of recent advancements in deep learning, in this paper we track and identify hockey players in broadcast NHL videos and analyze performance of several state-of-the-art deep tracking models on the ice hockey dataset. We also annotate and introduce a new hockey player tracking dataset on which the deep tracking models are tested.

2 Related work

There are a number of recent studies dealing with player tracking in basketball [19, 13, 27] and soccer [20, 9, 21, 7]. For basketball player tracking, Sangüesa *et al.* [19] demonstrated that deep features perform better than classical handcrafted features for basketball player tracking. Lu *et al.* [13] perform player tracking in basketball using a Kalman filter by making the assumption that the relationship between time and player's locations is approximately linear in a short time interval. Zhang *et al.* [27] perform basketball player tracking in a multi camera setting.

In soccer, Theagarajan *et al.* [20] track players using the deep SORT algorithm [26] for generating tactical analysis and ball possession statistics . Hurault *et al.* [9] introduce a self-supervised detection algorithm to detect small soccer players and track players in non-broadcast settings using a triplet loss trained re-identification mechanism, with embeddings obtained from the detector itself. Theiner *et al.* [21] present a pipeline to extract player position data on the soccer field from video. The player tracking was performed with the help of CenterTrack [29]. However, the major focus of the work was on detection accuracy rather than tracking and identification. Gadde *et al.* [7] use a weakly supervised transductive approach for player detection in soccer broadcast videos by treating player detection as a domain adaptation problem. The dataset used is generated with the help of the deep SORT algorithm [26].

3 Methodology

We experimented with five state-of-the-art tracking algorithms [3, 26, 28, 1, 4] on the hockey player tracking dataset. The algorithms include four online tracking algorithms [3, 26, 28, 1] and one offline tracking algorithm [4]. SORT [3], deep SORT [26] and MOT Neural Solver [4] are tracking by detection (TBD) algorithms. Tracktor [1] and FairMOT [28] are joint detection and tracking (JDT) algorithms.

Tracking by detection (TBD) is a widely used approach for multi-object tracking. TBD consists of three steps: (1) detecting objects (hockey players in our case) frame-by-frame in the video (2) calculating affinity between detected objects (3) inference - linking player detections using calculated affinities to produce tracks. Concretely, in TBD, the input is a set of object detections

Table 1. Tracking algorithms compared for hockey player tracking.

Algorithm	Description
SORT [3]	Kalman filter with simple IOU based re-id.
Deep SORT [26]	Kalman filter with deep CNN based re-id.
Tracktor [1]	JDT algorithm with separate detection and re-id networks.
FairMOT [28]	JDT algorithm with combined object detection and re-id network.
MOT Neural Solver [4]	Tracking using graph message passing with edge classification.

$O = \{o_1, ...o_n\}$, where n denotes the total number of detections in all video frames. A detection o_i is represented by $\{x_i, y_i, w_i, h_i, I_i, t_i\}$, where x_i, y_i, w_i, h_i denotes the coordinates, width, and height of the detection bounding box. I_i and t_i represent the image pixels and timestamp corresponding to the detection. Affinity calculation consists of calculating affinity between detections o_i by obtaining appropriate features. The features can be simple intersection over union (IOU) based [3] or using deep networks [25]. After affinity calculation, a set of trajectories $T = \{T_1, T_2...T_m\}$ is found that best explains O where each T_i is a time-ordered set of observations. This is done through an appropriate inference technique. Two widely used inference techniques are filtering [3, 25] and graphical formulation [4]. As an example of graphical formulation, the MOT Neural Solver [4] models the tracking problem as an undirected graph $G = (V, E)$, where $V = \{1, 2, ..., n\}$ is the set of n nodes for n player detections for all video frames. In the edge set E, every pair of detections is connected so that trajectories with missed detections can be recovered. The problem of tracking is posed as splitting the graph into disconnected components where each component is a trajectory T_i. After computing each node embedding and edge embedding using a CNN (affinity calculation), the model then solves a graph message passing problem. The message passing algorithm classifies whether an edge between two nodes in the graph belongs to the same player trajectory.

Joint detection and tracking (JDT) [1, 28] is the latest trend in multi-object tracking research. These methods either (1) Convert an object detector to a tracker by estimating the location of a bounding box in the adjacent frames [1] or (2) Perform detection and re-identification using a single network [28]. Bergmann *et al.* [1] use the bounding box regressor of a Faster RCNN [16] detector to regress the position of a person in the next frame. The re-identification is performed using a separate siamese network. Zhang *et al.* [28] perform object detection and re-identification with the same network using separate detection and re-identification branches. The differences and similarities between the five tracking algorithms are summarized in Table 1. We refer the readers to the publications of the respective tracking papers [3, 26, 28, 1, 4] for more detail.

4 Dataset

The player tracking dataset consists of a total of 84 broadcast NHL game clips with a frame rate of 30 frames per second (fps) and resolution of 1280×720 pixels. The average clip duration is 36 seconds. The 84 video clips in the dataset

Fig. 1. CVAT tool used for tracking annotations. The tool offers the ability to annotate bounding boxes with each box having one label - home or away team. Each player bounding box has player name and jersey number as attributes. CVAT also offers an interpolation mode which alleviates the need to draw bounding boxes multiple times for adjacent frames.

are extracted from 25 NHL games. The duration of the clips is shown in Fig. 2. Each video frame in a clip is annotated with player and referee bounding boxes and player identity consisting of player name and jersey number. The annotation is carried out with the help of the open source computer vision annotation tool (CVAT) [1]. An illustration of an annotation job using the CVAT tool is shown in Fig. 1. The dataset is split such that 58 clips are used for training, 13 clips for validation, and 13 clips for testing. To prevent any game-level bias affecting the results, the split is made at the game level, such that the training clips are obtained from 17 games, validation clips from 4 games and test split from 4 games respectively.
Table 2 compares the size of the dataset with other tracking datasets in literature. The hockey player tracking dataset is comparable in size with other tracking datasets used in literature. As compared to pedestrian datsets (MOT 16 [14] and MOT20 [6]), the bounding boxes per frame is less in our dataset since the maximum number of players on the screen can be 12, with usually less than 12 players actually in broadcast camera field of view (FOV). The NHL game videos used to create this dataset have been obtained from Stathletes Inc. with permission.

[1] Found online at: https://github.com/openvinotoolkit/cvat

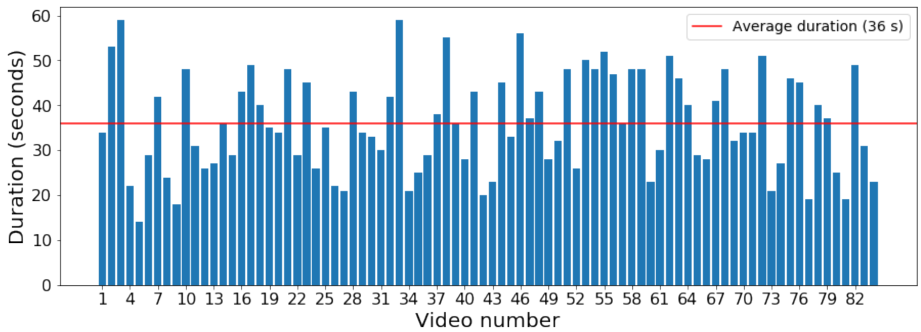


Fig. 2. Duration of videos in the player tracking dataset. The average clip duration is 36 seconds. The red horizontal line represents the average clip duration.

Table 2. Comparison of hockey tracking dataset with other tracking datasets in literature. Our hockey player tracking dataset is comparable to other multi-object tracking datasets commonly used in literature.

Dataset	Videos/sequences	Frames	Bounding boxes	Domain
MOT16 [14]	14	11, 235	292, 733	Pedestrians
MOT20 [6]	8	13, 410	2, 102, 385	Crowded pedestrian scenes
KITTI-T [8]	50	10, 870	65, 213	Autonomous driving
Ours	84	91, 807	773, 545	Ice hockey players

4.1 Annotation process

15 annotators annotated the whole dataset using the CVAT tool. The average time taken to annotate one minute of video is 10.45 minutes. The total time taken to annotate all 84 videos is 527 minutes. The manual annotation was done such that a bounding box as tight as possible was drawn around a player/referee. Linear interpolation was used to interpolate bounding box positions. Additionally, unlike other tracking datasets such as MOT16 [14] and MOT20 [6], the same ground truth identity was assigned to a player leaving camera FOV at a particular frame and re-entering after some time. If a player was occluded by board or another player, the bounding box was annotated based on the best guess of the tightest box enclosing the full body of the player. For quality control, all bounding boxes were checked to make sure each box has label-name(name of the player). When a player enters/exits the scene, his bounding box was labeled even if he was partially in camera FOV. Whenever players were occluded by other players, revision of annotations was performed to ensure high quality.

5 Results

Player detection is performed using a Faster-RCNN network [17] with a ResNet50 based Feature Pyramid Network (FPN) backbone [11] pre-trained on the COCO

Table 3. Comparison of the overall tracking performance on test videos of the hockey player tracking dataset. (↓ means lower is better, ↑ mean higher is better)

Method	IDF1↑	MOTA ↑	ID-switches ↓	False positives (FP)↓	False negatives (FN) ↓
SORT [3]	53.7	92.4	673	2403	5826
Deep SORT [26]	59.3	94.2	528	1881	4334
Tracktor [1]	56.5	94.4	687	1706	**4216**
FairMOT [28]	61.5	91.9	768	**1179**	7568
MOT Neural Solver [4]	**62.9**	**94.5**	**431**	1653	4394

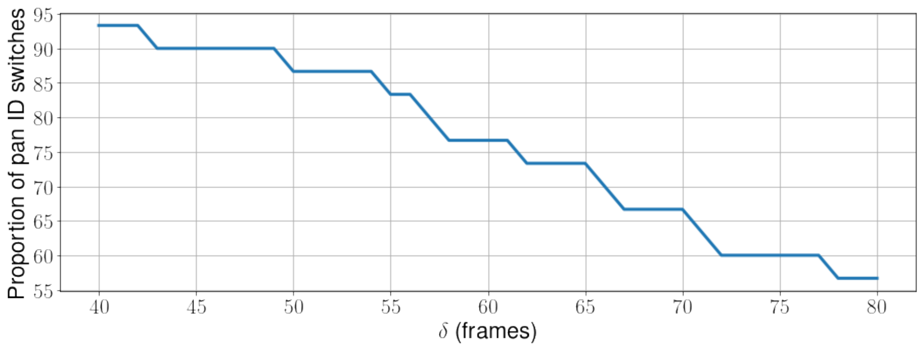


Fig. 3. Proportion of pan identity switches vs. δ plot for video number 9. Majority of the identity switches (90% at a threshold of $\delta = 40$ frames) occur due to camera panning, which is the main cause of error.

dataset - a large scale object detection, segmentation, and captioning dataset, popular in computer vision [12] and fine tuned on the hockey tracking dataset. The object detector obtains an average precision (AP) of 70.2 on the test videos (Table 4). The accuracy metrics for tracking used are the CLEAR MOT metrics [2] and Identification F1 score (IDF1) [18]. A ground truth object missed by the trackers is called a false negative (FN) whereas a false alarm is called a false positive (FP). For any tracker, a low number of false positives (FP) and false negatives (FN) are favoured. An important metric is the number of identity switches (IDSW), which occurs when a ground truth ID i is assigned a tracked ID j when the last known assignment ID was $k \neq j$. A low number of identity switches is an indicator of accurate tracking performance. For sports player tracking, the IDF1 is considered a better accuracy measure than Multi Object Tracking accuracy (MOTA) since it measures how consistently the identity of a tracked object is preserved with respect to the ground truth identity. The overall results are shown in Table 3. The best tracking performance is achieved using the MOT Neural Solver tracking model [4] re-trained on the hockey dataset. The MOT Neural Solver model obtains the highest MOTA score of 94.5 and IDF1 score of 62.9 on the test videos.

Table 4. Player detection results on the test videos. AP stands for Average Precision. AP_{50} and AP_{75} are the average precision at an IOU of 0.5 and 0.75 respectively.

AP	AP_{50}	AP_{75}
70.2	95.9	87.5

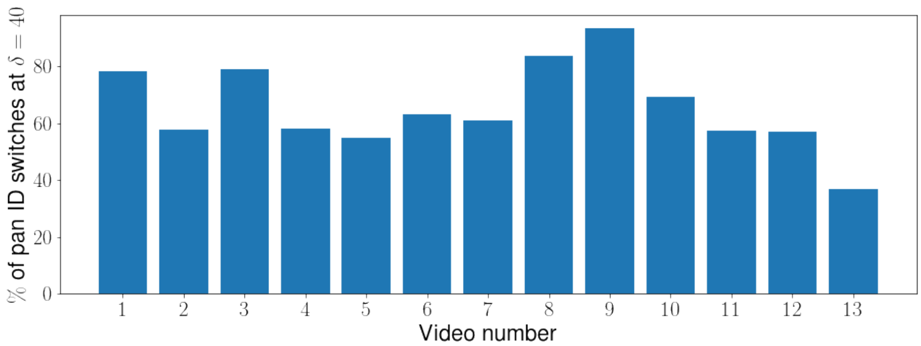


Fig. 4. Proportion of pan-identity switches for all videos at a threshold of $\delta = 40$ frames. On average, pan-identity switches account for 65% of identity switches.

6 Discussion

From Table 3 it can be seen that the MOTA score of all methods is above 90%. This is because MOTA is calculated as

$$MOTA = 1 - \frac{\Sigma_t(FN_t + FP_t + IDSW_t)}{\Sigma_t GT_t} \tag{1}$$

where t is the frame index and GT is the number of ground truth objects. MOTA metric counts detection errors through the sum $FP + FN$ and association errors through $IDSWs$. Since false positives (FP) and false negatives (FN) heavily rely on the performance of the player detector, the MOTA metric highly depends on the performance of the detector. For hockey player tracking, the player detection accuracy is high because of the sufficiently large size of players in broadcast video and a reasonable number of players and referees (with a fixed upper limit) to detect in the frame. Therefore, the MOTA score for all methods is high.

The SORT [3] algorithm obtains the least IDF1 score and the highest number of identity switches. This is due to the linear motion model assumption and simple IOU score for re-identification. Deep SORT [25], on the other hand uses features obtained from deep network for re-identification resulting in better IDF1 score and lower identity switches. For JDT based networks, performing detection and re-identification with a single network using a multi-task loss performs better than having separate networks for detection and re-id tasks, evident by better performance of FairMOT [28] compared to Tracktor [1]. JDT tracking

Table 5. Tracking performance of MOT Neural Solver model for the 13 test videos (↓ means lower is better, ↑ means higher is better).

Video #	IDF1↑	MOTA ↑	ID-switches ↓	False positives (FP)↓	False negatives (FN) ↓	Duration (sec.)
1	78.53	94.95	23	100	269	36
2	61.49	93.29	26	48	519	29
3	55.83	95.85	43	197	189	43
4	67.22	95.50	31	77	501	49
5	72.60	91.42	40	222	510	40
6	66.66	90.93	38	301	419	35
7	49.02	94.89	59	125	465	48
8	50.06	92.02	31	267	220	34
9	53.33	96.67	30	48	128	29
10	55.91	95.30	26	65	193	26
11	56.52	96.03	40	31	477	45
12	87.41	94.98	14	141	252	35
13	62.98	94.77	30	31	252	22

algorithms, however, [28, 1] do not not show any significant improvement over deep SORT evident by lower identity switches of deep SORT in comparison. The MOT Neural Solver method achieves the highest IDF1 score of 62.9 and significantly lower identity switches than the other methods. This is because the other trackers use a linear motion model assumption which does not perform well with the motion of hockey players. Sharp changes in player motion often leads to identity switches. The MOT Neural Solver model, in contrast, has no such assumptions since it poses tracking as a graph edge classification problem.

Table 5 shows the performance of the MOT Neural solver for each of the 13 test videos. We do a failure analysis to determine the cause of identity switches and low IDF1 score in some videos. The major sources of identity switches are severe occlusions and players going out of the camera FOV (due to camera panning and/or player movement). We define a pan-identity switch as an identity switch resulting from a player leaving and re-entering camera FOV due to camera panning. It is very difficult for the tracking model to maintain identity in these situations since players of the same team look identical with features such as, jersey color, helmet model, visor model, stick model, glove model, skate model, tape color etc unidentifiable from bounding boxes cropped from 720p broadcast clips. During a pan-identity switch, a player going out of the camera FOV at a particular point in screen coordinates can re-enter at any other point. We estimate the proportion of pan-identity switches to determine the contribution of panning to total identity switches.

To estimate the number of pan-identity switches, since we have quality annotations, we make the assumption that the ground truth annotations are accurate and there are no missing annotations in the ground truth. Based on this assumption, there is a significant time gap between two consecutive annotated detections of a player only when the player leaves the camera FOV and comes back again. Let $T_{gt} = \{o_1, o_2, ..., o_n\}$ represent a ground truth tracklet, where $o_i = \{x_i, y_i, w_i, h_t, I_i, t_i\}$ represents a ground truth detection. A pan-identity switch is expected to occur during tracking when the difference between timestamps (in frames) of two consecutive ground truth detections i and j is greater than a sufficiently large threshold δ. That is

$$(t_i - t_j) > \delta \tag{2}$$

Therefore, the total number of pan-identity switches in a video is approximately calculated as

$$\sum_{G} \mathbb{1}(t_i - t_j > \delta) \tag{3}$$

where the summation is carried out over all ground truth trajectories and $\mathbb{1}$ is an indicator function. Consider the video number 9 in Table 5 having 30 identity switches and a low IDF1 of 53.33. We plot the proportion of pan identity switches, that is

$$= \frac{\sum_{G} \mathbb{1}(t_i - t_j > \delta)}{IDSWs} \tag{4}$$

against δ, where δ varies between 40 and 80 frames. From Fig. 3 it can be seen that majority of the identity switches (90% at a threshold of $\delta = 40$ frames) occur due to camera panning. Visually investigating the video confirmed the statement. Fig. 4 shows the proportion of pan-identity switches for all videos at a threshold of $\delta = 40$ frames. On average, pan identity switches account for 65% of identity switches in the videos. This shows that the tracking model is able to tackle a majority of other sources of errors which include minor occlusions and lack of detections. The primary source or errors are pan-identity switches and extremely cluttered scenes.

7 Conclusion

In this paper, we test five state-of-the-art tracking algorithms on the ice hockey dataset and analyzed their performance. From the performance of trackers we infer that trackers with a linear motion model do not perform well on hockey dataset, evident by the high number of identity switches occurring in models with linear motion assumption. The best performance is obtained by the MOT neural solver model [4], that uses a graph based approach towards tracking without any linear motion model assumption. Also, the IDF1 metric is a better metric for hockey player tracking since the MOTA metric is heavily influenced by player detection accuracy. We find that the main source of error in hockey player tracking in broadcast video are pan-identity switches - identity switches results due to players going outside the broadcast camera FOV.

8 Acknowledgments

This work was supported by Stathletes through the Mitacs Accelerate Program and the Natural Sciences and Engineering Research Council of Canada (NSERC). We also acknowledge Compute Canada for hardware support.

References

1. Bergmann, P., Meinhardt, T., Leal-Taixé, L.: Tracking without bells and whistles. In: The IEEE International Conference on Computer Vision (ICCV) (October 2019)
2. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: The clear mot metrics. EURASIP Journal on Image and Video Processing **2008** (01 2008). https://doi.org/10.1155/2008/246309
3. Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B.: Simple online and realtime tracking. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 3464–3468 (2016). https://doi.org/10.1109/ICIP.2016.7533003
4. Braso, G., Leal-Taixe, L.: Learning a neural solver for multiple object tracking. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
5. Cai, Y., de Freitas, N., Little, J.J.: Robust visual tracking for multiple targets. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) Computer Vision – ECCV 2006. pp. 107–118. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
6. Dendorfer, P., Rezatofighi, H., Milan, A., Shi, J.Q., Cremers, D., Reid, I.D., Roth, S., Schindler, K., Leal-Taix'e, L.: Mot20: A benchmark for multi object tracking in crowded scenes. ArXiv **abs/2003.09003** (2020)
7. Gadde, C.A., Jawahar, C.: Transductive weakly-supervised player detection using soccer broadcast videos. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 965–974 (January 2022)
8. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361 (2012). https://doi.org/10.1109/CVPR.2012.6248074
9. Hurault, S., Ballester, C., Haro, G.: Self-supervised small soccer player detection and tracking. In: Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports. p. 9–18. MMSports '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3422844.3423054, https://doi.org/10.1145/3422844.3423054
10. IIHF: Survey of Players (2018), available online: https://www.iihf.com/en/static/5324/survey-of-players
11. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 936–944 (2017). https://doi.org/10.1109/CVPR.2017.106
12. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft *coco*: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. pp. 740–755. Springer International Publishing, Cham (2014)
13. Lu, W.L., Ting, J.A., Little, J.J., Murphy, K.P.: Learning to track and identify players from broadcast sports videos. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(7), 1704–1716 (2013). https://doi.org/10.1109/TPAMI.2012.242
14. Milan, A., Leal-Taixé, L., Reid, I.D., Roth, S., Schindler, K.: Mot16: A benchmark for multi-object tracking. ArXiv **abs/1603.00831** (2016)
15. Okuma, K., Taleghani, A., de Freitas, N., Little, J.J., Lowe, D.G.: A boosted particle filter: Multitarget detection and tracking. In: Pajdla, T., Matas, J. (eds.) Computer Vision - ECCV 2004. pp. 28–39. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)

16. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. p. 91–99. NIPS'15, MIT Press, Cambridge, MA, USA (2015)
17. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)
18. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Hua, G., Jégou, H. (eds.) Computer Vision – ECCV 2016 Workshops. pp. 17–35. Springer International Publishing, Cham (2016)
19. Sangüesa, A.A., Ballester, C., Haro, G.: Single-camera basketball tracker through pose and semantic feature fusion. ArXiv **abs/1906.02042** (2019)
20. Theagarajan, R., Bhanu, B.: An automated system for generating tactical performance statistics for individual soccer players from videos. IEEE Transactions on Circuits and Systems for Video Technology **31**(2), 632–646 (2021). https://doi.org/10.1109/TCSVT.2020.2982580
21. Theiner, J., Gritz, W., Müller-Budack, E., Rein, R., Memmert, D., Ewerth, R.: Extraction of positional player data from broadcast soccer videos. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 823–833 (January 2022)
22. Thomas, A.C.: The impact of puck possession and location on ice hockey strategy. Journal of Quantitative Analysis in Sports **2**(1) (2006)
23. Vermaak, Doucet, Perez: Maintaining multimodality through mixture tracking. In: Proceedings Ninth IEEE International Conference on Computer Vision. pp. 1110–1116 vol.2 (Oct 2003). https://doi.org/10.1109/ICCV.2003.1238473
24. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I (Dec 2001). https://doi.org/10.1109/CVPR.2001.990517
25. Wojke, N., Bewley, A.: Deep cosine metric learning for person re-identification. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 748–756. IEEE (2018). https://doi.org/10.1109/WACV.2018.00087
26. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 3645–3649. IEEE (2017). https://doi.org/10.1109/ICIP.2017.8296962
27. Zhang, R., Wu, L., Yang, Y., Wu, W., Chen, Y., Xu, M.: Multi-camera multi-player tracking with deep player identification in sports video. Pattern Recognition **102**, 107260 (2020). https://doi.org/https://doi.org/10.1016/j.patcog.2020.107260, https://www.sciencedirect.com/science/article/pii/S0031320320300650
28. Zhang, Y., Wang, C., Wang, X., Zeng, W., Liu, W.: Fairmot: On the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision pp. 1–19 (2021)
29. Zhou, X., Koltun, V., Krähenbühl, P.: Tracking objects as points. European Conference on Computer Vision (ECCV) (2020)

Where not to lose the puck

Erik Wilderoth, Ulf Johansson, Arsalan Sattari

Dept. of Computing, Jönköping University, Sweden
erik.wilderoth@gmail.com, ulf.johansson@ju.se, arsalan.sattari@ju.se

Abstract. In a fast-paced free-flow game as Ice Hockey the decision making of the players is crucial for the success of the team. A game in the Swedish Hockey League (SHL) has on average 244 possession changes where both teams play at full strength. Previous studies have shown that the most effective way to create scoring chances is by exiting and entering zones with the puck under control. On the contrary, this paper studies the question of risk and reward of different plays. Based on an extensive data-driven investigation of three full SHL seasons, the conclusion is that the best way not to concede goals is also by doing the transition plays with control. Specifically, a failed dump-out is 57% more likely to end up in the opponents scoring a goal than a failed outlet pass.

Keywords: Ice Hockey, Dumps, Controlled Entry, Controlled Exit

1 Introduction

Within sports there is a lot of conventional wisdom that has become truths, whether based on facts or not. Data analysis is now, sport by sport, tearing down these truths and creating new knowledge which is indeed well-grounded in facts and data. Using a data-driven approach, this paper will investigate risk and reward of different plays, and, consequently, what players should strive for and avoid.

Compared to other major sports like baseball, basketball, American football and soccer, ice hockey should be considered a sport where the results to a large degree are random. Weissbock [1] tried to quantify the randomness in sports, showing that in the NHL, the underdog wins more often than in any of the other major sports in the US. In fact, the favorite wins only 57% of the games in the NHL. In both the NFL (64%) and the NBA (64%) the favorite wins significantly more often. MLB (56%), finally, is very similar to the NHL.

Good teams of course try to increase that number and reduce the randomness. To minimize luck, teams need to calculate risk and reward for the actions in the game. Compared to baseball and American football, ice hockey is a “free flow 360 degree” game where a play (or an episode) in theory can last for a full period of 20 minutes. Players both attack and defend within the same play, in sharp contrast to baseball and American football where one team attacks (tries to score) and one defends. These fundamental characteristics of ice hockey create a lot of situations that cannot be planned for in advance. Players need to be quick thinkers and problem solvers in order to adopt to new and unique situations in this dynamic and high-speed game. To minimize randomness and achieve success the teams, however, set up some ground rules on how the coach wants the players to act in the different situations that occur frequently and in slight variations during the free-flowing plays.

2 Background

An Ice Hockey rink is divided into three zones. Defensive Zone (DZ), Neutral Zone (NZ) and Offensive Zone(OZ). To create scoring chances, teams need to transport the puck in some way from the DZ to the OZ. In fact, no goals the last three seasons in the SHL were scored from the the NZ or the DZ, when the teams both play at full strength and the goalkeeper has not been pulled. The combination of the rules offside and icing makes it almost impossible to go directly from the DZ to the OZ, so the NZ needs to be used for this transition. Here, the conventional wisdom says that players must be very careful not to lose the puck in the NZ, i.e., losing the puck in this zone increase the other team's scoring chance significantly.

Table 1: Terminology Entries and Exits

TYPE	SUB-TYPE	DESCRIPTION
Controlled	Carry	A Player transports the puck over the blue line
Controlled	Pass	A player passes the puck to another player over the blue line
Dump	Dump	A player shoots the puck to next zone without a direct receiver
Dump	Chip	A player shoots the puck in the air into next zone without intended receiver
Exit		Puck moves from Defensive Zone to Neutral Zone
Entry		Puck moves from Neutral Zone to Attacking Zone

Losing the puck - The term describes the next possession after the puck changes team. If Team A shoots and Team B collects the puck, it is a possession change. All situations where Team B touches the puck when Team A has it, count as a possession change and is therefore included in the term "Losing the puck".

On a risk/reward scale the *Dump-in-play* is generally considered to be low risk/low reward while *Controlled entries/exits* are associated with higher risk, but also higher reward.

This paper will focus on data from the SHL. Team wise the playing styles differ quite a lot when it comes to zone exit and zone entry strategies. For instance, Skellefteå AIK carries out the puck almost twice as often as they dump it out from the DZ, meanwhile Malmö Redhawks dumps it more often than they carry it.

Total average zone exit numbers for the SHL are:

– Dump Out 23%
– Carry Outs 25%
– Passing 51%

In Figure 1 the dump-out rates and dump-in rates are shown to highlight the different playing styles in SHL for the season 20/21. Malmö Redhawks was the team that used the "dump-out" as an exit strategy out of the DZ the most and "dump in" into the OZ most as well. On the opposite side, Skellefteå AIK makes the most controlled plays, both when exiting the DZ and entering the OZ. The differences in numbers are huge between the teams. Malmö Redhawks performed 41% more uncontrolled exit and entries during the season than Skellefteå AIK (3119 vs. 2208).

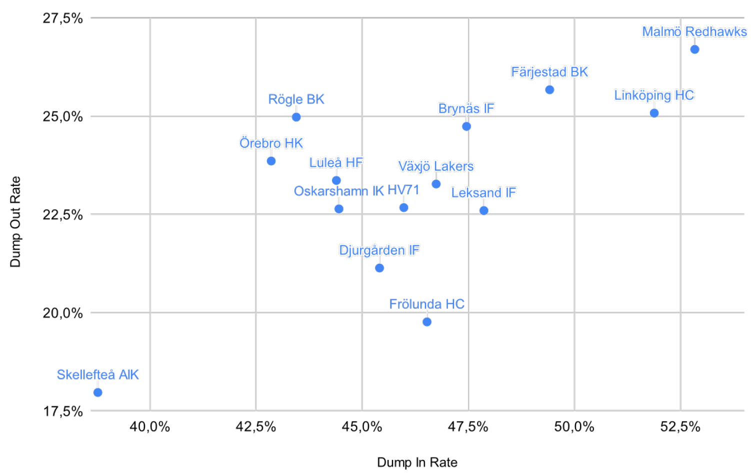


Fig. 1. SHL teams Dump out and dump in rates

3 Related Work

Chatel [2] presents base rates on how the different types of zone exits and entries are connected to expected goals (xG). By bringing the puck out of your own zone with control, the chance of scoring a goal increases dramatically. When entering the offensive zone, it is even more important. Actually, and as seen in Table 2 below, it is the chance of scoring a goal is almost doubled with a successful controlled entry compared to a successful dump-in. Other works concludes similar takes [6, 7] that carry-ins outperforms dump-ins by margin.

Stimson quantified [3] how the different breakout (exit) strategies were leading to shots for and against in the next play. He concluded that controlled exit

Table 2: Chatel's xG Contribution Figures

TYPE	SUB-TYPE	CONTRIBUTION TO XG
Zone Exit	Carry-Out	0.024
Zone Exit	Pass	0.026
Zone Exit	Dump-Out	0.016
Zone Entry	Controlled Entry	0.04
Zone Entry	Dump-In	0.022

had the best Net Shot Differential of all breakout types, meaning more shots for than against.

In the NHL, the entry strategy dump-in is getting more popular for the last couple of seasons [4]. Due to lower risk to get a turnover in the neutral zone, teams are more careful with the puck. Mike Kelly has earlier examined this [4] and concluded that dump-ins significant lower the number of odd man rushes against, which is one of the most efficient ways to score goals in ice hockey [8].

A study similar to this paper has been published present to this [5] and concludes that some existing results are in fact questionable when it comes to exiting strategies, the results presented show that neither of the exit strategies are superior to the other. The study, however, only targets successful plays with the motivation that it is reasonable to assume that a player on, in this case, the college level is generally successful in his attempts to play the puck. We argue that this assumption is incorrect, and consequently that the results have limited bearing on real-world ice hockey. In fact, there are a lot of "bad plays" in ice hockey resulting in turnovers to the defending team. As an example, teams in the SHL have on average only 57% successful entries into the offensive zone. The other 43% the defending team gets control of the puck.

4 Data Preparation

4.1 Data Collection

All data was extracted from SportLogiq[1] for the SHL regular season games 2018/19 to 2020/21. The dataset includes 4 160 282 events before filtering. There are 266 different ways to lose the puck possession to the other team in our data. Most of these are unusual, specifically 213 such events have occurred fewer than 500 times the last three seasons in SHL. A game in SHL averages 244 possession changes per game after filtering to both teams playing at full strength. 0.52% of all puck losses ends up in a goal against.

4.2 Data Preparation

The data was, as described above, filtered by removing all events occurring when not both teams play at full strength (5-vs.-5). In addition, all situations where

[1] http://www.sportlogiq.com

the goalie is the last player to touch the puck in a possession are also excluded since these situations, including e.g., rebounds from shots etc. are very specific. Furthermore, all situations where a team has been in possession of the puck for less than 1.5 seconds are also excluded.

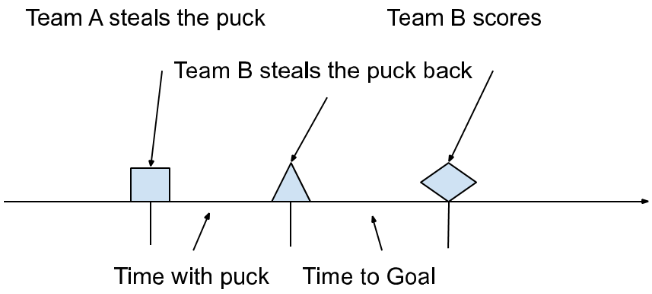


Fig. 2. Visual description of "Time with puck". Team A must have the puck in possession for at least 1.5 seconds for Team Bs goal to be included in the dataset.

The situation where a team has a possession for less than 1.5 second tends to be more of reactions than decision making and therefore creates noisy data, e.g., re-bounds from shots of the bodies of the defenders. It may be noted, though, that 13.6% of all goals in SHL are created in the possession after a "less than 1.5 second" possession.

Goals are created from possession changes in all zones as shown in Table 3. In our dataset 58% of all goals are created from possession changes in the DZ (seen from the team that did not score). SHL is a league where forechecking is an important part of the game and it is seen in the data. In total, 0.75% of all turnovers in the defensive zone is converted into goals against. It seems intuitive that the further away from your own goal, the safer you are. High level data confirms this, losing the puck in the offensive zone has a turnover rate to goal against at 0.38% which is lower than both the DZ, and the NZ (0.44%).

$$\Sigma \textit{ Goal Against} \;/\; \Sigma \textit{ Possession Drops} = \textit{Goal\%}$$

Table 3: Conversion rates to goal per zone

ZONE LOSING THE PUCK	NO OF GOALS	MEDIAN TIME TO GOAL	GOAL%
Defensive	837	5.7 Seconds	0.75%
Neutral	179	7.2 Seconds	0.44%
Offensive	427	8.3 Seconds	0.38%

5 Results

5.1 Location of Puck Drop

The results in Figure 3 are grouped in to 4x6 m quadrants. Each quadrant shows the Goal Conversion rate (Goal%) after puck loss. The number representing goals scored against after puck was lost at that quadrant. Focusing on the areas around the bluelines shows that the puck steals converted to goal does not increase in the transition phase between DZ and NZ. 0.4% of all lost pucks round defensive blue line is converted to goals against, which is close to the complete neutral zone (0.44%). On this high-level data, we do not know what the intention with puck was.

The offensive blue line, on the other hand, has an increased Goal% (0.5%) compared to the areas around it indicating that losing the puck on offensive blue line is a dangerous place to lose the puck. One area on the offensive blue line has close to 1% Goal% which is as high as losing the puck in the high slot.

The forechecks popularity is obvious, the highest total Goal% for data in x-axis is found behind the goal, winning back the puck when forechecking the opponent.

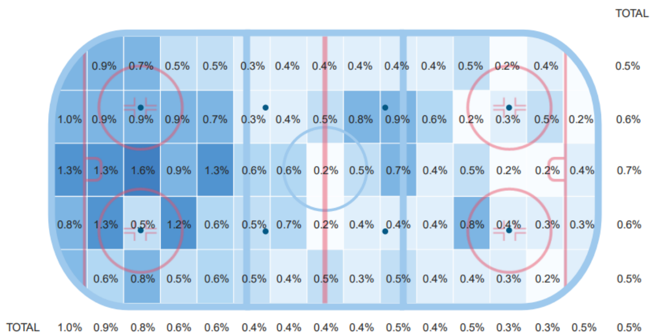


Fig. 3. Goal% per turn over location

5.2 Entries and Exits

Grouping data in the same way as Chatel [Table 1] did for the different types of exits and entries connected to xG-value, the actual outcome for these actions against is presented in Table 4 in Goal%.

So, based on these numbers, dump-outs are actually the most dangerous transition play in ice hockey. In particular, it is the failed ones that create these numbers. This key result of the paper is further broken down in Table 5. 1.63% off all failed dump-outs, that are not air bound (Flip Dump Outs) and fails to reach the NZ, turns in to a goal against and that is the highest Goal% for any sub-event of transitions plays.

Table 4: Goal% per transition type

TYPE	SUB-TYPE	Goal%
Zone Exit	Carry-Out	0.43%
Zone Exit	Pass	0.59%
Zone Exit	Dump-Out	0.65%
Zone Entry	Controlled Entry	0.43%
Zone Entry	Dump-In	0.29%

Table 5: Dump-outs breakdown

SUB-TYPE	Goals	Goal%
Dump Out-	36	1.63%
Flip Dump Out-	7	1.16%
Off Glass Dump Out-	57	1.02%
Flip Dump Out+	21	0.52%
Dump Out+	20	0.36%
Off Glass Dump Out+	31	0.36%
All Dump Out Attempts	**172**	**0.65%**

5.3 Risk/Reward

Plotting [figure 4] the result from Chatels's entry data [table 2] and comparing it to the result of this paper. setting xG gain equal to reward and goals against equal to risk. shows that making controlled plays when exiting the zone is better for both scoring more goals but also conceding fewer. Dump-Out has the highest risk of all plays and lowest reward. As the result implies this is due to the failed dump-outs. Entries is more complex with higher risk and higher reward for controlled plays. In the long run Controlled Entries beats Dump-Ins. The lower risk is worth to have in consideration when in lead and clock closing in.

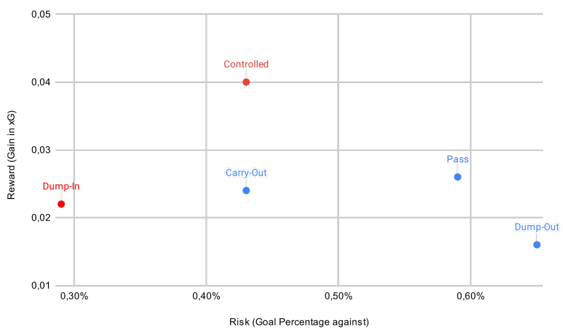


Fig. 4. Risk vs Reward. Red Dots = Entries. Blue Dots = Exits.

6 Conclusion

We have in this paper described risk/reward when moving the puck from the defensive zone to the offensive one. From the analysis, we have identified that moving the puck with control from the defensive zone is superior to dumping it. Controlled zone exits are better both for scoring goals and avoid conceding goals. In fact, a failed dump-out is one of the worst plays when looking into goals against in the next possession. The specific area with the highest conversion rate to goal. except from right in front of the net. is from behind the goal line. Teams in Sweden generally fore-check a lot and get goals from this specific situation.

The analysis also concludes that a dump-in is a safer option when entering the at-tacking zone, than doing it with control. Still due to the increased likelihood of scoring, when entering with control, a controlled entry is the best alternative, when not considering the scoreboard or the time left of the game.

7 Discussion and future work

We have discussed risk/reward of different type of plays and areas within the sport of ice hockey in this paper. When discussing controlled vs. uncontrolled exits and entries it's easy to regard it as a decision made by the player executing the play. But. the teammates/opponents positioning, coaching directives and the sequences building up the situation all have major implications on the final decision made by the player executing the play. A coach cannot just instruct the players to do more con-trolled plays but needs to change the overall structure to make it possible. While this is not considered within the paper, it should be kept in mind.

It should be noted that we are in this paper mixing data from the Swiss League NL (reward) with the Swedish league SHL (risk). While we have no reason to believe that the results would be significantly different if we had either studied the leagues separately, or combined both leagues, this remains to be verified.

To calculate risk, we did not use expected goals against but instead actual goals against. The reason was the data available. The xG-model for the reward uses sequences within the buildup of the figure. The data we have at hand does not provide us that level of information. Using goals against, we get the actual outcome over three seasons which should correlate well with an xG-model including sequences.

For future work we would like to use data (risk and reward) from the same league to verify our results from this paper, but also investigate other leagues to find and important differences between leagues.

References

1. J. Weissbock. "Forecasting Success in the National Hockey League using In-Game Statistics and Textual Data" 2014. [Online].

https://www.ruor.uottawa.ca/handle/10393/31553, last accessed 2021/06/18.
2. T. Chatel. "Introducing Offensive Sequences and The Hockey Decision Tree" 2020. [Online]. https://hockey-graphs.com/2020/03/26/introducing-offensive-sequences-and-thehockey-decision-tree/
3. R. Stimson. "How Much Support does a good Breakout Need?" 2017. [Online] https://www.allaboutthejersey.com/2017/2/27/14667642/breakout-piece
4. M. Kelly. "Dump-Ins: The offensive & defensive effects of a trend on the rise" 2019. [Online] https://www.thepointhockey.com/dump-ins-a-trend-on-the-rise/
5. A.C. Thomas "The Impact of Puck Possession and Location on Ice Hockey Strategy". Journal of Quantitative Analysis in Sports. Volume 2. Issue 1. Article 6. 2016.
6. Tulsky E. Detweiler G. Spencer R. Sznajder C. "Using zone entry data to separate offensive. neutral. and defensive zone performance". MIT Sloan Sports Analytics Conference. 2013.
7. A. Toumi. "From grapes and prunes to apples and apples: Using matched methods to estimate optimal zone entry decision-making in the NHL". Ottawa Hockey Analytics Conference 2019.
8. M.Elomo. T.Poikonen "Analyzing reasons behind the goals in ice-hockey" (Bachelor‘s Thesis) 2015
https://www.theseus.fi/bitstream/handle/10024/95439/Elomo%20Poikonen%20Final.pdf?sequence=1isAllowed=y