Evaluating the 2019-20 Erie Otters through Markov Models

Greg Ackerman

3/5/2021

Introduction

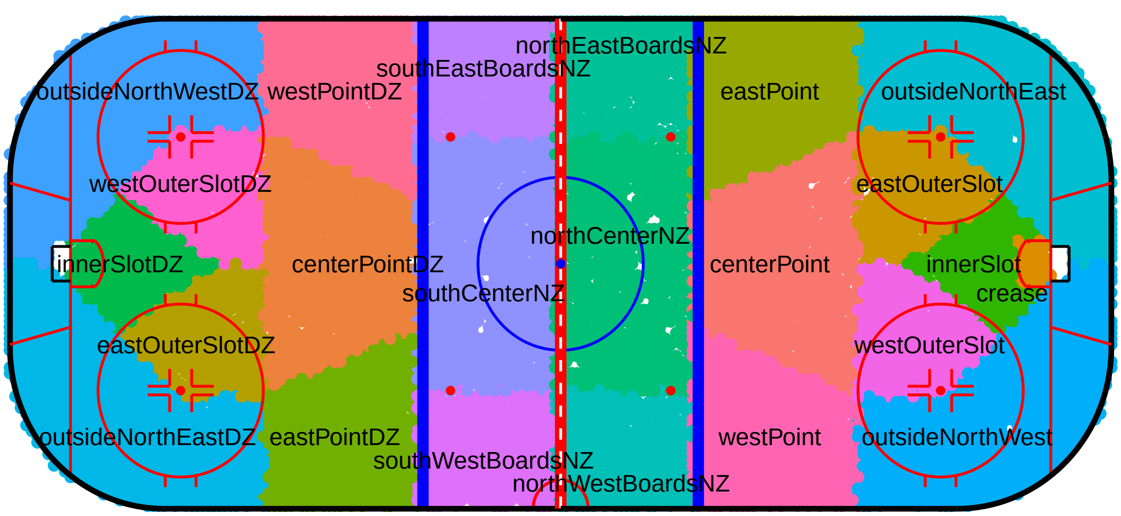
The hockey analytics community has obtained an excellent grasp of expected goals and possession metrics as an evaluation tool for hockey players from the professional to juniors ranks. While teams actively monitor these stats to improve hockey teams, there still remains a desire to understand the events proceeding shots and goals. Dividing credit on passes, entries, and other actions is a difficult task, and analysts and scouts alike struggle to objectively evaluate the incremental value of a play towards a goal.

The soccer analytics community has evolved its evaluation of players through Non-Shot Expected Goals Models. At the 2011 NESSIS conference, Sarah Rudd used touch-by-touch data from StatDNA and markov chains to weight the improved probability of a goal based upon deliberate attacking actions (passes, shots, etc.). In 2019, Derrick Yam built an attacking contributions markov model using StatsBomb's data.

Building off of the success from the soccer/football analytics community, this paper aims to use markov chains to evaluate how Erie Otters players contributed to 2019-20 goals scored, using Stathletes data.

Methodology

Using a similar framework to Rudd and Yam, I construct an OHL play-progression model through absorbing markov chains. An absorbing state in a markov chain is a state that, once entered, cannot be left. The two absorption states in this analysis are a Goal or Turnover. As Yam noted in his soccer analysis, inability to leave an absorption state makes using Goals much more favorable than raw or binned expected goals outputs. Leading up to a goal or turnover can be any number of transient states, in the case of this analysis are geographic zones of a pass, shot, or other non-goal event. As power play and penalty kill locations can be sparse with only 40 games of data, data is limited to just 5v5 play. I define the following geographic zones as transient states:



The geographic zones are based off of Sportlogiq's zone classifications from the 2020 Columbus Blue Jackets Hockey Analytics Competition. Transient states (zones) can transition between other transient states, and the transition probability (as later seen in markov chain results) is "memoryless" and depends only upon the current zone that the player is in, and thus is independent of previous states. The probabilities of transitioning from a zone to a Goal or Turnover are calculated using a transition matrix. The transition matrix has n transient states (23 play sections) and r absorbing states (2, goal or possession change). Q is the matrix of transition probabilities, Q = n x n. R is a matrix containing the absorption probabilities, R = n x r. The formula for the fundamental matrix, N, is N = (I-Q)^(-1), with I being the inverse of the n x n identity matrix. With this, the probability of a Goal or Possession Change is calculated for every play by calculating the product of N x R. The rest of this paper looks at just Goal probabilities, rather than looking at Turnover likelihood or expected possession length.

Markov Chain Results

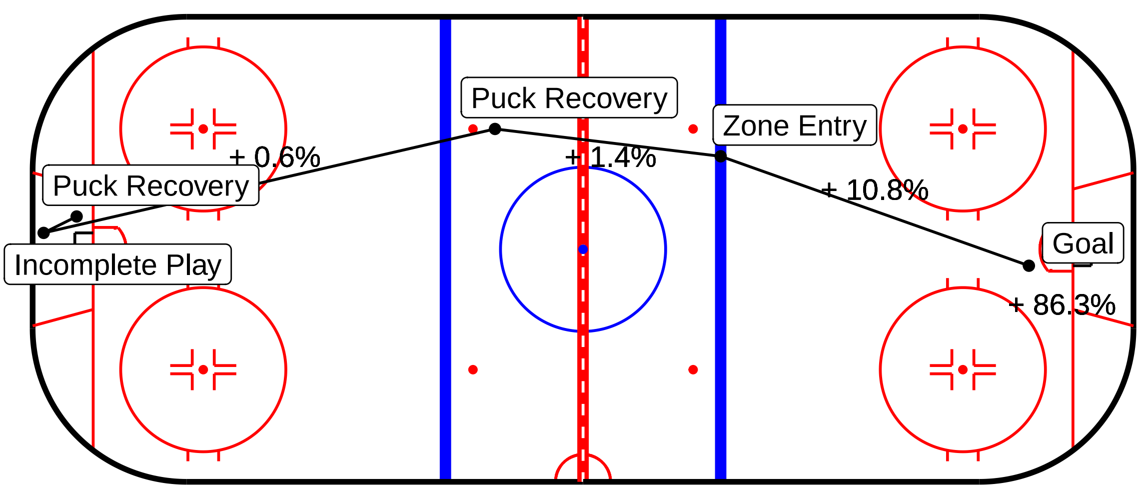
Unsurprisingly, the most likely geographic zones to result in a goal are the crease and slot areas, followed by the points and corners.

Top 10 Zones By P(Goal)

Play Section	P(Goal)
crease	31.8%
innerSlot	13.7%
eastOuterSlot	7.2%
westOuterSlot	7.0%
centerPoint	4.2%
westPoint	3.0%
outsideNorthEast	3.0%
eastPoint	2.9%
outsideNorthWest	2.9%
northCenterNZ	1.8%

I define a Goal Contribution as the incremental change in goal probability from a current state to the next state. Using the above table, a pass from the Outside North West to the West Outer Slot would result in a +4.1% Goal Contribution. To visualize these changes, I look at Danial Singer's goal on September 21, 2019.

Danial Singer Goal 9/21/2019 with Changes in P(Goal)

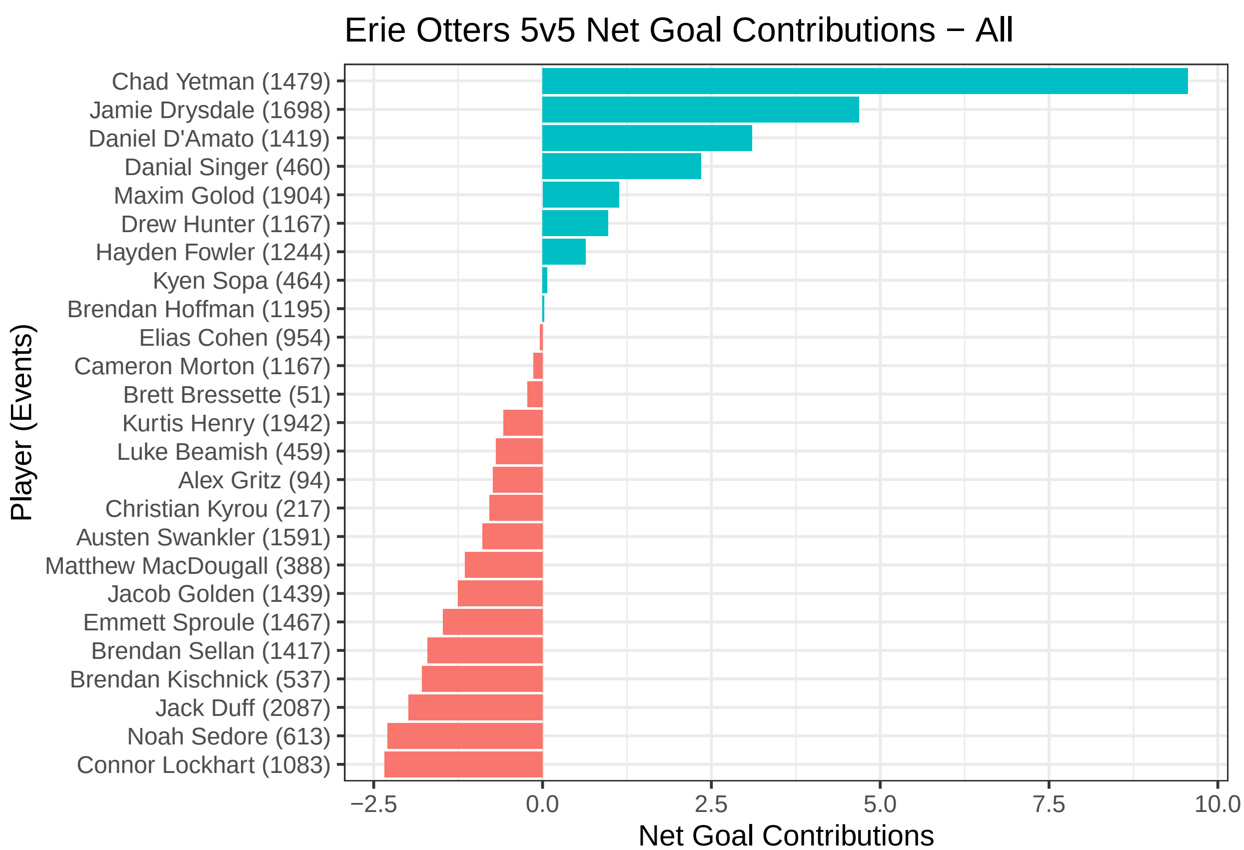


Jamie Drysdale recovers the puck behind his net, then attempts a pass off the boards that is incomplete, but recovered by teammate Danial Singer in the neutral zone, resulting in a (slightly) positive goal contribution. Singer carries the puck in for a successful zone entry, and then to the inner slot for a goal. The carry from the blue line to the inner slot changes the Goal Probability from 2.9% to 13.7% (+10.8%), and +86.3% represents the change from the inner slot goal probability (13.7%) to the absorption state of a goal (100%).

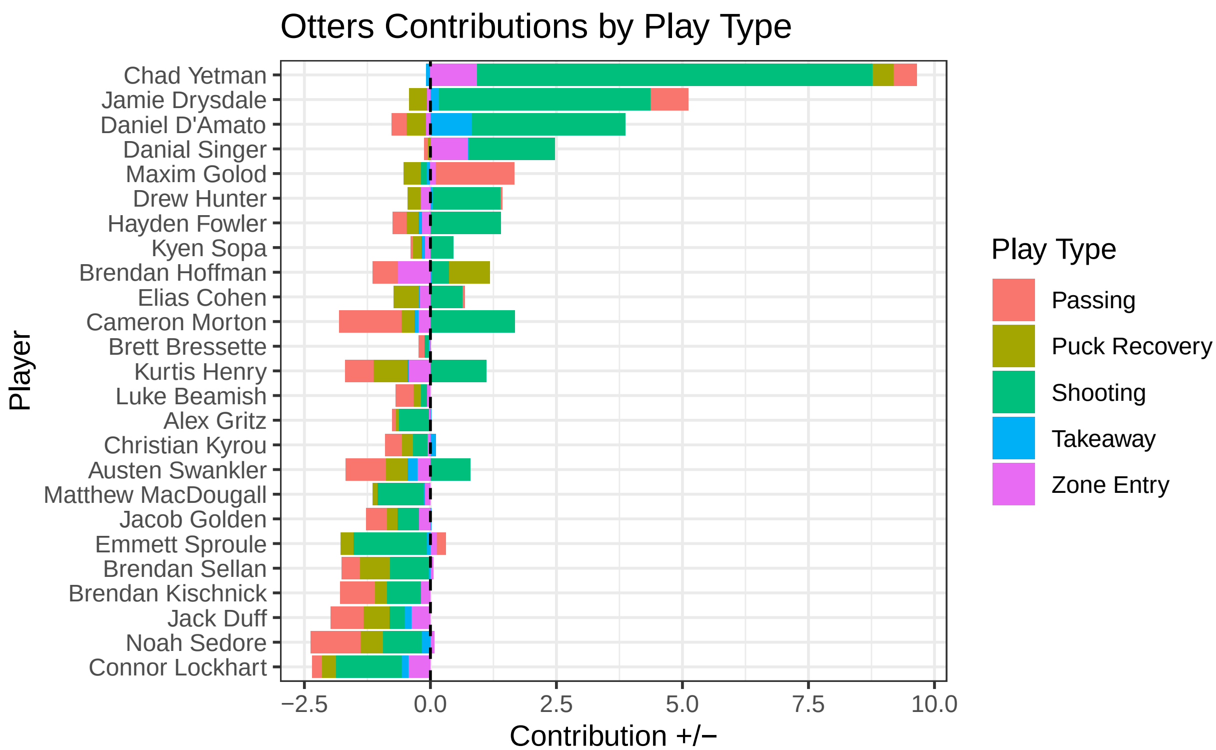
An advantage of using a markov chain is that it accounts for the fact that not all passes are created equally, which is not captured by simple completion percentages. An offensive zone pass to the slot is more valuable than a pass to the point. Additionally, a pass that might hit a skate and be labelled "incomplete" may still advance the puck to a likelier scoring state, and this method weights that positively where pass completion percentage would penalize the passer.

Scouting the Erie Otters

A player's net contribution is the sum of all individual contributions, positive or negative. This allows me to dig deeper into each Otters player's impact on goal probability as a whole and at a component level. Viewing all contributions, Chad Yetman, Jamie Drysdale, and Daniel D'Amato have the greatest net contributions, while Connor Lockhart, Noah Sedore, and Jack Duff had the lowest net contributions. Danial Singer performed decently well in a limited sample, while Noah Sedore played poorly in a small sample.



At the component level, we see top players Chad Yetman, Jamie Drysdale, and Daniel D'Amato, among others, who had were strong positives in their shooting (Goals + Non Goals). Chad Yetman had 43 goals in 61 games across all situations, so his shooting is unsurprisingly a driving force, but he is also a positive contributor in every aspect of attacking besides building off of takeaways. The Otters' scoring totals were top-heavy, so the distribution of total contributions is not overly surprising. However, it is important to remember that this is crude output, and context is needed to evaluate a prospect. While Connor Lockhart had the lowest net goal contribution, he was the second youngest player on the team and he was not a major negative in the other components, including passing. A scout can further evaluate if Lockhart was forcing bad shots because of bad teammates or if he was impacted by poor teammates.

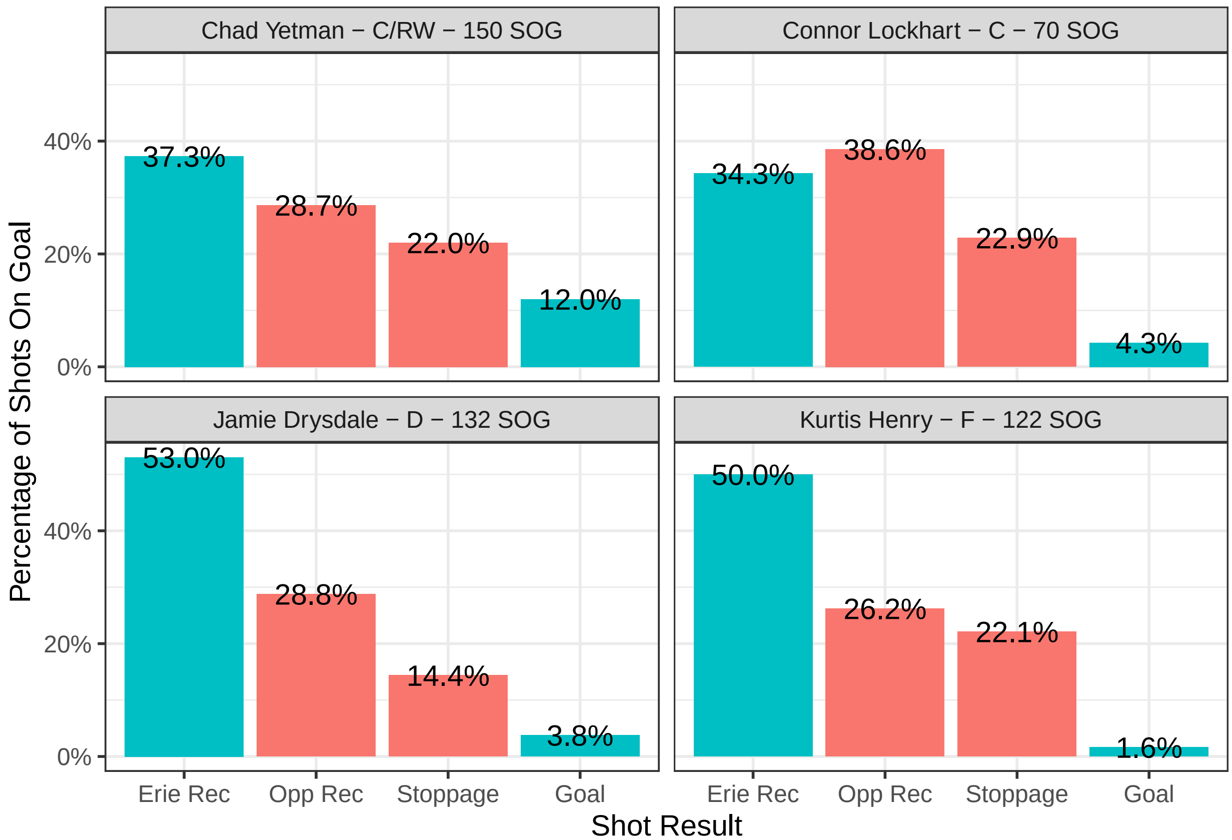


While the shooting component is heavily driven by Goals and thus some shooting luck, shots generating offensive rebounds in high probability zones are rewarded positively, while shots that result in a stoppages and opponent recoveries generate negative contribution values.

Otters Shooting Contributions
Shots On Goal, Min. 50

Player	SOG	Goal Contr	Shot Contr	Shooting Net Contr
Chad Yetman	150	16.0	−8.2	7.8
Jamie Drysdale	132	4.8	−0.6	4.2
Daniel D'Amato	124	8.3	−5.2	3.1
Danial Singer	80	4.9	−3.2	1.7
Cameron Morton	81	1.9	−0.3	1.7
Hayden Fowler	82	5.2	−3.8	1.4
Drew Hunter	80	1.8	−0.5	1.4
Kurtis Henry	122	1.9	−0.8	1.1
Austen Swankler	155	7.2	−6.4	0.8
Elias Cohen	56	4.0	−3.4	0.6
Brendan Hoffman	135	6.8	−6.5	0.4
Maxim Golod	181	8.2	−8.3	−0.1
Jack Duff	147	1.9	−2.2	−0.3
Jacob Golden	89	0.9	−1.4	−0.4
Brendan Sellan	122	5.2	−5.9	−0.8
Connor Lockhart	70	2.8	−4.1	−1.3
Emmett Sproule	115	5.0	−6.4	−1.5

Looking at shots on goal, Jamie Drysdale's shooting contributions are quite positive despite only shooting 3.8%. 53% of Drysdale's shots led to an Erie recovery, explaining why he rates so highly. Connor Lockhart as a forward does not shoot a very high percentage, and his shots resulted in fewer Erie rebounds. Yetman's shooting was strong and his rebounds resulted in more Erie recoveries than opponent. Kurtis Henry didn't shoot a very high percentage, but he was not heavily penalized for his shooting ability as a forward.



So what is Kurtis Henry doing differently than Lockhart to succeed as a forward? Shooting into traffic. 78.7% of Henry's shots on goal had traffic in front, compared to Lockhart's 37.1%. Corey Pronman's scouting report of Jamie Drysdale mentioned "His shot isn't that heavy, but he finds teammate's sticks often for deflections." With 81.8% of Drysdale's shots coming with traffic in front, it is clear he is trying to get the puck to the net and allow his teammates to finish the plays, and the markov chain rewards Drysdale's propensity for getting the puck to the net, even with a comparable lowly shooting percentage to most defensemen.

Conclusions

The goal contribution markov model succeeds in creating an easily digestible metric that can handle arbitrarily long hockey sequences. The model's simplicity is a feature, as it allows for a goal contribution of every single play from Stathletes' data, and non-shooting events are quite difficult to quantify. Derived from the chain are Goal Contributions, which weight every event on their increase or decrease in goal probabilities, based off of actual goals scored. The model does place a heavy emphasis on goal scoring, but also rewards shooters who shoot into generate and generate subsequent offense in high danger areas. In the case of the 2019-20 Erie Otters, the top-heavy nature as seen in their basic point totals is reflected in their individual contributions. Driving play is no simple task, and the non top-end players will be forced to pass to the points and make tough passes that are on average a negative. Thus, players with a negative offensive Goal Contribution can still be important players to a team, including those that are more defensively minded.

The markov chain operates as "memoryless", while in reality hockey is not a memoryless game and hypothetical "true goal contributions" would be conditionally dependent upon prior events. Additionally, the use of just Goals and Turnovers combined with binned geographic zones removes important context such as shot angle and distance that has an impact on goals, as well as passes that put the goaltender out of position. The model also treats a pass from zone A to zone B equally to a carry, while these events have differing difficulties in reality. Continued research is necessary to understand the repeatability of these metrics and further context of teammates and positions on their contributions. Regardless of these limitations, the markov chain does capture the potentially positive impact of an errant pass, the value of a shooter's rebounds, and the player's abilities to make plays after a takeaway or a successful zone entry.

With access to unlimited data and resources, I would want to know the defensive pressure for each play, and potentially obtain a strong classification of play type (on-rush vs. second chance, etc.). A blind spot of the model is that horizontal play will result in marginal goal contributions on certain passes, while Steve Valiquette and Clear Sight Analytics estimate 22% of all goals scored cross a "royal road". On those types of passes, the passer likely passes from an area of pressure, but finds a shooter with more space. By having some observed or estimated level of pressure, I believe the markov chain would more generously award Goal Contributions on horizontal plays. Additionally, I would like to have on-off data to record time-on-ice and put these net contributions on a per 60-minute basis. Lastly, as the majority of observations are from the Erie Otters, the inclusion of non-Otters data may generate better contribution values that reflect the OHL as a whole.

References

Albert, Jim. "Kershaw's Pitch Count Transitions." Exploring Baseball Data with R, 30 May 2016, baseballwithr.wordpress.com/2016/05/30/kershaws-pitch-count-transitions/.

Boyle, Chris. "The (Royal) Road to the Future of Goaltender Analytics." MSGNetworks.com, 27 Mar. 2017, www.msgnetworks.com/2015/03/06/the-royal-road-to-the-future-of-goaltender-analytics/.

Pronman, Corey. "Pronman's Scouting Report: What the Ducks Are Getting in Jamie Drysdale." The Athletic, The Athletic, 6 Oct. 2020, theathletic.com/1804475/2020/10/06/pronmans-scouting-report-why-jamie-drysdale-is-a-top-nhl-prospect/.

Rudd, Sarah. "A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains." nessis.org, www.onfooty.com, 2011, nessis.org/nessis11/rudd.pdf.

Yam, Derrick. "Attacking Contributions: Markov Models for Football." StatsBomb, 1 Apr. 2020, statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/.

Yurko, Ron. "NFL ScrapR Markov Model." GitHub, github.com/ryurko/nflscrapR-data/blob/master/R/markov_model.R.

Acknowledgements

Thanks to Ian Anderson of the Philadelphia Flyers for his help, his guidance led me to obtaining a better understanding of soccer Non-Shot Expected Goals Models, and opened up the idea for me to implement such a model in hockey. Thank you to Stathletes and the Erie Otters for the data provided in the Big Data Cup.