Shane Hauck

Big Data Cup 2022

Introduction

What makes a theoretically good offensive team on the powerplay is the team's ability to use the advantage of having an extra person on the ice to create the shot with the highest probability of being a goal. Shot quality has been a heavily researched topic in the realm of hockey analytics, however classifying defensive structure in terms of data analysis has been neglected. Rather than this paper focusing on the impact that individual players have during 5 on 4 powerplays, I have veered towards a strategical analysis of defensive shape and offensive possession tactics.

In order to make up for the deficit in players on the ice, penalty kill defenses need to work as a unit to cover the zones of the most danger. The four defensive players that are apart of the penalty kill essentially form a 4 sided polygon shape that is constantly adjusting in size and position to the offensive actions. To represent this shape, I used the tracking data from women's hockey games during the 2022 Winter Olympics in Beijing to calculate a **defensive centroid**, or the arithmetic mean position of all positions of players in the defense, to make a centralized point that generalizes the defensive unit. Minimizing the defense to an individual point allows for the position on the ice which has the closest thing possible to equivalent pressure from all angles of the defenders. To best represent the effect that the defensive centroid has on shots, I have plotted the positions of all available defensive players in the tracking data at the time that shots occurred according to the play by play data.

Previous research has shown that shot quality is enhanced on the power play. This could be due to shots of more dangerous distance or shot types. The same shot type from the same distance has a better chance of success on a power play (Ryder, 2004). It would seem that greater puck control gives rise to more dangerous shots (Ryder, 2004). Offensive teams are able to use their player surplus to create shots that feature less pressure from the defensive team. I believe that this is likely due to the fact that offenses are able to get closer to the goal than the centroid of the defense resulting in higher quality shots.

The initial goal of this study was to look at the patience of offensive teams during 5 on 4 power plays and what level of patience leads to the most consistent creation of high quality shots. To do this, I wanted to look at ways the teams opened up defenses to get in behind the defensive pressure in a way that created the most controlled shots from good angles. This paper will use two shot quality models, which compute probabilities that a shot will be a goal based on shot type (slapshot, snapshot, wristshot, and tip-in) and shot angle respectively. These models are used to ultimately answer the question of "Does taking a shot behind the centroid of the defense have an impact on the quality of shot created?"

Shot Quality Models

Both shot quality models were created by Ken Krzywicki (one from 2005 and the other from 2010). The main predictor in both of these models is the distance of the shot release point to the goal. This was calculated by:

$$distance = \sqrt{((189 - x_{coord})^2 + (49 - y_{coord})^2)}$$

The other calculation that was done outside of the dataset was the shot angle which is used in the second model. This was calculated by:

$$\theta = Tan^{-1}\left(\frac{y_{coord}}{89 - x_{coord}}\right)\left(\frac{180}{\pi}\right)$$

Variable	Range	Points
Intercept	Add to all records	-2.2369
Distance	Less than 10 ft	0.6884
	10 - 12 ft	0.6374
	13 - 14 ft	0.5564
	15 - 16 ft	0.5174
	17 - 22 ft	0.3654
	23 - 31 ft	0.0000
	32 - 36 ft	-0.3805
	37 - 38 ft	-0.4758
	39 - 44 ft	-0.8155
	45 - 57 ft	-1.0848
	58 ft or more	-1.3824
Shot Type	Wrap	-0.0742
	Slap	-0.0573
	Wrist	0.0093
	Snap	0.0130
	Backhand	0.0361
	Tip-In	0.1487
Rebound	Yes	1.3362
	No	0.0000
Situation	EV	-0.1244
	SH	0.0399
	PP	0.4007

Variable	Interval	Points
Intercept	Add to all records	-2.2899
Distance (adjusted)	Multiply points by adjusted distance	-0.0437
Shot angle	Multiply points by abs(shot angle)	-0.0162
Rebound	No	0.0000
	Yes	0.9948
Situation (shooting team)	EV	0.0000
	SH	0.0000
	PP	0.4370

The model score, or predicted probability of a goal, is defined using the above tables as follows:

$$P(GOAL) = \frac{1}{1 + e^{-\Sigma points}}$$

The reason that two models were chosen to be included in the research is due to the idea of representing two different factors of shot quality: control and position. The first model is representative of a shots control with its predictive variables including shot type. The second model is representative of a shots position as it uses shot angle for calculations. The use of the models is intended to be in accordance with one another and not necessarily for comparison. The mean probabilities for the models based on the shots in the dataset are given below:

Control Model Mean	Position Model Mean
0.103	0.038

Approach

The data used to conduct the research during this project was filtered to only include plays that occurred during 5 on 4 powerplays and in the offensive zone of the team with a player up. All research is focused on the shots taken in the positions during these powerplays and various factors that play into the quality of the shot.

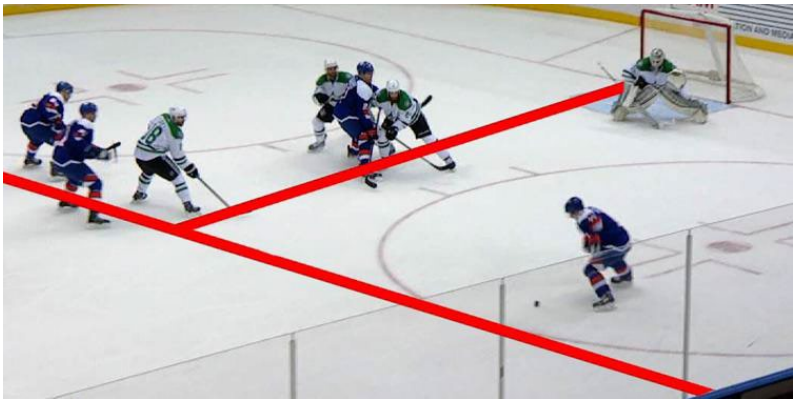
To investigate the patience of offenses, variables were created to represent the total number of consecutive successful passes and time that took to create a shot. The streak of consecutive successful passes and total time in between them were reset upon an event occurring that was a shot, unsuccessful pass, a puck recovery, or a zone entry.

To calculate the position of the centroid I converted the frames from the tracking data into the seconds that are equivalent to which it would be in the play by play data. I then calculated the x and y coordinates of the centroids based on the positions of the tracked defending teams' players. To eliminate the goalkeeper from the calculation of the centroid, if a player was closest to the goal and within a 6 foot radius of the goal center, their location was omitted. The calculations for the centroid's x and y coordinates were created following the formula:

$$centroid_x = \frac{\sum x_{coord}\ of\ defending\ teams\ players}{total\ number\ of\ defending\ teams\ players} \mid centroid_y = \frac{\sum y_{coord}\ of\ defending\ teams\ players}{total\ number\ of\ defending\ teams\ players}$$

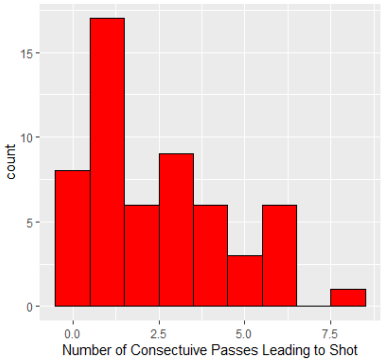
There were inconsistencies in the data where in certain instances, not all players were tracked. To best counter this deficiency, time of the positions used were selected by taking the frames with the most observations available.

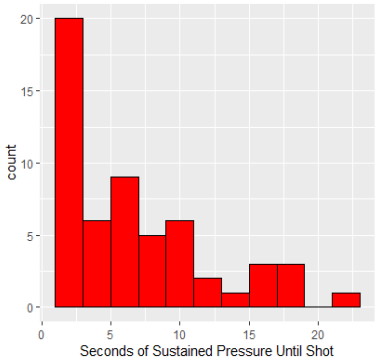
In depth analysis was conducted based on the location and probability of the shot with regards to its position relative to the defense's centroid. Throughout my research I stumbled upon the theory of passes across the Royal Road and the impact it has on the success rate of shots. The Royal Road is a line that goes directly through the middle of the ice from the goal center to a plane at the furthest perimeter of the faceoff circles in the offensive zone(Boyle, 2019). It has been studied in the National Hockey League(NHL) that passes across the Royal Road account for 22% of goals(Boyle, 2019). This type of movement is essential to goal creation because when the puck moves laterally with speed in this manner, it does not allow the goaltender to remain square because they struggle to set their depth and angle, making the save more difficult(Boyle, 2019).



For all plays directly leading to shots, a variable was created in the dataset to investigate the trajectory of the pass path and whether or not it intersected the royal road. The concept of the Royal Road in combination with my theory of defensive centroid position, investigates probability of successful shots.

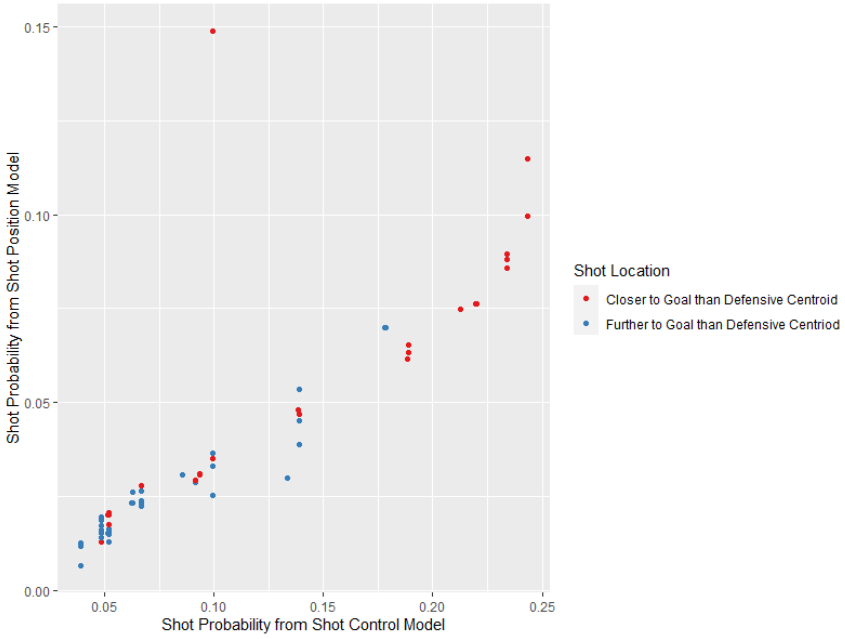
Findings



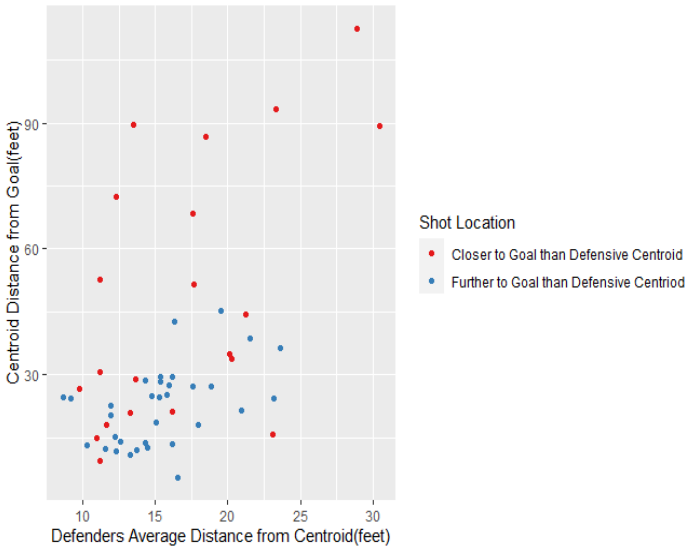


When investigating the patience of offenses during 5 on 4 power plays it became adamantly clear that teams generated the majority of their shots from quick possessions. Both of the histograms exploring the offensive zone build up have a right skew showing that it was a more common result for teams to take as many shots as possible after minimal passes and time rather than stringing together a high number of passes and possession during their 2 minute advantage. This is important to keep in mind as the rest of the data is investigated because it appears that the schemes of team's offensive power plays were centered around shot volume rather than shot quality.

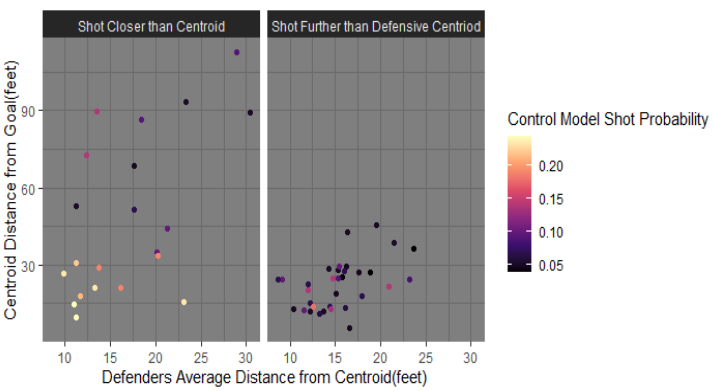
By evaluating the quality of shot with regards to the shots location to the defensive centroid, there appears to be a very clear association with the highest quality shots coming as a result of the shot location being closer to the goal than the location of the defensive centroid. When looking at the graph, a cluster of blue points takes shape towards the bottom left hand corner meaning that the majority of low quality shots taken during 5 on 4 powerplays were as a result of the shot location being ahead of the centroid of the defense. When shifting to the opposite corner where the shots are of higher quality, red points dominate the area due to the majority of high quality shots coming from in behind the defensive centroid. Although this data proves fruitful to my theory, it is important to keep in mind that shot distance from goal is a significant predictor in both shot quality models. It is likely to be true that for a shot to be

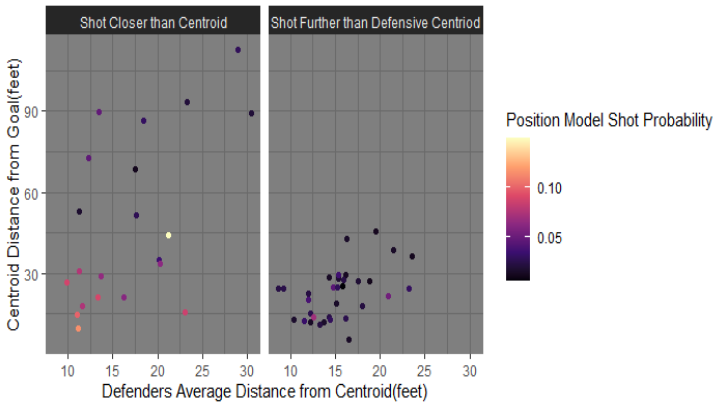


behind the centroid, it has to be relatively closer to the goal then a shot that is ahead of the centroid, hence resulting in a higher shot probability.



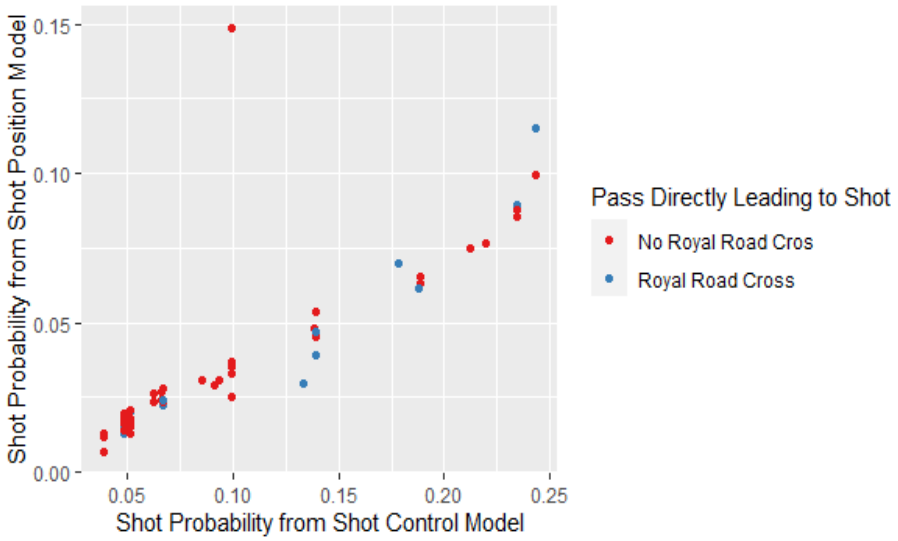
To attempt to help explain the reason for shots occurring behind the centroid of the defense, I decided to look at the overall structure of the defense. Variables were created to look at the centroids distance from the center of the goal and the average distance of the defenders from their team's centroid. The centroids distance from goal is intended to represent how "deep" a team are defending. The deeper a team's centroid is, the closer their centroids distance is to their goal. By looking at the defender's average distance from the centroid it can be interpreted how "spread out" a defense is. The more spread out defenses are, the greater the average distance each player is from the centroid. When looking at the graph above, when a defense is compact and close to its goal, the chances of the offense getting in behind their centroid are now far slimmer but still possible.

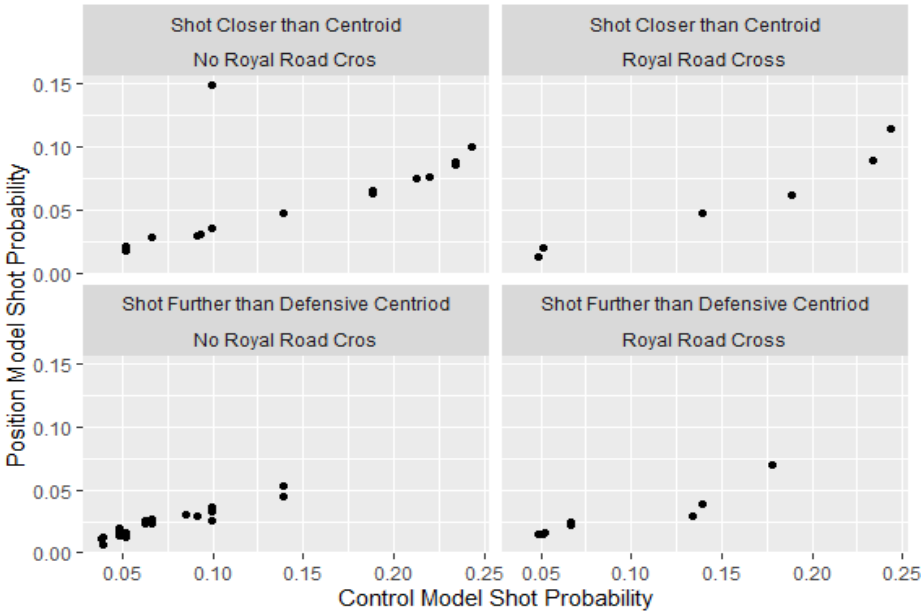




When looking at the shot probability models in comparison to the defensive position, high quality shots come at a premium when offenses are able to get a shot off closer to the goal than the location of the defending teams centroid. More specifically, when teams are able to break down a compressed defensive structure that is tight to goal, the goal probability is inflated.

When a pass successfully crosses the Royal Road directly before a shot, the majority of shots taken produce a high quality opportunity. However, these types of passes leading to shots appear to be quite rare in this dataset. From the data shown here, it would be hard to draw a conclusion verifying that passes that cross the Royal Road directly before a shot yield a higher goal probability than shots that do not.





Finally, when investigating the combination of the defensive centroid alongside the passes across the Royal Road theory, we can see clear trends supporting the most consistent high and low quality shots. When shots get taken from a position that is both further from goal than the defenses centroid and the previous pass failed to traverse the Royal Road, shot probability in both models was consistently low. Shots were more consistently of higher quality when the shot was either closer to goal than the defenses centroid or the prior pass crossed the Royal Road. However, the scenario that most efficiently yielded a high shot probability was a result of both conditions being met.

Conclusion

To best limit scoring opportunities during a 5 on 4 penalty kill, defenses need to best protect against penetration beyond their centroid and limit passes crossing the Royal Road.

References

Boyle, C. (2019). The Most Important Line on the Ice you've Never Heard of: The "Royal Road" and why it leads to more goals. *Omha*. https://www.omha.net/news_article/show/486107-the-most-important-line-on-the-ice-you-ve-never-heard-of

Krzywicki, K. (2005). Shot Quality Model: A logistic regression approach to assessing NHL shots on goal. *Hockey Analytics.* http://www.hockeyanalytics.com/Research_files/Shot_Quality_Krzywicki.pdf

Krzywicki, K. (2010). NHL Shot Quality 2009-10: A look at shot angles and rebounds. *Hockey Analytics. http://hockeyanalytics.com/Research_files/SQ-RS0910-Krzywicki.pdf*

Ryder, A. (2004). Shot Quality: A Methodology for the Study of the Quality of a Hockey Team's Shots Allowed. *Hockey Analytics.* http://hockeyanalytics.com/Research_files/Shot_Quality.pdf