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.

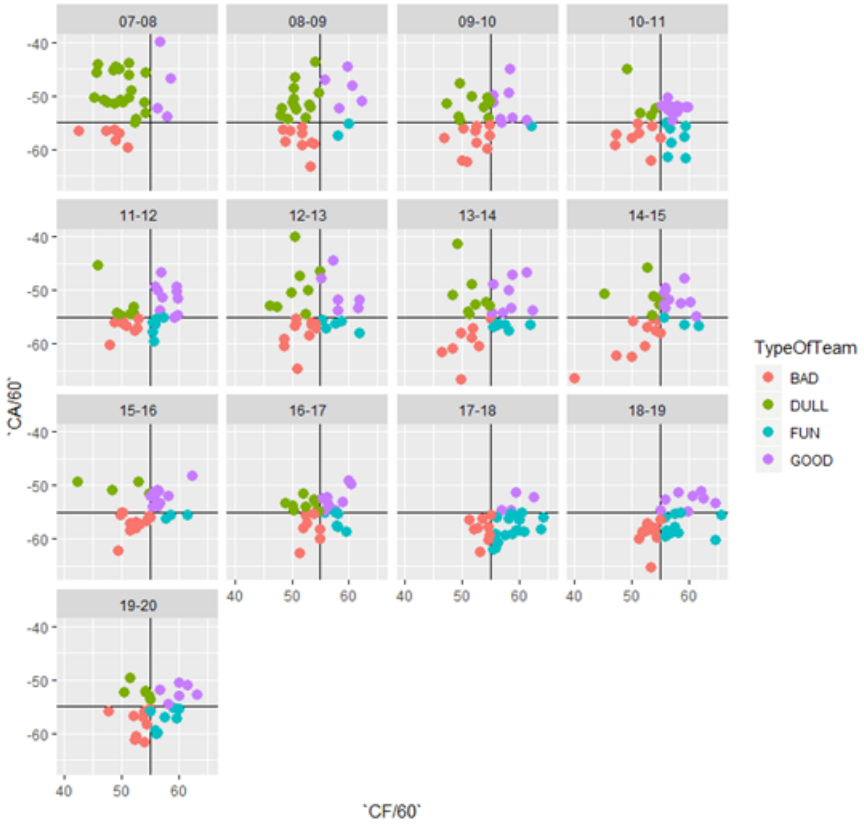

Fig. 2. NHL team development

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.

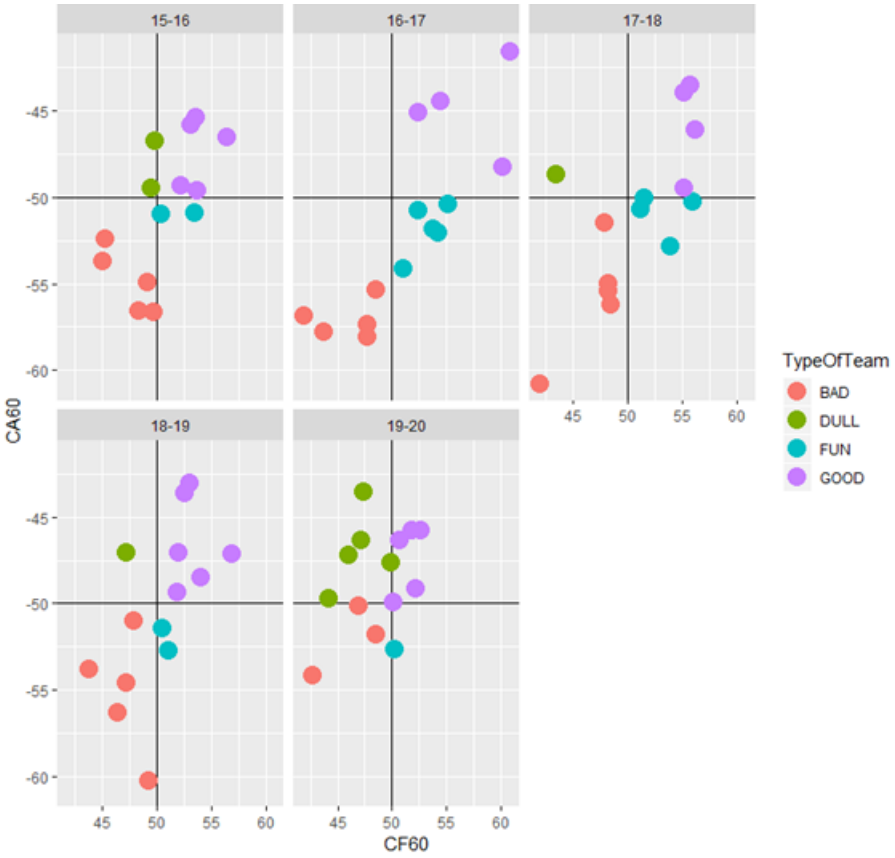

Fig. 3. SHL team development

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.

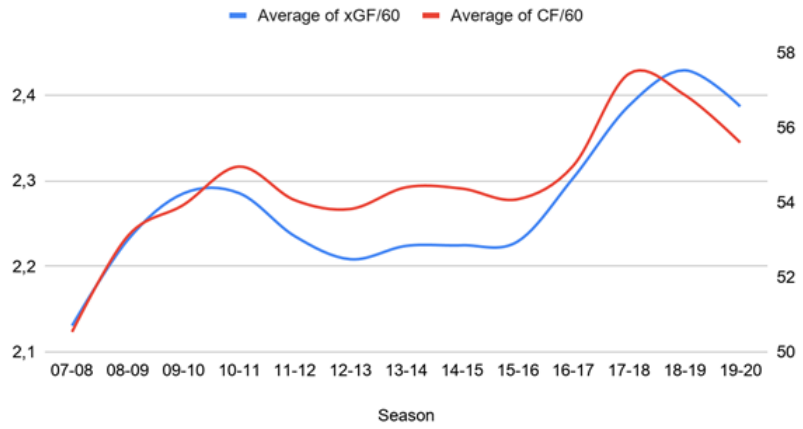

Fig. 4. Corsi vs. xG in NHL

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.

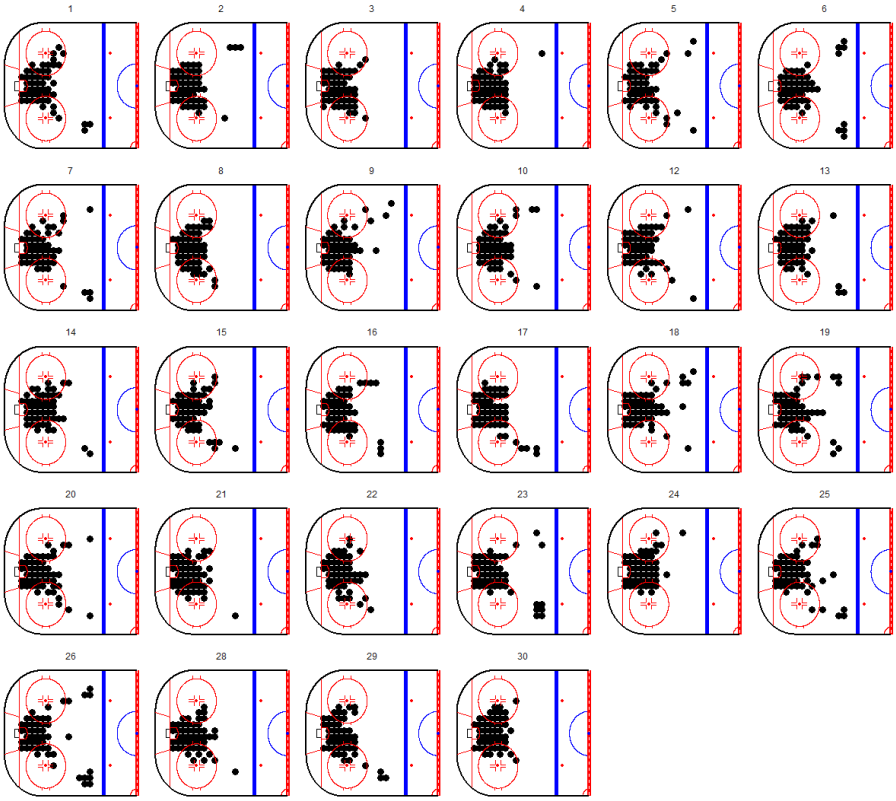

Fig. 5. NHL shot positions 2010

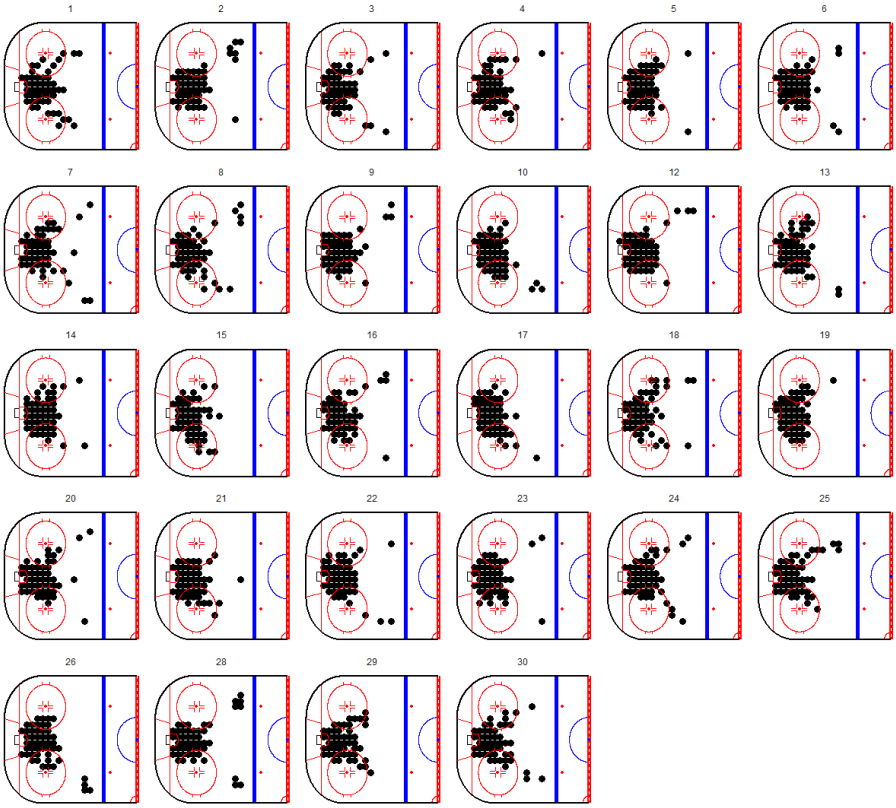

Fig. 6. NHL shot positions 2018

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)