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

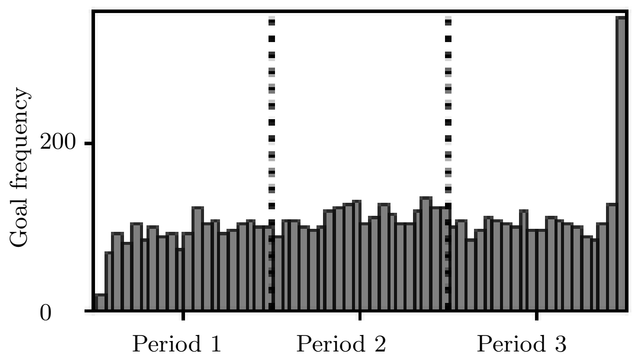

Fig. 1. Goal frequency for each minute of the first three periods in the NHL during the 2013-2014 season.

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:

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

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:

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

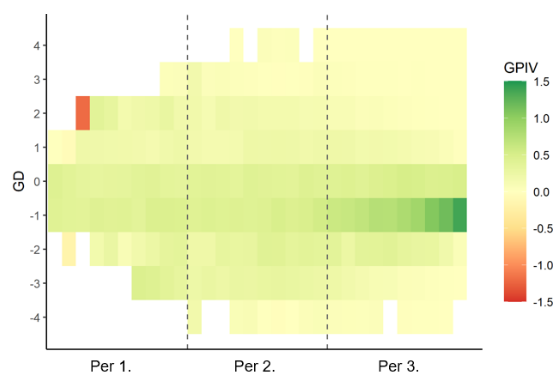

Fig. 2. GPIV versus GD for the 2013-2014 season. Each bin is two minutes. Less than two observations for each bin are left out.

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.

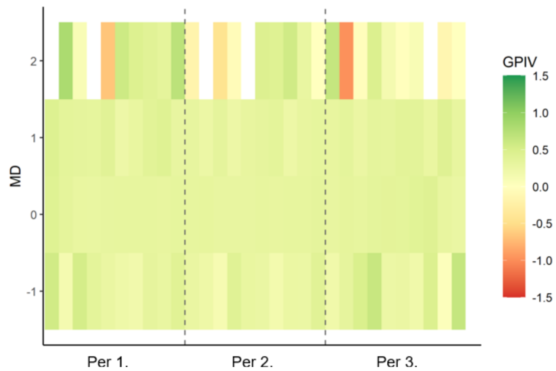

Fig. 3. GPIV versus MD for the 2013-2014 season. Each bin is two minutes. Less than two observations for each bin are left out.

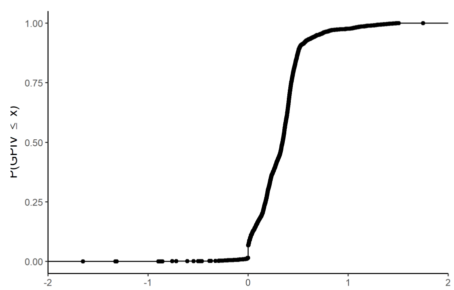

Fig. 4. Cumulative distribution function of GPIV.

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.

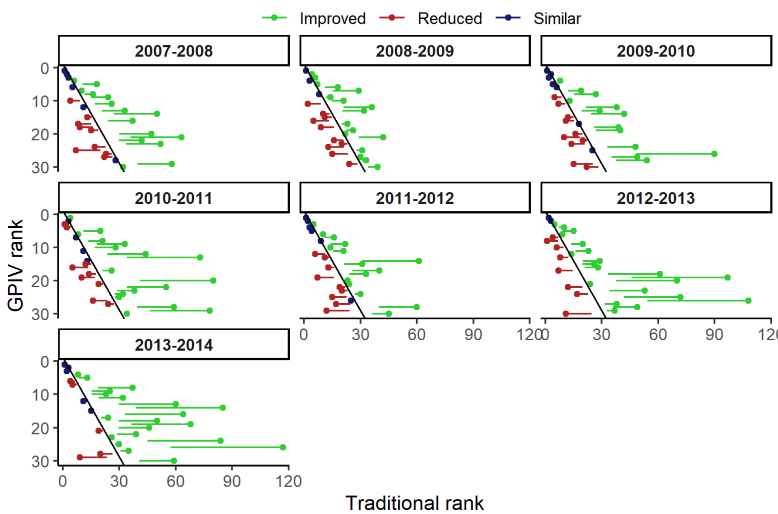

Fig. 5. Rank comparisons for traditional goals and GPIV-goals.

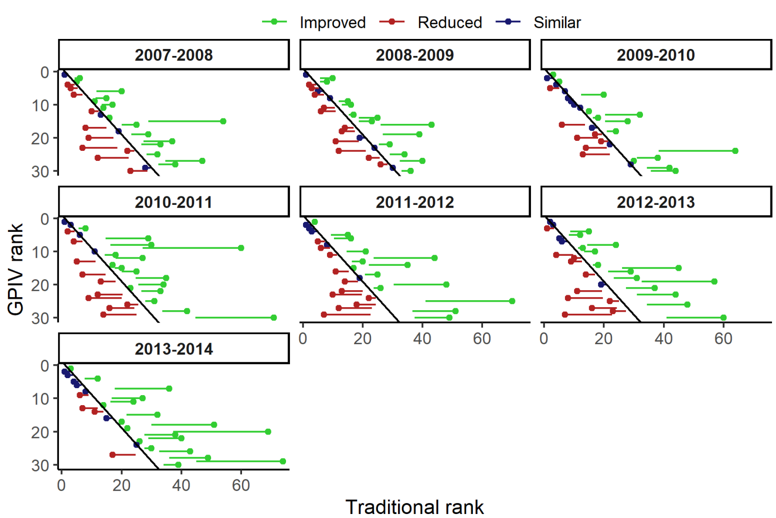

Fig. 6. Rank comparisons for traditional assists and GPIV-assists.

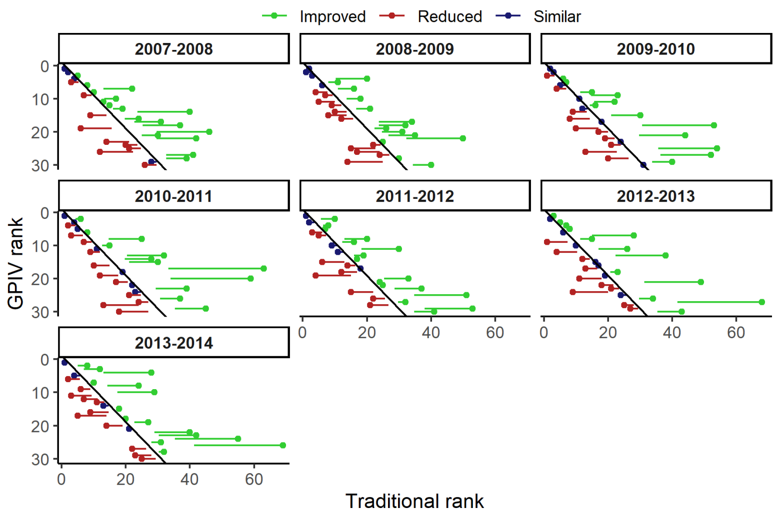

Fig. 7. Rank comparisons for traditional points and GPIV-points.

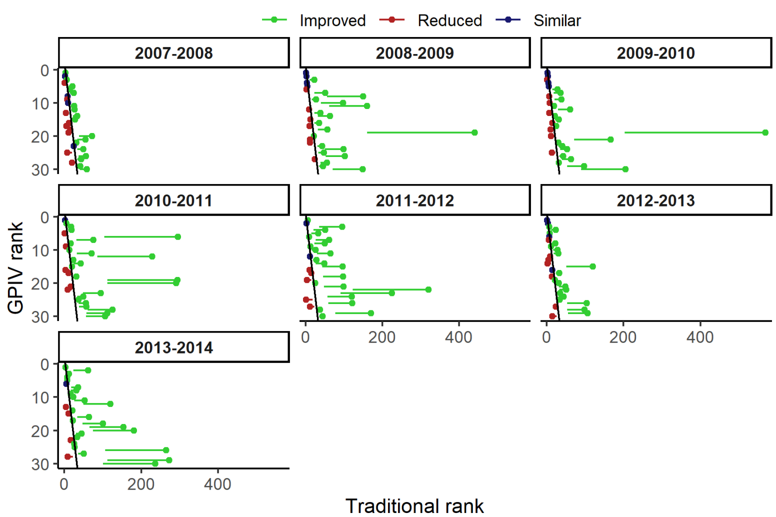

Fig. 8. Rank comparisons for traditional +/- and GPIV-+/-.

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-

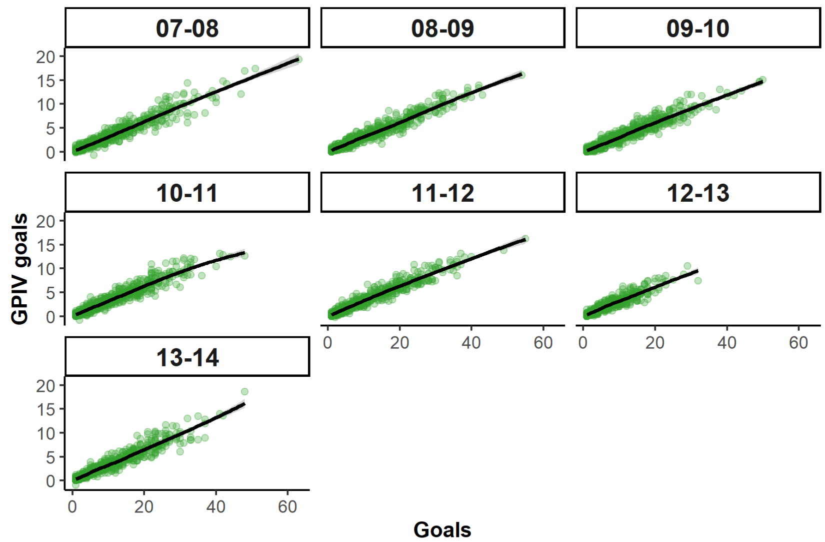

Fig. 9. Correlation traditional goals and GPIV-goals.

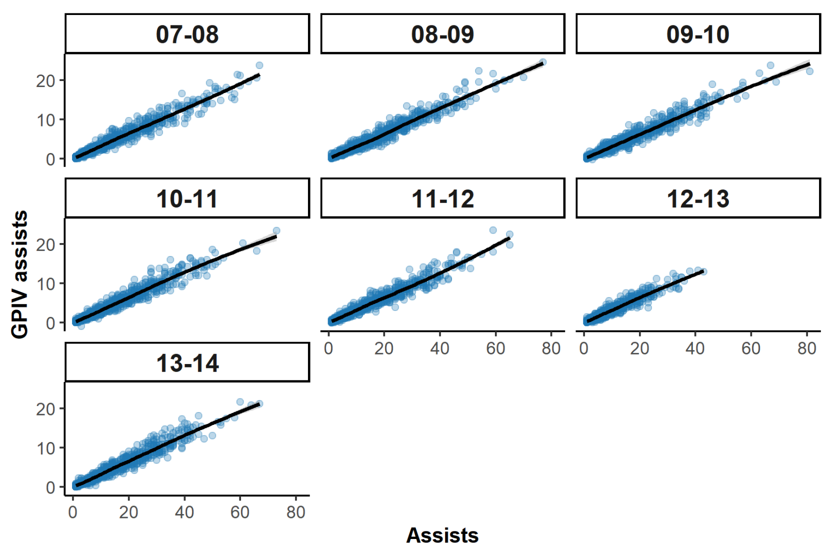

Fig. 10. Correlation traditional assists and GPIV-assists.

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.

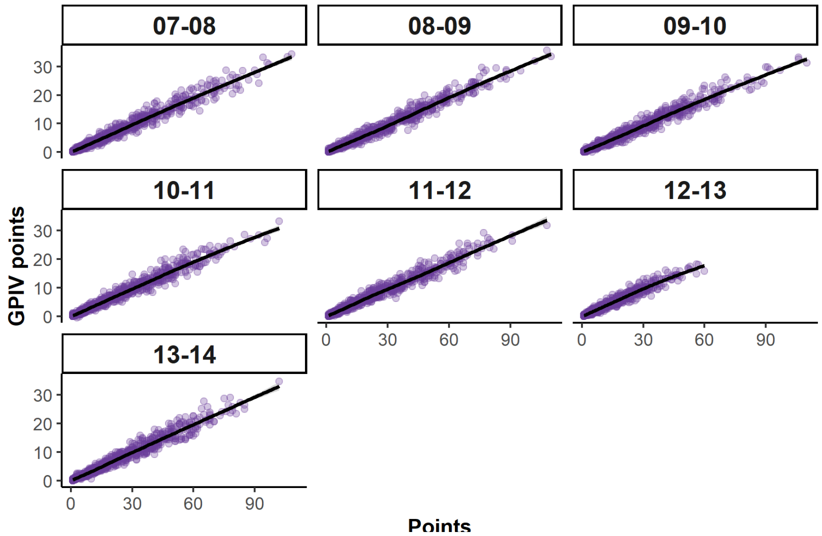

Fig. 11. Correlation traditional points and GPIV-points.

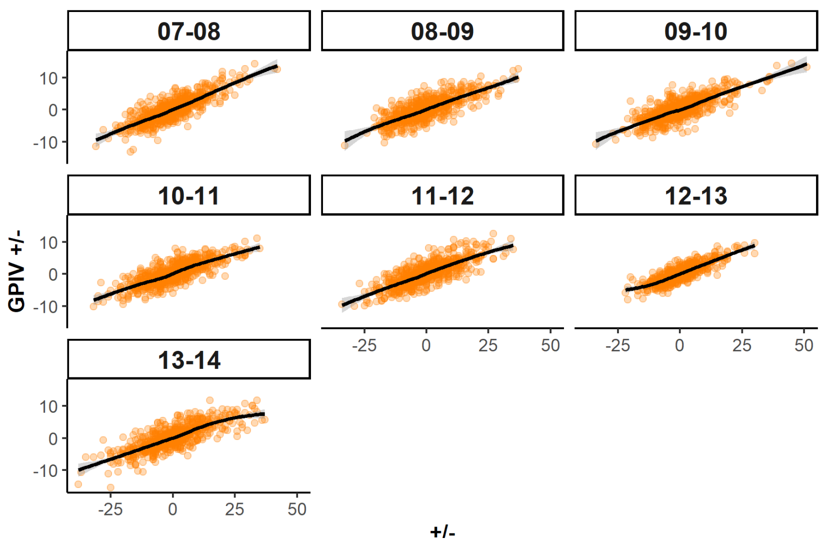

Fig. 12. Correlation traditional +/- and GPIV-+/-.

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.

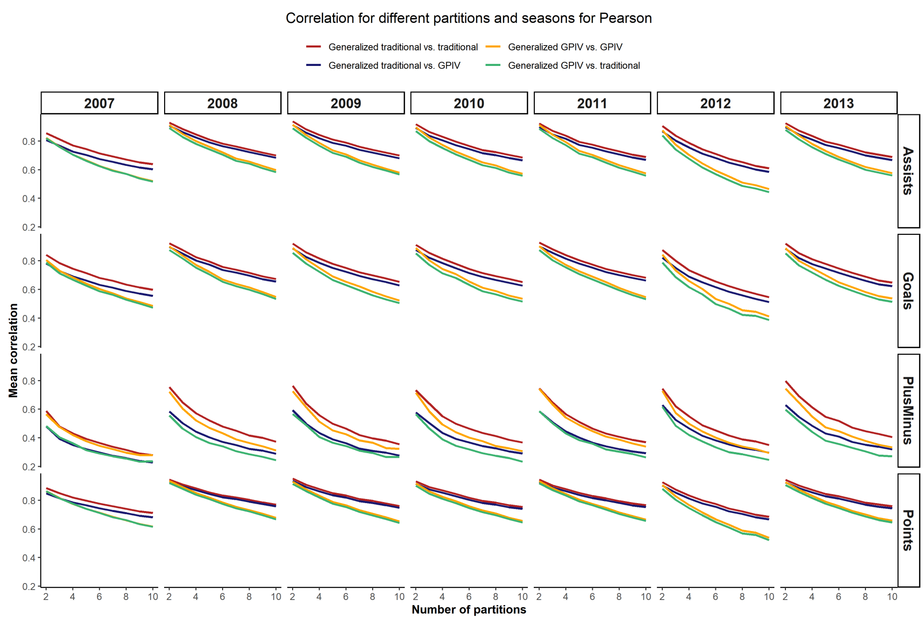

Fig. 13. Correlations for partitions for different metrics.

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 ingame 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.