Simple and Practical Goal Importance Metrics for Ice Hockey

Rasmus Säfvenberg, Niklas Carlsson, and Patrick Lambrix

Linköping University, Sweden
firstname.lastname@liu.se

Abstract. To capture that not all goals are of the same importance, a new performance metric called the Game Points Importance Value (GPIV) was recently proposed. While this metric takes into account the expected impact that a goal has on the outcome of a game based on the context when the goal was scored, it relies on a relatively fine-grained state space. To address this problem, this paper presents simplified and more practical variations of the GPIV metric. Motivated by our analysis of the relative importance of different dimensions of the state space, we present two metrics that capture the most important component(s) of GPIV. Our evaluation shows that the metrics are relatively stable and capture most of the relative differences between GPIV and traditional metrics (e.g., goals, assist, points, and +/-). These results suggest that these simple and practical metrics are intuitive, capture most of the desirable variations that GPIV captures, and that the value of a goal can be well estimated using GPIV data based on historic data.

1 Introduction

In ice hockey, not all goals are of equal importance or have the same impact on the outcome of a game. For example, a game-tying goal in the final minute of regulation has a greater impact on the game outcome than a goal scored while leading by seven goals, as in the latter case the outcome is all but decided. In recent work [7, 8], we proposed a metric to quantify the importance of a goal on the game outcome in the National Hockey League (NHL). This metric, referred to as Game Points Importance Value (GPIV), accounts for the current goal difference (GD), manpower difference (MD), and the time of the game when the goal was scored. For each such (goal) state, we then compute a GPIV value, quantifying the goal importance, as the estimated change in the weighted probabilities (before vs. after the goal) for winning, losing in overtime, and losing in regulation. As an example, a game-tying goal in the final minute of regulation will increase the probability of winning and losing in overtime while simultaneously reducing the probability of losing. Such goal will therefore obtain a relatively large GPIV. In contrast, a goal scored when leading by a large margin will have negligible impact on the expected outcome, resulting in a small GPIV.

One downside of GPIV and other complex metrics is that they rely on a relatively fine-grained state space. To address this problem, this paper presents and evaluates two simplified and more easy-to-use variations of the GPIV metric. To derive and motivate these metrics, we make use of a decision tree and estimates of the variations that each of the state parameters are responsible for when using the original GPIV metric, and then define approximate metrics based on the insights provided. The resulting approximations capture the most important component(s) of GPIV and provide practitioners with an easy and straightforward application of the metric to real-time situations.

Like the pure GPIV metric, instead of attributing each goal an equal value, the approximations assign each goal a value based on the current state. However, by using a much smaller set of states, the approximations provide a more intuitive description of which goals have the highest importance within a game. These simplified valuations of player performance can therefore provide fans, teams, and media with an easy-to-apply metric for evaluating and comparing players that account for goal importance. Our evaluation of the approximations also shows that the metrics are relatively stable, allowing past seasons (or past games played) to be used to estimate and apply the metrics on current and future games. As desired, the metrics also have stronger correlations with GPIV than with the corresponding traditional metrics, and the relative player ranking variations (compared to traditional metrics) capture most of the explainable variations that have been observed using GPIV. These results suggest that these simple metrics are practical, intuitive, and capture most of the desirable goal importance variations captured by GPIV.

Throughout the remainder of this paper, data from the 2013-2014 NHL regular season is used to illustrate the approach.

2 Two simplified GPIV metrics

The original GPIV metric is based on each goal being its own state, where a state is represented by a time (in seconds), GD, and MD. However, not all these components are of equal importance for the computation of the GPIV value.

To evaluate the importance of each variable (i.e., time, GD, and MD), a decision tree was fitted with GPIV as the outcome and time, GD, and MD as variables. These results are shown in Figure 1. From this decision tree, we observe that GD is the variable with the most splits, followed by time, while MD had no splits. This relative ranking is also echoed by the variable importance (summarized in the same figure): GD (216.177) being the most important variable, time (76.586) being the second most important variable, and MD (0.398) being the least important variable. The variable importance is computed by summing the contribution of each variable (either as a primary or surrogate splitter) with a higher value corresponding to a higher contribution.

Looking closer at the decision tree, we also note that the least important goals were scored while already in the lead (GD ≥ 1) within the first two periods, while the most important goals were goals in the final five minutes of regulation while trailing by one goal.

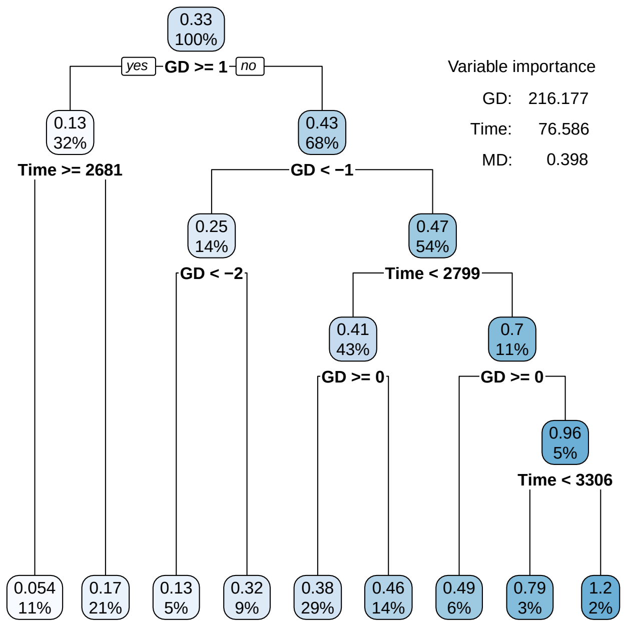

Fig. 1: Decision tree with GPIV as the outcome and GD, MD, and time as variables. Darker colors represent higher GPIV values, with the percentage describing the number of goals in each node.

These findings suggest that simplified metrics can be created by grouping some subsets of GD cases and time cases into a smaller set of categories, and that MD can be ignored as its importance is far smaller than the other factors.

First, for the GD dimension, we identified the following primary classes: reducing the deficit (GD ≤ -2), tying the game (GD $= -1$), taking the lead (GD $= 0$), and extending the lead (GD ≥ 1). For completeness, we also include the special case of an overtime (OT) winning goal, which exclusively occurs in overtime. Table 1 summarizes the average GPIV scores for each of these goal types (i.e., when considering GD only). Here, we note that the relatively big differences in the average GPIV of tying the game (0.594) and extending the lead (0.125), highlighting the value of such differentiating metric.

Second, for the time dimension, we selected to group goals according to the period a goal was scored but always considered them in combination with the above GD categories. The average GPIV values for each of these combined categories are shown in Table 2.

Based on the above categorizations, we can then define the two simplified GPIV metrics: *GD only* and *GD+Period*. For our notation, we let $\text{GPIV}^*_{\text{GD only}}$ be the approximated GPIV using GD only, and $\text{GPIV}^*_{\text{GD+Period}}$ be the approximated GPIV using both GD and period. In both cases, we simply assign each goal in a category the average GPIV value of all goals of that type. In the case

Table 1: Average GPIV per goal type when considering GD only.

Situation	Average GPIV	Goals
Reducing the deficit	0.249	887
Tying the game	0.594	1,263
Taking the lead	0.396	2,184
Extending the lead	0.125	2,102
OT winner	0.500	129

Table 2: Average GPIV per goal type when considering both GD and period. An OT winner has an average GPIV of 0.5.

	First period		**Second period**		**Third period**	
Situation	Average GPIV	Goals	Average GPIV	Goals	Average GPIV	Goals
Reducing the deficit	0.241	83	0.294	368	0.212	436
Tying the game	0.424	317	0.451	493	0.868	453
Taking the lead	0.370	1099	0.392	646	0.465	439
Extending the lead	0.180	397	0.162	789	0.070	916

that such GPIV values are based on prior games or seasons, these approximate GPIV values can therefore quickly be calculated at the time that a goal is scored.

When only considering GD, we note that tying the game is the most important goal, followed by OT winner, taking the lead, and reducing the deficit. Extending the lead is the least valuable goal. Similar conclusions can be drawn when also considering the period of the goal, although goals that tie the game or take the lead have an increased value as the game progresses. On the contrary, goals that reduce the deficit are most important in the second period while goals that extend the lead are most important in the first period, with the least important goals occurring in the third period for both situations. Another observation is that both simplifications lead to all goals having a positive value, which need not be the case in the full GPIV implementation [7, 8].

3 Stability of metrics

For previous estimations of the weights given to each goal to be useful, the metric should not change too drastically. Figures 2 and 3 visualize how the GPIV weights vary over time (on a season-per-season basis) for the approximate GPIV metrics based on GD only and GPIV based on GD+Period, respectively. When only considering GD, the GPIV weights exhibit low variability with stable weights over time. We also observe a strict order of the relative importance of the type of goals (matching the importance order from Table 1): goals that tie the game are the most important, followed by goals taking the lead, goals reducing the deficit, and goals extending the lead.

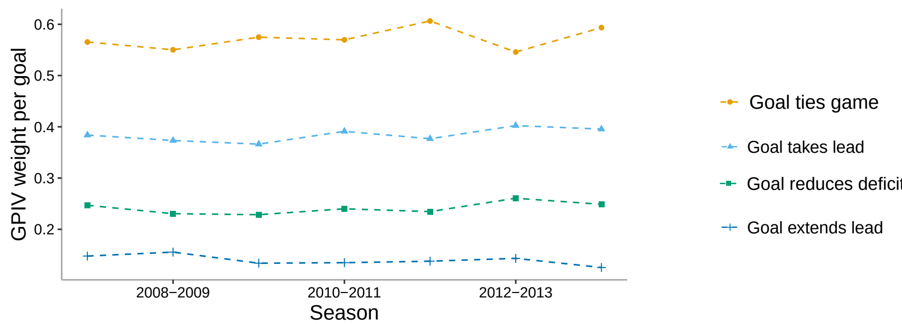

Fig. 2: GPIV weights by season when considering GD only.

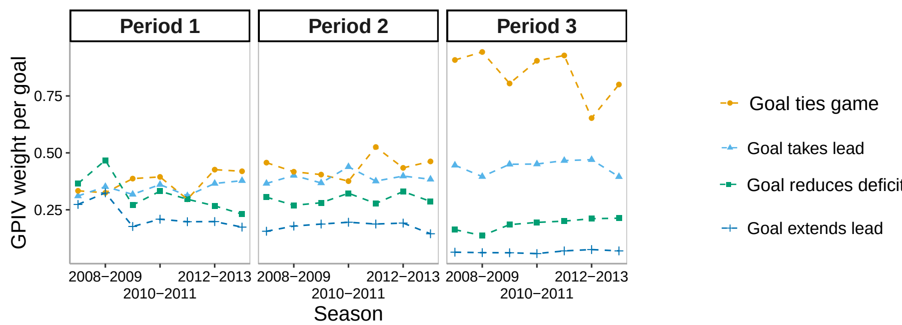

Fig. 3: GPIV weights by season when considering GD+Period.

If we also account for the period in which the goal occurs additional insights can be found. In general, regardless of the period, the least important goals are the goals that extend the lead, where the importance decreases with time. Although goals that reduce the deficit are the third most important goal in both the second and third periods as well as most seasons, they had the highest importance in the first period in both the 2007-2008 and 2008-2009 seasons. The importance of these goals also decreases in the third period. Another overall trend present is that goals taking the lead are the second most important goal type, with a mainly stable weight across all periods. The most important goals are found when tying the game in the third period, with the importance of a game-tying goal increasing as the game progresses.

4 Evaluation

Ideally, we would like the approximate metrics to capture most of the performance variations observed with GPIV. To determine if the metrics behave in a similar way as GPIV, we performed correlations comparisons and a rank-based analysis in which we compare with both GPIV and the corresponding traditional metrics. Some of these results are presented next.

Table 3: Spearman correlation between GPIV, simplified GPIV, and traditional metrics. Blank cells indicate a correlation between the same metric.

			Simplified GPIV metrics	
Class	Prior metric	Full GPIV	GD only	GD+Period
Goals	Traditional	0.971	0.979	0.975
Goals	Full GPIV		0.989	0.993
Assists	Traditional	0.979	0.986	0.983
Assists	Full GPIV		0.993	0.996
Points	Traditional	0.988	0.992	0.990
Points	Full GPIV		0.996	0.998
+/-	Traditional	0.775	0.843	0.802
+/-	Full GPIV		0.924	0.962

4.1 Correlation comparisons

As a point of comparison, we compute the Spearman rank correlation between the GPIV-based and the traditional metrics of goals, assists, points, and +/-. The correlations can be found in Table 3. Here, the class indicate what type of action the metric is calculated (i.e., the goals scored by a player, the assists made by the player, the sum of the first two, and whether a player was on the ice or not when a goal was scored during even strength). For each class we then present the correlation between the traditional metrics of that class and the three corresponding GPIV metrics (first row of each class), as well as between the full GPIV metric and the two approximations (second row of each class). As an example, we compute the correlations between Traditional Goals and Full GPIV by considering the total seasonal values for all players (Goals and GPIV-G).

First and most importantly, we note that the correlations between the full GPIV and the simplified GPIV metrics (second row for each class) are higher than the correlation with the traditional metrics (first rows). This suggests that the simplified metrics capture the most important variations of the full GPIV.

The table also highlights that GPIV has the lowest correlation of the considered methods, with both approximate methods having a higher correlation for all metrics. In particular, the simplified GPIV based on GD only was observed to have the highest correlation of all methods. This can be explained by GPIV considering a larger number of possible states, where some goals receive little to no value, which in turn lowers the correlation as the contrast between a goal of value one and close to zero is far larger than the approximate methods where the lowest value is 0.125 (simplified GPIV based on GD only) and 0.07 (simplified GPIV based on GD+Period).

Similarly, Figure 4 depicts the correlation for each pair of metrics across the analyzed seasons. We observe that goals, assists, and points all exhibit similar patterns over time, while +/- differs from the rest, particularly in the 2012-2013

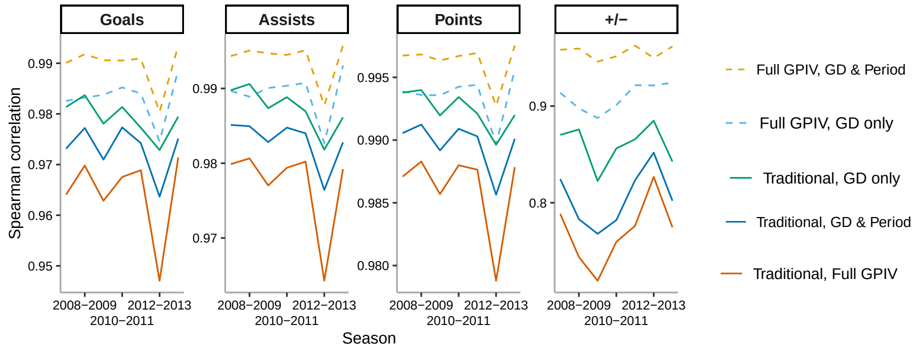

Fig. 4: Spearman correlation for each pair of metrics by season.

Table 4: Top-10 players for GPIV-P for the 2013-2014 season. Change is the difference in rankings for traditional and GPIV.

Rank							
P	GPIV-P	Change	Player	Position	P	GPIV-P	GPIV-P/P
1	1	0	Sidney Crosby	C	104	36.360	0.351
8-11	2	6	Alex Ovechkin	R	79	30.415	0.385
8-11	3	5	Nicklas Bäckström	C	79	29.199	0.370
19-22	4	15	Blake Wheeler	R	69	29.114	0.422
8-11	5	3	Joe Pavelski	C	79	27.995	0.354
4	6	-2	Tyler Seguin	C	84	27.614	0.329
3	7	-4	Claude Giroux	C	86	27.440	0.319
19-22	8	11	Kyle Okposo	R	69	26.951	0.391
16-18	9	7	Anze Kopitar	C	70	26.327	0.376
6-7	10	-4	Phil Kessel	R	80	26.225	0.328

season, which can be attributed to the lockout. Moreover, the correlations for +/- have a larger range, between 0.72 and 0.96, while goals, assists, and points all have values between 0.94 and 0.99. As desired, the strongest correlation was consistently observed between the full GPIV metric and its approximations. Among the five pairs, the correlation between the traditional and full GPIV metrics was the lowest, although the correlation is still high.

4.2 Player rankings

GPIV can also be used in the context of player valuation. In this section, we provide the top-ten rankings for GPIV points (P) for each of the three methods for comparison. The top-ten players according to the GPIV method can be found in Table 4, while Tables 5 and 6 contain the results for the simplified GPIV based on GD only and (simplified GPIV based on GD+Period), respectively. Overall, eight players are present in all three tables: Sidney Crosby, Alex Ovechkin, Nick-

Table 5: Top-10 players for simplified GPIV*-P based on GD only for the 2013-2014 season. Change is the difference in rankings for traditional and GPIV*.

Rank								
P	GPIV*-P	GPIV-P	Change	Player	Position	P	GPIV*-P	GPIV*-P/P
1	1	1	0	Sidney Crosby	C	104	36.520	0.351
8-11	2	5	6	Joe Pavelski	C	79	28.546	0.361
19-22	3	4	16	Blake Wheeler	R	69	28.514	0.413
8-11	4	2	4	Alex Ovechkin	R	79	27.742	0.351
2	5	11	-3	Ryan Getzlaf	C	87	27.406	0.315
4	6	6	-2	Tyler Seguin	C	84	27.300	0.325
8-11	7	3	1	Nicklas Bäckström	C	79	27.254	0.345
6-7	8	10	-2	Phil Kessel	R	80	27.185	0.340
3	9	7	-6	Claude Giroux	C	86	26.643	0.310
6-7	10	13	-4	Taylor Hall	L	80	26.399	0.330

Table 6: Top-10 players for simplified GPIV*-P based on GD+Period for the 2013-2014 season. Change is the difference in rankings for traditional and GPIV*.

Rank								
P	GPIV*-P	GPIV-P	Change	Player	Position	P	GPIV*-P	GPIV*-P/P
1	1	1	0	Sidney Crosby	C	104	36.485	0.351
8-11	2	2	6	Alex Ovechkin	R	79	28.877	0.366
8-11	3	5	5	Joe Pavelski	C	79	28.129	0.356
19-22	4	4	15	Blake Wheeler	R	69	28.072	0.407
8-11	5	3	3	Nicklas Bäckström	C	79	27.985	0.354
6-7	6	10	0	Phil Kessel	R	80	26.787	0.335
4	7	6	-3	Tyler Seguin	C	84	26.751	0.318
13	8	12	5	Joe Thornton	C	76	26.749	0.352
3	9	7	-6	Claude Giroux	C	86	26.687	0.310
2	10	11	-8	Ryan Getzlaf	C	87	26.264	0.302

las Bäckström, Blake Wheeler, Joe Pavelski, Tyler Seguin, Claude Giroux, and Phil Kessel. Although the rankings of players differ between the tables, Sidney Crosby remains atop all three tables with similar GPIV values. A possible explanation is the large difference in points between him and the lower-ranked players. If we consider the simplified GPIV based on GD only, simplified GPIV based on GD+Period, and GPIV in order of complexity, we can also see that both Alex Ovechkin and Nicklas Bäckström gain ranks with increased GPIV complexity. Their ascent is likely a result of the number of goals scored in close games, losing by one or tied, at the end of the game or in overtime with Ovechkin as the likely goalscorer and Bäckström with the assist. Similarly, Blake Wheeler also gains ranks in all cases while also having the highest average importance per point. On the contrary, Ryan Getzlaf, who ranked second in total points, loses ranks as he is ranked fifth for the simplified GPIV based on GD only, tenth for the GD

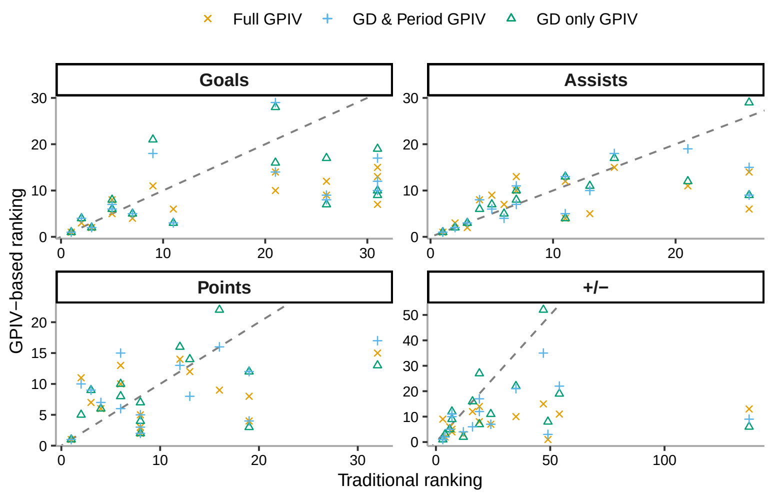

Fig. 5: Top-15 ranked players per full GPIV metric. Note that some players share the same rank on the x-axis.

and Period simplification, and eleventh for the full GPIV. These observations are consistent with the correlation results.

To investigate the impact that each GPIV-based method has on play ranking, Figure 5 shows ranking using each GPIV-based metric (y-axis) vs. the traditional ranking (x-axis) for the top-15 ranked players using the full GPIV metric. For both GPIV-Goals and GPIV-Assists it appears that the top-ranked players are mainly unaffected by the different methods as their ranks remain stable. The variability in ranking becomes more prevalent for somewhat lower traditional rankings, as the highest climbing players may have higher variability in ranking across the GPIV-based methods. As an example, Anze Kopitar was ranked 28th and 29th when considering the GPIV-based approximations based on GD only and GD+Period, respectively, but for the full GPIV he was ranked 14th. For GPIV-Points some players also manage to maintain a stable rank, for instance, the top-ranked player Sidney Crosby. The GPIV-based ranking for points also had the lowest range, between 1 and 22. On the contrary, +/- had the highest range, between 1 and 52, while also having the largest spread for traditional ranking with Matt Duchene climbing from 137th in the traditional +/- to top 15 in GPIV-+/-. The variability of rankings for the GPIV-based +/- metrics is also the highest, as some players have a larger discrepancy between their rankings of the different methods. For instance, Sidney Crosby had rankings of 15th (full GPIV), 35th (GPIV based on GD+Period), and 52nd (GD only).

Another way to illustrate the difference in ranking between the traditional ranking and the GPIV ranking can be found in Figure 6, where the top-15 players, for each full GPIV metric, are visualized. Here we note that the largest rank increases are found for +/-, with Matt Duchene gaining over 100 ranks

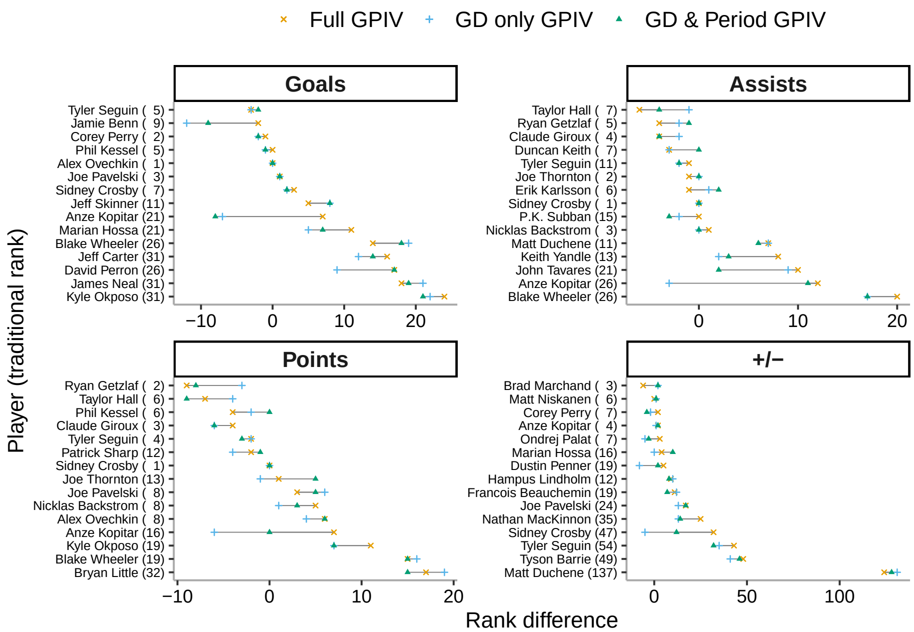

Fig. 6: Rank difference between the traditional and GPIV-based ranking for the top-15 per full GPIV metric.

while e.g., Tyler Seguin and Tyson Barrie also gain approximately 50 ranks. Another thing to note is that for GPIV-+/-, few players lose ranks compared to the traditional +/-. As for goals, assists, and points, players losing ranks is more prevalent. For GPIV-Goals we also observe that the different GPIV-based methods result in mostly similar rankings, although Jamie Benn and Anze Kopitar have a larger difference between the full GPIV and the approximations. In addition, for the simplified GPIV-based methods, Anze Kopitar loses ranks for GPIV-Goals while he gains ranks for the full GPIV. The same results can also be seen for Anze Kopitar in GPIV-Assists and GPIV-Points. In general, GPIV-Assists and GPIV-Points also have a higher difference in rankings when comparing the GPIV methods than goals. We also note some of the big climbers, e.g. Bryan Little (Points) and Blake Wheeler (Assists and Points), while players who lose ranks, e.g., Ryan Getzlaf (Assists and Points) and Taylor Hall (Assists and Points), also have variability between the different GPIV-based methods, with the full GPIV tending to assign them the lowest ranking.

5 Related work

The most used performance metrics in ice hockey are the total number of goals, assist, and points accumulated over a season or some other time period or set of games. Like these metrics, the GPIV metric and the GPIV-based approximations presented here are calculated as the sum of all goals. The main difference is in the weight given to each goal and that these traditional metrics do not account for the potential impact a goal may have on the game outcome.

Some extensions to these traditional metrics have been proposed (e.g., for the +/- metric [11, 4]) and combined metrics have been proposed (e.g., based on principal component analysis [5]). Others have proposed player performance metrics take game context into account (e.g., the probability that an event leads to a goal in the subsequent 20 seconds [15]) or that incorporate the game models using Markov games where two opposing sides (i.e., the home team and the away team) try to reach states in which they are rewarded (e.g., scoring a goal) [18, 6, 13, 16, 17, 9, 14, 10]. 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. In this work, we aim to present such simpler and more practical metrics that still differentiate between the potential impact that a goal may have on the game outcome.

Prior works have also considered the importance of scoring the first goal [1], a two-goal lead [2], and late-game reversals [3]. For example, it was found that teams that take a two-goal lead win in 83% of games, while having the lead after two periods leads to a win in 84% and 80% of games for the home and away team, respectively.

Except our prior work defining the original GPIV metric, the only other work that considers the importance of goals is the added goal value (AGV) metric presented by Pettigrew [12]. The importance of a goal is based on GD and time and is defined using win probabilities for that context and neighboring contexts (with GD one higher and one lower). The AGV is then defined for a player by comparing the importance of the player's goals to the importance of all other players' goals.

Some players can have a positive (or negative) impact even when they are not the player scoring the goal or assisting to the goal. Perhaps the most used metric to estimate the value a player brings to team performance (during 5-on-5 play) is the +/- metric. While the metric has been criticized due to its disregard of contextual information [18], alternative approaches typically also ignore the importance of individual goals. Interesting examples falling in this category include works based on hazards models [18], regularized logistic regression for predicting player impact on scoring [4], or models that sum over all actions performed by a player [13] or set of players [10] when on the ice at the same time.

6 Conclusions

This paper has presented two approximate GPIV metrics: GD only and GD+Period. The design of the metrics was motivated by our analysis of the relative importance of different dimensions of the state space, and our evaluation demonstrated that the metrics are relatively stable and capture most of the relative differences between GPIV and traditional metrics (e.g., goals, assist, points, and +/-). The presented metrics are practical, intuitive, capture most of the desirable variations that GPIV captures, and show that the value of a goal can be well-estimated using GPIV data based on historic data. These properties should make it desirable for fans, teams, and media that want an easy-to-apply metric for evaluating and comparing players that account for goal importance.

References

1. Brimberg, J., Hurley, W.: A note on the importance of the first goal in a National Hockey League game. International Journal of Operational Research **6**(2), 282–287 (2009)
2. Brimberg, J., Hurley, W.: Do professional hockey teams with a two-goal lead lose more often than they should? International Journal of Operational Research **15**(2), 226–233 (2012)
3. Gill, P.S.: Late-game reversals in professional basketball, football, and hockey. The American Statistician **54**(2), 94–99 (2000)
4. Gramacy, R.B., Jensen, S.T., Taddy, M.: Estimating player contribution in hockey with regularized logistic regression. Journal of Quantitative Analysis in Sports **9**(1), 97–111 (2013)
5. Gu, W., Foster, K., Shang, J., Wei, L.: A game-predicting expert system using big data and machine learning. Expert Systems with Applications **130**, 293–305 (2019)
6. Kaplan, E.H., Mongeon, K., Ryan, J.T.: A Markov Model for Hockey: Manpower Differential and Win Probability Added. INFOR Information Systems and Operational Research **52**(2), 39–50 (2014)
7. Lambrix, P., Carlsson, N.: Performance metrics for ice hockey accounting for goal importance. In: Linköping Hockey Analytics Conference. pp. 11–25 (2022)
8. Lambrix, P., Carlsson, N., Säfvenberg, R.: Goal-based performance metrics for ice hockey accounting for goal importance (2023), submitted
9. Liu, G., Schulte, O.: Deep reinforcement learning in ice hockey for context-aware player evaluation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. pp. 3442–3448 (2018)
10. Ljung, D., Carlsson, N., Lambrix, P.: Player pairs valuation in ice hockey. In: Machine Learning and Data Mining for Sports Analytics. MLSA 2018. pp. 82–92 (2019)
11. Macdonald, B.: A Regression-Based Adjusted Plus-Minus Statistic for NHL Players. Journal of Quantitative Analysis in Sports **7**(3) (2011)
12. Pettigrew, S.: Assessing the offensive productivity of NHL players using in-game win probabilities. In: MIT Sloan Sports Analytics Conference (2015)
13. Routley, K., Schulte, O.: A Markov Game Model for Valuing Player Actions in Ice Hockey. In: Uncertainty in Artificial Intelligence. pp. 782–791 (2015)
14. Sans Fuentes, C., Carlsson, N., Lambrix, P.: Player impact measures for scoring in ice hockey. In: MathSport International 2019 Conference. pp. 307–317 (2019)
15. Schuckers, M., Curro, J.: 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)
16. Schulte, O., Khademi, M., Gholami, S., Zhao, Z., Javan, M., Desaulniers, P.: A Markov Game model for valuing actions, locations, and team performance in ice hockey. Data Mining and Knowledge Discovery **31**(6), 1735–1757 (2017)

17. Schulte, O., Zhao, Z., Javan, M., Desaulniers, P.: Apples-to-apples: Clustering and Ranking NHL Players Using Location Information and Scoring Impact. In: MIT Sloan Sports Analytics Conference (2017)
18. Thomas, A., Ventura, S.L., Jensen, S., Ma, S.: Competing Process Hazard Function Models for Player Ratings in Ice Hockey. The Annals of Applied Statistics **7**(3), 1497–1524 (2013)