Player performance in ice hockey

Dennis Ljung, Carles Sans Fuentes,
Niklas Carlsson, Patrick Lambrix

Linköping University, Sweden

Parts of the content of this presentation is published as:

Ljung D, Carlsson N, Lambrix P, Player pairs valuation in ice hockey, *5th Workshop on Machine Learning and Data Mining for Sports Analytics*, CEUR Workshop Proceedings Volume 2284, 14-24, Dublin, Ireland, 2018.

Sans Fuentes C, Carlsson N, Lambrix P, Player impact measures for scoring in ice hockey, *MathSport International 2019 Conference,* 307-317, Athens, Greece, 2019.

Outline

- Motivation
- Method
- Results
- Conclusion

Outline

- Motivation
- Method
- Results
- Conclusion

Motivation

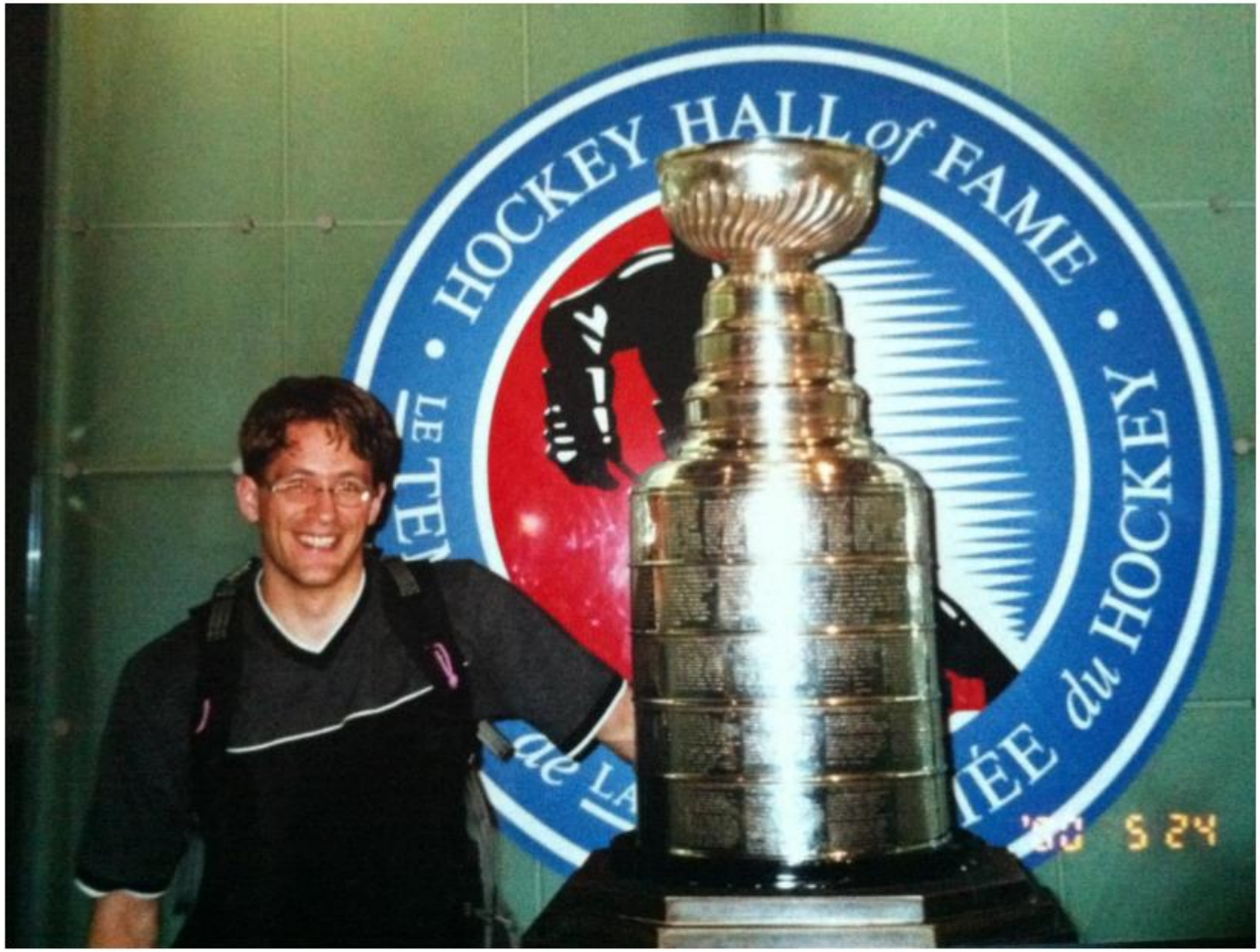
Niklas
(and many others)
dream of:

Motivation

First try

Motivation

A bit easier …



Motivation

Not completely given up first method …

Motivation

- 1. Performance of players
  - → need performance metrics

- 2. Ice hockey is a team sport
  - → important to identify players that play particularly well together (or not).

Motivation

- On ice: usually two defenders, three forwards, and a goaltender
- Performance metrics for individuals and pairs:
  - defender pairs are natural
  - more data on forward pairs than triplets
  - mixed pairs not studied

Performance metrics - traditional

RANK	SPELARE	NR	LAG	POS	GP	G	A	TP	PIM	GWG	PPG	SOG	HITS	BKS	+	-	+/-	TOI/GP
1	Ryan Lasch	81	Frölunda	F	37	10	25	35	16	3	2	66	9	2	23	31	-8	19:02
2	Joakim Lindström	10	Skellefteå	F	38	13	21	34	14	3	4	132	4	8	29	25	4	17:15
3	Derek Roy	9	Linköping	F	39	5	29	34	22	0	2	74	4	20	29	20	9	17:15

- *Offensive:* G: goals, A: assists, TP: points, GWG: game winning goals, PPG: powerplay goals, SOG: Shots on goal
- *Defensive:* HITS: hits, BKS: blocked shots
- +/-: plus-minus
- PIM: penalty minutes
- *Time:* GP: games played, TOI: time on ice

Performance metrics - advanced

- **Corsi:** (Shots on goal FOR + missed shots FOR + *blocked shots FOR*) – (Shots on goal AGAINST + missed shots AGAINST + *blocked shots AGAINST*)
  - "+/- for shots"
  - Better predictor for future goal differential than past goal differential
  - Critique: what if shots are not of good quality?

- **Fenwick** (usat) unblocked shots: (Shots on goal FOR + missed shots FOR) – (Shots on goal AGAINST + missed shots AGAINST)
  - Interesting to look at when blocking is an intentional strategy

Performance metrics - advanced

- **xG** (Expected Goals): assigns a value to each shot, based on the likelihood of the shot resulting in a goal.
  - “not all shots are equal”
  - Uses unblocked shots
  - Incorporates shot location, rebounds, rushes
  - Prediction for future goals similar to Corsi
- **Corsi/Fenwick/xG For percentage**
  - X For / (X For + X against)
  - > 50 % is good outcome

Performance metrics - advanced

- **Corsi/Fenwick/xG Relative to teammates**
  - Compare value for team when player is on the ice with value when player is not on the ice

- **Score adjusted metrics**
  - Uses league-average shots in different situations (leading/trailing with x, or tied)
  - Favors teams in the lead (raises value)
  - Better predictability

Performance metrics - advanced

Critique on advanced metrics: context

Some new approaches:

- Schulte group (this presentation)
- Schuckers group: THOR (Total Hockey Rating)
- Thomas et al.: hazard function models
- Gramacy et al: uses regularized logic regression

Outline

- Motivation
- Method
- Results
- Conclusion

Action Impact Model

- Based on the work by Routley and Schulte 2015*
- Idea:
  - Define state $s = < c, ps >$

    where c is a context and ps is a play sequence
  - Actions are performed in states
  - Define impact of action in a state
  - Define player impact based on action impacts

*Schulte's group presented a more extended model at IJCAI 2018.

Action Impact Model

Context

Notation	Name	Range
GD	Goal Differential	[-8,8]
MD	Manpower Differential	[-3,3]
P	Period	[1,7]

Action Impact Model

Events

Action Event	Start/End Event
Faceoff	Period Start
Shot	Period End
Missed Shot	Early Intermission Start
Blocked Shot	Penalty
Takeaway	Stoppage
Giveaway	Shootout Completed
Hit	Game End
Goal	Game Off
	Early Intermission End

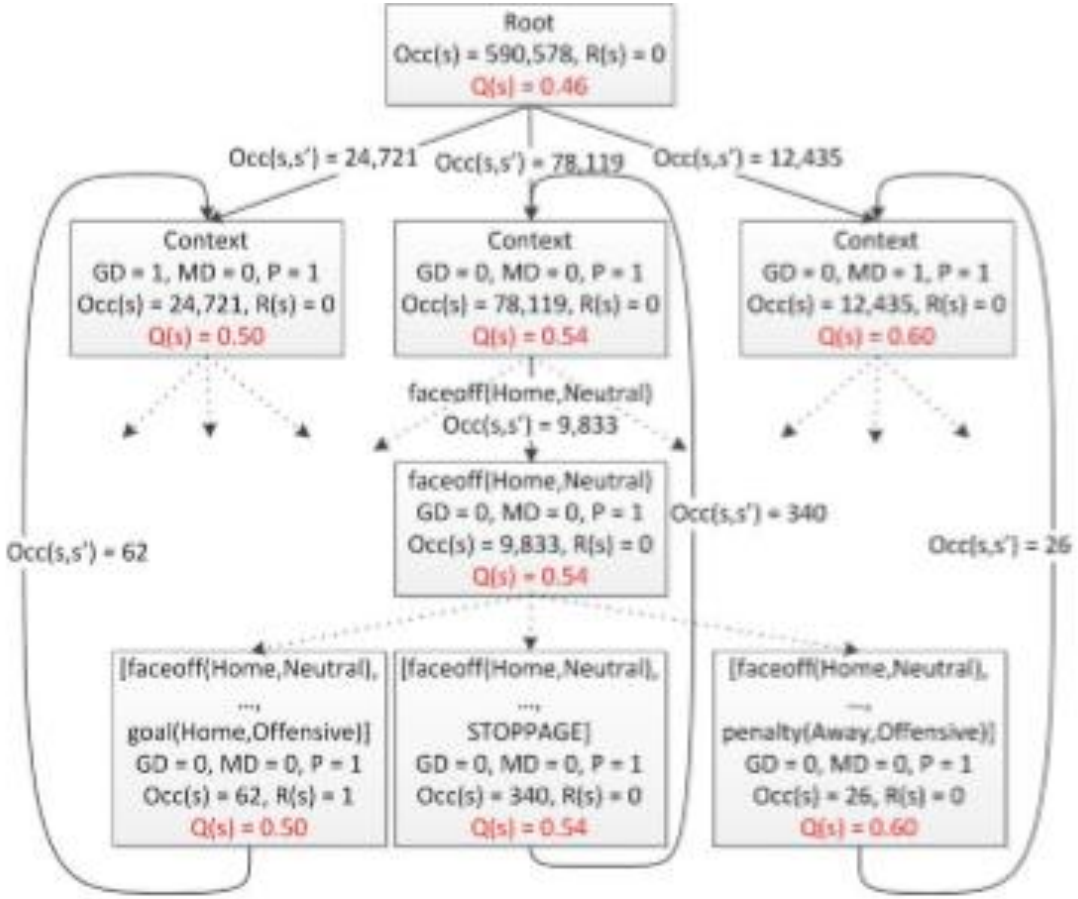
Action Impact Model

A play sequence is defined as

- the empty sequence or
- a sequence of events
  - first event: start marker
  - (possible) next events: action events
  - (possible) last event: end event

(→complete sequence)

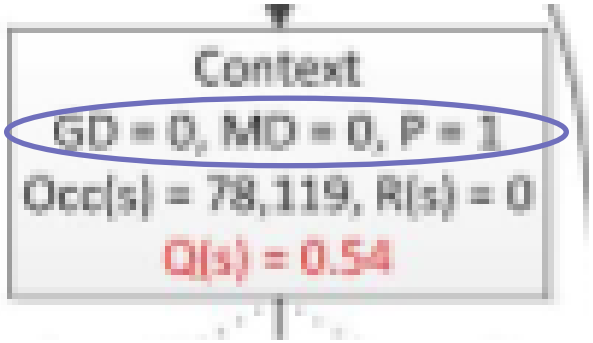
Action Impact Model

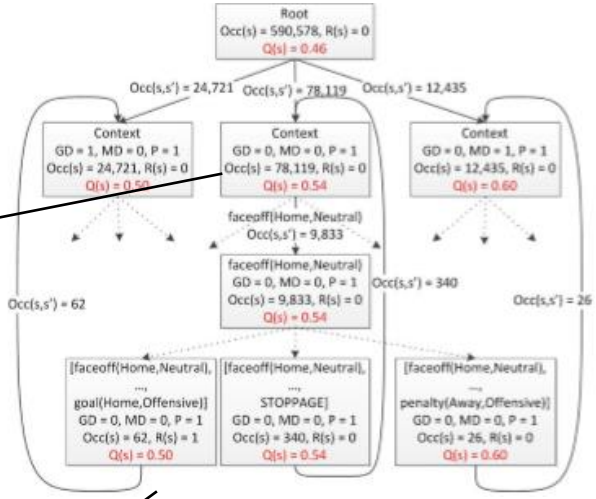


Routley and Schulte, 2015

Action Impact Model

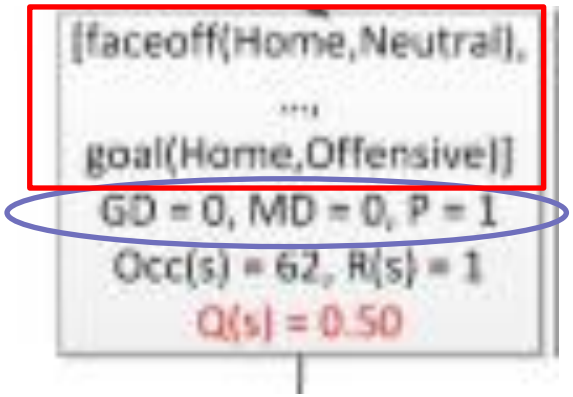
State $s = < c, ps >$





Context

Play sequence



Routley and Schulte, 2015

Action Impact Model

- Actions are performed in states

$< c, ps > * a =$

$< c, append\ (ps,a) >$ if state has no end event

(add action to play sequence, e.g., shot)

$< c', \ empty\text{-}set >$ if state has end event

(change context, e.g., after a goal)

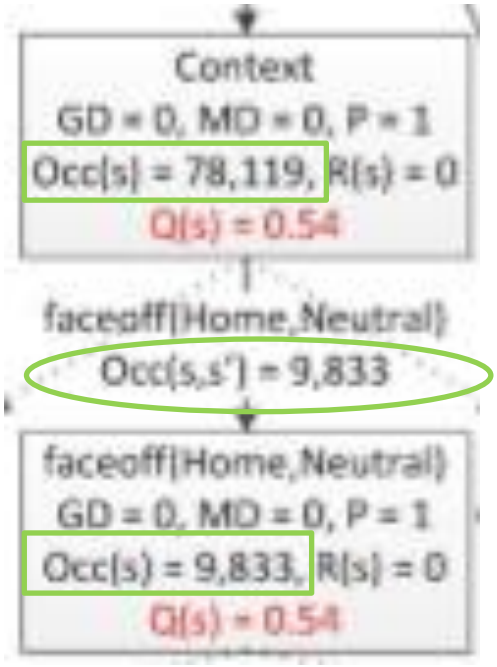
Action Impact Model

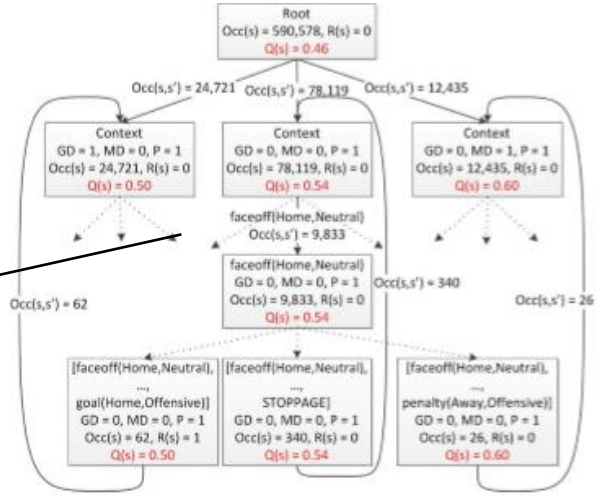
Based on play-by-play data:

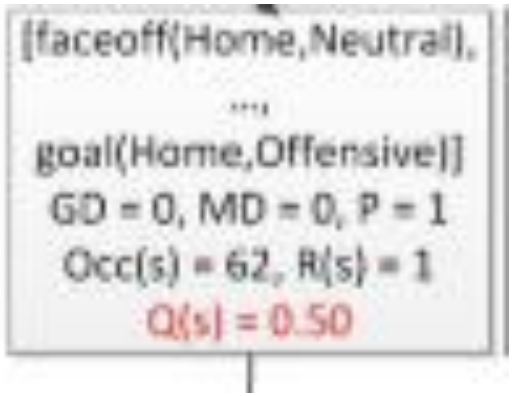
- Occurrences of state *s*: *Occ(s)*
- Occurrences of state *s* immediately followed by state *s'*: *Occ(s,s')*

- Transition probability $T(s,s') = Occ(s,s') / Occ(s)$

Action Impact Model







Occurrences

Occurrences

Routley and Schulte, 2015

Action Impact Model

Value iteration algorithm → Q-values

Reward function: goal states receive reward 1

(In single player experiments
also goal against reward -1)

- Impact of action *a* in state *s*: $Q_T(s * a) - Q_T(s)$

Action Impact Model

Algorithm 1 Dynamic Programming for Value Iteration

Require: Markov Game model, convergence criterion c, maximum number of iterations M

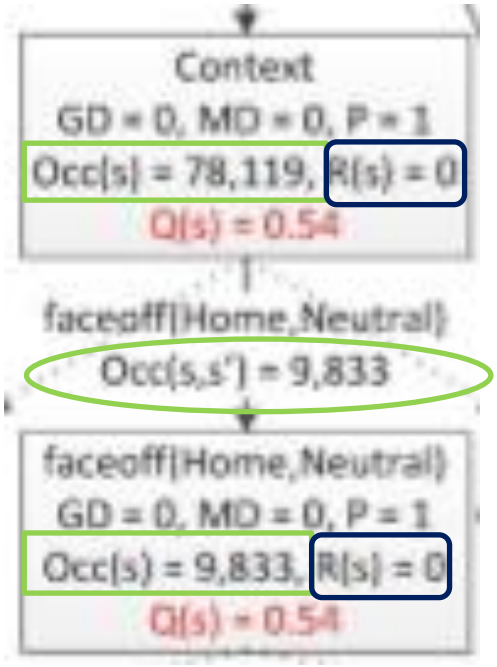
$lastValue = 0$
$currentValue = 0$
$converged = false$
for $i = 1; i \leq M; i \leftarrow i + 1$ **do**
  for all states s in the Markov Game model **do**
    if $converged == false$ **then**
      $Q_{i+1}(s) = R(s) + \frac{1}{Occ(s)} \sum_{(s,s') \in E} (Occ(s, s') \times Q_i(s'))$
      $currentValue = currentValue + |Q_{i+1}(s)|$
    end if
  end for
  if $converged == false$ **then**
    if $\frac{currentValue - lastValue}{currentValue} < c$ **then**
      $converged = true$
    end if
  end if
  $lastValue = currentValue$
  $currentValue = 0$
end for

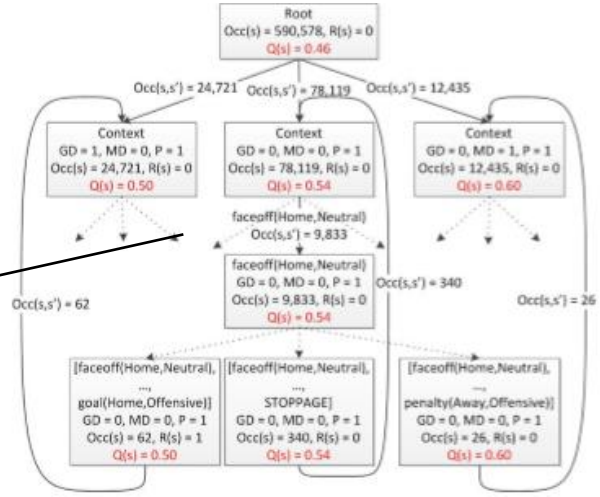
Action Impact Model

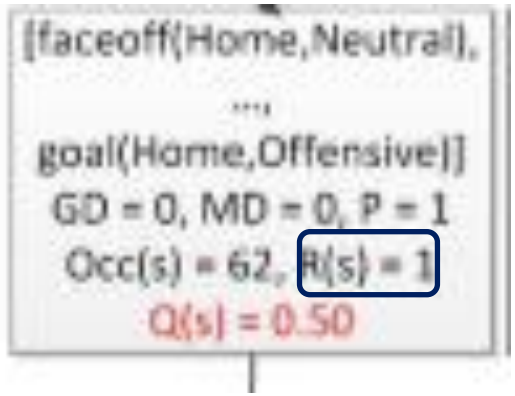
7: $$Q_{i+1}(s) = R(s) + \frac{1}{Occ(s)} \sum_{(s,s') \in E} (Occ(s,s') \times Q_i(s'))$$

Compute separate Q-values for Home and Away teams

Action Impact Model







Occurrences

Occurrences

Reward

Q-value

Routley and Schulte, 2015

Player Impact

Sum of action impacts

1. Based on all actions performed by the player (direct impact)
2. Based on actions when the player is on the ice ((collective) impact)

Variants normalized by time

Player Pair Impact

Sum of action impacts when both players are on the ice

Variants normalized by time

Definitions of impact

Table 1. Basic action sets.

A is the set of all state-action-pairs $< s, a >$ where action a is performed in state s
$\mathrm{A}_i(\mathrm{p}_k)$ is the set of state-action-pairs when player p_k is on the ice $\mathrm{A}_i(\mathrm{p}_k,\mathrm{p}_l)$ is the set of state-action-pairs when players p_k and p_l both are on the ice $\mathrm{A}_i(\mathrm{p}_k,\mathrm{p}_l) = \mathrm{A}_i(\mathrm{p}_k) \cap \mathrm{A}_i(\mathrm{p}_l)$
$\mathrm{A}_p(\mathrm{p}_k)$ is the set of state-action-pairs where the action is performed by player p_k $\mathrm{A}_p(\mathrm{p}_k) \subseteq \mathrm{A}_i(\mathrm{p}_k)$

Definitions of impact

Table 2. Player and player pair impact.

The direct impact of a player is the sum of the impact values of the actions performed by the player: D-impact(p_k) = $\Sigma_{<s,a> \in A_p(p_k)}$ impact(s,a)
The impact of a player is the sum of the impact values of the actions when the player is on the ice: impact(p_k) = $\Sigma_{<s,a> \in A_i(p_k)}$ impact(s,a)
The impact of a player pair is the sum of the impact values of the actions when both players are on the ice: impact(p_k,p_l) = $\Sigma_{<s,a> \in A_i(p_k,p_l)}$ impact(s,a)
The impact of a player without a second player is the sum of the impact values of the actions when the player is on the ice and the second player is not on the ice: impact-without(p_k,p_l) = $\Sigma_{<s,a> \in (A_i(p_k) \setminus A_i(p_k,p_l))}$ impact(s,a)

”Impact” called ”collective impact” in some experiments

Outline

- Motivation
- Method
- Results
- Conclusion

Experiments

Data:

NHL play-by-play data from the 2007-2008 through 2013-2014 NHL season*

* As provided by Routley and Schulte

Top players 2007-2008 and 2008-2009 for direct impact

PlayerName	Position	Age	Salary	GP	G	GA	PlusMin	Points	**Direct**	Directh	Collective	Collectiveh
2007												
Alex Ovechkin	F	22	3.83	82	65	47	28	112	**71.96**	182.65	232.56	588.85
Dion Phaneuf	D	22	0.94	82	17	43	12	60	**59.22**	134.05	246.12	559.67
Rick Nash	F	23	5.50	80	38	31	3	69	**59.01**	181.80	158.82	485.99
Jarome Iginla	F	30	7.00	82	50	48	27	98	**58.94**	161.92	204.12	560.88
Dustin Brown	F	23	1.18	78	33	27	-13	60	**53.78**	156.41	171.40	501.48
Brenden Morrow	F	28	4.10	82	32	42	23	74	**51.15**	146.62	171.59	504.57
Zdeno Chara	D	30	7.50	77	17	34	14	51	**50.74**	117.69	203.78	468.89
Trent Hunter	F	27	1.55	82	12	29	-17	41	**50.31**	167.65	153.36	508.27
Mike Green	D	22	0.85	82	18	38	6	56	**48.26**	122.63	219.72	545.08
Pavel Datsyuk	F	29	6.70	82	31	66	41	97	**48.22**	134.68	198.44	559.41
2008												
Alex Ovechkin	F	23	9.00	79	56	54	8	110	**75.93**	194.34	239.89	612.23
Dustin Brown	F	24	2.60	80	24	29	-15	53	**59.76**	177.60	178.34	540.84
Shea Weber	D	23	4.50	81	23	30	1	53	**53.14**	136.10	201.19	511.36
Evgeni Malkin	F	22	3.83	82	35	78	17	113	**50.76**	134.92	220.41	591.75
Dion Phaneuf	D	23	7.00	79	11	36	-11	47	**50.34**	122.64	240.57	532.49
Vincent Lecavalier	F	28	7.17	77	29	38	-9	67	**49.46**	143.99	188.17	549.37
Sheldon Souray	D	32	6.25	81	23	30	1	53	**49.38**	125.86	203.08	514.73
Jeff Carter	F	24	4.50	82	46	38	23	84	**48.88**	141.78	189.35	548.30
Rick Nash	F	24	6.50	78	40	39	11	79	**48.88**	145.11	171.59	498.26
Martin St. Louis	F	33	5.00	82	30	50	4	80	**47.82**	135.55	204.19	569.06

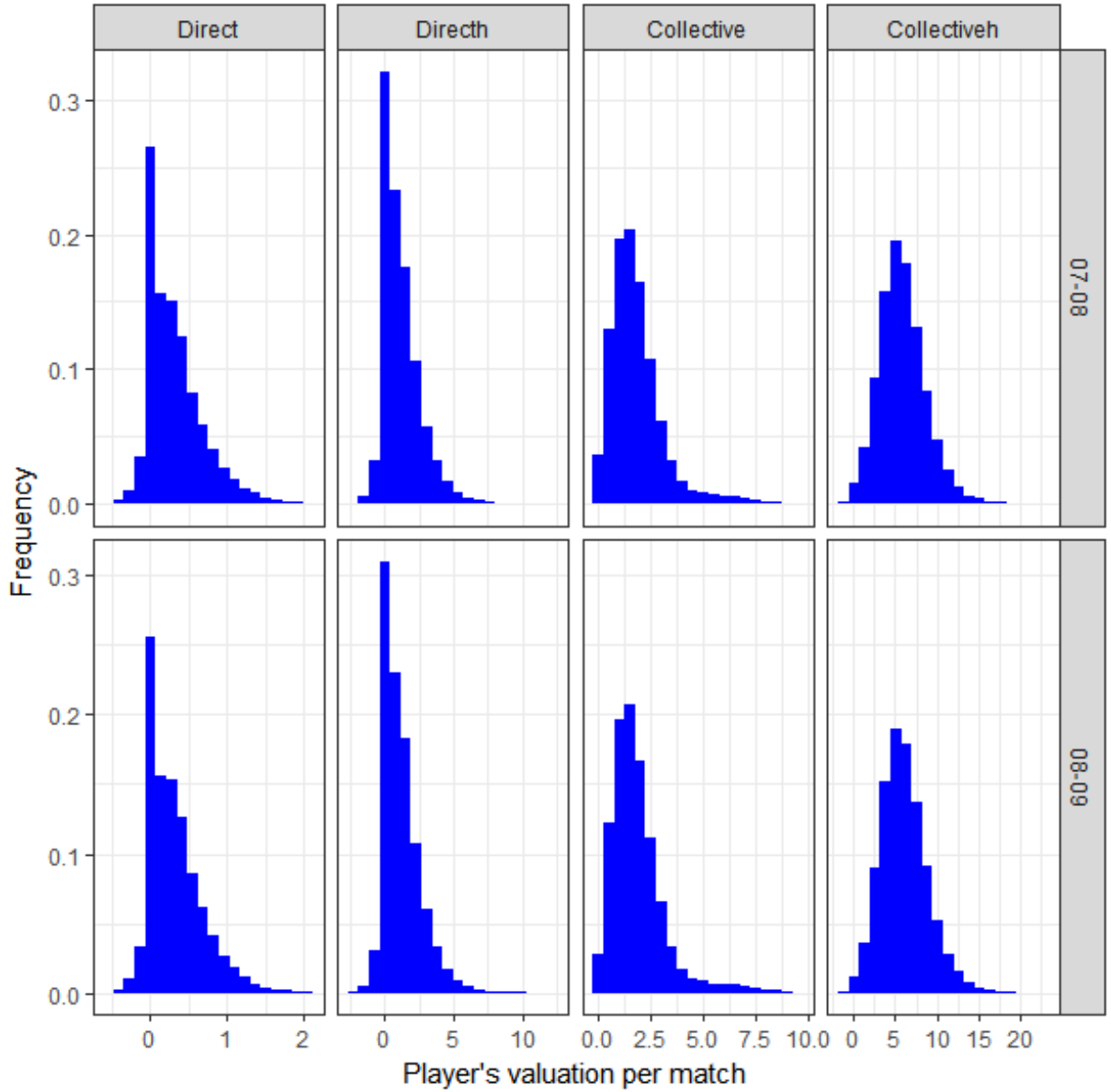
Table 5.1: Top 10 Players performance for 2007-2008 and 2008-2009 for the Direct metric.

Top players 2007-2008 and 2008-2009 for impact

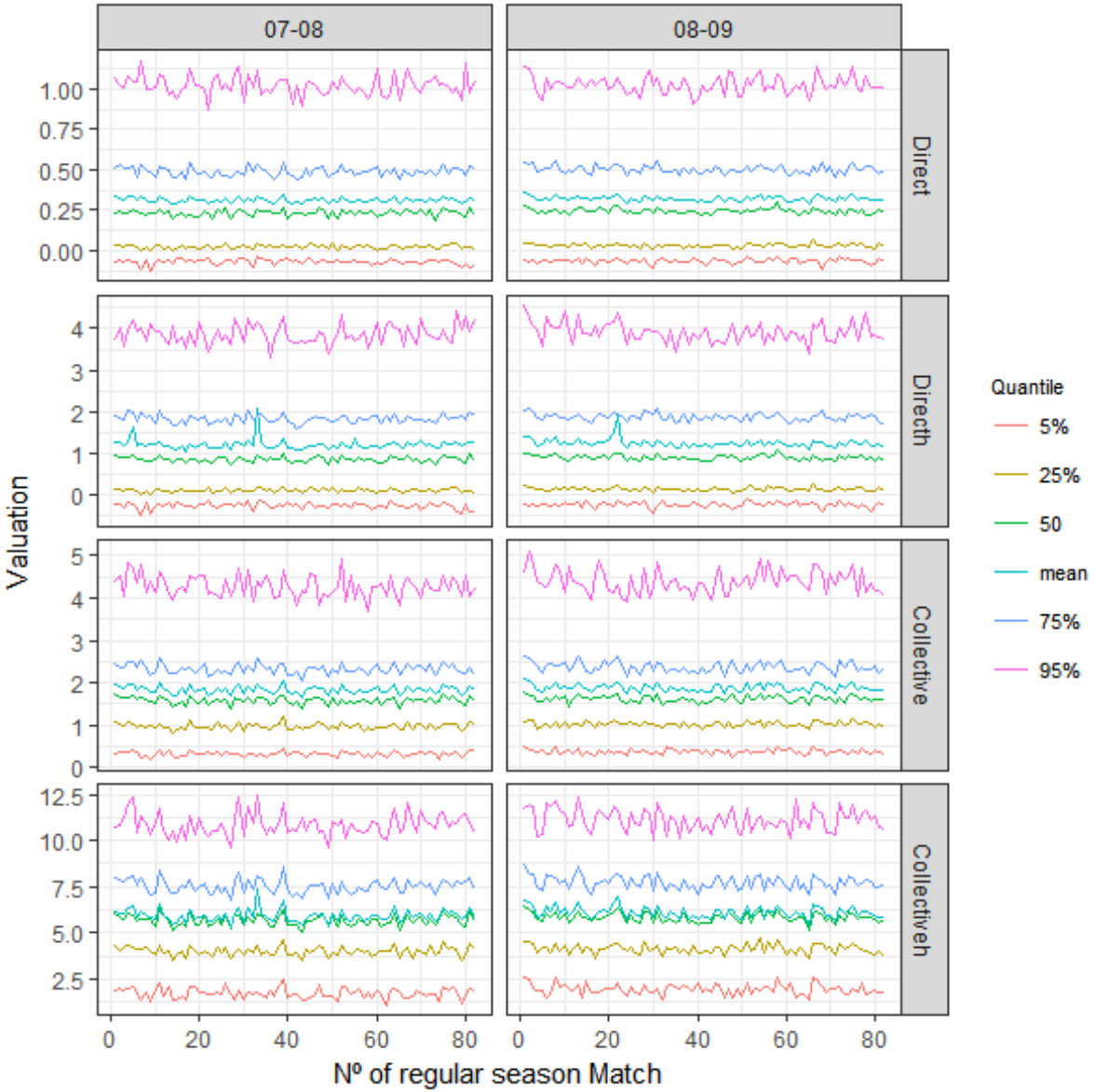
PlayerName	Position	Age	Salary	GP	G	GA	PlusMin	Points	Direct	Directh	**Collective**	Collectiveh
2007												
Dion Phaneuf	D	22	0.94	82	17	43	12	60	59.22	134.05	**246.12**	559.67
Alex Ovechkin	F	22	3.83	82	65	47	28	112	71.96	182.65	**232.56**	588.85
Tomas Kaberle	D	29	4.25	82	8	45	-8	53	38.32	93.36	**221.93**	551.72
Mike Green	D	22	0.85	82	18	38	6	56	48.26	122.63	**219.72**	545.08
Andrei Markov	D	29	5.75	82	16	42	1	58	42.37	105.18	**213.81**	530.37
Nicklas Lidstrom	D	37	7.60	76	10	60	40	70	29.04	66.41	**205.68**	480.18
Jarome Iginla	F	30	7.00	82	50	48	27	98	58.94	161.92	**204.12**	560.88
Zdeno Chara	D	30	7.50	77	17	34	14	51	50.74	117.69	**203.78**	468.89
Lubomir Visnovsky	D	31	2.05	82	8	33	-18	41	32.64	83.52	**201.34**	523.00
Roman Hamrlik	D	33	5.50	77	5	21	7	26	37.79	93.89	**201.29**	509.39
2008												
Dion Phaneuf	D	23	7.00	79	11	36	-11	47	50.34	122.64	**240.57**	532.49
Alex Ovechkin	F	23	9.00	79	56	54	8	110	75.93	194.34	**239.89**	612.23
Evgeni Malkin	F	22	3.83	82	35	78	17	113	50.76	134.92	**220.41**	591.75
Dan Boyle	D	32	6.67	77	16	41	6	57	36.11	88.65	**219.94**	539.81
Chris Pronger	D	34	6.25	82	11	37	0	48	43.40	99.89	**217.92**	503.72
Mike Green	D	23	6.00	68	31	42	24	73	46.41	106.62	**214.33**	493.09
Nicklas Backstrom	F	21	2.40	82	22	66	16	88	37.12	111.83	**214.19**	630.43
Braydon Coburn	D	23	1.20	80	7	21	7	28	40.78	100.10	**211.64**	516.12
Andrei Markov	D	30	5.75	78	12	52	-2	64	38.03	96.17	**209.18**	527.62
Mark Streit	D	31	4.10	74	16	40	6	56	39.38	97.60	**206.59**	504.31

Table 5.4: Top 10 players performance for 2007-2008 and 2008-2009 for the Collective metric without goalkeeper positions

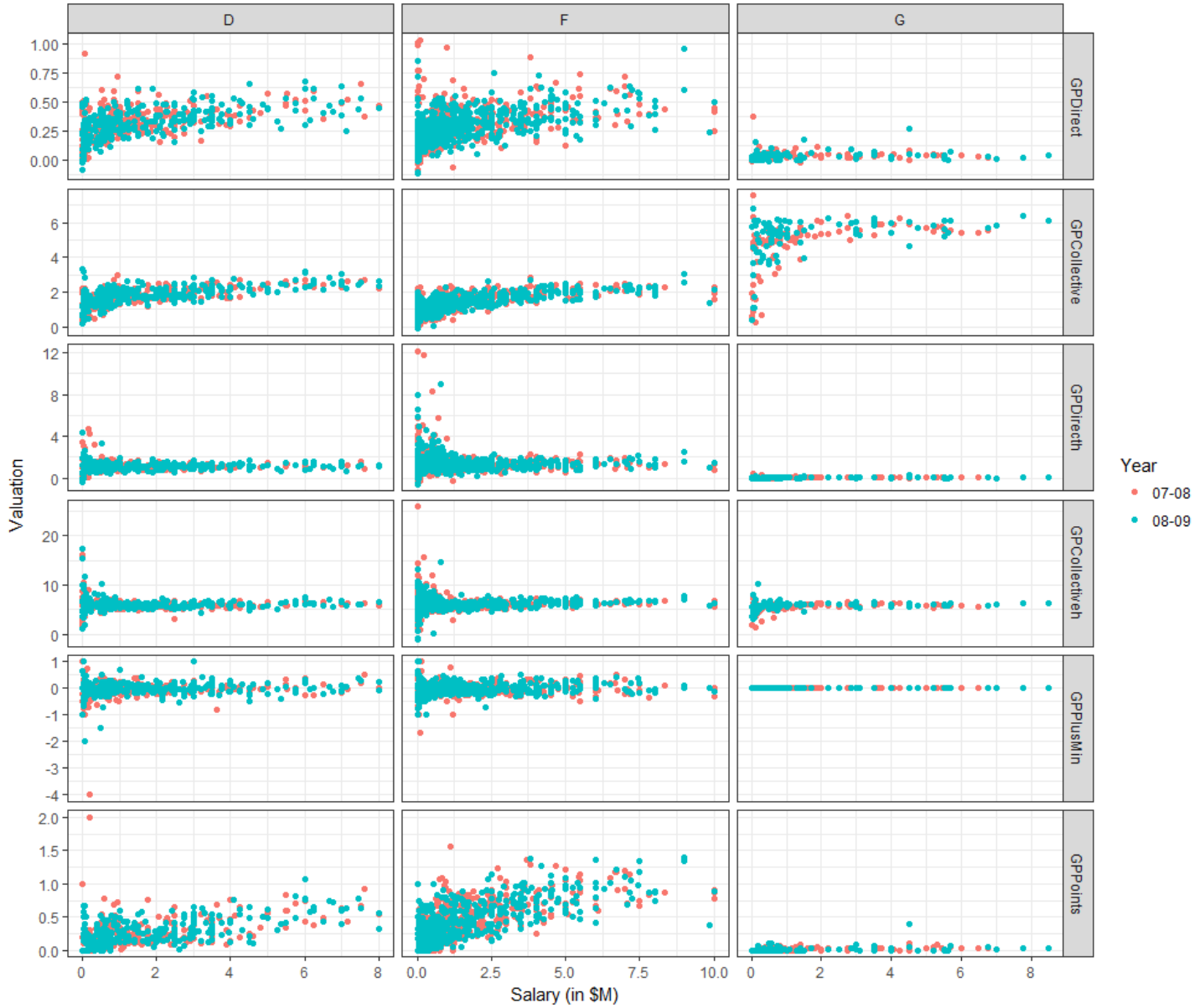
Distribution of impact values



Quantiles per game



Impact vs salary per position

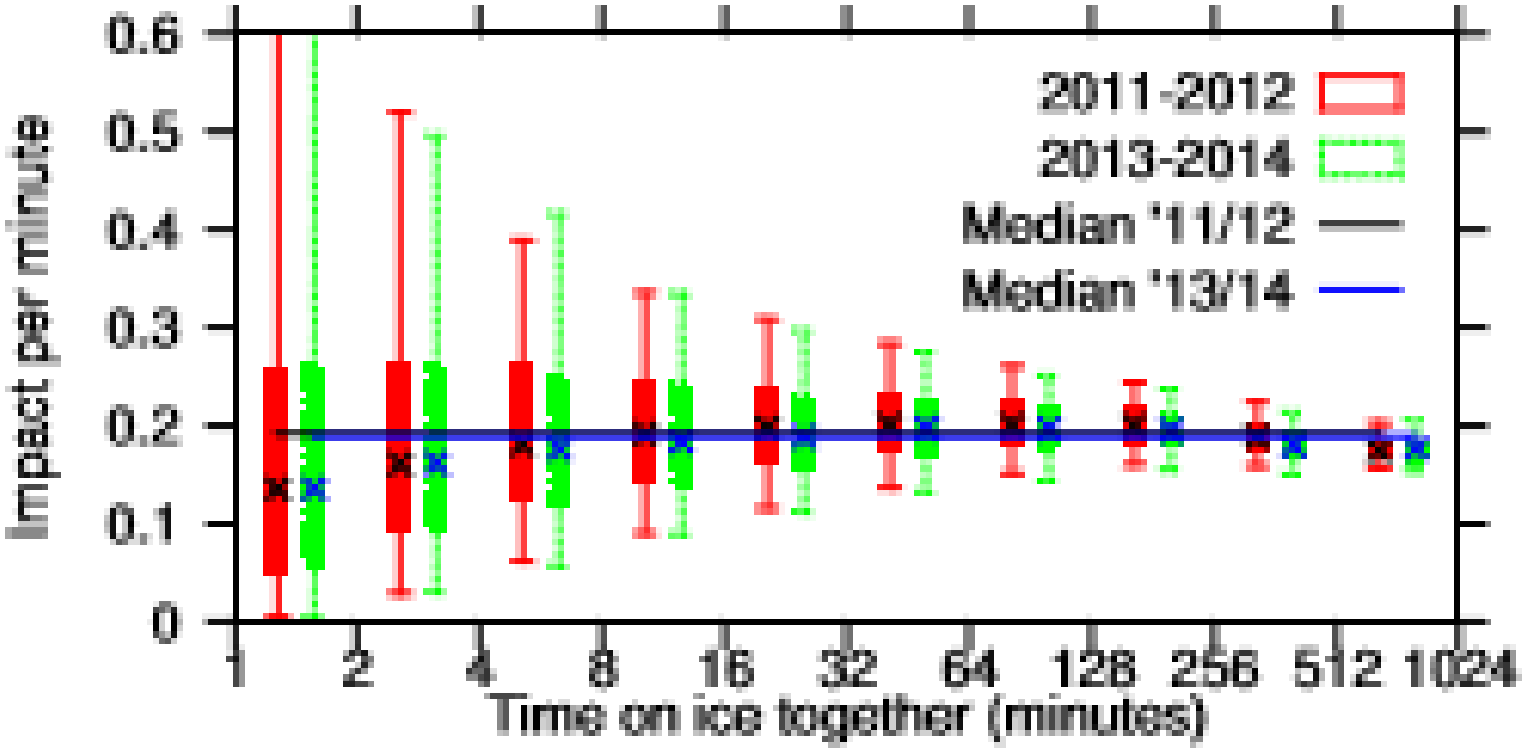


Top pairs 2011-2012

Table 3. Top pairs 2011-2012 according to total impact.

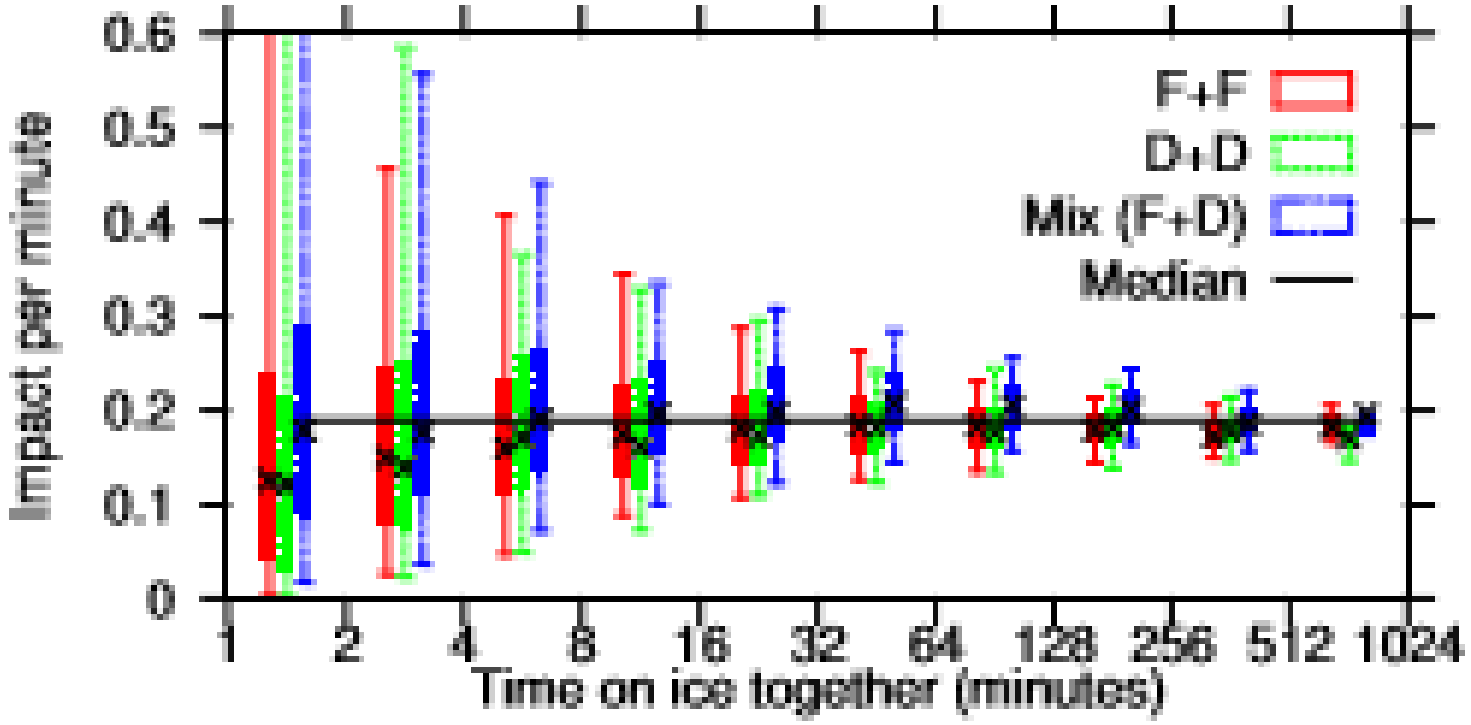
	Player 1					Player 2					Pair stats		
	Name	Pos	G	A	+/-	Name	Pos	G	A	+/-	Team	Impact	TOI
Forwards	Ilya Kovalchuk	R	37	46	-9	Zach Parise	L	31	38	-5	NJD	121.17	40,163
	Ryan O'Reilly	C	18	37	-1	Gabriel Landeskog	L	22	30	+20	COL	115.74	39,021
	Joe Pavelski	C	31	30	+18	Joe Thornton	C	18	59	+17	SJS	112.65	39,353
	Steven Stamkos	C	60	37	+7	Martin St. Louis	R	25	49	-3	TBL	111.77	35,941
	Milan Michalek	L	35	25	+4	Jason Spezza	C	34	50	+11	OTT	111.73	36,689
Defenders	Dan Girardi	D	5	24	+13	Ryan McDonagh	D	7	25	+25	NYR	155.28	55,911
	Filip Kuba	D	6	26	+26	Erik Karlsson	D	19	59	+16	OTT	134.74	47,985
	Francois Beauchemin	D	8	14	-14	Cam Fowler	D	5	24	-28	ANA	125.54	45,795
	Josh Gorges	D	2	14	+14	P.K. Subban	D	7	29	+9	MTL	125.16	44,390
	Carl Gunnarsson	D	4	15	-9	Dion Phaneuf	D	12	32	-10	TOR	123.06	36,181
Mixed	Jason Spezza	C	34	50	+11	Erik Karlsson	D	19	59	+16	OTT	110.58	35,990
	Joe Pavelski	C	31	30	+18	Dan Boyle	D	9	39	+10	SJS	106.04	35,612
	Joe Thornton	C	18	59	+17	Dan Boyle	D	9	39	+10	SJS	102.96	35,160
	Tomas Fleischmann	L	27	34	-7	Brian Campbell	D	4	49	-9	FLA	98.08	31,804
	Stephen Weiss	C	20	27	+5	Brian Campbell	D	4	49	-9	FLA	96.79	32,995

Impact per minute



Variation decreases when more joint TOI
Medians highest in 16-256 minutes joint TOI

Impact per minute



Mixed pairs may have higher impact

Outline

- Motivation
- Method
- Results
- Conclusion

Conclusion - summary

- Investigated ways to define impact of (pairs of) players in ice hockey

Conclusion – ongoing work in ice hockey

- Alternative reward functions for the performance model
  - not all goals are equally important
- Game prediction and season simulation