A Look at the Statistical Evaluation of Hockey Goalies

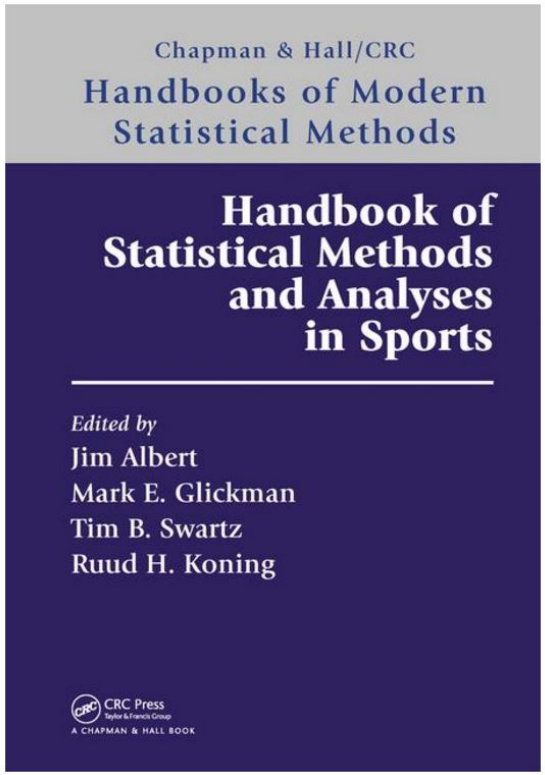
Michael Schuckers

St. Lawrence University
Statistical Sports Consulting, LLC
schuckers@stlawu.edu
@SchuckersM
@EmpiricalSports

October 1, 2016

Sharing images from this talk is permitted.

Chapter from New Book



Handbook of Statistical Methods and Analyses in Sports,
Edited by Jim Albert, Mark E. Glickman, **Tim B. Swartz**, Ruud H. Koning

Hockey Chapters in Handbook . . .

- *Poisson/Exponential Models for Scoring in Ice Hockey*
  by Andrew Thomas
- *Hockey Performance via Regularized Logistic Regression*
  by Robert Gramacy, Matt Taddy, Sen Tian
- *Statistical Evaluation of Ice Hockey Goaltending*
  by Michael Schuckers
- *Educated Guesswork - Drafting in the National Hockey League*
  by Peter Tingling

Can be pre-ordered from Amazon

This Talk

- Incomplete Overview (Too much out there)

This Talk

- Incomplete Overview (Too much out there)
- Focus on work in new chapter in *Handbook of Statistical Methods and Analyses in Sports* Editors: Jim Albert, Mark E. Glickman, **Tim B. Swartz**, Ruud H. Koning

This Talk

- Incomplete Overview (Too much out there)
- Focus on work in new chapter in *Handbook of Statistical Methods and Analyses in Sports* Editors: Jim Albert, Mark E. Glickman, **Tim B. Swartz**, Ruud H. Koning
- My chapter is *Statistical Evaluation of Ice Hockey Goaltending*

This Talk

- Incomplete Overview (Too much out there)
- Focus on work in new chapter in *Handbook of Statistical Methods and Analyses in Sports* Editors: Jim Albert, Mark E. Glickman, **Tim B. Swartz**, Ruud H. Koning
- My chapter is *Statistical Evaluation of Ice Hockey Goaltending*
- Data from 2009 to 2013. Ugh

This talk - More

- Rob Vollman has a book! Have you heard?

This talk - More

- Rob Vollman has a book! Have you heard?
- There will be some Stats and Math

This talk - More

- Rob Vollman has a book! Have you heard?
- There will be some Stats and Math
- Update on SLU Hockey

This talk - More

- Rob Vollman has a book! Have you heard?
- There will be some Stats and Math
- Update on SLU Hockey
- American talk so if I screw up it's cause I "care too much"
- . . . and consequently I might be done early.

Outline

- Traditional Metrics

Outline

- Traditional Metrics
- Save Percentage/Adjusted Save Percentage

Outline

- Traditional Metrics
- Save Percentage/Adjusted Save Percentage
- Rebounds Rates

Focus only on 5v5, Nonempty Net

Traditional Metrics of Goaltending

- Wins
- Goal Against Average (GAA)
- Save Percentage (SVP)

Basic Notation

For shot i, $\mathbf{z}_i$ is vector of attributes including location, shot type, ...

n total shots with $i = 1, \ldots, n$

Define Φ_i by

$$\Phi_i = \begin{cases} 1 & \text{if shot } i \text{ is saved,} \\ 0 & \text{if shot } i \text{ is not saved.} \end{cases} \quad (1)$$

Denote time on ice for goalie by T

Subscript of j, to distinguish between goalies $j = 1, \ldots, J$.

Wins

Wins: A goaltender receives a win if he is on the ice when his team scores the game-winning goal.

For 2015-16: Top Wins Goalies

Holtby(48), Quick(40), Jones(37), Bishop(35), Crawford(35)

Wins

Wins is highly correlated with team performance
because involves scoring (not under goalies's control)

For 2015-16: Top Wins Goalies

Holtby(WAS), Quick(LAK), Jones(SJS), Bishop(TBL), Crawford(CHI)

2015-16: Rinne(NSH) won 34/66 with 0.908 SVP, 6^{th}

League mean was 0.915.

Wins in 2009-10

For 2009-10: Marc-André Fleury won 37/65 games 0.905 save proportion, 6^{th}.

Jonathan Quick won 39/71 games with 0.907 save proportion, 8^{th}.

League mean was 0.908.

Goals Against Average

Goals Against Average (GAA) is the average number of goals that a goalie concedes per 60 minutes.

GAA

In our notation

$$GAA = 60\left[\frac{n - \sum_{i=1}^{n} \Phi_i}{T}\right]. \quad (2)$$

GAA

In our notation

$$GAA = 60\left[\frac{n - \sum_{i=1}^{n} \Phi_i}{T}\right]. \tag{2}$$

$$GAA = 60\left[\frac{n}{T}\left(1 - SVP\right)\right]. \tag{3}$$

GAA

In our notation

$$GAA = 60\left[\frac{n - \sum_{i=1}^{n} \Phi_i}{T}\right]. \tag{2}$$

$$GAA = 60\left[\frac{n}{T}(1 - SVP)\right]. \tag{3}$$

n and T are not under goalies direct control and is function of SVP

I'll drop subscript j unless I'm talking about multiple goalies

Save Percentage

This is the percentage of shots faced that are saved by the goalie
In our notation:

$$SVP = \frac{1}{n}\sum_{i=1}^{n}\Phi_i.$$

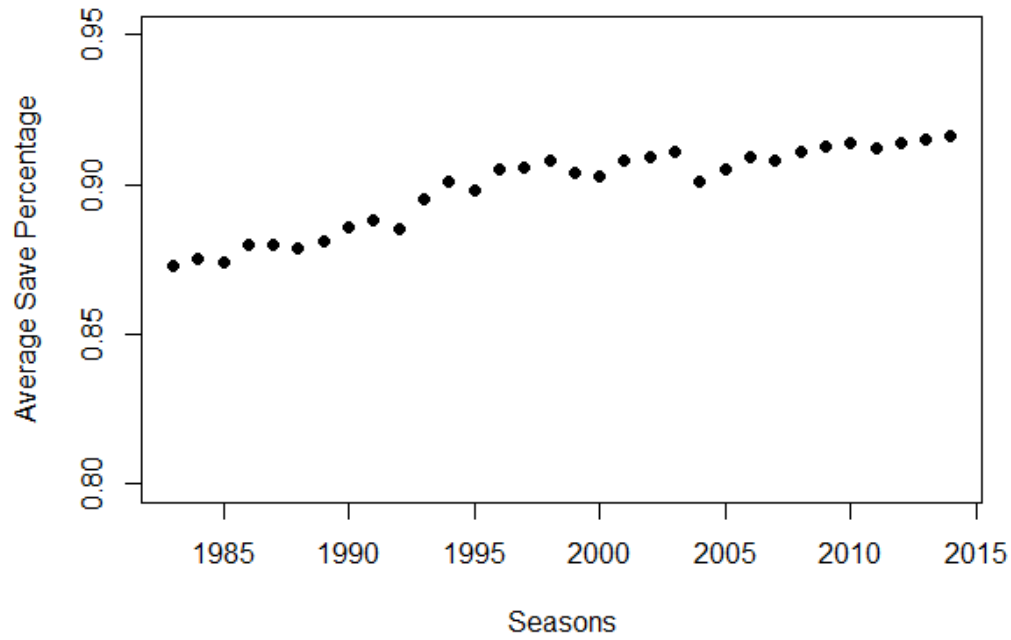
This is our primary traditional metric.

Goalie has 'control' over metric in that they face every shot.

Save Percentage

Been changing over time

Figure 1: NHL League Average SVP by Season 1983 to 2015



Source: Hockey-reference.com

SVP

Inherently SVP is composed of two parts:

1. Goalies ability to stop a shot of type u,
2. How often shots of type u are faced,

SVP Notation

$$\begin{aligned} SVP_j &= \frac{1}{n_j}\sum_{i=1}^{n_j}\Phi_{ij} \\ &= \sum_{u=1}^{\mathcal{U}} G_j(u)S_j(u). \end{aligned}$$

where

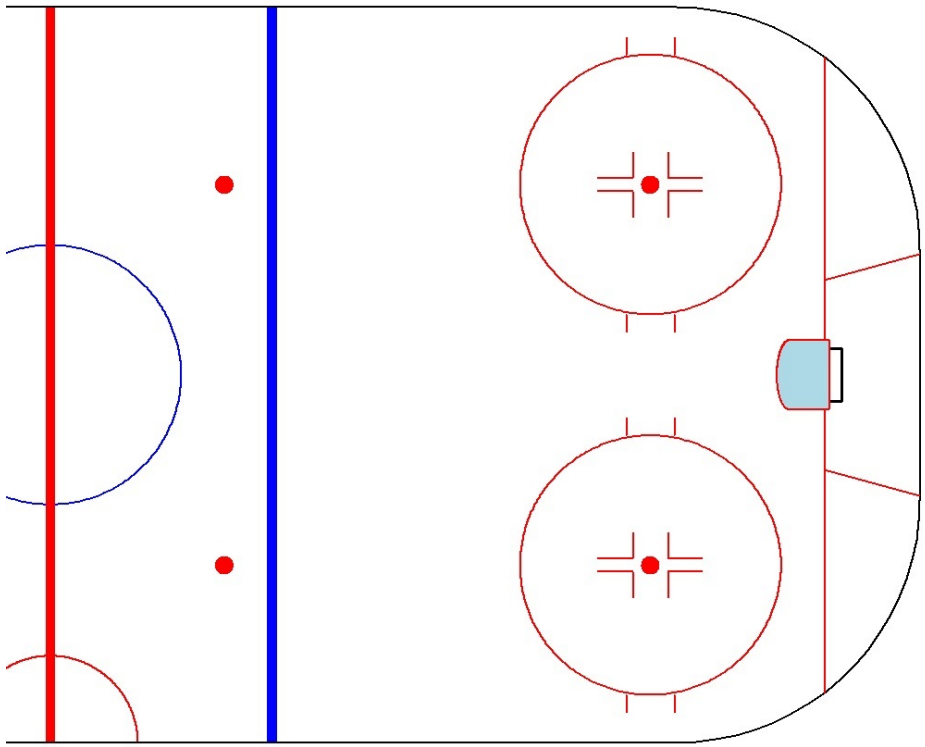
- u is type of shots
- $G(u)$ is a probability that a goalie saves a shot with attributes $\mathbf{z}_u$
- $S(u)$ is the probability of facing a shot with attributes $\mathbf{z}_u$
- $\mathbf{z}_u$ is a vector of descriptors for the attributes of a shot
- $\mathbf{z}_u = (x, y, d, \theta, w, \ldots)^T$ where $u = 1, \ldots, \mathcal{U}$

SVP Notation

Shot attributes as recorded by NHL RTSS system:

- X location
- Y location
- Shot Type (Backhand, Slap, Snap, Wrist, Tip, Wrap)
- Angle of shot from center ice
- Distance from net
- Rebound or not, Rush or not

Attributes



SVP

Inherently SVP is composed of two parts:

1. Goalies ability to stop a shot of type u, $G(u)$
2. How often shots of type u are faced, $S(u)$

SVP Notation-Simplified

Shots Attributes, $\mathbf{z}_u$		Goalie A		Goalie B		Goalie C	
Shot Type	Distance	$G(u)$	n_u	$G(u)$	n_u	$G(u)$	n_u
Slap	Near	0.900	1200	0.880	100	0.880	500
Slap	Far	0.930	200	0.920	700	0.920	500
Wrist	Near	0.920	500	0.900	300	0.900	500
Wrist	Far	0.930	100	0.910	900	0.910	500
Overall		0.9095	2000	0.9105	2000	0.9025	2000

Here, $\mathcal{U} = 4$ and $S(u) = n_u/n$.

(Average) Shot Quality

Long discussions in 'hockey analytics' circles about this.

Argument: Differences in shot distributions for different goalies $S_j(u)$'s tend to even out

In our framework this mean that $S_j(u) \to \bar{S}(u)$ and n_j increases
where $\bar{S}(u)$ is an average across the league.

Data Issues

There are issues with NHL Data (x,y)'s

- Data from NHL's RTSS feed has some issues
- x,y coordinates often far from ground truth (video analysis) especially in certain rinks (MSG)
- Spatial tracking data is coming (cf. Basketball)
- I hope it is public

See Schuckers and Macdonald (2014) for more on these issues and what to do with/about them . . .

Adjusting X,Y shot coordinates

From Schuckers and Curro (2013)

Take league average CDF for X and Y separately, $F_X(x)$ and $G_Y(y)$

Calculate CDF's of X and Y for rink R, for shots by away teams A and for shots by away teams at rink R, RA.

For shot i from rink R,

$$x_i' = round(F_X^{-1}(F_{XR}(x_i) + (F_A(x_i) - F_{RA}(x_i))) \text{ and}$$
$$y_i' = round(G_Y^{-1}(G_{YR}(y_i) + (G_A(y_i) - G_{RA}(y_i)))).$$

Image warping, see e.g. Pishchulin *et al* (2012) deal with bivariate nature of data (future work?)

Evaluation of Metrics

Predictability/Reliability: Year to year correlation

Consistency: Within year correlation (even and odd)

Validity: Correlation to Winning or goal differential

Evaluation of SVP

Table 1: Intraseason correlation of SVP

Season	NHL Average SVP	More than 500 shots Correlation	J	More than 750 shots Correlation	J
2009/10	0.9198	0.203	48	0.072	33
2010/11	0.9221	0.282	46	0.385	32
2011/12	0.9215	0.180	45	0.165	35
2012/13	0.9203	0.145	24	0.252	9

J is number of goalies

Table 2: Correlation in year to year SVP

Seasons	More than 500 shots		More than 750 shots	
	Correlation	J	Correlation	J
2009/10 v 2010/11	0.186	34	0.120	22
2010/11 v 2011/12	0.018	37	-0.038	28
2011/12 v 2012/13	0.060	20	0.909	8

Rebounds: a digression

What is the impact of rebounds on SVP?

Are rebounds something goalies can control?

Rebound Rate $= \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} \Phi_i}$

where δ_i is a 1 if shot i results in a rebound, zero otherwise.

Myrland (2009), Pettapiece (2013)

Rebounds

Table 3: Intraseason correlation of Rebound Rate for even and odd shots

Season	NHL Average Rebound Rate	More than 500 shots faced Correlation	J	More than 750 shots faced Correlation	J
2009/10	0.0823	0.066	46	0.327	33
2010/11	0.0825	0.329	48	0.395	32
2011/12	0.0842	0.512	45	0.642	35
2012/13	0.0838	0.268	24	0.336	9

Rebounds

Table 4: Correlation between a goalie's rebound rate and their save proportion

Seasons	More than 500 shots		More than 750 shots	
	Correlation	J	Correlation	J
2009/10	-0.164	48	-0.114	33
2010/11	-0.440	46	-0.210	32
2011/12	-0.261	45	-0.252	35
2012/13	-0.066	24	-0.195	9

Rebounds

Table 5: Correlation between a goalie's rebound rate in one year and their rebound rate in subsequent years.

Seasons	More than 500 shots		More than 750 shots	
	Correlation	J	Correlation	J
2009/10 v 2010/11	-0.086	61	0.217	22
2010/11 v 2011/12	0.483	37	0.474	28
2011/12 v 2012/13	-0.261	20	0.067	8

Adjusted SVP

Two ways to adjust SVP

1. Adjust for the goalie
2. Adjust for the shots

Adjusted SVP

Two ways to adjust SVP

1. Adjust for the goalie
2. Adjust for the shots

Each answers different question

Adjusted SVP

Two ways to adjust SVP

1. Adjust for the goalie
2. Adjust for the shots

Each answers different question

1. What would an average goalie do with shots faced by goalie j
2. What would goalie j do with shots faced by average goalie

First historically adjusted SVP, second Shot Neutral SVP (A. Ryder)

Adjusted SVP

What does this look like in our notation

Adjusted SVP

What does this look like in our notation

$$\begin{aligned} SVP_j &= \sum_{u=1}^{\mathcal{U}} G_j(u) S_j(u) \\ \text{aSVP}_{\overline{G}_j} &= \sum_{u=1}^{\mathcal{U}} \overline{G}(u) S_j(u) \\ \text{aSVP}_{\overline{S}_j} &= \sum_{u=1}^{\mathcal{U}} G_j(u) \overline{S}(u) \end{aligned} \tag{4}$$

where $\bar{G}$ represents average goalie and $\bar{S}$ represents average distribution of shots.

Adjusted SVP

How to get $\overline{G}(u)$ and $\overline{S}(u)$?

For $\overline{G}(u)$, find league SVP at each u
Smoothing makes replace $\overline{G}(u)$ with $\hat{G}(u)$

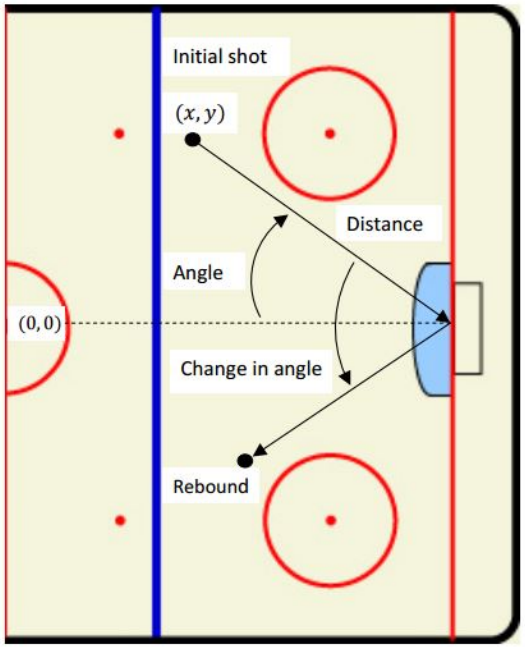
For $\overline{S}(u)$, find league average distribution of u
Smoothing replace $\overline{S}(u)$ with $\hat{S}(u)$

Many smoothing (regression) attempts

Adjusted SVP

Smoothing Regression based shot probability models (usually logistic)

Krzywicki (2005, 2009, 2010), Macdonald (2012)



Logistic regression with x, y, distance, shot type, angle, change in angle rebound covariates

From Macdonald *et al* (2012). Used by permission.

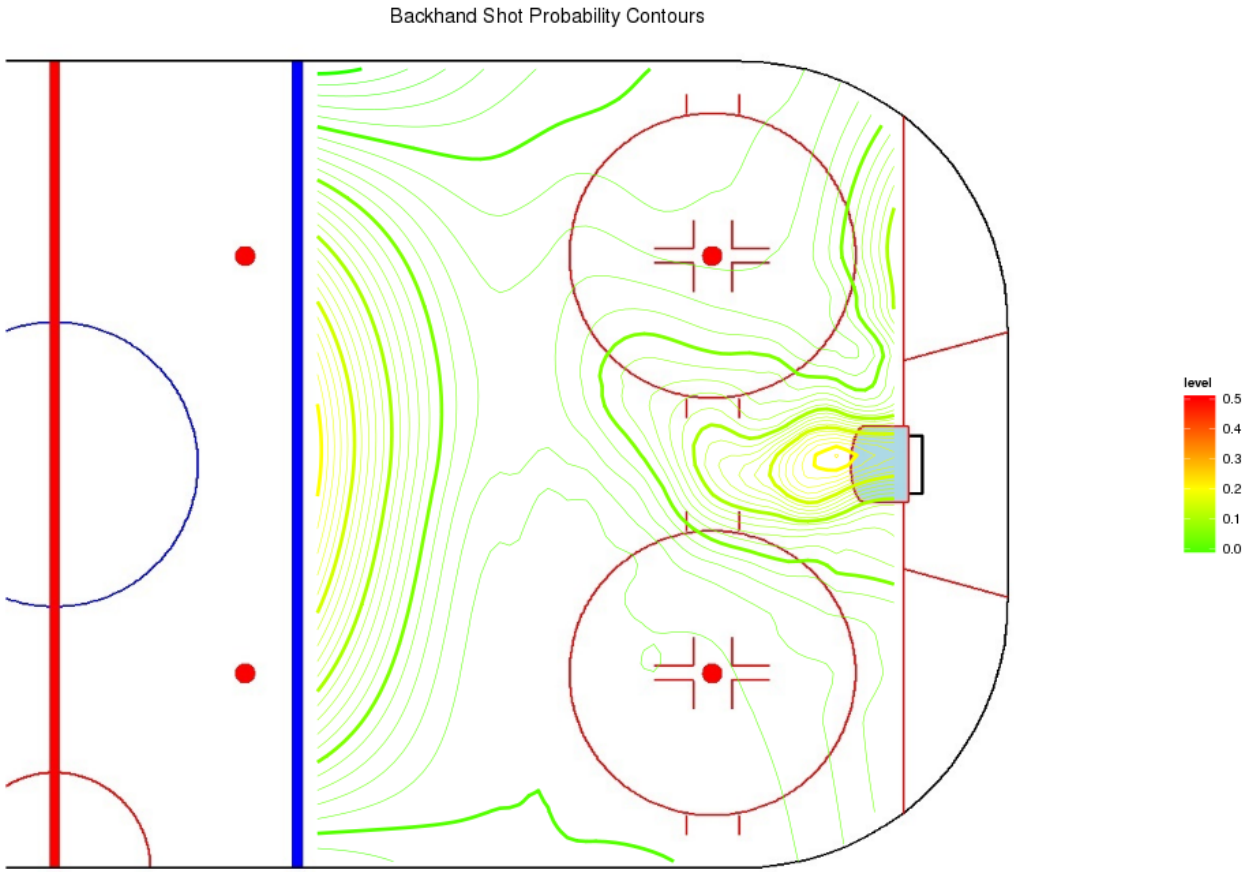
Adjusted SVP

More smoothing: Schuckers (2011, 2016)
Spatial smoothing by shot type
NB: in images to follow:

- Data from 2009 to 2013
- 5v5 Non-empty net
- By NHL shot type + rebound + 'rush'(Johnson, 2014)
- Eliminated shots based upon unusual recording (7 on ice, etc.)
- Kept locations with errors (clearly some artifacts)
- Adjusted for rink

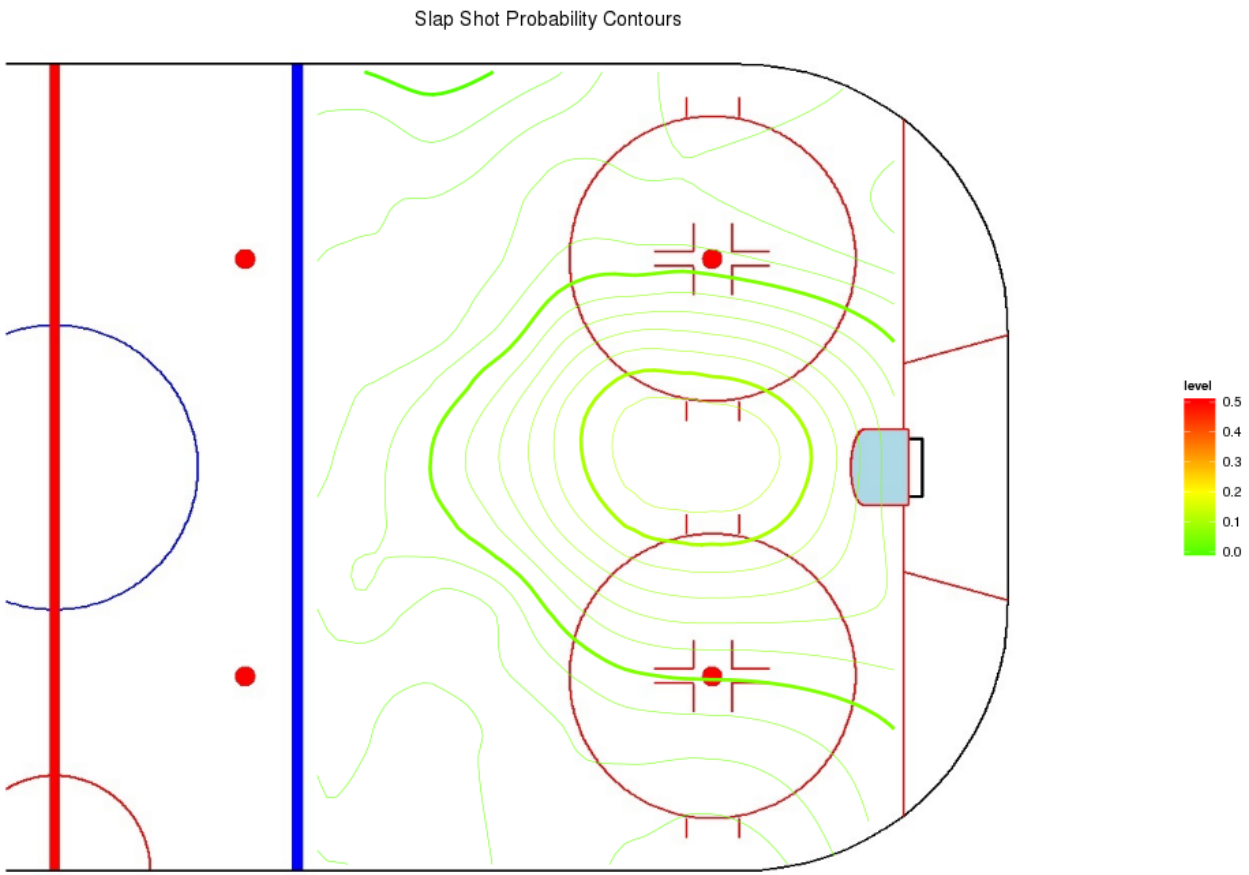
Adjusted SVP

Backhand Shots



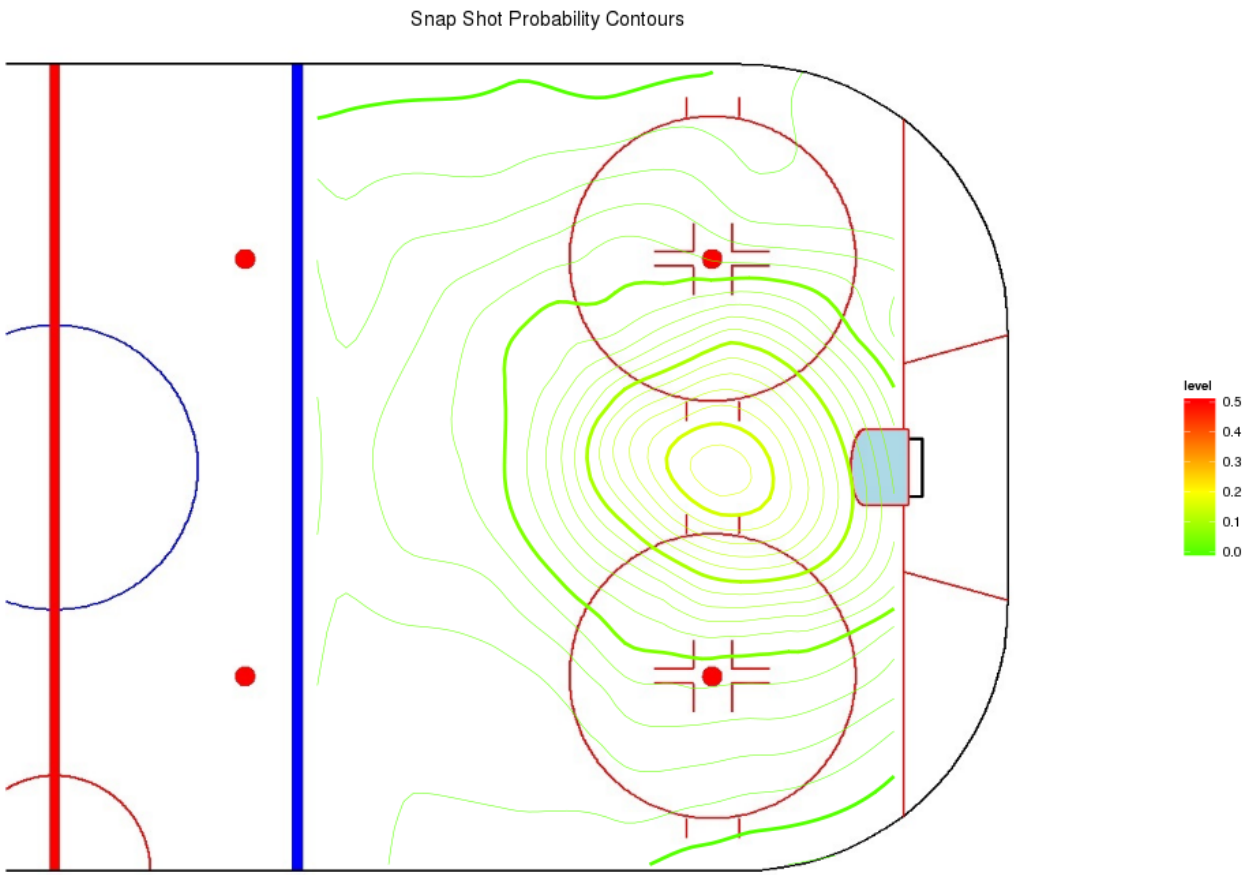
Adjusted SVP

Slap Shots



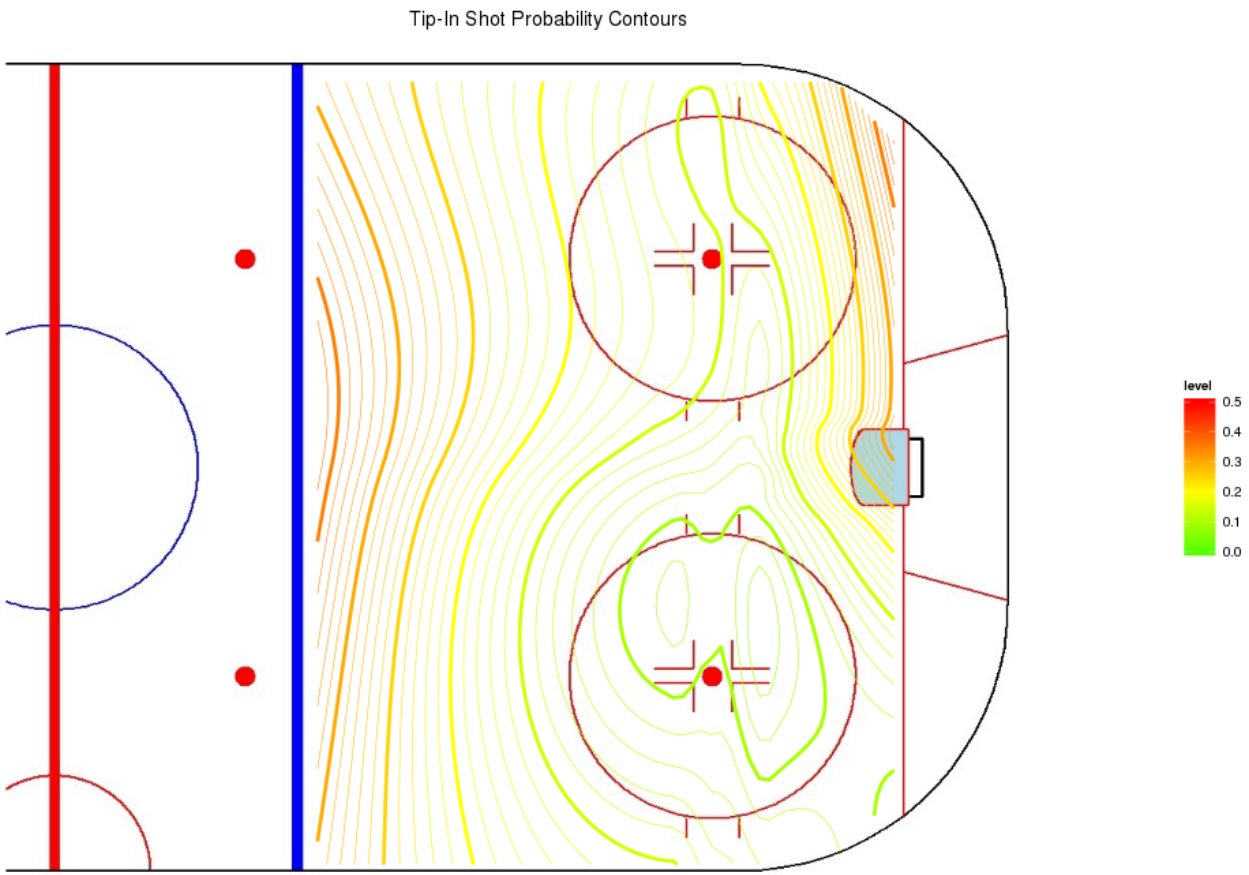
Adjusted SVP

Snap Shots



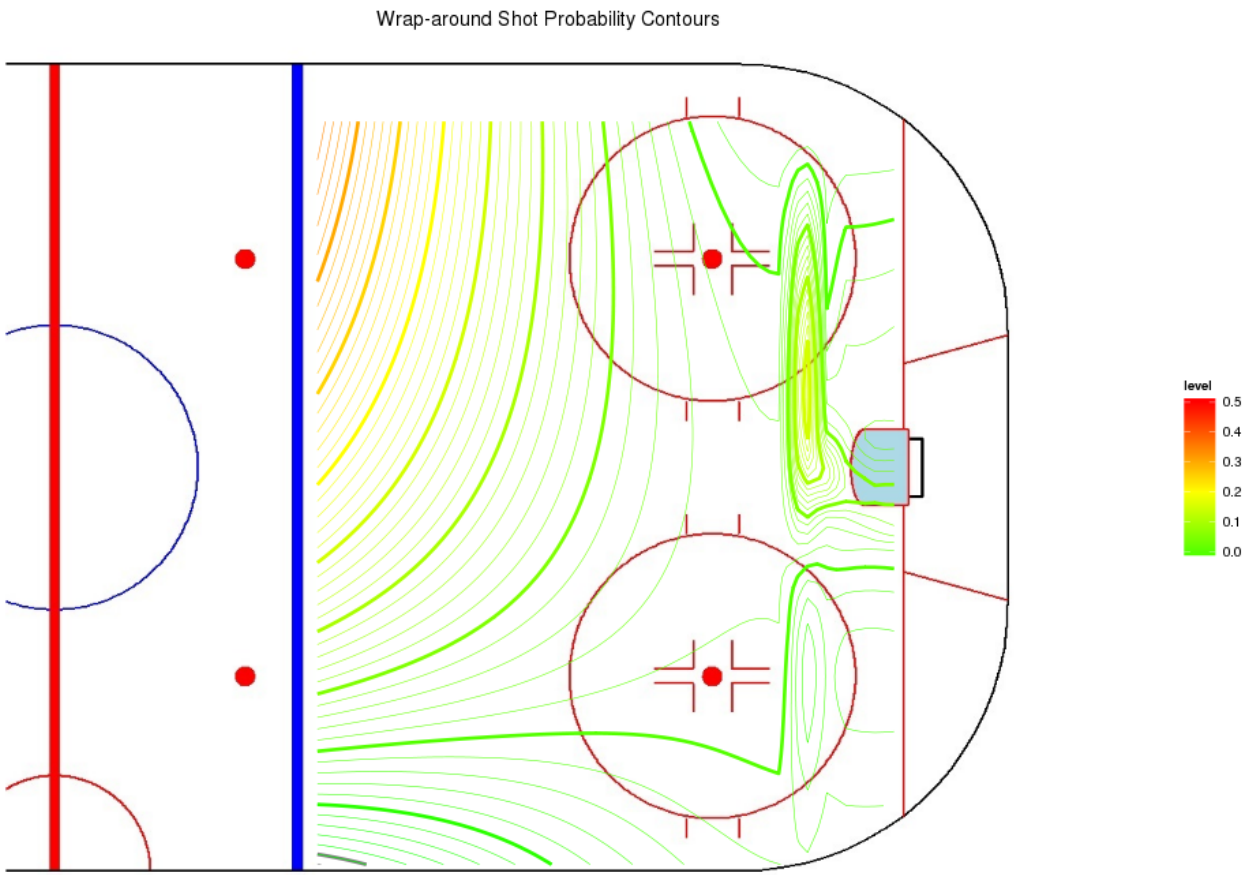
Adjusted SVP

Tip-In Shots



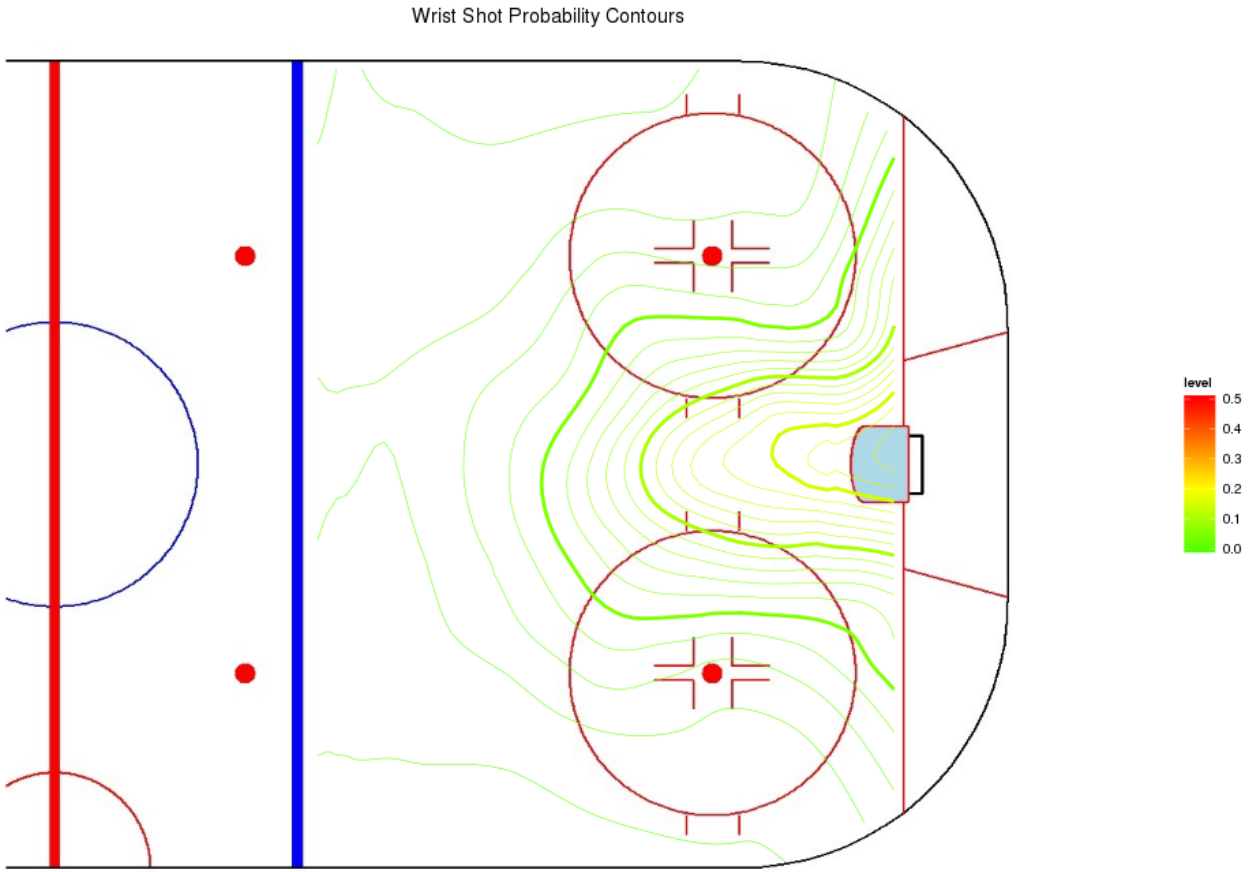
Adjusted SVP

Wraparound Shots



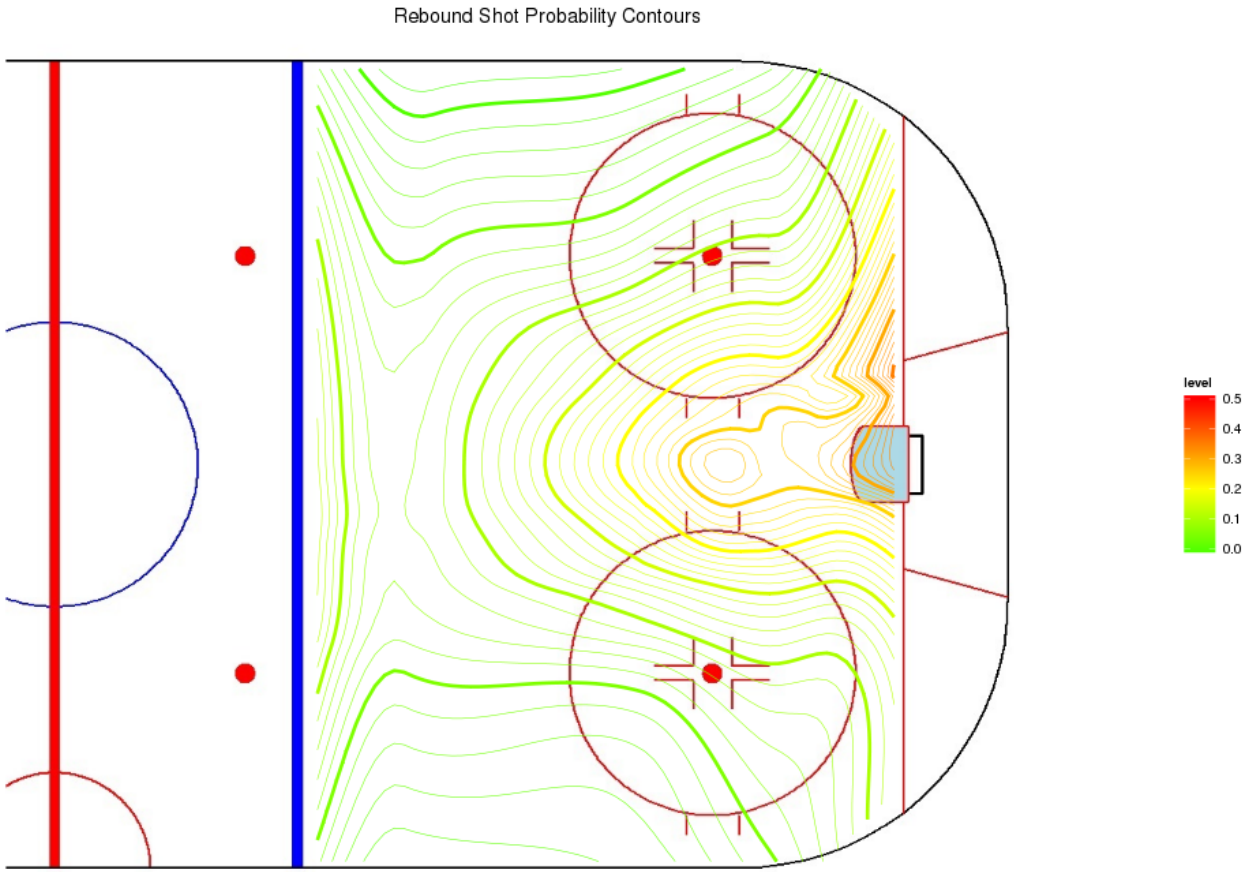
Adjusted SVP

Wrist Shots



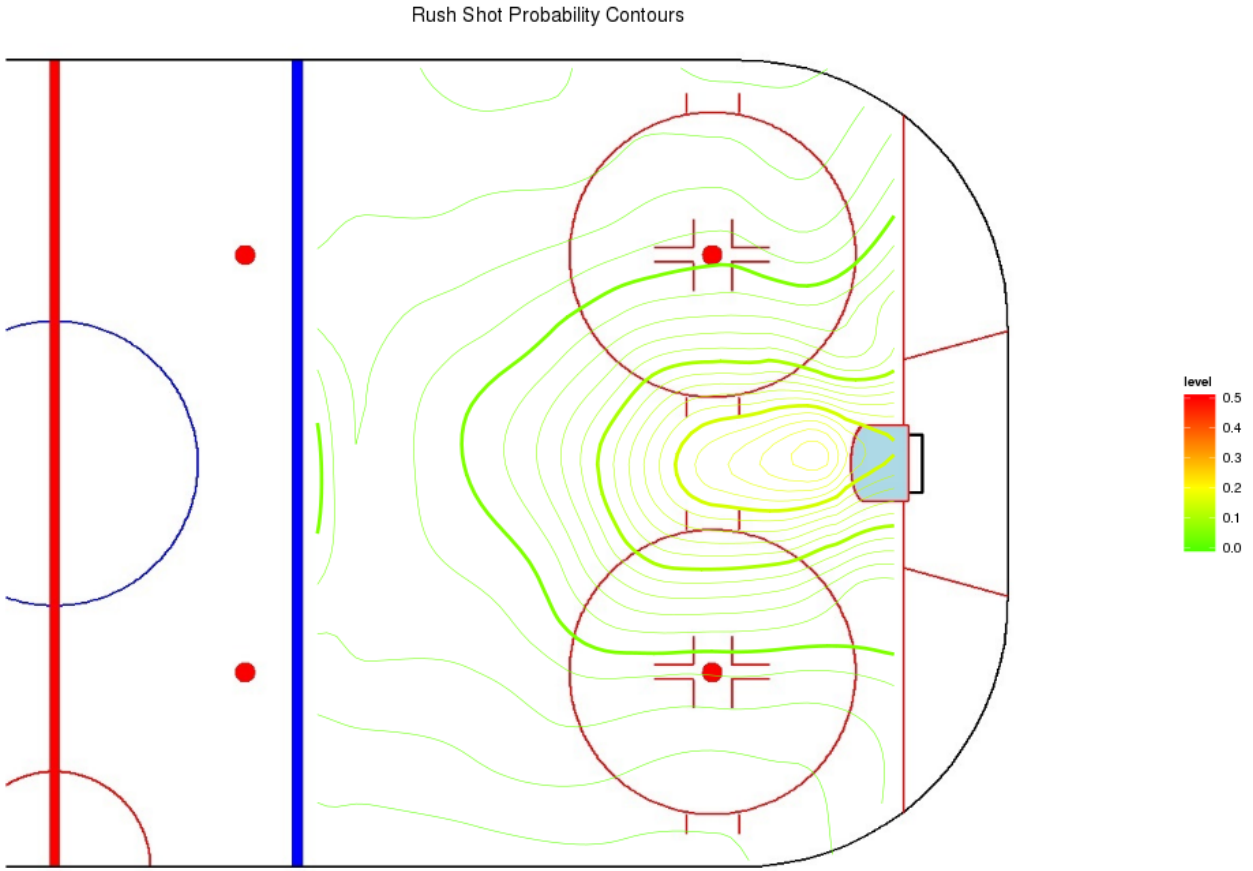
Adjusted SVP

Rebound Shots



Adjusted SVP

Rush Shots



Shot Neutral SVP

Updating: Defense Independent Goalie Rating (DIGR), Schuckers (2011)

$$\begin{aligned} DIGR_j &= \sum_{u=1}^{\mathcal{U}} \widetilde{G}_j(u)\overline{S}(u) \\ &= \sum_{u=1}^{\mathcal{U}} \left[\alpha(u)G_j(u) + (1-\alpha(u))\overline{G}(u)\right]\overline{S}(u) \end{aligned}$$

where $\alpha(u) = \frac{n_u}{n_u+n^\dagger}$,
n_u is the number of shots faced by goalie $j, j = 1, \ldots, J$
of shottype w at strength s and $n^\dagger$ is a 'shrinkage' constant.

DIGRing around

Table 6: Intraseason correlation of DIGR for even and odd shots using $n^{\dagger} = 1000$

Season	More than 500 shots faced		More than 750 shots faced	
	Correlation	J	Correlation	J
2009/10	0.288	48	0.309	33
2010/11	0.644	45	0.751	36
2011/12	0.655	45	0.621	33
2012/13	0.251	24	0.178	9

More DIGRing around

Table 7: Correlation between a goalie's DIGR in one year and their DIGR in subsequent years.

Seasons	More than 500 shots		More than 750 shots	
	Correlation	J	Correlation	J
2009/10 v 2010/11	0.237	34	0.393	22
2010/11 v 2011/12	0.647	37	0.665	28
2011/12 v 2012/13	-0.174	15	0.677	6

Goals Saved

What about the other adjusted save percentage? $aSVP_{\overline{G}}$? Compare this to SVP.

$$\begin{aligned} SVP_j &= \sum_{u=1}^{\mathcal{U}} G_j(u) S_j(u) \\ \text{aSVP}_{\overline{G}j} &= \sum_{u=1}^{\mathcal{U}} \overline{G}(u) S_j(u) \end{aligned} \tag{5}$$

Goals Saved

Goals gained (or lost) based upon shots faced by goalie j, $S_j(u)$ relative to league average

$$= n_j \sum_{u=1}^{\mathcal{U}} \left[G_j(u) - \overline{G}(u) \right] S_j(u) \tag{6}$$

Or Goals Saved Above Average Per 60 (Mercadante, 2015)

$$\frac{60 n_j}{T} \sum_{u=1} \mathcal{U} \left[G_j(u) - \bar{G}(u) \right] S_j(u) \tag{7}$$

Incomplete list of things I haven't talked about

- Royal Road, pre shot puck movement (Valiquette)
- Shot Quality Project (Boyle)
- Career Trajectories (Tulsky and others)
- League Projections (Vollman and others)
- Quality Starts (Vollman)
- Pulling the Goalie (Beaudoin & Swartz, Thomas & Ventura)

Future

- Player tracking data
- Better shot attributes (e.g. speed)
- Better defensive positioning (e.g. obstruction)
- Better information for non-NHL data
- Exorcize the Voodoo

Summary

In general, we know

- Goalies are hard to predict
- Adjusting SVP can improve prediction
- Rebounds impact SVP but not super consistent
- Framework for SVP and aSVP
- So many missing covariates/attributes