Edgar

Micah Blake McCurdy
micah@hockeyviz.com

Rochester NY
RIT Hockey Analytics Conference
October 21, 2017

Models, As Abstractly As Possible

A *model* is a way to gather some important aspects of an interesting thing, so that we can benefit.

Friendly, Familiar Models

- Model cars for children
- Models from physics:

Friendly, Familiar Models

- Model cars for children
- Models from physics:
  - Planetary motion
  - Friction
  - Relativity

Friendly, Familiar Models

- Model cars for children
- Models from physics:
  - Planetary motion
  - Friction
  - Relativity
- Wind-tunnel models from aerodynamics design
- Mental models for everyday life
  - Excitement
  - Danger
- Statistical models from the social sciences:

Friendly, Familiar Models

- Model cars for children
- Models from physics:
  - Planetary motion
  - Friction
  - Relativity
- Wind-tunnel models from aerodynamics design
- Mental models for everyday life
  - Excitement
  - Danger
- Statistical models from the social sciences:
  - Behavioural economics
  - Criminology

Friendly, Familiar Models

- Model cars for children
- Models from physics:
  - Planetary motion
  - Friction
  - Relativity
- Wind-tunnel models from aerodynamics design
- Mental models for everyday life
  - Excitement
  - Danger
- Statistical models from the social sciences:
  - Behavioural economics
  - Criminology
  - Hockey!

What Makes A Model Good?

- Accuracy
- Efficiency
- Interpretability

Relevance

Model Inputs $\longrightarrow$??? $\longrightarrow$ Model outputs

Relevance

$$\text{Model Inputs} \longrightarrow ??? \longrightarrow \text{Model outputs}$$

$$\text{(mostly relevant)} \longrightarrow ??? \longrightarrow \text{(always relevant)}$$

Model Classification

- When the inputs are combined by minimizing some function, that makes the model *statistical*. Most of the best hockey models (Luszczyszyn,Perry) are statistical.

Model Classification

- When the inputs are combined by minimizing some function, that makes the model *statistical*. Most of the best hockey models (Luszczyszyn,Perry) are statistical.
- When the inputs are combined by some encoding of the mechanics of the thing being modelled, that makes the model *phenomenological*.

Model Classification

- When the inputs are combined by minimizing some function, that makes the model *statistical*. Most of the best hockey models (Luszczyszyn,Perry) are statistical.
- When the inputs are combined by some encoding of the mechanics of the thing being modelled, that makes the model *phenomenological*.
  - When the combining is done systematically with oversight, that makes the model *scientific*.

Edgar

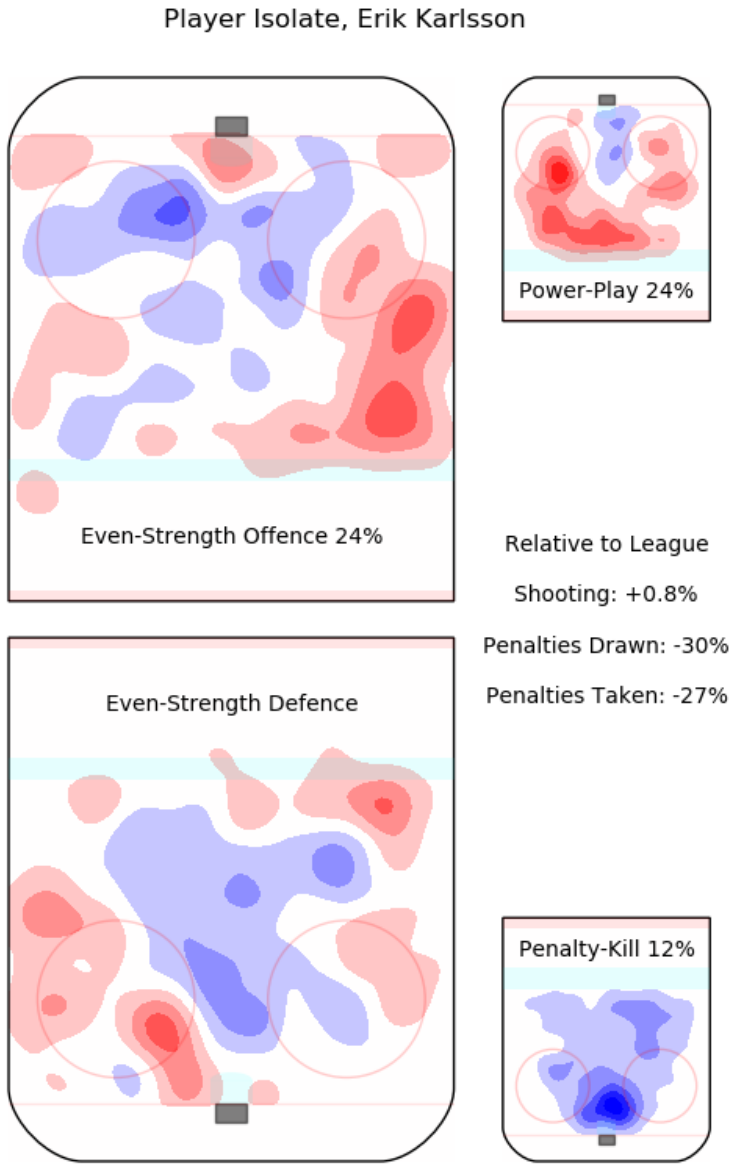
I made what I think is a "scientific" model in this sense and I called it: Edgar

Which aspects are important?

- Unblocked shot rates and their locations.
- Penalty rates
- Shooting talent
- Goaltending talent
- Who takes the shots
- Rest

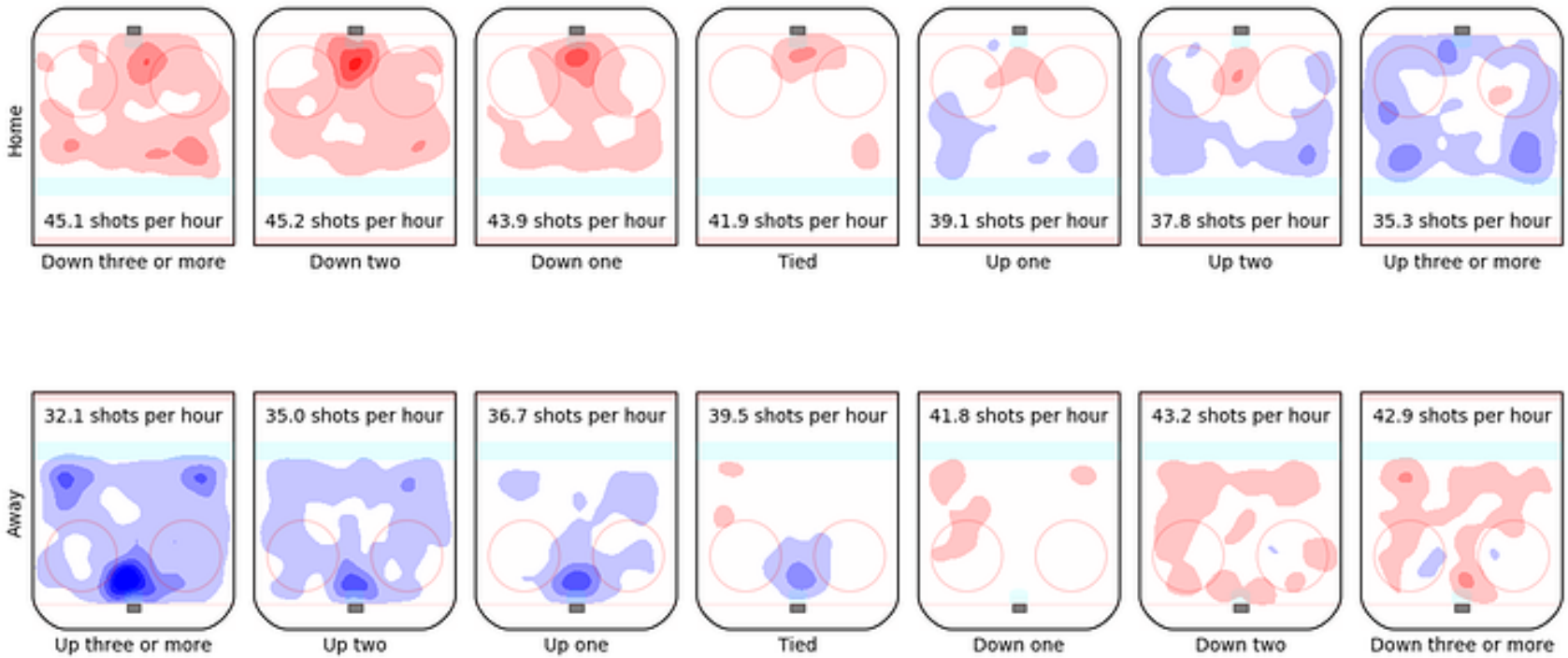
I estimate them all with statistics.

Isolating Player Ability



Isolating Player Ability: Shot Maps

Shot rate maps are adjusted for score



Some day: adjusted for teammates (and then maybe zone usage and competition)

Shooting and Goaltending Abilities

All measured *relative to where shots are taken*.

Shooters

Excellent shooters:

Jake Guentzel	+6.5%
Patrik Laine	+6.2%
Sven Baertschi	+5.1%
Jannik Hansen	+4.5%

(Also Ho-Sang, Malkin, Barkov, Gourde, Athanasiou, Oshie)

Regressed 2/3 of the way to the mean

Goalies

Excellent goalies:

Brayden Holtby	+1.3%
Matt Murray	+1.2%
Carey Price	+1.2%
Henrik Lundqvist	+1.2%

(Also Smith, Grubauer, Crawford, Gibson, Reimer, Bobrovsky)

Regressed 1/2 of the way to the mean

Additional Sneakiness

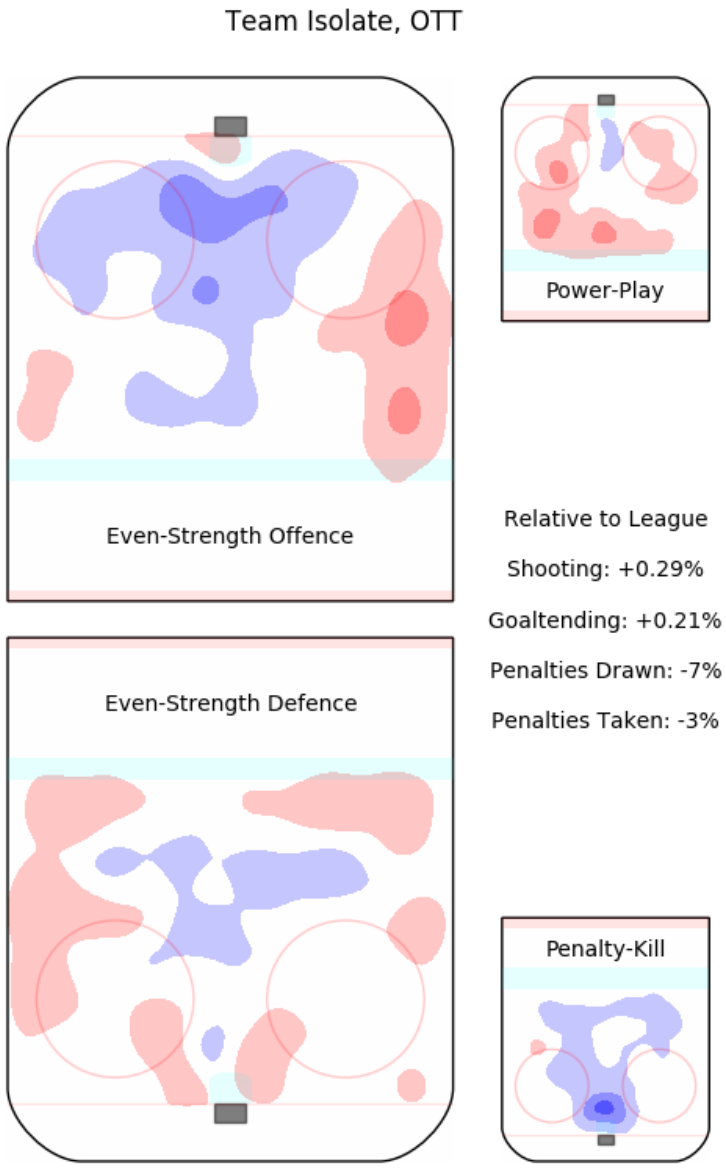
Penalty Rates and Shot Propensities: Untouched! (for now)

Rookies treated as league average, except for a chosen few.
(estimated by Hannah Stuart)

Some ad-hoc regression for people with very little relevant icetime.

Isolating Team Abilities

Estimate team rosters



Adjust for rest

Simulation Mechanism

Model shots and penalties as Poisson processes with the measured rates.

For every shot taken:

- Choose a location
- Choose a shooter
- Adjust for shooting talent
- Adjust for goalie
- See if it's a goal

And so on, for sixty or perhaps sixty-five minutes.

Information in Excess of Guessing

My preferred measure of accuracy for single games:

$$100 \log_2 2p$$

where p is the probability for the outcome that happened.

Really just log-loss, scaled onto 0 (guessing) and 100 (perfection).

Results From 2016-2017

Creator	**Model**	**Information In Excess of Guessing (per game)**
Perry	Salad	5.03
Nandakumar	Feline Frenzy	4.65
Luszczyszyn	Preszczyszyn	4.36
M.	**Edgar**	4.13
Sprigings	DTMAH	4.10
M.	Cordelia	2.11

Thanks!