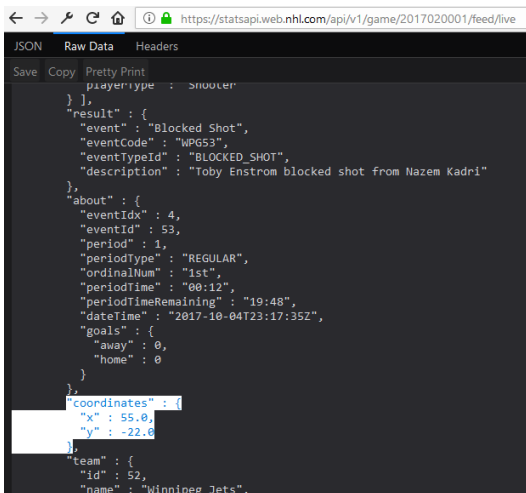# Hockey in Space!

Characterizing team-wise differences in shot locations with spatial point processes

Devan Becker
The University of Western Ontario
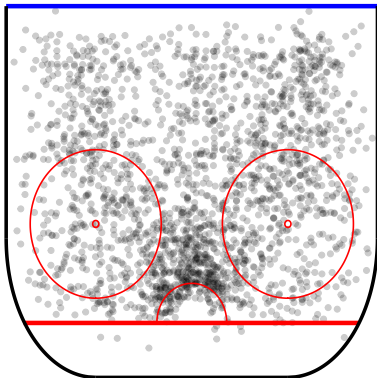dbecker7@uwo.ca

September 16, 2018
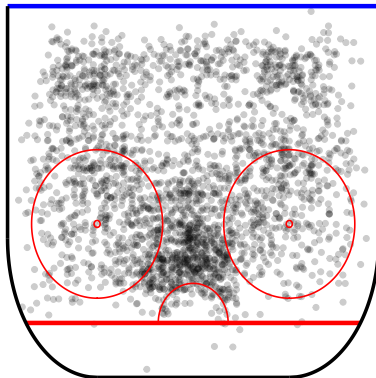
Introduction

The Data

2017-2018 Season Shot Locations

All Shots, Detroit

All Shots, Tampa Bay

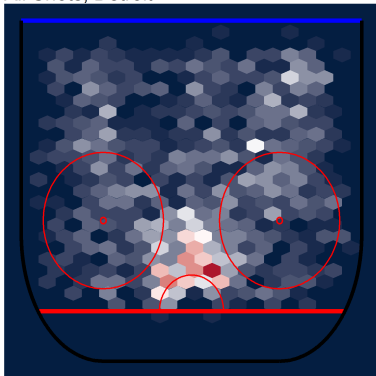# 2017-2018 Season Shot Hexbins
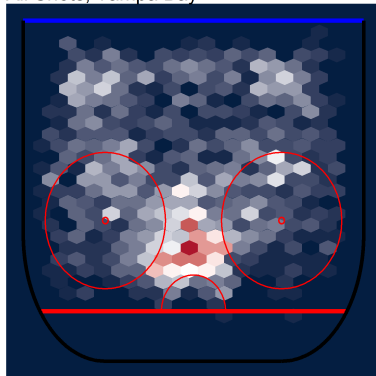
```
## Coordinate system already present. Adding new coordinate
## Coordinate system already present. Adding new coordinate
```
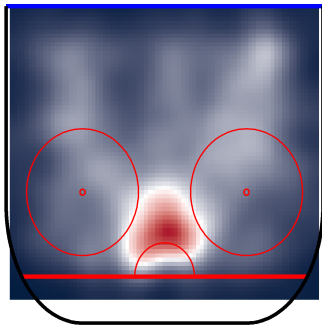

All Shots, Detroit

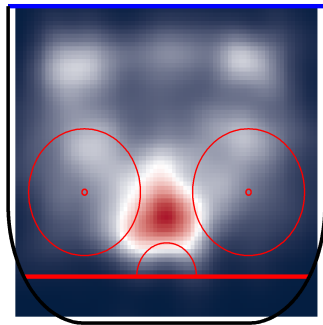All Shots, Tampa Bay

2017-2018 Season Shot Density

Shot Density, Detroit

Shot Density, Tampa Bay

Objective

Fit a parametric statistical model to determine:

- Where do different teams shoot from?
- Are the patterns consistent? (Variance!)
- ~~Which shots go in?~~
- What can goalies expect?

LGCP

Log-Gaussian Cox Processes (Yay math!)

$$\log(\Lambda(x, y)) = \mu + \beta C(x, y) + S(x, y)$$

The log of the rate of points in a given location is modelled as an intercept plus a spatial covariate plus a random process.
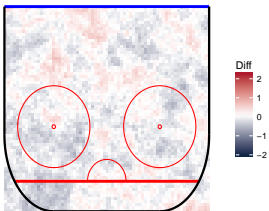
Log-Gaussian Cox Processes (Yay math!)

$$\log(\Lambda(x,y)) = \mu + \beta C(x,y) + S(x,y)$$

The log of the rate of points in a given location is modelled as an intercept plus a spatial covariate plus a random process.
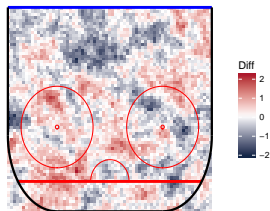
- "Random" doesn't mean unstructured!
- The random process is a smooth function based on the normal distribution.
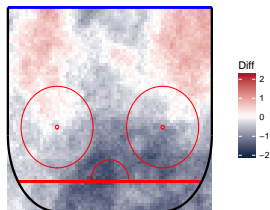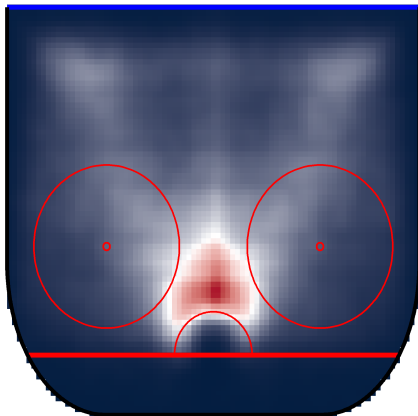
Variance and Clustering (Simulated)

League Average as a Spatial Covariate

Density Estimate - 2017/2018 Season

League Average as a Spatial Covariate

Our interpretation becomes:

$$\log(\Lambda(x, y)) = \mu + C(x, y) + S(x, y)$$

The log of the rate of shots is modelled as an intercept plus the league average plus a team specific deviation.

League Average as a Spatial Covariate

Our interpretation becomes:

$$\log(\Lambda(x, y)) = \mu + C(x, y) + S(x, y)$$

The log of the rate of shots is modelled as an intercept plus the league average plus a team specific deviation.

- $S(x, y)$ can be seen as the *intended* strategy difference.
- The variance and range illustrate the team's *consistency*.
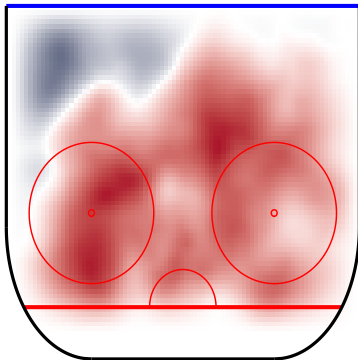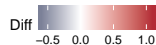
Estimation of LGCP
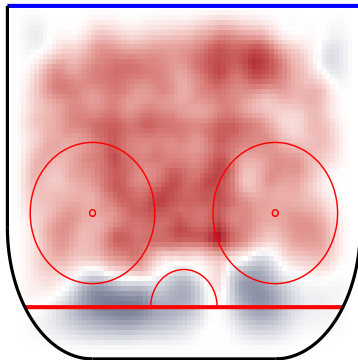
Estimation of LGCP

oh no

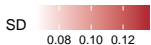Results
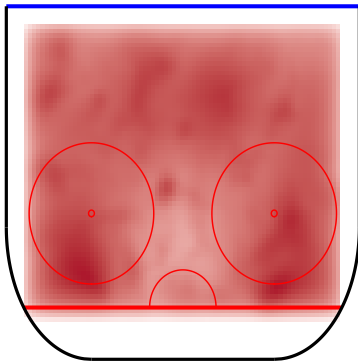
Random Processes - S(x,y)

Diff. from League – Detroit

Diff. from League – Tampa Bay

Variance at Every Location

SD – Detroit

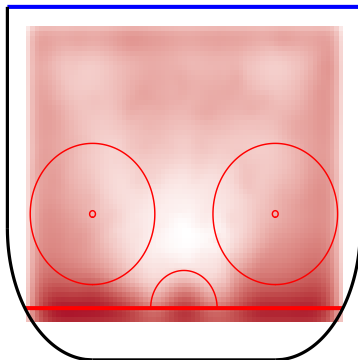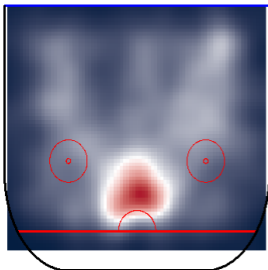SD – Tampa Bay
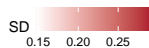
All Results - Shiny App
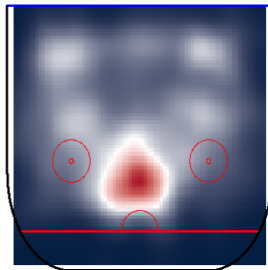
Limitations

- Differences from league are often minor

- One dense cluster means small clusters estimated everywhere

- Needs a lot of data
  - Can't just look at one team's goals

Final Notes

Conclusions

- This method can indicate the manifested strategy
- Variance and cluster size indicate a teams consistency
    - Careful interpretation

Future Work

- Different play types (e.g. first shot after possession)
- Statistical comparison of teams
- Spatially varying range parameter
- All the things that the audience suggests

Acknowledgments

Thank you to my supervisors: Doug Woolford, Charmaine Dean, and W. John Braun

Thanks for listening! Any further questions: dbecker7@uwo.ca