Rambling Notes to Accompany the SEAHAC Presentation:

Evaluation of Game Event Data to Assess Team Performance

Will Fields, PhD
3 December 2022

Title Slide

I'll begin with a more extended introduction of myself because I think it helps give some context for the general philosophy I have about analytics. I've been a hockey fan since I was kid, but when I was young I was also interested in ecology. I wanted to pursue a career in research, and I eventually started a PhD program at North Carolina State University. At the time I was interested in rare amphibians, and I wanted to do research to understand how local populations were affected by what went on in the surrounding landscape. This was complicated by the fact that places where we did surveys were wetlands that didn't flood every year, so we needed to explore new ways to analyze data where species might be present even if they were undetected because they were rare or because conditions weren't favorable for them to be present. My research questions were ecological in nature, but relied heavily on mathematical and statistical models of data.

After finishing my doctorate I went on to do postdoctoral work with the U.S. Geological Survey's Patuxent Wildlife Research Center. There my work focused on quantifying the spatial structure of stream networks using graph theory and using that to assess variation the distribution and abundance of animals. The dendritic nature of stream networks has implications for our thinking about terrestrial environments as well because as habitat is lost to human activities, we see linear corridors develop in landscapes that are more likely to facilitate movement of animals between populations.

Following my postdoctoral work I had a short stint as a visiting faculty member at Antioch University New England, where I advised people on the use of statistics and maps in conservation research, and then I took a research position with the US Army Corps of Engineers. I started by applying my skills to research on conservation planning for Army installations, but as funding for programs shifted I began working with other researchers in two other areas: regularization methods for tensor decomposition and the development of map-based decision support systems.

I'm doing freelance work now, but I'm still curious about ways to tell stories with data. I think there's a lot of interesting work being done in different sports, and there seem to be several opportunities to apply analytics in interesting ways to hockey. As a fan, analyses like what I worked on for this presentation help me think more critically about how I watch games. While I'll continue to appreciate both the skill and competitive nature of players from across the league, analyzing data challenges me to think more critically about how I understand the game. I hope you'll find some of what I go through here to be interesting too.

The Essential Nature of Sports

It's a little dark, but it's not completely wrong. Each year when the Cup Final ends, 31 of the 32 fanbases are left with thoughts of what might have been rather than what was.

The Essential Nature of Analytics

I’m really not trying to be cynical here. We advance knowledge and understanding of topics when we probe the things that don’t make sense. When someone says they’ve figured everything out, they’re telling you they’re not interested in learning anything else. If you’re new to sports analytics and things aren’t making sense….well, get used to it. It’s part of the process. To paraphrase something Richard McElreath said more eloquently in one of his lectures, sometimes when you’re trying to learn something and it makes you confused, it’s just a sign that you’re paying attention.

Which Teams Are the Best?

Each season we see the end of the regular season, and then all that gets set aside and the playoffs start. Standings at the end of the regular season don’t seem to really matter in a consistent way beyond qualifying for the playoffs. Each year there are a few teams that are perceived as weak playoff teams or Cup contenders, but it’s hard to put numbers to things. We usually have some informed opinions about how teams might do, but we frequently see surprises once things get underway.

Some people have put a lot of careful thought into how to assess individual players, and assessing player skill can be useful. There’s a lot going on in the world, and, tragically, I will not have time to watch the 51+ hours of hockey from all 51 games scheduled this week! My deficit of games missed will only continue to mushroom as season goes along. Analytics help me cope with the ongoing frustration of not being able to watch ALL of the hockey that’s out there.

In the past there’s been some tension between “analytics people” and “hockey people”. I think it’s fair to say that this debate has been mostly settled by acknowledging that there are different domains of knowledge that are very useful for thinking about the sport, and there’s benefit in trying to understand where things can be quantified that represent tangible things related to hockey knowledge. For me personally, this conflict was always a little odd. In my previous professional life as an ecologist, I worked in labs where lots of people who were ecologists collaborated with statisticians and mathematicians. There were a lot of very productive relationships between people who had a thorough understanding of something like a rare butterfly, and people who could work out the math to model how to estimate their populations. Careers were made by resolving issues where models had unreasonable assumptions or fell short of what was needed in other ways to describe natural phenomena. Whatever your level of comfort is with hockey analytics, I hope it can be something that provides another way to go about understanding the game, and everyone should recognize that feedback will result in changes in the scope and focus of analytics as time goes on.

While there’s been a decent bit of work done with individual players, there’s less being done with publicly available data to evaluate teams. Frequently this is limited to quantifying shot quantity or shot quality. However, there’s other play by play data that’s available to us. It gets used for player metrics. In theory, it could be applied to teams as well. Developing models for team performance would be a good complement to our knowledge about individual players from other analyses.

The player-focused analyses are incredibly valuable, however, their utility in playoffs can be a little hit and miss. Sure there are people like Leon Draisaitl who are obvious talents, and it’s not surprising to see someone like that continue to find success in playoffs when they’re essentially playing on one leg. However, we also see instances where teams deploy particular styles of play that can really disrupt the typical game plan for an opponent. It’d be great to see the sport continue to develop perspectives on a range of scales, including analyses of individual players, forward lines, defensive pairs, 5-man units on the ice, and teams as a whole.

A Different Approach

Here I'm proposing a different way to think about using publicly available data. Rather than focusing on how individual players perform, we can examine team performance. We can do this in two stages. First, we can try to quantify the way teams play by looking at events recorded on the ice during games. Then we can try to link the variation in those events to team performance. If we do this on a continuous basis, fitting models to new data as it becomes available, we should be able to detect when new approaches begin to be used that give teams advantages.

Let's take a moment to consider the public data available to us. Game event data includes things like shots on goal, blocked shots, missed shots, giveaways, takeaways, faceoffs, hits, and penalties in addition to goals. It also identifies line changes, and it tells us who's in goal when an event happens. We know that there are lots of different strategies and tactics used by different teams. A 1-3-1 might be used to try to choke off an opponent's transition from the defensive zone to the offensive zone. Some teams might generate more chances off of rush plays. Other teams might emphasize puck possession within the offensive zone. All of these various approaches might result in particular patterns of events happening on the ice.

I took the play by play data from the 2021-2022 NHL regular season and applied principal components analysis to try to identify patterns in game event data. Principal components analysis is a method for reducing the dimensionality or complexity of a dataset. This sort of approach is commonly used in ecology to describe the variation in what species occur in different habitats. Suppose someone is working with plants here in Washington. There could literally be hundreds of species! It would be tedious to try to describe what grows in a particular place by running down a list of everything that occurs there. However, techniques like principal components analysis can help uncover underlying gradients that are associated with where plants grow. Instead of listing all of the species at a site, one might simply observe characteristics about a site that are associated with it. Things like elevation, slope, and aspect actually do a pretty good job of describing the world from a perspective of a plant, and if you told a plant ecologist you were at a low elevation forest in a ravine flowing west towards the Pacific Ocean in the Olympic peninsula they would have a good understanding of what that meant.

The goal with describing hockey is something similar. It would be tedious to recount every shift of a game. It's more efficient to talk about the styles of play teams used and how those translate into scoring chances. The first part of this analysis is a preliminary attempt to put some numbers on the sorts of things we might read about in a recap of a game or hear during an intermission report. We want to put numbers to this because if we can do that successfully it could help us see what's associated with teams scoring goals and winning games.

The second part of this is fitting models to the game event data to predict the margin of victory for the home team. I really feel like this sort of approach is undervalued. Predicting the margin of victory rather than just a win or a loss means that you can evaluate models more critically. It's a slight hedge against the risk of being right for the wrong reasons. This type of model considers the scores of two teams in a game, and uses that to estimate the offensive and defensive strength of teams as well as the home ice advantage. The model requires some assessment of initial team strengths, and then allows for random variation as the season progresses. This gets around the challenge of trying to do simulations based off of roster and then wondering what to make of things when someone gets injured for a month, or gets moved at the trade deadline, or a coach gets fired. Initially it might seem weird to be starting with a model that just considers game scores, but I'll point out that this sort of thing has been used in other situations – consider the Pythagorean winning percentage developed by Bill James and the Elo rating system developed for chess. Philosophically it makes sense to have a simple model to compare to more complex models, and this sort of thing has a new wrinkle with a Bayesian approach, but the general concept is pretty well established.

How Do Teams Play?

The principal components analysis of the event data was disappointing, but not really surprising. Typically when people use these sorts of techniques with ecological data, they might capture 70-80% of the variation within a dataset in the first 3 components. Here that number was closer to 15%. The red vectors on the plot of the first two components show the direction in which things are increasing. For example, moving right along the x-axis means that we see games that have more shots and missed shots for the home team. Games in the lower left quadrant of the graph would have more penalties for both the home and away teams. This is sort of interesting, but with so little of the variation in the data explained, it wouldn't be very useful in an analysis. Put another way, this means that the game events don't have a strong correlation with one another **in their counts alone**. However, we could look at the timing, location, and players associated with events to try to better quantify patterns in game event data. More on that later, but for now just understand that someone will need to come back to stir things again and see if clearer results can be obtained.

Score Differences Look Better

I don't have time to get into the details of the Bayesian workflow in the 5-minute talk, so I want to highlight the general process here briefly. We begin by thinking about a type of statistical model to capture the phenomenon we're trying to understand. This class of statistical models was intended to predict game outcomes where scores of two teams might not necessarily be independent of one another. We compile data showing the home and away teams and their scores for each game. We then fit a model using Markov Chain Monte Carlo methods. In this case I'm using Stan in R. After a model is fit to the data, we examine output to see if there are reasons to think the estimates are unreliable. We can assess model fit through a few different metrics, and if we wanted to compare models and generate averaged predictions from a set of models there's a process for doing that. In this case I kept things simple fitting the same types of models that are in the footBayes R package to the 2021-2022 regular season game data. I'll have a more technical summary of things later.

The general results from predictions about the score differences look reasonably good. We see a yellow prediction interval that covers 95% of the predictions from the model. In theory this would cover about 95% of the observed data, but in this case it's more like 74%. When we examine the figure carefully we see that blowouts aren't really captured very well by the model. There's also a slight upward trend in the prediction interval, but it's not as much as we'd like. Ideally the blue lines would be near the center of the yellow across the whole figure.

Projected Points

Once we've predicted the margin of victory, we can move on to predict the point projections for teams across the season. The colors here are as before: yellow projections from the model, which are expected to contain the blue lines representing the observed data. You'll notice the scale on the y-axis seems suspicious. This is because I used a function developed for soccer league data where regular wins result in 3 points. This means the total possible points are 246 instead of 164. However, if we rescale things by points percentage, the best teams look to be finishing with around 115 – 120 points, which is in the right ballpark for how teams like Florida, Carolina, and Colorado did last year. I'll have more details on this in a more technical writeup later.

Problem: Goodhart's Law

This concept, "when a measure becomes a target it ceases to be a good measure" was introduced to me as Goodhart's Law, but it shows up in different fields, and the true origin of it is unclear to me. I think this represents what is perhaps one of the most significant challenges for the appropriate use of analytics. It would be easy to see that a particular way of playing seems to result in wins and then seek to keep emulating that approach going forward. For example, a team might recognize that teams with a favorable share of 5 on 5 shot attempts have had good success in recent years. However, if this is how a team decides to optimize their style of play, it may be possible that they'll eventually find it difficult to enjoy continued success against teams that focus on shot quality rather than shot quantity.

What can we do? Well, the approach I've outlined here should be a good hedge against this. First, we ask the data about patterns of game event data. Then we ask the data about the association between those patterns and the margin of victory. By continually updating these types of models we could detect when a change in style of play produces competitive advantages for teams. We're not picking one thing and sticking with it indefinitely; we're continually seeing what the data can tell us. This sort of thing is potentially a source for spicy takes and animated discussions, but the point of doing data driven decision making is to actually consider the data, not simply adopt analyses that favor whatever the current conventional wisdom might be.

Too Many Men Tribute

In honor of a certain podcast I wanted to close with a short prospectus for further work in this area. In the interest of having something accessible for hockey fans of all ages I've modified the usual 3-letter acronym to TEB for Trade, Extend, and Buyout. If you're trading for a player you might anticipate immediate benefits and you're excited to see how this will fix some roster issue that's been plaguing your team. This is immediate fun. Whether it works well in the long run with the draft picks you've given up…..who knows, that's not the point. Extend refers to contract extensions. When you're signing your star to that long-term deal, it's a commitment, especially if the player get trade protection. On balance you think this should be good, but there are always ups and downs to work through in any long-term relationship. Finally, the buyout. This brings to mind the Concorde fallacy about not tying yourself to a situation where there are costs that have already been incurred and can't be recovered. If you're going nowhere fast with a player who's not meeting expectations anymore, sometimes this is the best move to make.

Too Many Men Tribute – Trade

Tensors are an interesting option for analyzing hockey play by play data. A tensor provides flexibility to handle varying levels of complexity in data. Tamara Kolder has done some great work providing overviews of this sort of thing, and it's a really promising area. A quick search on Google Scholar will turn up some things, and I'll say more about this when I have a more technical writeup of what I did. One thing that might be interesting would be to combine information from shot pressure charts with the game event data (see https://hockeyviz.com/howto/tide).

Too Many Men Tribute - Extend

This class of statistical models that was developed for soccer data can be refined by incorporating variables that actually quantify something intelligible and meaningful about the variation in the data. In the long run, having some sort of time series analysis of team performance feels like the right sort of thing to focus on. Again, I'll say more about this in a more technical writeup.

Too Many Men – Buyout

Finally, the principal components analysis, as noted before, is not committed to playing this game of predicting game outcomes the right way. We thank it for its hard work here, and wish it well in other areas. It may enjoy renewed success with a change of scenery.

Questions? Comments? Feedback?

Wow, did you really read through all of this? You probably deserve something for your trouble. I hope this tea-fueled brain dump makes some sense and provides some broader context to the talk itself. I've tried to make things more accessible to a general audience, but I'm sure there are some things that may remain unclear. Feel free to reach out to me via email at wfields7 at gmail dot com, or on Mastodon at @openfields@mastodon.skrimmage.com. I'll post more things about this under my github account (openfields) sometime in the next few weeks. If you're interested in hockey, data visualization, statistical analysis, machine learning, data science, data engineering, reproducible research, version control, or DevOps practices….I'm happy to nerd out with you, and I always appreciate the opportunity to learn from people with diverse perspectives that don't neatly align with my own. Enjoy the hockey!