Mathematics and Sports

Predictive Modeling and Analysis of Hockey Using Markov Chains

Harvey Campos-Chavez[1], Will deBolt[1], Miles Mena[1], Jacob Prince[1], Alia Alramahi[2], Robert Dudzinski[2], Soren Thrawl[1], Anthony DeLegge[2], Amanda Harsy[1,*]

[1]Lewis University, Romeoville, IL, USA, [2]Benedictine University, Lisle, IL, USA

[*]Corresponding Author E-mail:
harsyram@lewisu.edu

Abstract

The fast-paced, volatile nature of hockey makes it a challenging sport to analyze and predict the final outcome. This paper presents two continuous-time Markov process models that predict the probability that a team will win a hockey game given particular states during the game. These states incorporate shot and goal differential relative to the opposing team and are used to approximate the probability that the home team would win depending on the state they are currently in at a given time in the game. We also provide several examples of using this model to predict National Hockey League games.

Keywords: sports analytics, predictive modeling, Markov chains, hockey

1 Introduction

Ice hockey, like basketball, is a very action-packed, entertaining sport with multiple variables that change fluidly as the game progresses. These variables must be considered when attempting to model any potential outcomes. Any sort of sport predicting model is difficult to develop, mainly due to the unpredictability of human nature; however, hockey extends these common difficulties with its own distinctive specialties: the penalties that result in the loss of a player, the high intensity and physical nature each team endures, the continuous substitution of players during play, and the fact that the game is played on ice.

While some measurements exist to help quantify a player's impact on the game such as the plus/minus rating (number of goals scored by that player's team minus the number of goals scored by the opposing team while the indicated player is on the ice), they tend not to be very predictive as a whole for how a player or team will do from game to game. As Shea simply puts it, "currently available hockey statistics do not respect the complexity of the game" [15]. Shea furthers his point as he describes the current state of hockey analytics for organizations as stuck in a paradox [15]. The data available is extremely limited in scope which leads to research and results that do not impress the big organizations, which in turn prevents them from funding new methods of recording data (i.e. spatial tracking).

Nevertheless, at the professional (National Hockey League, or NHL) level, there are data repositories like the website created by Peter Tanner called `www.MoneyPuck.com` that provide offensive and defensive statistics on every player and every team over the course of a season. Some of these statistics are relatively simple like the number of goals scored or the number of penalty minutes, and some are more complicated like the number of shots taken classified by the probability of leading to a goal and the Corsi number (the percentage of face-offs won by player's team while that player was on the ice).

Although advanced models are limited, they can still provide a great foundation upon which different problems can be explored with the given data. In 2011, Wang and Zhang sought to create their own spatial tracking system using mixture hidden Markov models for video analysis of different events during games [18]. In 2015, Pettigrew sought to create a new statistic that would be more useful than the outdated and limiting plus/minus, Corsi, and Fenwick systems [14]. He developed a new win probability metric that was used to find a player's individual effect on a game through their scoring which Pettigrew referred to as their Added Goal Value. In 2019, Czuzoj et al. created an expected events model for different players' face-off win percentages [2]. More recently, in 2020, Kari used machine learning and neural networks to calculate the effects of random chance on different events in a game [6]. These models are great explorations into what is currently possible in hockey analytics, but our specific focus for this part of our research will be the models that were derived from earlier Poisson models.

In 1977, Mullet first employed the use of a Poisson distribution with regards to determining the number of goals scored in an NHL game [12]. Since then, there have been multiple Poisson models that have built upon Mullet's work to implement hockey analytics to optimize and determine the likelihood of winning. For example, one characteristic of hockey is that when a player on a team commits a penalty, they are momentarily taken out of play. This can result in more effective scoring opportunities as it gives the opposing team a player advantage. This is referred to as a *power play*. Teams also have the option to remove (pull) the goalie from play in exchange for an extra skater. This can be used to increase scoring opportunities, especially late in a close game, at the risk of losing by a larger margin. Later, Morrison and Wheat found that the functions used to determine the optimal time to pull the goalie depend on the number of minutes left in the game before the coach pulls the goalie assuming there are 5 attackers and the number of minutes remaining once the goalie is pulled and there are then 6 attackers [11].

To expand upon the model, in 1987, Erkut pointed out that an unfair assumption was made by Morrison and Wheat. [11] They assumed that "both teams have their goalies on the ice" and "both teams are of equal ability and score at a constant rate of L per minute." To improve the model with its limiting assumptions, Erkut included a weight for each variable [3]. Then, in 1989, follow-up work by Nydick, Weiss, and Morrison highlighted the effects of giving values of scoring rates depending on the team and whether or not they decide to pull a goalie to determine the optimal time for a goalie to be pulled during a normal even strength scenario (5-*on*-5) [13]. With these probabilities, the model would determine the likelihood of the team winning. Washburn's 1991 work built upon the Poisson process in order to create a state-spaced model for goal differential and time remaining [19].

Markov chains are another example of linear algebra models used in hockey analytics. Markov chains can be used to build a stochastic model describing a sequence of possible events in which the probability of every event depends solely on the state it was previously in. Many existing sports models apply Markov chain methods to the predictive modeling of baseball, basketball, and football [1, 4, 7, 8, 16]. One model developed by Paul Kvam and Joel Sokol used a combination of logistic regression, Markov chains, and steady state vectors to rank NCAA Basketball teams [8]. In 2014, Kaplan et al. built upon Washburn's work in [19], and developed a model utilizing Markov states dependent upon how many goals the home team has relative to the opposing team (goal differential) and how many players they have on the ice relative to the opposing team (manpower differential) [5]. Due to the nature of hockey, there are many instances that there would be a differing amount of players on ice per team. This is known as *manpower*. Utilizing a similar process of a Markov chain-based model, they found a win probability graph that considered both a manpower differential of ± 2 (the difference between the number of players on the ice for the home team against those on the ice for the away team) and a goal differential of ± 4 (the difference between goals for the home team against goals given up by the home team). Markov chain-based models are solely dependent on the state they were previously in. This allows us to model a win probability graph utilizing these features based on the number of times a certain state changes to another over the time spent in the previous state. In our project, we decided to explore two different types of states: shots on goal differential's impact on the win probability of a team and unique manpower scenarios' impact on the win probability of a team.

In this research, we attempt to estimate the probability winning a game of professional (NHL) ice hockey by building upon Kaplan et al.'s work in [5]. Our first model incorporates shots on goal and goal differentials in order to give a win percentage based upon the state of the home team at any point in a regular season game. Our second model uses unique manpower scenarios and goal differential in order to give a new win percentage based upon the state of the home team at any point in a regular season game. This model splits Kaplan et al.'s manpower parameters into more descriptive states. Thus, we can compare the different approaches to discover the strengths and weaknesses of each approach.

2 Methods

Our models utilized Markov processes: state-based stochastic models that describe the sequences of possible events where the probability of each event depends solely on the state at which it was previously in. With a source of game data, the probability of switching between specific state conditions can be found.

2.1 Model 1: Shots On Goal Transition Rates

Our first approach focused on the relationship between shots on goal and goals. In hockey, a *shot on goal* is a shot that would have gone into the net unless the goalie made a save. For example, if the puck hits the crossbar, it would not count as a shot on goal. In order to replicate Kaplan's original model [5], this meant we needed to create our states in the form (x,y,t), where x and y represent two variables that will be used to estimate the probability of winning. In our model, relative to the home team, the x variable represents the shots on goal differential, in terms of the number of standard deviations. This means that the home team is up or down x amount of shot standard deviations. y represents the goal differential, meaning the home team is up or down y amount of goals. t represents the time in the game in seconds (a full game, not including overtime periods, is 60 minutes, or 3600 seconds). Our model deviates from Kaplan's original model as we analyzed shots on goal differential rather than manpower differential. In hockey, shots on goal occur often; in the 2021-2022 NHL season, each team averaged between 26-37 shots on goal per game, or roughly 1 shot every 2 minutes on average [17]. Because of these high numbers, the number of possible differentials would create too many states for our model, so we

divided the number of shots to be categorized as above or below one and two standard deviations of the average number of shots on goal. To do so, we used a data set of every shot taken in the 2013-2014 NHL season from www.moneypuck.com. From this data set, we found the average shot differential for each of the 30 NHL teams and computed the average and standard deviation of these. Thus, we found the average shot differential to be -0.003 and found the standard deviation to be about 3.5 shots. This meant that we would change standard deviation at the following shot differentials: $\{-7,-3.5,0,3.5,7\}$ and if we substituted for standard deviations, would get $\{-2,-1,0,1,2\}$. For example, if the home team found themselves with 7 or more shots on goal than the away team, the current state would be in the form $(2,y,t)$ in the model. Likewise, if the away team has 4 more shots on goal than the home team, the current state would be at $(-1,y,t)$ in the model.

In order to be consistent with Kaplan's original model [5], the y value is assigned to represent goal differential with a range of -4 to 4. To simplify our model, we made the assumption that a team's likelihood of winning will be roughly the same in score differentials greater than or equal to 4. So, a team that is up by 4 will have the same probability of winning as a team up by 6 with the same shots on goal differential.

With these parameters, we set out to identify all the possible states the home team could be in at any point of the game. To simplify the model, as was done in [5], we restricted the total goal differential for the following range $(\pm 2,\pm 4,t)$. This meant that x would have 5 possible states $\{-2,-1,0,1,2\}$ for shot differential and y would have 9 possible states $\{-4,-3,-2,-1,0,1,2,3,4\}$ for goal differential. We then have 45 unique combinations for possible states.

Next, we must find whenever there would be a change in a state, meaning a goal was scored or a new standard deviation of shots was reached. For example, if the home team is up by 4 shots and a goal, the state is $(1,1,t)$. Suppose the away team then breaks out into transition and gets a shot off that is trapped by the goaltender. The shot would bring down the shot differential past the standard deviation threshold, so the game will move to the new state $(0,1,t)$.

To help us narrow down which states a previous state could change to, we made some restrictive assumptions. First, we made some assumptions about impossible goal scenarios that we would not count as state changes. These were if a team scores two goals at the same time or if a team shoots a shot that results in a goal for the other team. Secondly, we had assumptions about the extremes of our parameters. For example, if a game's shot differential is at -2 or 2, or if a game's goal differential is -4 or 4, we assumed that the state could not go beyond those extremes, meaning a change will never occur to a state not included in our 45 possible options.

Note that not all changes in the shot differential results in a change of state. For example, if we assume the home team is currently up 9 shots, an x state of 2, and the away team gets a shot off, then the shot differential becomes 8, which is still an x state of 2. Despite this being a meaningful event, it did not cause a change in state since we considered a state change to occur at the standard deviation markers. The code used to filter this data so we could find events which resulted in state changes is available for further exploration in [10]. These states were then used to create our system of differential equations (see Equation 1 in Section 2.3).

2.2 Model 2: Incorporating Manpower Transition Rates

Our second approach sought to build upon the original base model's exploration into manpower's effect on win probability in [5]. However, Kaplan et al. used manpower differential as its main variable, which meant that it only looked at the difference between the number of men on the ice for both teams [5]. This meant that a 5-on-4 situation was treated the same as a 4-on-3 situation. To clarify, throughout this discussion, we will refer to manpower scenarios in the following form: the number of the home team men on ice *on* the number of the away team men on ice. We wanted to differentiate between each possible scenario, which means that, noting it is not possible to have less than 3 attackers or more than 6 attackers for either team by NHL rules, we would have 15 different manpower states:

6-*on*-5, 6-*on*-4, 6-*on*-3,

5-*on*-6, 5-*on*-5, 5-*on*-4, 5-*on*-3,

4-*on*-6, 4-*on*-5, 4-*on*-4, 4-*on*-3,

3-*on*-6, 3-*on*-5, 3-*on*-4, 3-*on*-3.

As with our "Shots on Goal" model, in order to be consistent with Kaplan's original model [5], the y value is assigned to represent goal differential with a range of -4 to 4. We will continue to make the assumption that a team's likelihood of winning will be roughly the same in score differentials greater than or equal to 4. So, a team that is up by 4 will have the same probability of winning as a team up by 6 with the same manpower differential.

With our parameters set, we again wrote the states in the form (x,y,t), where x represents our specific manpower scenario, y represents our goal differential, and t represents the time in the game. With 15 different possible manpower scenarios and the 9 possible goal differentials, we will now have 135 possible states. For example, if the home team is on a power play with a two-man advantage and down one goal, we would write this state as (5-*on*-3,-1,t).

To find our state changes for manpower, we needed to add all of the possible penalties to our Excel sheet of shots and goals. We accomplished this by taking the penalty summaries for each game from www.hockeyreference.com and adding the times and

manpower effects to our data. To try and prevent adding unnecessary data, we made the decision to use only the penalties that would affect manpower. So, we excluded any game misconduct penalties and fighting penalties, which normally do not cause man advantages. This caused some slight problems in very specific situations (a double minor or major penalty) where goals scored on a power play did not end the power play. Additionally, if a player is penalized while trying to stop a breakaway player on the opposing team, the other team may be granted a penalty shot instead of a power play. This involves the player getting one shot against the goalie in a 1-on-1 scenario. According to `www.hockeyreference.com`, in the 2018-2019 season, there were merely 43 penalty shots when compared to 7409 power play opportunities. With such a small portion of the penalties resulting in penalty shots, we did not want to truncate the model to fit these specific instances. Thus, for this model, a penalty shot goal is considered a shot to have occurred at the moment and manpower advantage of the initial stoppage of play.

With our final set of penalty data, we then went through to find the possible state changes for our differential equation model. We considered a state to change if its goal state or its manpower state changed. For example, a team can draw or cause a penalty, which would change their manpower state to $(x \pm 1, y, t)$. We made the assumption that these manpower changes can only change by moving one player at a time. For example, the home team can only move from a 3-*on*-5 to a 4-*on*-5 then a 5-*on*-5, but, not directly from a 3-*on*-5 to a 5-*on*-5. This is consistent with NHL rules; even if two players on the same team are penalized at the same time, one is designated to come out first. The other individual state change occurs when a team scores or gives up a goal $(x, y \pm 1, t)$. Like the shots on goal model, a team can only change its goal differential by one, as it is impossible to score two goals at the exact same time. We again made the assumption that a team's probability of winning will remain the same for goal differentials greater than or equal to 4. This was done as a simplifying measure to decrease the number of states. We provide some examples of state transitions in Table 1.

Previous State	Event	New State
(5-on-5,0,t)	home team scores a goal	(5-on-5,1,t)
(5-on-5,0,t)	away team scores a goal	(5-on-5,-1,t)
(5-on-5,0,t)	away team loses a player	(5-on-4,0,t)

Table 1: A table of examples of some state transitions.

The system of differential equations, discussed in Section 2.3, requires a count of the number of times one particular state move to another particular state, which we call λ. For example, we would need to count the number of times the game changed from (6-*on*-4,0,t) to (6-*on*-3,0,t). While writing a code to automate this process, there were some state changes that never occurred in this season, corresponding to a λ value of 0. Therefore, instead of having 135 equations that describe the probability of winning at each state, we found 107.

2.3 Differential Equations Used in Model

We followed the methods used by Kaplan et al. to set up our system of differential equations [5]. We then used the ODE45 differential equation solver in MATLAB to numerically solve this equation and plot the solution [9]. Each differential equation in the system takes the summation of a state's neighboring states along with either respected transition rate λ as their coefficients minus the initial state and summation of all the λ values in the first summation as the coefficient:

$$\frac{dw(x,y,t)}{dt} = \sum_{(x'y')\neq(x,y)} \lambda_{xy}^{x'y'} w(x',y',t) - \left(\sum_{(x'y')\neq(x,y)} \lambda_{xy}^{x'y'} \right) w(x,y,t) \tag{1}$$

$$x = -2,-1,0,1,2$$

$$y = -2,-1,0,1,2$$

$$0 < t \leq 3600$$

To find the transition rate λ, we use:

$$\lambda_s^{s'} = \frac{n(s,s')}{\tau(s)}, \tag{2}$$

where λ will represent the number of changes from an initial state to a neighboring state, $n(s,s')$, over the total amount of time spent in that initial state, $\tau(s)$.

For example, with the shots on goal model, suppose we use the initial state $(0,0,t)$. We must then find the λ values of $(0,1,t)$, $(1,0,t)$, $(-1,0,t)$, $(0,-1,t)$, $(1,1,t)$, and $(-1,-1,t)$ as those are the only possible states $(0,0,t)$ can change to. If the respective λ

values are .5, .25, .2, .3, .7, and .6, we can use our differential equation for our initial state $(0,0,t)$ as follows:

$$\begin{aligned}\frac{dw(0,0,t)}{dt} &= 0.5(0,1,t)+0.25(1,0,t)+0.2(-1,0,t)+0.3(0,-1,t)+\\ &\quad 0.7(1,1,t)+0.6(-1,-1,t)-(0.5+0.25+0.2+0.3+0.7+0.6)(0,0,t)\\ &= 0.5(0,1,t)+0.25(1,0,t)+0.2(-1,0,t)+0.3(0,-1,t)+0.7(1,1,t)+\\ &\quad 0.6(-1,-1,t)-2.55(0,0,t)\end{aligned}$$

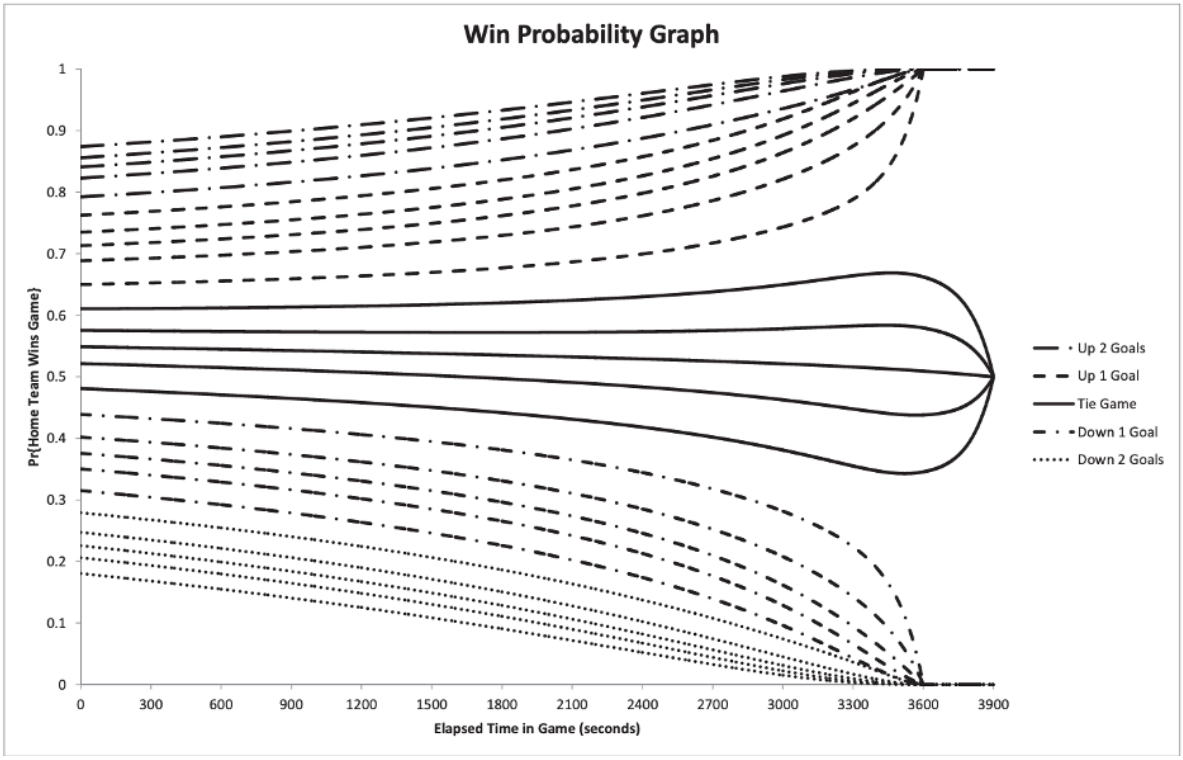

Figure 1: Sample graph from [5], Shows win probability based on goal differential (the different textured lines) and manpower differential (Different lines within the goal differential. From top to bottom up 2 players to down 2 players)

We do this for all possible initial states. We would then take all of our differential equations and use the ODE45 differential equation solver in MATLAB [9]. We can then create a graph similar to the one found by Kaplan et al. as shown in Figure 1 [5].

3 Results and Visualization

After solving our system of differential equations using ODE45, we plotted our solutions, which can be found in Figure 2 (Model 1) and Figure 3 (Model 2). Looking at Figure 2 for our shots on goal model, our graph has a strong correlation with Kaplan's model [5]. Each line represents one of the 45 equations and the probability of winning a team has at any point in time during the game. There are only 3 outcomes: the home team is up by a goal or more and wins (goes to 1), the home team is down by a goal or more and loses (goes to 0), or the match is tied and goes into overtime (goes to 0.5), and our graph clearly plots that. Ideally, we would like to add a legend to make this graph easier to interpret as there are many lines or format it in the way the base model describes the lines rather than each individually. Even so, we can see that states which have the home team starting with at least a 2-goal advantage over the away team start with a high probability of winning and continue towards the 100% probability of winning. States that start between 0.4 and 0.8 have a smaller margin of goal differential between the home and away team and states will either go towards a win, a loss, or a tie. Notice that there are more states starting with a probability of winning above 60% since states in which the games start at 0-0, reflect a slight home team advantage. In general, we did not find shots to be as impactful as goal differential changing the win probability of the home team which reflects similar results to Kaplan [5] (see Figure 1).

For our manpower model, shown in Figure 3, our graph becomes less correlated with Kaplan's model in [5]. We believe this to have occurred by the nature of the differences in the models. Our model breaks up the specific manpower differentials in the original model into separate occasions that will happen less individually than in the original combined state. This has created fewer occasions for the

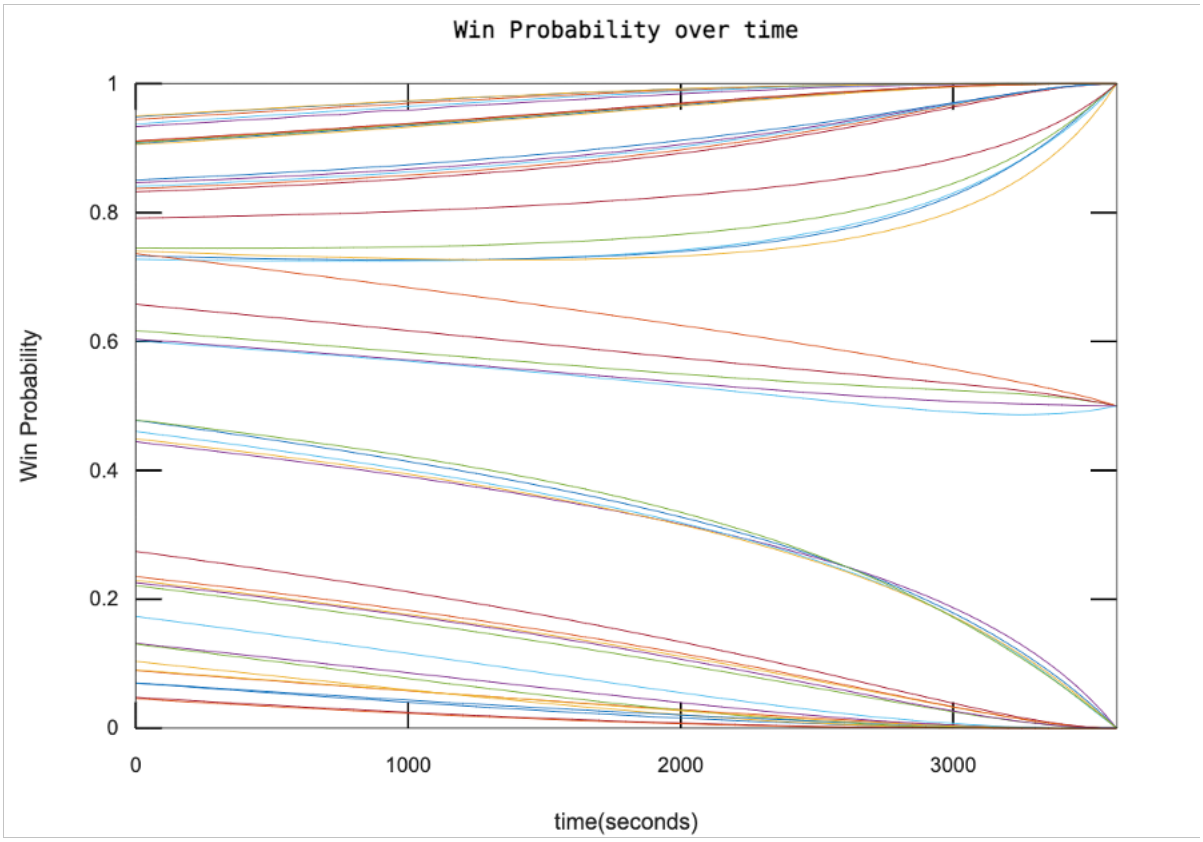

Figure 2: Shots on Goal Model Based on Figure 1 (without manpower)

state transitions, which in turn has caused the probabilities that the home team will win to cluster together into the three possible ending conditions: the home team is up a goal or more, the home team is tied with the away team, or the home team is down a goal or more. Again, a legend would be ideal, but with over 100 equations, we found it not very useful to include.

Using our predictive model, we can take a single game and find the probability that the home team wins at any given time. We take the state that the game is in with the time and find the probability that the Markov Model gives us. For example, for any given game, if it were in the state (5-*on*-5,1,t), the home team up one goal with both teams at normal manpower, at time 748 seconds, the probability that the home team wins is approximately 0.98444 according to our model.

By analyzing single games, we can test the effectiveness of each of our models. We used our models to analyze multiple games from 2013. Figures 4, 5, and 6 show the probability that Chicago beats Minnesota, Arizona beats Vancouver, and Pittsburgh beats Toronto, respectively, as the games progress. Notice that, when any team goes up by any amount of goals at any time in the game, the probability that this team wins is nearly 1 with the manpower model. In contrast, the shots on goal model doesn't weigh scores nearly as heavily. Take, for example, Figure 4, where the orange line is the shots on goal model and the blue line is the manpower model. Once Minnesota scores around time 750, our manpower model (in blue) predicts Minnesota will win, while our shots on goal model (in orange) still gives Chicago a greater chance of not losing. The reason for this is that penalties, and thus manpower state changes, are expected to occur much less frequently than shots on goal state changes; in the 2021-2022 NHL season, teams averaged between 3 and 5 penalties per game but between 26 and 37 shots on goal per game [17]. With more possible states in the manpower model, this means that state changes are more likely to occur with goals, thus goals are weighted higher in this model.

The other two games represented in Figure 5 and Figure 6 also demonstrate how heavily our manpower model weighed goals. In the game represented in Figure 6, Pittsburgh takes an early lead, which is quickly snatched by Toronto who, according to the graph, was sitting comfortably with a win. However, in the closing seconds, Pittsburgh rallies to tie the game. On the other hand, the game in Figure 5 demonstrates the effect of manpower situations in a game. During this game between Arizona and Vancouver, there were several specific manpower situations that can be seen after time 1000. These situation changes had a small but noticeable affect on the probability of the home team winning.

One of the motivations behind the parameters of the second model, specific manpower instead of manpower differential, is that the game of hockey is completely different with a 5 on 5 state as opposed to a 3 on 3 state. However, the model in Kaplan would predict both of these states as the same win probability. Our model was aiming to determine if there was a difference in the win probability for specific man powers at the same manpower differential. We applied our model to several games as shown in Figure 7, Figure 8, and

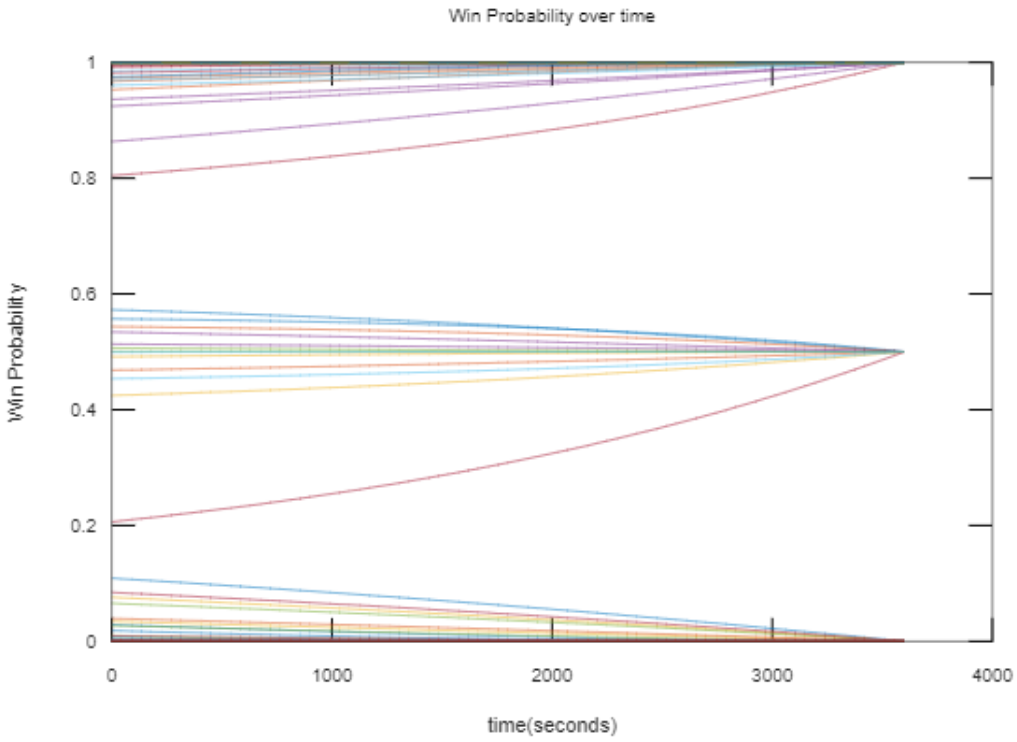

Figure 3: Manpower Scenario Model Based on Figure 1

Figure 9. These figures show some of our predictive curves based on states within the game which would have been treated as the same state in [5]. From the graphs, we see that while we can distinguish these states, the probability does not seem to change much more than 5% when the manpower differential is ± 1 (Figure 8). We also noticed not much of a difference when the manpower differential is 2. We believe this is due to the fact that these situations did not occur often enough to make a huge impact on the model (except for the $(0,5,6,t)$ state. Notice that the state $(0,6,4,t)$ is hardly distinguishable from $(0,5,3,t)$ in Figure 9). This seems to suggest that Kaplan's decision to combine these states was a good decision [5].

4 Future Work

This model was built using the 2013-2014 NHL season, so in the future, we would like to add data up through the 2019 season. Furthermore, we found a better database which would make adding all possible play by play events possible on `www.nhl.com`. This website provides in-depth play by play documentation that would need a program to process the data into a usable form. The first approach, which incorporates shot differential, may be improved by looking at the quality of the shots (like high danger shots - shots that have a 20% or higher probability of scoring). Furthermore, with the second approach (exploring different manpower situations), Kaplan's model often predicted similar states incorrectly by 5%-30% for a score differential of 0, but that these simplifications were necessary for the model to make sense [5]. Therefore, to make the second approach more accurate, we could combine similar states within a threshold of win probability into the same bin. For example, Figure 9 demonstrates that a manpower differential of 2 with a goal differential at 0 has the same home team win probability as a manpower differential of -2 with the same score, so we could combine those into one state. To do this, we would combine the count of times that these states occurred to find the shared λ value and combine the equations in the system of differential equations so that states that move to either $(0,6,4,t)$ or $(0,5,3,t)$ are computing the shared win probability for that state. Another extension could be to merge the two models to incorporate more states which depend on manpower, shots, and goal differential. Finally, adding an overtime factor would, just as the base model does, make our model more accurate and meaningful as not all games end in regulation.

5 Conclusion

In this paper, we provided two differential equation models which can be used to determine the probability that a team will win a hockey game at a given time in the game and particular game state. These states were based on goal differential, and the first approach

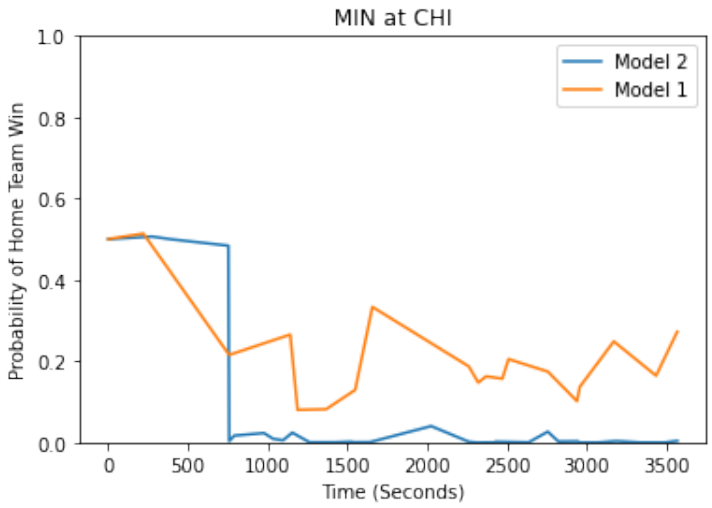

Figure 4: Both Models CHI vs MIN

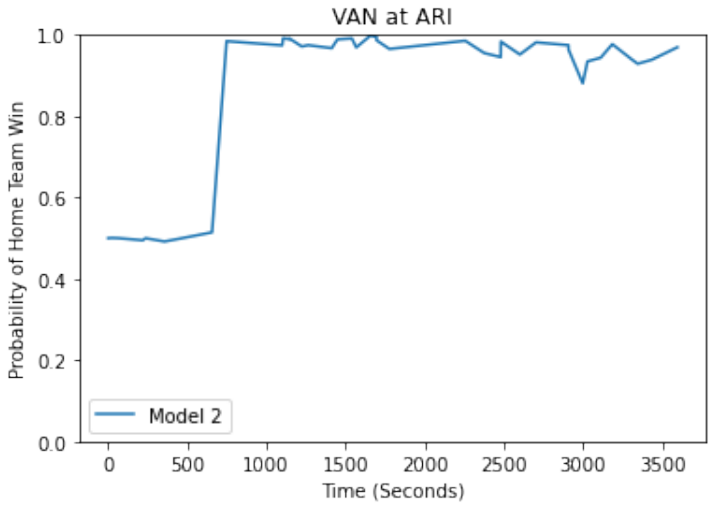

Figure 5: ARI vs VAN

incorporated shot differential while the second approach incorporated specific manpower states that occur during the game. Although our models were not able to perfectly predict the results of individual games, this is not surprising as hockey is a very fast-paced, volatile game that has not lent itself well to the same analytics as other sports (e.g., baseball). Thus, while there is still more work to be done to improve these models, our models seem to suggest that knowing a team has a manpower advantage does not significantly impact the probability that the advantaged team wins the game, but also knowing that a team has a shots on goal advantage can impact their probability of winning the game. This suggests that a team having a power play does not significantly impact the outcome of a game as one might think, especially given the manpower advantage and seemingly increased opportunities for scoring a goal. However, in the NHL, only about 20% of goals scored are during a power play, and each team scored a power play goal between 12% and 27% of the time they were on a power play during the 2021-2022 NHL season [17]. Thus, the vast majority of goals scored occurred when neither team had a manpower advantage, and even when a manpower advantage occurred, a goal was scored less than 30% of the time, so data from the NHL also seems to bear out that a manpower advantage does not necessarily significantly impact a game's outcome.

Naturally, however, those teams with a higher shots on goal differential are getting more opportunities to score, which would logically lead to more goals, so recognizing a team has a high advantage in shots on goal is definitely a good indicator of a potentially winning team. In fact, the top team of the 2021-2022 NHL season, the Florida Panthers, also had the highest number of shots-on-goal of all NHL teams [17], so the model demonstrating this correlates with what is observed in the NHL as well.

In each of the models, the variable with the most significant impact on the win probabilities is the goal differential, which is not surprising since it is the goal differential that determines the game-winner. Thus, of all of the quantities studied, those who want to predict mid-game who is likely to win should look at the goal differential above all else. While this may seem like an unsatisfying

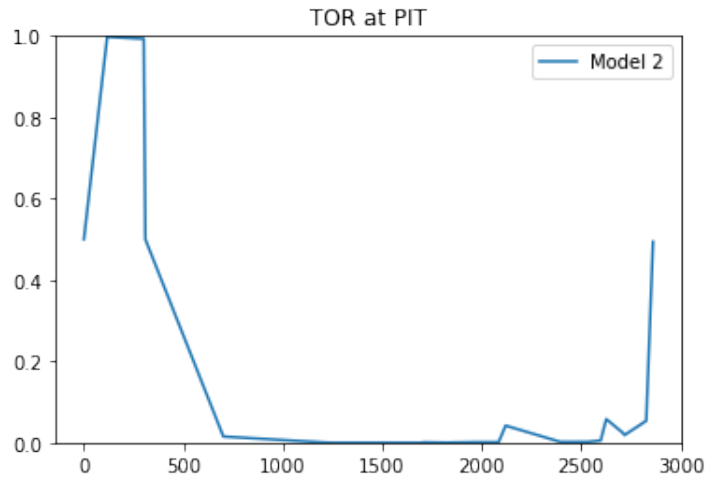

Figure 6: PIT vs TOR

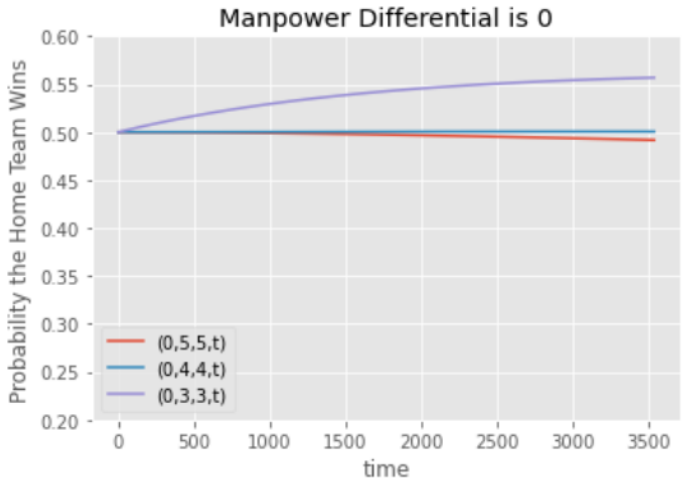

Figure 7: Manpower Differential of 0

overall result given the complexity of the game of hockey, concluding that arguably the simplest in-game statistic to keep track of is the most predictive of a winner means our model may be suggesting that analyzing hockey may not be as complex after all.

Acknowledgements

This project was funded in part through the Center for Undergraduate Research in Mathematics Minigrant (funded by the National Science Foundation DMS awards 0636648, 1148695, and 172256), The Promotion of Underrepresented Minorities in Academic STEM (PUMA-STEM) Alliance (supported by the National Science Foundation through the LSAMP Program under Award Number 1911271), and Lewis University's Summer Undergraduate Research Experience. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors are grateful to these organizations for their support of undergraduate research.

References

[1] Bruce Bukiet, Elliotte Rusty Harold, and José Luis Palacios. A markov chain approach to baseball. *Operations Research*, 45(1):14–23, 1997.

[2] Nick Czuzoj-Shulman, David Yu, Christopher Boucher, Luke Bornn, and Mehrsan Javan. Winning is not everything: A contextual analysis of hockey face-offs. *arXiv preprint arXiv:1902.02397*, 2019.

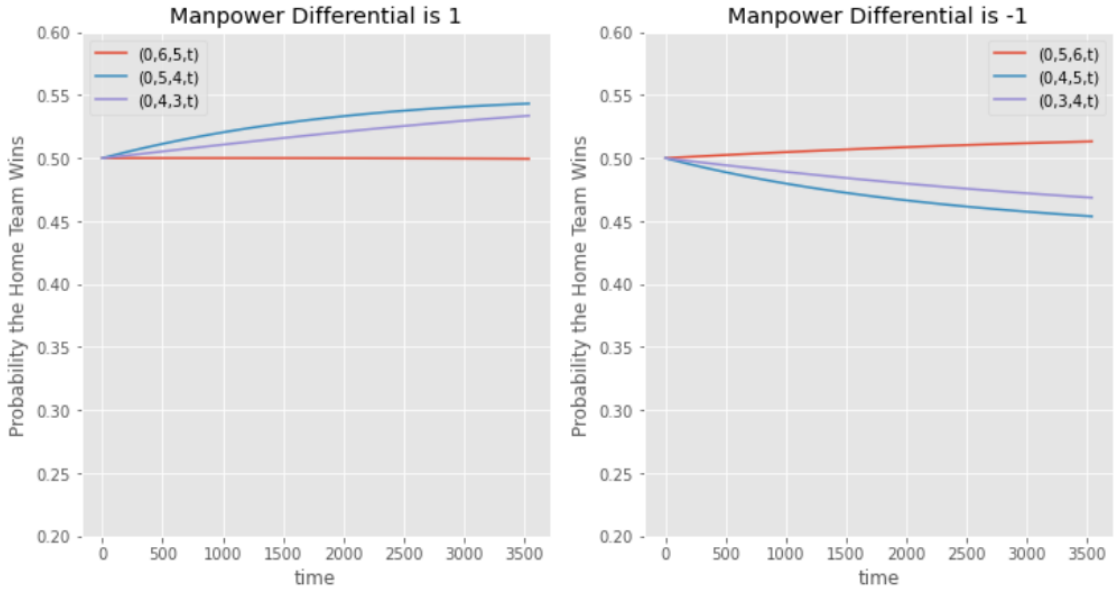

Figure 8: Manpower Differential of 1 and -1

[3] Erhan Erkut. Note: more on morrison and wheat's "pulling the goalie revisited". *Interfaces*, 17(5):121–123, 1987.

[4] Nobuyoshi Hirotsu and Mike Wright. A markov chain approach to optimal pinch hitting strategies in a designated hitter rule baseball game. *Journal of the Operations Research Society of Japan*, 46(3):353–371, 2003.

[5] Edward H Kaplan, Kevin Mongeon, and John T Ryan. A markov model for hockey: manpower differential and win probability added. *INFOR: Information Systems and Operational Research*, 52(2):39–50, 2014.

[6] Daniel Kari et al. Understanding how random chance affects the outcome of an ice hockey game. *HELDA - Digital Repository of the University of Helsinki*, 2020.

[7] Jason Kolbush and Joel Sokol. A logistic regression/markov chain model for american college football. *International Journal of Computer Science in Sport*, 16(3):185–196, 2017.

[8] Paul Kvam and Joel S Sokol. A logistic regression/markov chain model for ncaa basketball. *Naval Research Logistics (NrL)*, 53(8):788–803, 2006.

[9] MathWorks. Mathworks ode45 help center website.

[10] M. Mena, H. Campos Chavez, J. Prince, and W. deBolt. Milesmena hockey-research. `https://github.com/MilesMena/Hockey-Research`, 2022.

[11] Donald G Morrison and Rita D Wheat. Misapplications reviews: Pulling the goalie revisited. *Interfaces*, 16(6):28–34, 1986.

[12] Gary M. Mullet. Simeon poisson and the national hockey league. *The American Statistician*, 31(1):8–12, 1977.

[13] Robert L Nydick Jr and Howard J Weiss. More on erkut's "more on morrison and wheat's 'pulling the goalie revisited'". *Interfaces*, 19(5):45–48, 1989.

[14] Stephen Pettigrew. Assessing the offensive productivity of nhl players using in-game win probabilities. In *9th annual MIT sloan sports analytics conference*, volume 2, page 8, 2015.

[15] Stephen Shea and Christopher Baker. *Hockey Analytics: A Game-Changing Perspective*. Advanced Metrics, LLC, 2017.

[16] Zachary James Smith. *A Markov chain model for predicting major league baseball*. PhD thesis, University of Texas at Austin, 2016.

[17] StatMuse. Statmuse website.

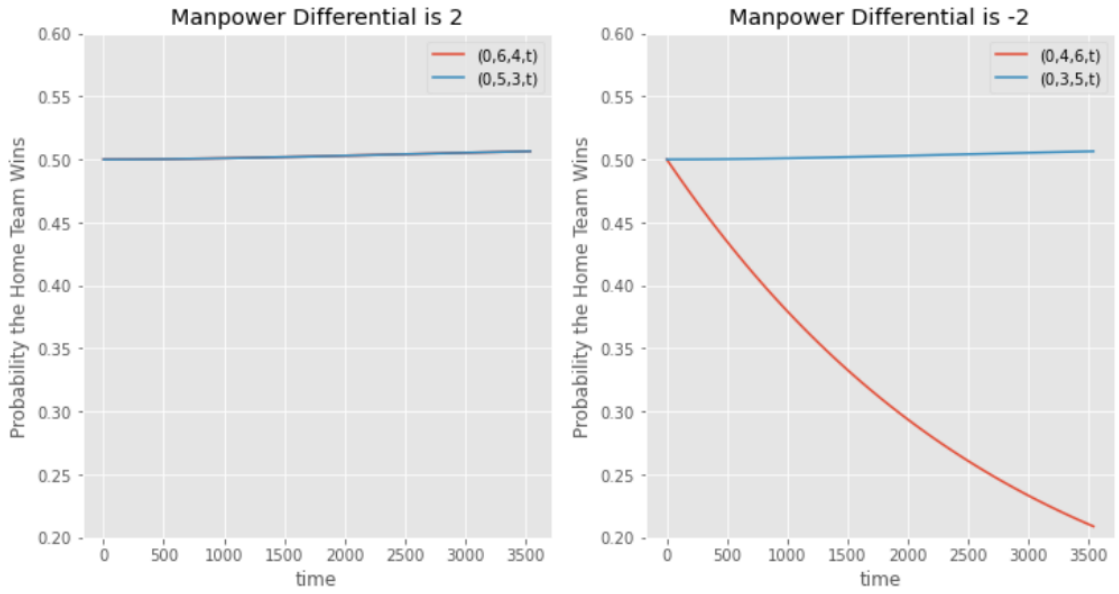

Figure 9: Manpower Differential of 2 and -2

[18] Xiaofeng Wang and Xiao-Ping Zhang. Ice hockey shooting event modeling with mixture hidden markov model. *Multimedia tools and applications*, 57(1):131–144, 2012.

[19] Alan Washburn. Still more on pulling the goalie. *Interfaces*, 21(2):59–64, 1991.