AN ANALYSIS OF THE GROWTH OF MAJOR LEAGUE SOCCER

By

DANIEL ESTEBAN MARULANDA

A SENIOR RESEARCH PAPER PRESENTED TO THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE OF STETSON UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE

STETSON UNIVERSITY

2012
ACKNOWLEDGMENTS

First of all I would like the entire Mathematics and Computer Science department including Dr. Hale and Dr. Williams. Throughout my four years of education, someone was always willing to help guide me through my concerns. Thank you all for taking the time to teach me the simple things that made my experience one that I will never forget. At this point, I would like to give special thanks to Dr. Miles and Dr. Coulter. Thank you for guiding me through a tough semester of research on a very interesting topic, but I mostly want to thank both of you for contemplating through a topic that is of great interest to me.

I would also like to thank Dr. Vogel. After taking my first theory class with him, I got the motivation to really step up my game and start doing better. All of the three classes I took with him were extremely challenging but I am glad I got to live the experience. Dr. Vogel, thank you for helping me “build character.”

I would have never been able to get through my education if it wasn’t for the special group of people in Elizabeth 205. If it wasn’t for everyone that joined me those numerous nights of losing sleep, I would have spent many hours doing the overrated activity of “sleeping,” instead of doing my work. I am glad I was able to share many safe and humorous moments with this group of friends. College would have not been the same without them. Thank you

Lastly, I would like to thank my family for giving me the opportunity to attend Stetson University. When times got rough and I might have not been doing as well as I wanted, I always had my parents and little brother to lift me up and give me advice. When I felt like it wasn’t worth it, I would always push myself that extra hour and say, “You have made it this far, you can’t give up now.” Thank you mom and dad for raising me with the best morals and teaching me that giving up is not an option and bro, thanks for being the best role model any big brother can ever ask for.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ............................................................ 2
LIST OF FIGURES ........................................................................ 4
ABSTRACT .................................................................................. 5

CHAPTERS

1. INTRODUCTION........................................................................ 6
   1.1. Background................................................................. 6
   1.2. Possible Factors That Affect The Growth Of The MLS........... 7
   1.3. Mathematical Model....................................................... 9

2. CONSTRUCTION AND ANALYSIS OF A PREDATOR-PREY MODEL
   2.1. Applying the Predator-Prey Model..................................... 10
   2.2. The Predator-Prey Model.................................................. 12
   2.3. Manipulating Our Model to Find Appropriate \( b_{MLS} \)......... 15

3. CONCLUSION/FURTHER ANALYSIS............................................ 16

APPENDIX .................................................................................. 18
REFERENCES ............................................................................. 20
LIST OF FIGURES

1. Demonstrates that the higher the number of sports teams within the state of the MLS team, the lower the stadium attendance of the team.

2. The behavior of the leagues when they do not interact with the population and decay at a rate equal to average death rate of the United States.

3. Revenue per team in the MLB from 2006 to 2011.

4. Revenue per team in the NFL from 2006 to 2011.

5. Revenue per team in the NHL from 2006 to 2011.

6. Revenue per team in the NBA from 2006 to 2011.

7. Model manipulated with $b_{MLS}=0.04$.

8. Model manipulated with $b_{MLS}=0.05$. 
ABSTRACT

AN ANALYSIS OF THE GROWTH OF MAJOR LEAGUE SOCCER

By

Daniel Esteban Marulanda

Fall 2012

Advisor: Dr. Lisa Coulter and Dr. William Miles

Department: Mathematics and Computer Science

Over the past few years, soccer’s popularity in the United States has grown significantly, but not enough to compete with such major sports as baseball, basketball, football and hockey. In this project, our goal is to investigate some of the possible variables that affect the growth of soccer in the United States. We will analyze factors such as high school sports participation, television coverage, Major League Soccer’s (MLS) stadium attendance, growth of competitive youth soccer clubs, aspects of the social environment related to soccer, and possibly other variables. Once correlations between these variables have been identified, a model will be constructed in terms of ordinary differential equations; more specifically a predator-prey model where the five leagues are the predators and the population of the United States is the prey. This model will explain how various factors can impact the popularity of the world’s most beloved sport within the United States.
CHAPTER 1
INTRODUCTION

1.1. BACKGROUND

In 1994, the FIFA World Cup was held in the United States. The U.S.A. had to play against top teams in the qualifying round such as Switzerland, Romania, and Colombia, who was ranked fourth in the world at the time. After beating Colombia and tying Switzerland, the United States qualified to the second round. Most importantly, this was the first time they had qualified to the next round after 64 years. Two years later, Major League Soccer (MLS) was established with ten teams, and it seemed like soccer’s popularity was starting to grow in the United States. In the first season, the stadium attendance for all of the teams in the league was 2,785,001. As the MLS grew so did the teams’ stadium attendance [1]. In 2011 the stadium attendance had grown to 5,468,849 and television rights had also increased by 15 percent compared to the previous year. Even though these figures seem like a great start for any league, other national sports leagues have significantly higher attendance. If we compare the MLS to the National Hockey League (NHL), which is now ranked as the fourth most popular league in the United States, the MLS’s stadium attendance does not come close to that of the NHL. For example, after reviewing data from ESPN stadium attendance, if we compare the worst total stadium attendance for the NHL in the past ten years which was 19,146,510 to the best total stadium attendance in the past ten years for the MLS which was 5,468,849, we can obviously see that there is a huge difference [7]. This shows us that the MLS has a long way to go in order to compare themselves with the popularity of the four major sports in the US. This is somewhat surprising because after the U.S. hosted the FIFA World Cup in 1994, soccer seemed to be exploding and many people thought that, by today, it would be one of the most beloved games in the United States. Some even thought that it would become as popular as the top three sports: baseball, basketball, and football, but as we can see this is not the case, and the popularity of soccer is far from being at the level of these top three sports. There are many variables affecting why the United States does not seem to be attracted to this game. In this paper, we take the time to analyze which variables affect the popularity of soccer in hopes of gaining insight into ways in which the popularity of the game
may be increased. Doing this could also help the MLS target the areas of marketing that would help the game grow. For these reasons, making a mathematical model that would explain why the United States does not gravitate towards soccer and especially the MLS is extremely valuable to soccer fans all over the country.

1.2 POSSIBLE FACTORS THAT AFFECT THE GROWTH OF THE MLS

Before we can start to construct our model, we need to figure out what factors might be effecting the popularity of the MLS. Some Americans might say, “Soccer is just too boring” or “Soccer has no contact,” but after talking to Alan Hancock, director of the Miami Fusion Marketing Program, one could conclude that there is a lot more to the picture. It actually has to do with “Getting your foot in.” In other words, the MLS just needs to be more exposed. It needs to start signing more television agreements. Also, there needs to be more research done about the cities in which the clubs are going to be established [10]. Just because a city has a large population of people from other soccer loving countries, that does not imply that the MLS team will be successful and that it will fill stadiums. When the MLS decides to establish another team, they also need to make sure that they select a city where the team will not have to compete with many other sports teams from other leagues. According to Hancock, “since, the MLS is still trying to become popular it would not be an intelligent idea to place a team where the population of the city would rather pay to watch the other sports that have been there for years” [10]. Instead, the MLS need to select a city where there are not many sports teams and this way they can have a higher probability of grabbing all the population that has not yet been taken by the other leagues. In order to support this proposition, we decided to check for any correlation between the total number of teams from other leagues in a state and the attendance of the MLS team in that same state. First we searched for the corresponding states of all of the MLS teams and made a list of how many sports teams were in their state. For example, for the state of California, there are a total of fifteen other sports teams, not including the three from the MLS. As we can see in Figure 1, we plotted the total number of teams as a function of the quotient of the teams’ attendance and the population of the state and noticed that when using a linear
fit there was a strong negative correlation between the two. The correlation coefficient, r, was 0.7398 and the r-squared value was 0.5473. This means that 54.73% of variance in the attendance is explained by the linear regression model [12].

![Graph](image)

Figure 1. Demonstrates that the higher the number of sports teams within the state of the MLS team, the lower the stadium attendance of the team.

Now to see how significant it was, by using a significance level of 0.01, and a one tailed test and a critical value of $t = 2.6245$ [12], we were able to say our correlation was significant because we observed a value of 2.8471 which is greater than the critical value.

In addition to the research that needs to be done, MLS teams need to play in a stadium that is the right size for them, or they just need to build their own. There are many teams that are currently playing in a stadium that is just too big and as a result, the fans that attend their games feel like they are wasting their time. According to Alan Hancock, marketing director of Miami Fusion FC, “The club was forced to play in a football stadium that was just too big. Even though the club was averaging a good stadium attendance when it was first established, once the fans started to feel like the environment was empty, they stopped coming back” [10]. Even with the rowdiest fans, if there are sixteen-seventeen thousand
fans (average stadium attendance from 2007 to 2011) in a stadium that holds seventy thousand for the NFL, the soccer environment will just seem empty.

Most importantly, according to Melissa Pitts, writer of Forbes Magazine, “MLS needs to keep young athletes engaged to recruit them here in the U.S., not send them off to play for another country” [9]. Pitts idea is obvious, but if the MLS concentrates a bit more on getting kids to stick to soccer, its popularity can grow significantly in the next couple of years. This might be a bit audacious to say, but it seems that every kid in America started out playing soccer, but for some reason after they started to become more athletic they decided to play another sport. Right now seems to be the time where the MLS needs to step into play and start to interact with more of these kids. According to Bet MLS Soccer, a news website for the MLS, “Among Americans age 12 -24, pro soccer now ranks as their second favorite sport (behind only the NFL)” (Bet MLS Soccer) [2]. This is huge news, and if the MLS does not take advantage of all of these new possible fans to increase their popularity, they will soon start to lose them to other sports leagues.

1.3 MATHEMATICAL MODEL

We now turn our attention to addressing the popularity of soccer from a more mathematical perspective. One possibly is to consider a compartment model similar to the model used by Mike Jones for his senior research: Modeling a potential spread of H5N1 Influenza for the United States [6]. His project seemed to have many similarities with the one we are trying to construct to model the growth of the MLS. First of all, both of the models would be constructed with boundaries inside the United States. Also, the idea that Mike Jones used of setting the initial spread of the H5N1 influenza in a region of the country and then letting this spread throughout the country as people migrated from one region to the other is also along the lines of how we could model our research because the popularity of the MLS could be spread by fans moving around the country [6]. His project seemed to be a great idea, but there were also certain parts that did not fit well with what our research aims to demonstrate. For example, in our
model when the population interacts with a league, we have assumed that they have become a fan of the 
league; while in the H5N1 model, when the population interacts with the disease, there are multiple stages 
the population can be at to spread the disease [6]. Most importantly, the H5N1 influenza paper only had 
one disease infecting the entire population while for our model we need to implement the fact that there 
are five powerful sports leagues trying to interact with the entire population, and the population is at risk 
of being affected by more than one league.

After analyzing the H5N1 influenza paper, we noticed that there were several concepts that could be 
applied to our research but we still had to introduce the idea that the population was interacting with more 
than one league. Technically, the five sports leagues would be competing for the entire population of the 
United States, and if they do not interact with much of the population, sooner or later they would die out 
because the other leagues would take advantage of their disability and steal their fans. In summary, all of 
these characteristics seem appropriate to set up our research with a predator-prey model. It would make 
sense to set up our model as a system of equations where each equation represents a league. Each one of 
the five dependent variables relies on the other because if one league grows it is taking possible fans away 
from the other leagues, which will cause that league to die out if it doesn’t interact with more of the 
population [4]. As a result, we will also have to find the appropriate value at which each league is 
interacting with the entire population to fully understand how the leagues are becoming more popular.

CHAPTER 2
CONSTRUCTION AND ANALYSIS OF A PREDATOR-PREY MODEL

2.1. APPLYING THE PREDATOR-PREY MODEL

To find the possible variables that explain why the MLS does not grow to be as popular as other 
sports leagues in America, we will analyze how the most popular leagues in the nation interact with some 
of their fans and how rapidly some of these leagues have increased or decreased in popularity over the 
past ten years. According to ESPN’s stadium attendance and the National Federation of High School
Sports Association these are the most popular sports in the United States: Major League Baseball (MLB), National Football League (NFL), National Hockey League (NHL), National Basketball League (NBA) and Major League Soccer (MLS) [8]. For that reason those are the leagues that will be analyzed in our model. Using the regular predator prey model found in Marco Di Francesco’s *Mathematical Models in Life Science*, we develop a model of how the leagues will grow or decay in the future using a Differential Predator-Prey Model, which is the system of equations (1) [4]. For our case, the predators will be each league and the popularity of the United States (all possible fans) will be the prey

\[
\frac{d(MLS)}{dt} = MLS(-a_{MLS} + b_{MLS}(Pop - (MLS + MLB + NFL + NHL + NBA)))
\]

\[
\frac{d(MLB)}{dt} = MLB(-a_{MLB} + b_{MLB}(Pop - (MLS + MLB + NFL + NHL + NBA)))
\]

\[
\frac{d(NFL)}{dt} = NFL(-a_{NFL} + b_{NFL}(Pop - (MLS + MLB + NFL + NHL + NBA)))
\]

\[
\frac{d(NHL)}{dt} = NHL(-a_{NHL} + b_{NHL}(Pop - (MLS + MLB + NFL + NHL + NBA)))
\]

\[
\frac{d(NBA)}{dt} = NBA(-a_{NBA} + b_{NBA}(Pop - (MLS + MLB + NFL + NHL + NBA)))
\]

where \(a_{MLS}, a_{MLB}, a_{NFL}, a_{NHL},\) and \(a_{NBA}\) are negative values indicating decay rates of each league and the \(b\)’s represent the strength of the interaction between a sport and the total fan base, including that of other sports. Notice that for our model, the coefficient values, \(a\), are negative because the leagues would decrease if it wasn’t for the fans attending some of their games. The dependent variables MLS, MLB, NFL, NHL, and NBA, represent the number of fans for each respective sport and \(Pop\) is the entire population of the United States, which we approximated to be around 300 million. For the fan variables we mention above, we will use the total stadium attendance of each league from the 2011 season.
2.2 THE PREDATOR-PREY MODEL

Using the representation we made for the predator-prey model, we need to search for possible growth and interaction coefficients to understand how the popularity of the MLS will behave in the next couple of years. Before we can manipulate our model and search for possible values for growth and interaction, we have to make a couple of reasonable assumptions. First we have assumed that the number of fans for each league (MLB, NFL, NHL, NBA, and MLS) is the total stadium attendance of the 2011 season. We know that this number is a bit off because not all fans have an equal opportunity to attend the stadium, but this seems to be a decent place to start. Choosing the correct values for $a$ and $b$ is not so easy. Initially, we have decided to set our decay values of $a$, for all sports, equal to the average death rate of the United States, which was found in the Census Bureau, multiplied by the quotient of the stadium attendance of each league and the total population of the United States [3]. Doing this allows us to mutually normalize each decay rate $a$. If we assume that the league will not interact with anymore fans, then $b_{\text{league}}=0$. Hence, we should see that all of the leagues decay exponentially towards extinction. Notice that this is exactly what Figure 2 displays.

![Diagram](image)

**Figure 2.** The behavior of the leagues when they do not interact with the population and decay at a rate equal to average death rate of the United States.
Now that we can observe how the model would behave if the leagues did not interact with the population, we need to complicate the model and add the interaction coefficients to make it more realistic. Here is where the research actually starts to become interesting and a bit more intense because there are many unknown variables that affect the ultimate value of $b_i$ which make these coefficients difficult to determine. A decent guess would be to measure the interaction (popularity) between the league and the population by analyzing the rate at which its total yearly revenue is increasing or decreasing over time. We were able to find the yearly revenue of the following four leagues: MLB, NFL, NHL, and NBA from the 2006 to 2011 seasons; however, we could not find the revenue of the MLS [13]. Once we found the revenue of each league we took the quotient of the yearly revenue and the number of teams in the league to get the approximate revenue per team.

After finding these results for the four leagues, we plotted the revenue per team as a function of time from 2006 to 2011. Then we added an exponential trend line to the plot of each league and used the exponential coefficient as the interaction values between the leagues and their fans, which are represented by $b_{MLB}, b_{NFL}, b_{NHL}$ and $b_{NBA}$. But there is now a problem: we have values for all of the interaction values except for that of the MLS. This is a huge problem because in order to successfully model our project we need initial values for all of our constants.
Figure 3. Revenue per team in the MLB from 2006 to 2011.

\[ y = 2E^{31}e^{0.0446x} \]

Figure 4. Revenue per team in the NFL from 2006 to 2011.

\[ y = 7E^{44}e^{0.0591x} \]

Figure 5. Revenue per team in the NHL from 2006 to 2011.
2.3 MANIPULATING OUR MODEL TO FIND AN APPROPRIATE $b_{MLS}$

There are several ways of attacking the problem of not having a value for $b_{MLS}$. The most straightforward procedure is to compare all of the $b$ values that we have found from the other leagues and conclude a possible range for the value of $b_{MLS}$. Since the MLB, NFL, NHL and the NBA have the following $b$ values 0.0446, 0.0596, 0.0468, and .0287 respectively, we can observe that an accurate range is somewhere between 0.01 and 0.1; but we are not sure of where exactly. By running our Matlab file and not manipulating anything besides the $b_{MLS}$ value from 0.01 to 0.1, we can observe what value makes a significant difference in our model. According to our results, for $b_{MLS}$ from 0.01 to 0.03, the popularity of the MLS decreases to zero. On the other hand, as we can see, in Figures 7 and 8, when $b_{MLS}$ is between 0.04 and 0.05 the popularity of the MLS increases significantly. If the interaction value were closer to 0.05 than 0.04, the popularity would even start to become larger than the MLB, NHL, and the NBA in the next twenty-five years.
Figure 7. Model manipulated with $b_{MLS} = 0.04$.

Figure 8. Model manipulated with $b_{MLS} = 0.05$.

CHAPTER 3

FURTHER ANALYSIS /CONCLUSION

At this point, all of the values for our coefficients have been estimated, and we still do not have a precise estimation for $b_{MLS}$. Until now, we have assumed that the death rate of the United States has been a proper way to measure the growth coefficients $a$, but for future work we need to find a way to test if these values are accurate for our model. Also, it seems that we have found a reasonable way to interpret the interaction values of each league so for future work we need to more closely examine the range of 0.04 to 0.05 to find a more exact value of interaction. Hopefully, we can find accurate estimations for the model parameters because finding them will be key to understanding the primary factors that influence the popularity of MLS.

Even though our predator-prey model seems to explain the behavior of the leagues for the next couple of years, for future research we need to add another equation that explains the behavior of the population. Just like any other predator-prey model, we need to include the equation for the prey and at this point we have only been working with the equations for all of the five leagues (predators). Adding an
equation that explains how the population of the United States behaves is essential to our model because this equation will let us analyze the possible fans that have not yet interacted with a league. We hope that using the equation below will help us implement the model and make it more realistic.

\[
\frac{d(\text{Pop})}{dt} = \text{Pop}(d_{\text{pop}} - c_{\text{MLS}}\text{MLS} - c_{\text{MLB}}\text{MLB} - c_{\text{NFL}}\text{NFL} - c_{\text{NHL}}\text{NHL} - c_{\text{NBA}}\text{NBA}) \tag{2}
\]

In equation (2), the \( d_{\text{pop}} \) value represents the rate at which the population is growing and the \( c \)'s represent the rate at which the fans of each league’s fans are being taken by other leagues. Most importantly, for next semester, we wish to work on the following three improvements: First, we need to find a way to model how some fans can interact with more than one league. Secondly, we need to find data for the MLS revenue using accurate numbers from Nielsen Ratings. Lastly, we would like to model the spread of each league with the discovered growth and interaction coefficients.
Just like any other coding program, MATLAB lets us initialize our variables before we can run the program. After we initialize all of our variables, we will use the built-in MATLAB command ode45, which is used to solve the Ordinary Differential Equations predator-prey model’s system of equations (1). The system of equations is then called from the process.m file and executed by the ode45 command, which is then plotted.

```matlab
function dydt = predatorpreymodel(t,y)

%Total population size
M=3000;

%The growth rate, a, of the individual leagues. We calculated the growth rates to be the Total Stadium Attendance/Percentage of total United States Population
a1 = 8.419*.001253;
a2 = 8.419*.003087;
a3 = 8.419*.007185;
a4 = 8.419*.001746;
a5 = 8.419*.001727;

%The interaction coefficients, b, of the individual leagues. We calculated them to be the Total Revenue/Total Number of teams in the league
b1 = 0.05;
b2 = 0.0446;
b3 = 0.0596;
b4 = 0.0468;
b5 = 0.0287;

%System of equations for the Predator-Prey Model. Each equation represents the popularity of each league with respect to time. MLS, MLB, MFL, NHL, NBA, respectively.
dydt(1,1) = y(1)*(-a1+b1*(M-(y(1)+y(2)+y(3)+y(4)+y(5))));
dydt(2,1) = y(2)*(-a2+b2*(M-(y(1)+y(2)+y(3)+y(4)+y(5))));
dydt(3,1) = y(3)*(-a3+b3*(M-(y(1)+y(2)+y(3)+y(4)+y(5))));
dydt(4,1) = y(4)*(-a4+b4*(M-(y(1)+y(2)+y(3)+y(4)+y(5))));
dydt(5,1) = y(5)*(-a5+b5*(M-(y(1)+y(2)+y(3)+y(4)+y(5))));
```
%----------------------------------
%File: process.m
%----------------------------------

%In this m files, we call the ode45 command to solve the system of
%equations' Predator-Prey Model. We also initialize all of the values for
%our system.

%To solve the system, we called the ode45 method. The first set of closed
%brackets contains time from 0 to 50 and the second set of closed brackets
%initializes the fan base for the MLS, MLB, NFL, NHL, NBA were set to be
%the total stadium attendance of the 2011 season.
[t y]= ode45(@predatorpreymodel, [0 50], [3.75877 9.26509 21.55452 5.23650
5.18194]);

%Closes the plot after each trial.
close;

%Displays the plot of the leagues for all t from 0 to 50.
plot(t,y(:,1)); %MLS
hold on;
plot(t,y(:,2),'r'); %MLB
plot(t,y(:,3),'g'); %NFL
plot(t,y(:,4),'m'); %NHL
plot(t,y(:,5),'c'); %NBA
xlabel('Years')
ylabel('Popularity (Thousands)')
title('The Popularity of Each League In The Next 50 Years')
REFERENCES


