SM-4290

Final Year Project Proposal

Proposed Title: Building a Pitch Control Model and Understanding the Mathematics behind it

Supervisor: Dr. Haziq bin Haji Jamil

Name: Abdul Haziq Ikhwan bin Abdullah Rudiman

Registration Number: 18B9001

Date of Submission: 15th May 2022

ABSTRACT

Football has always been a game of numbers that is played between two teams. Numbers, on the other hand is in essence the baseline of mathematics. Football and mathematics have been co-existing with one another as there is a lot of statistics that is involved in football. This project will attempt the replication of a model called Pitch Control Model that looks into the likelihood that a player can control a ball when the ball is passed to a certain area of the football field. The model will be made in two software programmes, namely Spyder and R. The model can help in football analysis development in the country to help improve the football industry and community.

Table of Contents

1. INTRODUCTION

a. General Background

The application of mathematical approaches to assess performances, recognise trends and patterns, and anticipate results has been one of the most significant developments in the sports industry. The use of mathematics in football is not something that is new in the world of the sport. Statistics and data science play huge roles in analysis of football matches. Sumpter (2016) believed that mathematics can be used to comprehend football, and vice versa. Passing networks between players can be illustrated by geometry, Messi and Ronaldo's otherworldly records can be perceived as statistical extremes, and last-minute winners in a football match can be justified by probability and randomness. It is not to say that mathematics is the most important aspects in football, after all, there is a tactical and psychological side to football. However, it can be said that mathematics and football are intertwined.

An example of the importance of statistical and data analysis is understanding the importance of possession in football. Possession in this case, is the percentage of how long a team is in control of the ball over a period of time. Where, the longer the team has the ball, the higher their possession. Bloomfield et al. (2005) found that the top three teams in the English Premier League during the 2003-2004 season (Chelsea FC, Manchester United FC, and Arsenal FC) had more possession time compared to their opposition. In addition, ball possession has been increasingly important in many football clubs' offensive game plans in recent years (Casal et al., 2019)

This report will discuss the plans of action that will be taken for my Final Year Project for next semester. The proposed title for the project is: Building a Pitch Control Model and Understanding the Mathematics behind it. Spearman et. al. (2017) defined pitch control at a specific area of the football field is the probability that a player will be able to control the ball if it is at that area. The probability is determined by the time it takes each player to get to that area, which is determined by their location and velocity. Event and tracking data sets of a football match (explained in 2.b.ii) are used in creating a pitch control model. A pitch control model can help in match analysis to evaluate the best passing option at any given moment during a football match. It can also tell players where they should be during a match to give them the highest pitch control possible.

b. Significance of the Project

This project can help in providing necessary variables that can improve the on-field performance of a team. Projects like this can help in bridging the local or national football teams toward the importance in data analysis in football. Furthermore, it can help pique an interest in someone who enjoys a blend of both football and mathematics.

c. Aims and Objectives

This project will look into the mathematics that is involved in the creation of the pitch control model. To add to that, the pitch control model will also be replicated in different programming softwares (Spyder and R).

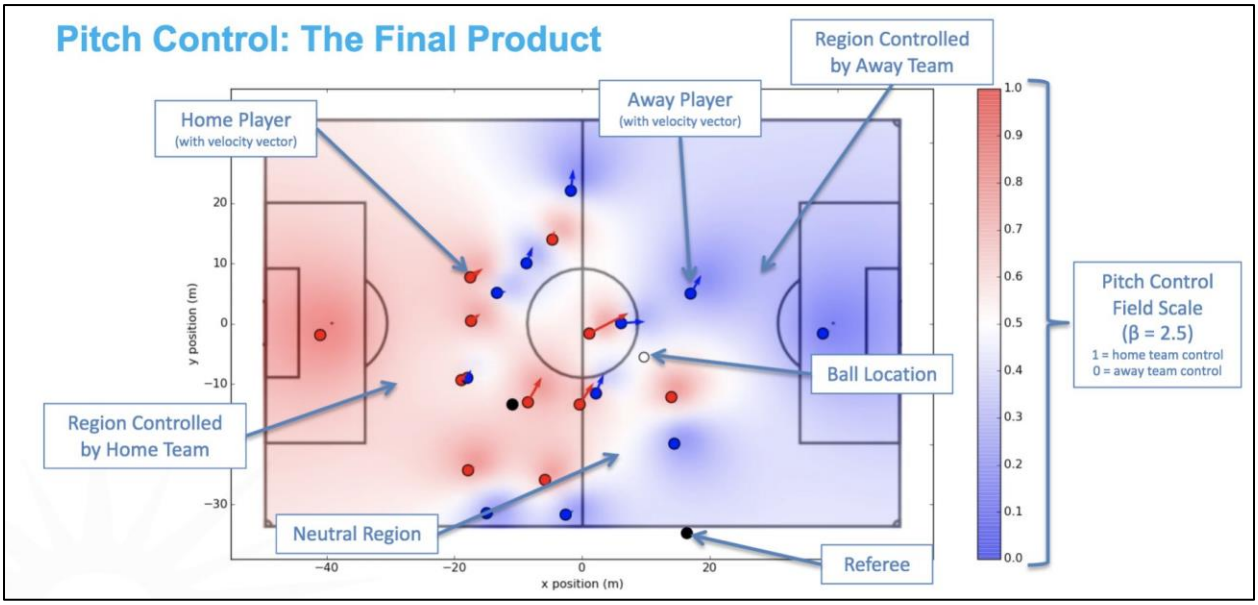

Figure 1: An example of Pitch Control Model

The figure above is a pitch control model that was created by William Spearman. The photo was taken from Spearman's public Twitter account.

2. METHODOLOGY AND MATERIALS

a. Methodology

i. Replication in Spyder

The pitch control model will be created in Spyder, which can be downloaded online. It is a free programming software that writes programming codes in Python. The Spyder software combines a comprehensive programming tool's powerful editing, analysis, debugging, and profiling functionality with a scientific package's data exploration, interactive execution, deep inspection, and stunning visualisation capabilities (Spyder Team, 2018). Python, on the other hand, is one of the most well-known programming languages that is used by programmers. Its design philosophy prioritises code readability, and its language enables programmers to express concepts with fewer lines of code compared to other programming languages (Srinath, 2017).

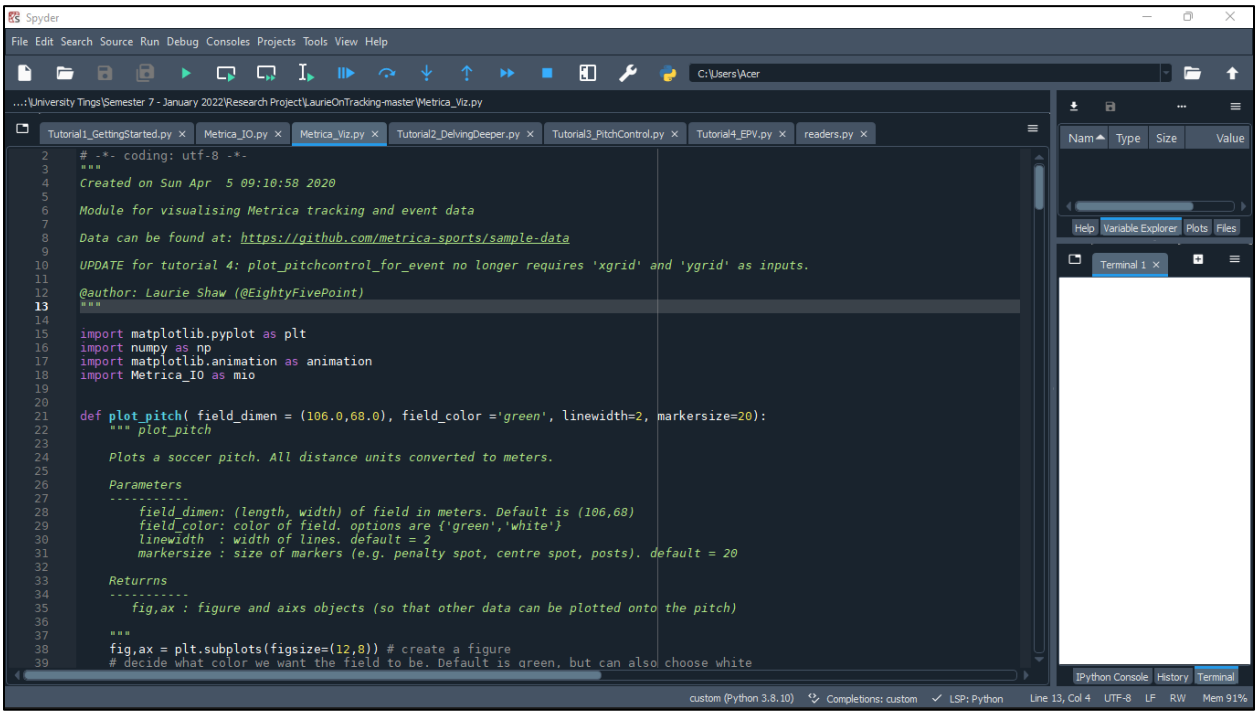

Figure 2: Spyder Interface

ii. Replication in R

The replication of the pitch control model will also be attempted in R. It is also a free programming software that can be downloaded online. R is a statistical computing and graphics programming language. Comparing the two software, R has a simpler interface.

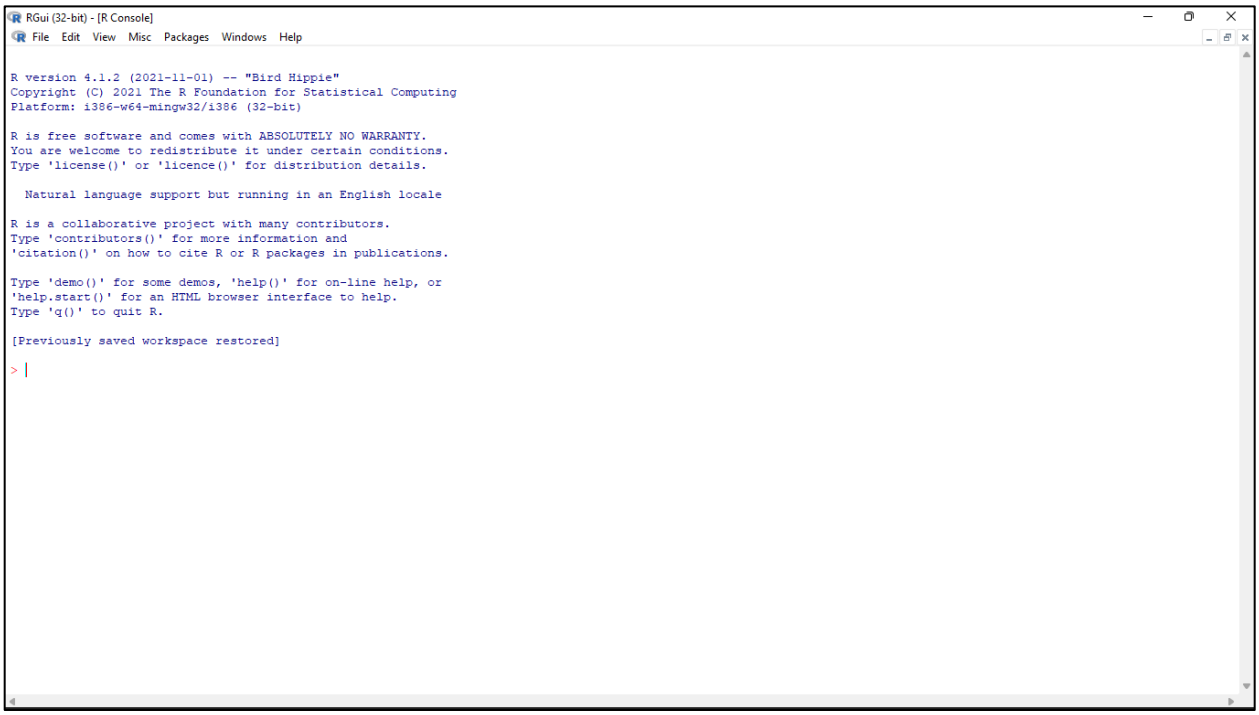

Figure 3: R Interface

b. Materials

i. Video Tutorials

The project will mainly look into five videos which can be found on YouTube from the channel called Friends of Tracking. The channel primarily uploads videos that educate anyone who is interested in the field of football analytics. The first video that will be looked into is title ‘Liverpool FC data scientist William Spearman’s masterclass in pitch control’. The video is a lecture on pitch control presented by William Spearman who is the Lead

Data Scientist at Liverpool Football Club. The other four videos are from a series of videos teaching the use of Python codes in Spyder for football analysis. The videos are presented by Laurie Shaw who is currently the Lead A.I. Scientist at City Football Group.

ii. Data Sets

The data sets are available on the github.com website. They were published by Metrica Sports, a software company based in Amsterdam, Netherlands. The data set is a collection of event data and tracking data that are compiled from a football match. Event data is the log of each on-ball event such as passes and shots. Tracking data on the other hand, looks at the position of player and the ball which are sampled at 25Hz. The data set is compiled in a spreadsheet in Microsoft Excel. Metrica Sports have uploaded three different data sets from three different football matches. These data sets will be used to create the pitch control model in Spyder.

iii. Coding Scripts for Spyder

The coding scripts for football analysis is also available on github.com. The scripts were also written and published by Laurie Shaw. The scripts are divided into four parts which relate to the four tutorial videos presented by Laurie Shaw. Each collection of coding scripts will be used for different aspects of football analysis using the Metrica data set.

iv. Coding Scripts for R

On the other hand, for R, the coding scripts for creation of the pitch control model were written and uploaded by Tony ElHabr on the itsmetoeknee.app website. The scripts also make use of the Metrica data set.

3. LIMITATION TO THE PROJECT

Since this will be my first time coding in Python and Spyder, I am concerned that it will take a lot of time for me to get used to them, especially when errors arise during the building of the pitch control model.

4. PEOPLE INVOLVED

This project will involve myself, Abdul Haziq Ikhwan bin Abdullah Rudiman as the project creator, proposal maker, the writing of the journal article report and presentation slideshow. The project will be under the supervision of Dr. Haziq Jamil throughout the whole semester.

5. EXPECTED OUTCOMES

It is expected of me to be able to replicate and build the pitch control model in both Spyder and R while understanding the codes that were written by Laurie Shaw and Tony ElHabr. In addition, understanding the mathematics that happen behind the building of the model will also be one of the main expected outcomes.

6. TIMELINE OF THE RESEARCH PROJECT

Month	Date	Task
August	1st August – 14th August 2022 (Week 1 – Week 2)	Literature reviews and watching the aforementioned videos
	15th August – 15th September 2022 (Week 3 – Week 6)	Replicating the models in Spyder and R
September	16th September – 30th September 2022 (Week 7 – Week 8)	Understanding the codings that are used for the pitch control model
October	1st October – 14th October 2022 (Week 9 – Week 10)	Discussions and reviews of the pitch control model and subsequently, the project
	15th October – 31st October 2022 (Week 11 – Week 12)	Project report write up
November	1st November – 14th November 2022 (Week 13 – Week 14)	Finishing up the project report and presentation slideshow

Table 1: Expected Timeline on the Development of the Project

The table above outlines the timeline of the project from the beginning until the end of the semester. However, I plan to start doing some of the early tasks above before the semester starts, especially the tasks that involve Spyder. This will help to free a lot of my time during next semester to allow me to focus on my other modules.

7. ETHICAL CONSIDERATIONS

Since the video tutorials, data sets, and coding scripts are publicly available online, it can be presumed that there will not be any ethical matters that can surface.

8. SUMMARY

To summarise, there is without doubt that mathematics plays a huge role in football and pitch control is one that involves some form of mathematics in it. Creating a pitch control model can help in understanding the values and importance of passes in any moment of a football match. The model will be made in Spyder and R. Although, being an unexperienced Spyder user can be a limitation to the project. This project will start on the 1st of August 2022 and is expected to end in November 2022.

9. REFERENCES

Bloomfield, J., Polman, R., & Donoghue, P. G. (2005). Effects of score-line on team strategies in FA Premier League Soccer. *Journal of Sports Sciences*, *23*(2), 192-193.

Casal, C. A., Anguera, M. T., Maneiro, R., & Losada, J. L. (2019). Possession in football: more than a quantitative aspect–a mixed method study. *Frontiers in psychology*, *10*, 501.

Spearman, W., Basye, A., Dick, G., Hotovy, R., & Pop, P. (2017, March). Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*.

Spyder Team. (2018). *Spyder Website*. Spyder-Ide.org. https://www.spyder-ide.org/

Srinath, K. R. (2017). Python–the fastest growing programming language. *International Research Journal of Engineering and Technology (IRJET)*, *4*(12), 354-357.

Sumpter, D. (2017). *Soccermatics: Mathematical adventures in the beautiful game.* Bloomsbury Publishing Ltd.