IT 19 024
Examensarbete 45 hp
Juni 2019

Modelling the Collective Movement of Football Players

Francisco José Peralta Alguacil

Institutionen för informationsteknologi
Department of Information Technology

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Modelling the Collective Movement of Football Players

Francisco José Peralta Alguacil

Since the development and rise of different methods for obtaining tracking data from sports games, data analysis has played a very important role in the world of football. With many studies and models having been created for carrying out a posteriori analyses of football games, there is a necessity to take a step forward and include collective motion simulations in these analyses. With this thesis, a new model for computing the optimal positions of football players during an attacking situation is proposed.

Three different computational models that quantify paramount aspects of football such as pitch control, pass success probabilities and pass impact will be used as the basis of our simulations. The new developed model built on top of these is able to look forward one or two seconds into the future and predict the optimal positions of the attacking players according to certain simple rules that are set to maximise or minimise different combinations of the mentioned algorithms.

Handledare: David Sumpter
Ämnesgranskare: Matteo Magnani
Examinator: Jarmo Rantakokko
IT 19 024
Tryckt av: Reprocentralen ITC

Acknowledgements

First of all, I wish to express my gratitude to my supervisor **David Sumpter**. His continuous encouragement and support, together with his noticeable enthusiasm for maths and football have made all these months of work much more exciting and enjoyable. Without him it would not have been possible to successfully carry out this project that combines two of my main passions: football and science.

Special thanks to **Fútbol Club Barcelona** and **Javier Fernández de la Rosa** for their close collaboration and help along the whole thesis. I will always be grateful to them for their warm welcome during my two-week stay in Barcelona and their eagerness to cooperate with us as much as they could, including the allowance to use their implementation of the pitch control model used in our simulations.

I would also like to thank **Hammarby IF Fotboll** and the Sweden-based company **Signality** for letting us use the tracking data of their matches and for their interest on exploring the new possibilities that the implementation of our models could offer to their football performance.

And last, but not least, thank you to my thesis reviewer **Matteo Magnani** for his advice and corrections on this report.

Contents

1 Introduction

Self-propelled particles models have been a paramount method for analysing collective behaviour in the last years, especially in the field of computational biology, where the movements of flocks of birds or schools of fish have been modelled and simulated in different ways.

In 2002, Couzin et al. modelled the behaviour of a group of animal individuals by establishing three zones around each one: repulsion, orientation and attraction area [2]. Depending on the zone where the rest of the individuals are, different rules are applied regarding crash avoidance, movement coordination or isolation prevention. In this study, different parameters are also modified in order to analyse the way physical or motivational variations affect how animals organise themselves when moving together. They found out that the aggregation patterns that they obtained in their simulations clearly resembled collective movements in animals. Mosquitoes or midges tend to group together with little or no global alignment, which results in groups not moving along long distances [13]. On the other hand, most of birds and fish tend to form highly aligned groups with lower density of individuals, allowing them to move through the flock or school [11]. They also found a more unnatural torus-like formation that could match movements sometimes observed in groups of barracudas or tuna [14]. After this, many papers appeared improving and taking this model further away. It is the case of the one published by Wood and Ackland in 2007 [21], where they introduce evolutionary algorithms to the simulated flock in order to study the effects that predator risk have on the adopted shape of the group.

When it comes to human movements simulations, there have been some studies regarding pedestrian behaviour being carried out by Mehdi Moussaïd. In his 2011 paper [12], together with Helbing and Theraulaz, they modelled the movements of the pedestrians using simple rules to avoid collisions with others and reach their destination point following the most direct trajectory. This was achieved with the assumption that people can estimate other pedestrians trajectories and possible collision times [9] [15]. Other studies follow a different approach to crowd simulations, considering them a fluid, where individual movement decision making have little or no importance in the collective flow [10] [8].

The fact that a game of football is composed by only 22 players makes it different from simulating a crowd of people, since individual movements have much more importance since they are not ”covered” by the collective motion, as studied in Silverberg et al. [16] or in Huges [10].

However, even though many professional sports clubs already have research groups working on data analysis, these kind of models have not yet been completely extrapolated to football due to the high complexity of the games, where players' main objective might not always be to catch the ball at any cost (as birds would do with preys). This is where other factors as ball location on the pitch, match current score, individual roles or counter-attack risk evaluation among others come on the scene. Nevertheless, in an era when computer-assisted tactics and statistics-based team-planning are becoming of utmost importance in professional football, these self-propelled particles models could certainly be a great basis of useful tools for coaches and analysts when it comes to teamwork improvement and tactics enhancement.

Due to these intrinsic complexities, developing a programme that simulates a game of football with all its different situations (passing, ball conduction, defensive movements, shooting, offsides, dribbling, etc.) from the perspective of data analysis would be unachievable at the moment and is out of the scope of this project. This is why the main problem should be subdivided in smaller ones, and we will first focus on one of them, which could be considered to be the main pillar of collective sports: passing.

Since the undeniable dominance of F.C. Barcelona and the Spanish national team between the years 2008 and 2012, with both teams achieving overwhelming victories in almost every tournament they played, many of the best football teams in the world tried to copy their playing style and started to base their offensive strategy on what are known as "static" or "positional" attacks. These are situations on which the attacking team continuously passes the ball from side to side, between the opponent's penalty area and the midfield line, waiting for the other team to make a marking mistake which allows the midfielders to send the ball behind the defensive line, when one the attackers makes a disruptive run. Our objective is to understand and simulate these kind of game situations. We consider that the player who has the possession of the ball wants to make a pass as soon as possible so that the ball is continuously moving in order to create flaws in the opponent's defensive block and be able to successfully send it into an open space behind them.

With the help of GPS tracking data from real matches that will by provided to us by F.C. Barcelona, Hammarby IF and the Swedish company *Signality*, we will perform our analyses on passing situations by applying a compound of motion and probabilistic laws. Furthermore, two other models for computing pitch control (developed by F.C. Barcelona) and pass impact (created by the Swedish company *Twelve*) will be implemented and glued together for developing a completely new algorithm that will allow us to analyse and make in-depth studies on the players' positioning during certain in-game situations.

Nevertheless, all of this is not a straightforward task to do, since aspects like decision making or risk bearing are non-physical qualities that cannot be perfectly modelled with common parameters for all the players. For example, it is clear that Luka

Modric or Iniesta would make "smarter" decisions than most of the rest of the players when it comes to handling risky situations and that their passing accuracy and movement patterns would be significantly better than the average. This is why we will focus on making our simulations with "general players" that are equally skilled, both from a physical and mental point of view, with them always making the decision that leads to the optimal situation depending on their role on the pitch.

2 Background

Statistical studies and data analysis have increased their presence in professional sport during the last years thanks to the existence of video recordings of almost every played game, and many studies regarding spatio-temporal analysis have become easier to be carried out due to the irruption of high precision GPS data and other similar position-tracking systems.

The first studies were carried out with a focus on pitch control using both simple and deformed Voronoi diagrams [20], with Fujimura & Sugihara improving them in 2005 by adding a new motion model for the players [5]. In these very first researches they focused on computing what they call "dominant areas" for the players, i.e., all the field points where a certain player can get faster than anyone else when they move according to a motion model. The motion model introduced by Fujimura & Sugihara has been of great utility in subsequent research papers, as its parameters can be easily tuned using real tracking data for estimating the players' maximum speed and acceleration in different sports. Furthermore, a primitive pass model is developed in [5] making use of the dominant regions in a game of field hockey. However, given a certain pass direction and speed, this model gives discrete results, i.e., either a player can or cannot receive the pass and does not take into account the possibility of disputing the ball with an opponent in a 50-50 pass with both of them having almost equal chances to get it.

After these, many more papers which perform spatio-temporal analysis on sports and football games have been published. Regarding passing models, in 2014 Gudmunsson and Wolle used again the same discrete principle based on dominant regions to compute the receiver of a pass made with certain speed and direction [7]. However, it was not until 2017 when Spearman et al. introduced a passing model with a probabilistic basis [18]. In their model, which will be explained in detail in the upcoming sections, they take into account aspects like players agility or the necessary time to control a ball, but most importantly, they introduce smooth transitions between player dominant regions thanks to their probabilistic approach to the problem. The passing model that we will implement and apply in our study will be based on this one.

Since the publication of this research paper, some others have followed this probabilistic approach to model and simulate football passes [6] or for dominance and tactics analysis [17], and great improvements have been made towards the simulation of a football game from the perspective of data analysis.

But not only passing models have been developed and published, other tactical aspects of football have been taken into account in this journey to a computational-driven analysis of football games. Fernández and Bornn made one of the last improvements in this area [3], with the development of algorithms to compute pitch control and pitch value in a game of football. The fact that they focus on studying off-ball situations rather than on-ball events like passes or shots is of utmost importance because, as Johann Cruyff once said and as they reproduce in their paper: "*It is statistically proven that players actually have the ball 3 minutes on average. So, the most important thing is: what do you do during those 87 minutes when you do not have the ball? That is what determines whether you are a good player or not.*"

Our main aim is to follow this trend towards a computational analysis and simulation of a football game. The starting point will be to implement a pass model for analysing different real in-game passing situations and, after that, an attempt to simulate and analyse offensive plays with the presence of disruptive runs will be made, combining this pass model with F.C. Barcelona's pitch control model and *Twelve*'s pass impact algorithm.

3 Pass Probabilities model

In a game of football, and mainly during attacking situations, passes are usually made in risky conditions, which results in the fact that given a pass with a certain ball speed and direction, it is almost impossible to assure with no error which player will intercept it. Although the sender of the ball tries to maximise the probabilities of his team mates to receive it, there are always chances for the defenders to intercept the pass; this is why Spearman et al. [18] proposed their probabilistic approach to model these paramount in-game situations. An implementation of their model, with some changes that we will introduce, will be included in this project, as we consider it to be of utmost importance to have a physics-and-probabilistic-based model as the basis of our simulations.

3.1 Ball dynamics.

In order to make realistic and reliable simulations and analyses it is necessary to include a physical model in our algorithms to reproduce the movements of the players and the ball in the best possible way. In Spearman et al. the ball movement is modelled using a pure ballistic approach, i.e., only aerodynamic drag is considered to be responsible of the natural deceleration of the ball during a pass. This simplification works for their studies, mainly because they implement a model that include long passes, which are supposed to be intercepted before the ball touches the pitch, so the friction force between the grass and the ball can be considered to be negligible. However, since the main focus of our study will be on ground passes, during which the ball is almost all the time in close contact with the grass, friction force needs to be added to our model.

That being said, we will consider two external elements acting on the ball that are responsible for its deceleration: the air and the grass. Nevertheless, finding the exact way of modelling this friction force with the grass is, for sure, almost impossible in general. Aspects like grass length, wetness and type (natural, artificial or mixed), as well as the materials, pressure and weight of the ball should be taken into account for the simulations. All of these influence the ball dynamics, the way it rolls and the magnitude of its deceleration along its trajectory on the pitch, so some simplifications should be assumed in order to compute this force.

The first deceleration, caused by the air, is easier to compute, though. For modelling its influence on the ball, the general formula of aerodynamic drag can be used and no Magnus force[1] is included because of the lack of data about the spinning movement of the ball. So, in accordance to this, the first part of the equation of motion for the ball is

$$\ddot{\vec{r}}_{aero} = -\frac{1}{2m}\rho C_D A \dot{r}\dot{\vec{r}} \tag{1}$$

Where m is the mass of the ball, ρ is the density of the air, C_D is the so-called drag coefficient and A is the cross section area of the ball. The following values are used for these constants:

- The **mass of the ball** is set to $0.42\,kg$, which is the minimum weight accepted by FIFA for a size 5 ball for professional games [4].
- We use a **density of the air** of $1.225\,kg/m^3$, value used in the *International Standard Atmosphere* model [1].
- Finally, we set the **drag coefficient** to 0.25 and the **cross-sectional area** is $0.038\,m^2$, as used in Spearman et al.

[1]The Magnus effect or Magnus force is a physical phenomenon that provokes changes in the trajectory of an object inside a fluid when it spins, due to the differences in pressure that it creates on opposite sides of the object.

The second term in the ball's equation of motion will be the friction between the ball and the grass, This term, although being mathematically simpler, is more likely to introduce errors in the simulations, because of the ignorance of the exact value of the friction coefficient between the ball and the grass. It is well known among football players and fans that the condition of the pitch is highly influential in the deceleration of the ball: a pass done on a field with short and wet grass will be faster than another made when it is dry and long. The equation of motion with this force is the following one:

$$\ddot{\vec{r}}_{friction} = -\mu g \hat{r} \tag{2}$$

Due to the changing conditions of the pitch and the unavailability of data, we decide to take a value of $\mu = 0.55$, which is a value in the middle of the interval that FIFA recommends for high quality artificial grass surfaces [19].

Once the two forces that will act on the ball are clear, we have to answer two relevant questions: Do both forces act along the whole trajectory of the ball? Would any of them be negligible in certain situations?. We decided to make the assumption that at the beginning of the pass the only force that significantly acts on the ball is the aerodynamic one, because, when the ball is kicked, it tends to "float over the grass" almost without touching it and without rolling. On the other hand, in the last part of the trajectory, the friction force is much stronger than the aerodynamic, so this one can be neglected. With these considerations, the decision to split the trajectory of the ball in two parts was made, but we still needed to determine the span of them.

With the tracking data that F.C. Barcelona gave us of one of their matches, we performed some trial and error experiments with passes in which the ball did not get too high over the pitch (at most 10 or 20 cm), so that they can be considered to be ground passes. Using this method we took the real trajectory of the ball and compared it to the one obtained with our simulations, determining that, for most of the analysed passes, the option that resulted in the most similar dynamics was to stablish that the deceleration of the ball for the first two thirds of its trajectory was mainly caused by the aerodynamic drag and that ball-grass friction was the main force acting during the last third of the pass. So, putting together both forces with their respective time intervals, the final equation of motion for the ball to be used in this project is:

$$\ddot{\vec{r}} = \begin{cases} -\frac{1}{2m}\rho C_D A \dot{r} \dot{\vec{r}} & \text{if } t \leq \frac{2t_{max}}{3} \\ -\mu g \hat{r} & \text{if } t > \frac{2t_{max}}{3} \end{cases} \tag{3}$$

3.2 Interception times. Player dynamics.

When it comes to the players' dynamics and obtaining the time that it takes them to intercept the ball in a certain pass, Spearman et al. make use of a minimization procedure to obtain the fastest trajectory of a player to intercept the ball during a certain pass. Here, the following equation of motion is solved for all the players, with two constraints which set limits to the maximum players' speed and acceleration.

$$\vec{r} = \vec{r_0} + \dot{\vec{r}}t + \frac{1}{2}\ddot{\vec{r}}t^2 \tag{4}$$

Solving this equation with the final position of the player ($\vec{r}$) set to be the current ball position ($\vec{r_b}$) for all the ball positions along the trajectory of the pass, they can determine whether the player can intercept the ball or not if his minimum arrival time to that point is smaller than the arrival time of the ball, i.e, if the player can reach at least one point of the ball's trajectory before it gets there, interception occurs.

This method is reliable and provide good results to them; however, having to solve a minimization problem for all the possible points along the trajectory of the pass would be too computationally expensive and, as we will see once we get to the upcoming sections, any computing time that we can reduce will be of utmost importance. So, the taken decision was to change this method for the one introduced by Fujimura and Sugihara in 2005 in which they consider the players as objects whose movements are described by an equation of motion with a driving force (which represents the force exerted by the players' legs) and a drag force (which bounds their maximum possible speed). In this way, the problem we are solving is still faithful and physics-based but less computationally demanding.

Thus, the equation of motion that we will use is the following one,

$$m\frac{d}{dt}\vec{v} = \vec{F} - k\vec{v} \tag{5}$$

Whose solution (see Appendix I for step-by-step explanation) is given by:

$$\vec{x} - \vec{x_0} = V_{max}\left(t - \frac{1 - e^{-\alpha t}}{\alpha}\right)\vec{e} + \frac{1 - e^{-\alpha t}}{\alpha}\vec{v_0} \tag{6}$$

Where $V_{max} = {}^{F}\!/_{k} = 7.8\,m/s$ is the maximum velocity that a player can reach, $\alpha = {}^{k}\!/_{m} = 1.3$ is the magnitude of the resistance force and $\vec{e}$ is the unit vector that denotes the direction of the acceleration of the player. The values for these constants are the same ones as Fujimura and Sugihara used in their paper and were obtained by performing a study with several field hockey players.

With this result, it is possible to see that all the points that a player with starting position $\vec{x_0}$ and initial velocity $\vec{v_0}$ can reach are enclosed inside the circle with centre

$$\vec{x_0} + \frac{1 - e^{-\alpha t}}{\alpha} \vec{v_0} \tag{7}$$

and radius

$$V_{max} \left(t - \frac{1 - e^{-\alpha t}}{\alpha} \right) \tag{8}$$

This makes the finding of the interception times for the players much easier than with the minimization problem that Spearman proposed. In order to obtain them the time must be discretised (we do that in steps of 0.04 seconds) and, for each time step, we check the already computed position of the ball and calculate the reachable area of the player. If the current ball position falls outside the circle, we advance to the next time step and repeat the process until the ball is in inside the player's reachable area; that moment determines the interception time.

3.3 Passes as probabilistic events.

Once the physical models behind the ball and the players have been established, the probabilistic model proposed by Spearman et al. can be used. A detailed description of it can be found in their paper [18], but its main feature is the usage of a logistic distribution to determine the probability of a player getting the ball at time T knowing his arrival time t.

$$P_{int} = \frac{1}{1 + e^{\frac{T - t_{int}}{\sqrt{3}\sigma/\pi}}} \tag{9}$$

Note that this function does not compute the probability for a certain player to get the ball during a pass, but the probability of him being able to intercept the ball after T seconds (without considering the rest of the players). Furthermore, another consideration that is made is that a player has to be in the vicinity of the ball for a certain time in order to have control over it, this is modelled with the term:

$$P(t) = 1 - e^{-\lambda t} \tag{10}$$

With the combination of these two, the final system of differential equations that gives the probability for each player to receive the pass is built as follows,

$$\frac{dP_j}{dT}(T) = \left(1 - \sum_k P_k(T)\right) P_{int,j}(T)\lambda \tag{11}$$

System that will be solved computationally and will result, when adding all the reception probabilities of the team mates of the sender of the ball, on the total probability of the pass being successful.

3.4 The simulation of a pass.

Gluing together all the aspects mentioned in the previous sections we finally obtain a probabilistic and physics-based model for analysing passing possibilities in a game of football that can be easily adapted to certain situations (faster or slower players, conditions of the grass, measures of the ball, etc.) by changing its parameters. The next step to be taken now is to take this theoretical model into practise and implement it in an algorithm that can be used in the rest of our simulations.

In order to analyse the pass possibilities at a certain in-game moment a, total of 750 "virtual passes" are simulated, distributed over 50 angles all around the ball and 15 pass speeds between 1 and 20 m/s. For each of these passes the probabilities of pass success together with its respective interception point, i.e., the location on the pitch where it is most likely to be received, are computed and a heatmap is generated with them.

3.5 Last interception points.

However, there is one main limitation to this approach and it is the fact that we are only considering ground passes. This means that there will be some areas on the pitch (mainly all of the points that lay behind a player) that cannot be reached with one of these passes, either because the ball is always intercepted before it gets there or because a really strong pass is needed for the ball to get there and there is no possibility of interception due to its speed. The points that bound this area around the ball are what we call the "last interception points".

In figure 1, a frame of the match between F.C. Barcelona and Real Betis is used to show this issue. In it, all the interception points for the simulated passes are marked with a blue cross and the plotted heatmap is a result of interpolating the probability of pass success for each of those points. It is clear to see how the different players (teammates or not) act as a barrier for the model and no ground passes are possible to be sent behind them.

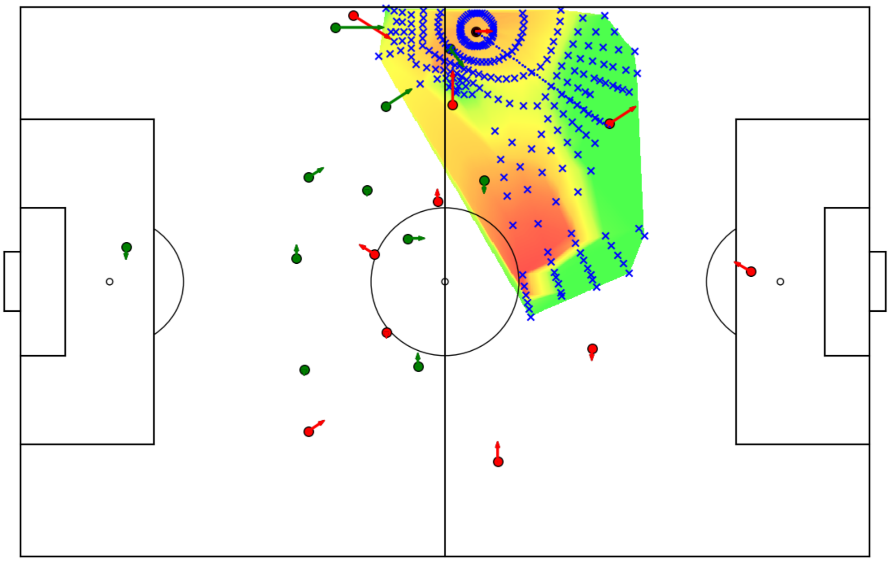

Figure 1: Interception points inside the possible ground-pass area around the ball.

But does this mean that, in a real in-game situation there is no possibility to send the ball further away? Certainly not. What the model is telling us is that, if the ball gets outside of the marked area after a ground pass it should always be **under total or partial control of a certain player**. This means that a player can take the decision of letting the ball roll for longer time instead of intercepting it as soon as he can, maintaining a certain positional dominance over it.

These decisions are, nevertheless, quite difficult to model into a computer algorithm because of the fact that many non-physical and decision-making factors that depend on the individual mentality of each player come into play. That is why a method of extrapolation was developed, a method that allows us to create a trustful heatmap of pass possibilities over the whole pitch and that is based on another paramount aspect of football: pitch control.

3.6 Extrapolating the model.

Using pitch control is a way of mimicking long passes in an easier and, possibly, more trustful way than simulating the trajectory of a long ball in the air because two main reasons: the first one is that factors that are unknown with the datasets that we use like wind speed and direction or ball spin play an important role in these passes and, even if we could perfectly model the flight of the ball, not all the players have the same skills when sending high passes and modelling this "player accuracy factor" properly would be almost impossible.

Pitch control in the area we want to extrapolate the model to is computed, following Spearman's proposal [18], in a very similar way as the pass probabilities model, but substituting the ball with static points on the pitch. A grid of points is created outside the zone that is already covered and, for each of them, the same algorithm as in the passes model is followed. The arrival times for all the players to the desired point is computed and they substitute the interception times in the original model. Then, once these times are obtained for all the players, the system of differential equations (11) is solved. In this way we manage to have a smooth transition between the two zones, without the need to develop a different algorithm for high passes. Figure 2 shows the extrapolation of the model for the same situation as figure 1 following this procedure.

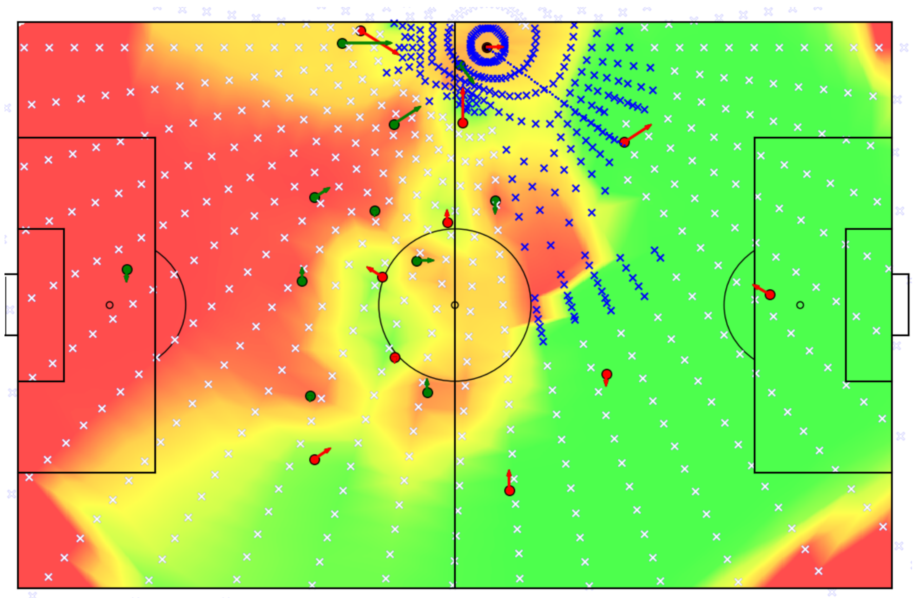

Figure 2: Interception points inside (blue) and outside (white) the possible ground-pass area around the ball.

As can be seen in the image, in this way the ground passes model can be extrapolated smoothly, being consistent with the physical model introduced before and maintaining the desired probabilistic approach towards pass success. Furthermore, a reliable method for computing pitch control has been created that could be potentially used for the upcoming simulations. However, as will be shown in the next section, the usage of this model for pitch control will not be possible at the moment due to computational limitations and a different one will have to be chosen.

4 Pitch control model

Pitch control is a factor that needs to be taken into account in every in-game situation during a football match. Being able to have positional control over a certain area of the pitch is key for any team that wants to have dominance over the possession of the ball, as it helps both to make safer and more dynamic attack transitions and drastically improves the team capacity and speed to recover the ball right after a loss or a deflection during the offensive situation. Since our simulations will try to reproduce the optimal way of playing that teams like Barcelona have, implementing a pitch control model is essential in order to obtain trustful results.

As described in the previous section, the way that we use for extrapolating the ground passes model is, in fact, a way of computing pitch control because it is based on the time that it takes for the players to reach a certain location on the field. This means that, if we use this method for the whole pitch creating an evenly distributed grid, we would have a pitch control model that gives smooth, probabilistic and physics-based results. Figure 3 shows an example of this method for a frame in the Hammarby vs Djurgårdens game, correspondent to the 6th match day in the Swedish Allsvenskan.

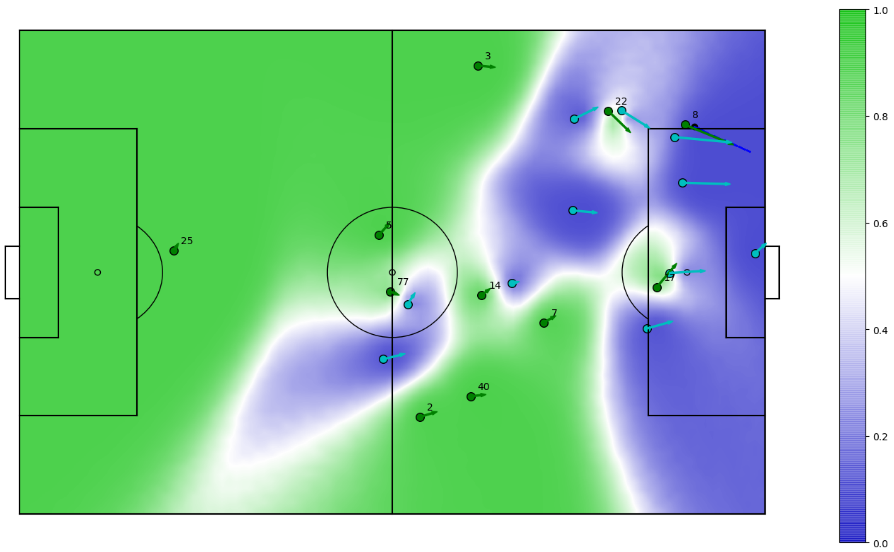

Figure 3: Pitch control in a frame of the Hammarby (green) vs Djurgården (blue) match.

However, we face a major problem with the model: it is very computationally expensive due to the necessity of solving the system of differential equations (11). Running the model for obtaining the pitch control on a normal situation (just for one frame, 22 players and a grid of 105 x 68 points) takes more than four minutes. This time is not extremely bad if we want to compute only one frame but, as will be shown later, the simulations that we are going to perform require recomputing the pitch control around 4500 times, which makes the usage of this model unviable.

For this reason, we will instead use a different model that was proposed by F.C. Barcelona, Javier Fernández and Luke Bornn in [3]. In their work they follow a different approach than Spearman's, basing it on what they call "player influence areas" instead of arrival times. The player influence at a certain point on the pitch p at time t is determined by the position and speed of the player and defined by:

$$I_i = (p, t) = \frac{f_i\,(p, t)}{f_i\,(p_i(t), t)} \tag{12}$$

Where,

$$f_i\,(p, t) = \frac{1}{\sqrt{(2\pi)^2\, det\,[COV_i(t)]}} exp\left(-\frac{1}{2}\,(p - \mu_i\,(\vec{s}_i(t)))^T\, COV_i(t)^{-1}\,(p - \mu_i(t))\right) \tag{13}$$

A full description of all the parameters and the rationale behind these equations can be found in their paper [3]. They managed to GPU-parallelise their algorithm so that it takes less than one tenth of a second to compute a frame in the same situation as in figure 3, which is, obviously, much faster than our previous implementation. Consequently, even though we will keep Spearman's pitch control model for computing the extrapolation part of the pass success probabilities algorithm, once it comes to purely computing pitch control, Fernández and Bornn's model will be used.

Figure 4 down below shows a comparison between both pitch control models, for the same case of study as before.

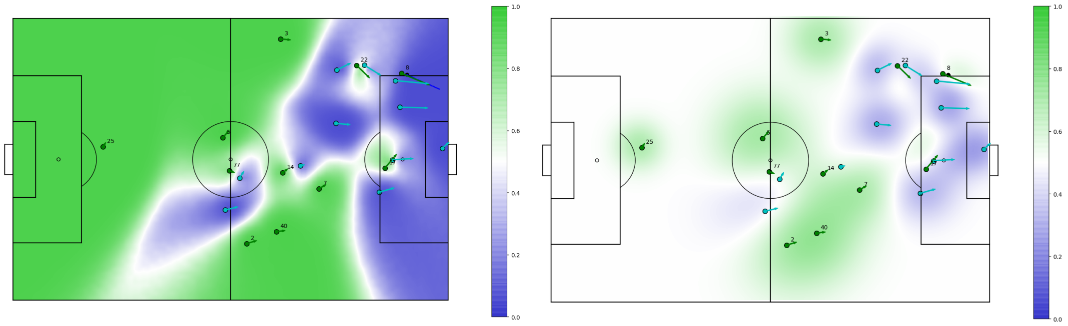

Figure 4: Pitch control comparison between our implementation of Spearman's model (left) and Fernández & Bornn's model (right).

Although both of them give reasonable and trustful results, the differences between both methods are noticeable from the plots. The main difference appears when it comes to computing pitch control in areas that are far away from all the players, like those around the corners. Fernández & Bornn's model limits the area of influence of the players, resulting in large areas that are not controlled by any of the teams, having pitch control value of 0.5. Meanwhile, Spearman's model does not show these "uncontrolled areas", since it is not based on the total amount of time that it takes to the players to get there, but the differences in arrival time between them, which results in a completely "filled in" heatmap.

5 Pass Impact model

The pass impact (or expected goals) model developed by *Twelve* is the last algorithm that we will have as the basis of our simulations. This model is mainly used by them to analyse whole matches and evaluate both teams' performance to determine whether the final result of the games was deserved or not. However, it can also be used to compute the probability of individual passes to result in a goal at the end of the play during which they are made, and this is the part of the algorithm that we will take advantage from.

The algorithm consists basically in two regressions that are fitted with the help of the historical data of several seasons, which includes tens of thousands of passes. All the matches used to train the model are broken down into sequences of possession, i.e., fragments of the game during which one of the teams holds the possession of the ball without losing it and without any stops in the play (due to fouls, throw-ins, offsides, etc.).

The first regression is obtained by assigning each chain a value between 0 (if the play ends without a shot) and 1 (if the sequence finishes with a goal); a value between 0 and 1 is also assigned if the play ends with a shot but without scoring, in this case a value given to the shot (computed with a different algorithm) is used. Thus, they use a first logistic regression to obtain the probability of a certain pass (defined by its starting and ending position on the pitch) leading to a shot and a second regression to compute the probability of a shot leading to a goal, having a final probability of:

$$P_{pass}(goal) = P(goal|shot) \cdot P(shot) \tag{14}$$

In this way we have a model that, given the starting and ending coordinates of a pass, gives us the probability of the team scoring a goal before the play which that pass belongs to ends. The following figure 5 shows two examples of the pass impact algorithm applied to two different starting coordinates of the pass.

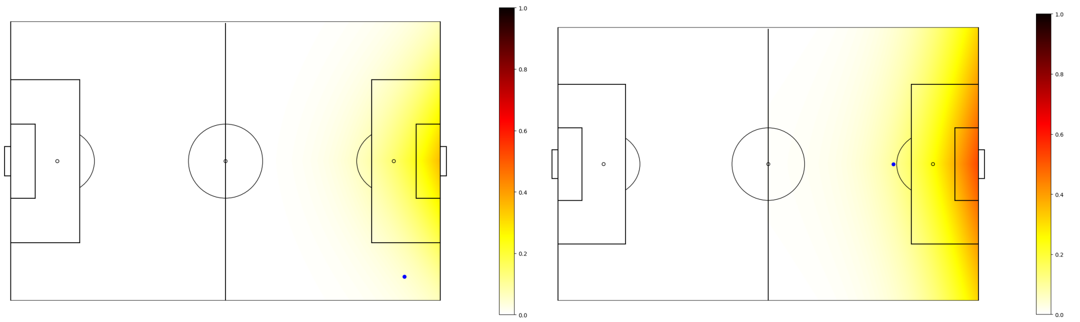

Figure 5: Pass impact for two different starting coordinates of a pass (blue dot). The figure is generated by applying the expected goals algorithm for the starting coordinates and a grid of points on the pitch as the ending ones to generate a heatmap.

6 The PIC model for an attacking play in football

Simulating a play of football resembling exactly how a certain team should perform according to its current players and the opponent team, requires not only modelling the players' physical abilities, ball dynamics and other aspects like pitch control or pass success probabilities as we have done during this project, but also being able to reproduce each player's individual decision-making, risk-bearing abilities and mentality, which is something that cannot be done with the data that we have and is out of the scope of this project.

However, what can be done is to create a model that computes the optimal final position of the attacking players according to certain rules and the individual roles that they have in the play. Using the models that we have already developed and implemented, we can obtain the best possible final situation for different cases of study and compare it to the actual positions that the players had at the end of the play; in this way, we have developed what we call the PIC maximisation model, standing for "Probabilities, Impact and Control".

6.1 Description of the model.

The kind of attacking situations we want to study and analyse with our model are those in which players have very few seconds (two or three at most) to reposition themselves in order to gain advantage over the opponent's defensive line. This usually happens when a player does a disruptive run and receives the ball behind the defenders, or when the player that has the possession of the ball dribbles an opponent and gets to a better position to send a dangerous pass. For these kind of situations, we want to have a model that, given an initial state of players positions and velocities, ball position and the starting and ending frames of the play to analyse, returns the corrected position of certain attacking players, which should be selected and categorised by hand into different roles after watching the real play in the match.

One thing that the simulated attacking players will have in common is that they will always know where the defenders are going to be at the end of the play. At first, this might sound too unrealistic if we think about long attacks where players have enough time to move almost anywhere on the pitch if they wanted to; however, the situations we are going to study will last for, at most, 2 seconds, which will make this assumption more reasonable. In fact, in a real football (or any other sport) game, one of the things that make some players better than the rest is their ability to be continuously predicting the next movements of their opponents in order to gain advantage over them. This model will make use of this essential skill to determine the best possible movements for the attackers.

6.1.1 Attacking roles. The five-rules behaviour.

The model we propose is based on the fact that each player, depending on his intentions during a play (receive a pass, fix the defenders, be in a good position in the case of a rebound, etc.) will behave in one way or another, but always having something in common with the rest of his teammates: he is maximising a certain function which is constructed using the Pitch Control, Pass Probabilities and Pass Impact models. The five different rules that will be used in our simulations are the following ones:

- ***Rule 1: Stay on ball.*** This rule apply to the player that has the possession of the ball or that is going to imminently receive a pass in the real in-game situation. In our case, if we are considering a passing situation, this rule applies both to the sender and the receiver of the pass and, in practice, it means that we are not going simulate their position, so they follow their real trajectories according to the tracking data.

- ***Rule 2: Move to a dangerous position.*** Usually, the player that is in the most centred position and closest to the opponent's goal at the moment of the play is the one that has this role. This basic rule consists on him running towards the point of maximum Pass Impact, i.e., the most dangerous position for receiving a pass, which normally is situated at the centre of the goal.
- ***Rule 3: Gain pitch control.*** Most of the attacking players will follow this rule, which consists on moving to the point in their reachable area that gives the maximum value of pitch control while also having possibilities of receiving a pass, i.e., the one that maximises the multiplication $PitchControl * PassProbabilities$.
- ***Rule 4: Disruptive run.*** This rule is applied to the player that runs towards the back of the defence in order to open up space behind it. He moves to the point with the maximum value of $PitchControl * PassProbabilities * PassImpact$. This means that he is the most complete player, who creates spatial dominance for his team while being a menace for the defenders, since he also takes into account being able to receive a pass in a dangerous position.
- ***Rule 5: Defend.*** This is the only rule that is applied to a defender of the attacking team and it consists in staying close to the defending team's striker, mainly for avoiding a possible fast counter-attack if the possession of the ball is lost.

Determining which player or players should have each of these described roles and which rules they should follow is the key part of the model that, at the moment, should be done manually. For sure, there are many different opinions about what could be considered to be an "optimal" movement or position for a player. Nevertheless, there is a way that F.C. Barcelona uses for determining which role should each player have, based on their position on the pitch during the attacking situation. This model they developed is based on three zones around the ball: intervention, mutual help and cooperation area. We will use this idea of having three zones as an assistance when choosing the role of each player in the play. However, we will slightly modify them and manually select the area that these zones cover depending on the situation that we are analysing, rather than having fixed radii for them:

- ***Intervention zone:*** This is the zone close to the ball where players are more likely to intercept the ball or receive a pass. This zone is the one that covers the immediate points around the ball. It usually includes the player with the ball and those ones that could touch or intercept it imminently.

- ***Mutual help zone:*** Here, players are in a relatively close position to the ball, but further away than the players in the intervention area. Some players inside this zone could be the receivers of the pass in the next seconds, while others are mostly expected to maintain or search for an advantageous position for the upcoming situations, rather than having a direct intervention on the ball. For us, most of the attacking players will be included in this area, having different types of helping roles.
- ***Cooperation zone:*** Players in this zone are further away from the play and not expected to receive the ball during it. They will remain far away from the ball but controlling the situation around them in case the opponent recovers the possession. In our model, the players included here will mainly be the defenders of the attacking team.

As an example for showing these zones we have figure 6, which is a frame of the match F.C. Barcelona vs Real Betis, played during the first match day of the 2017/2018 La Liga season. This play will be also used as an example situation for our optimal positions model in the upcoming section.

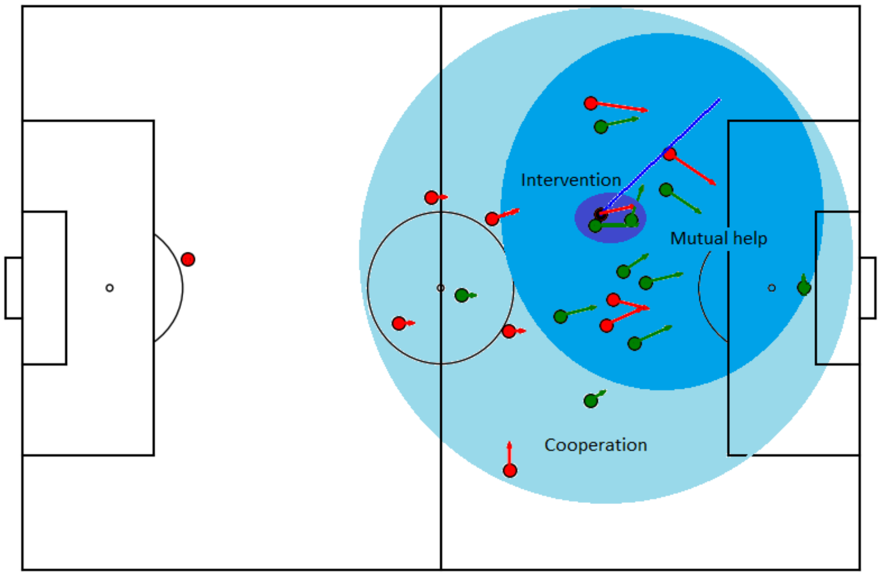

Figure 6: Tracking data view of the match situation in minute 45:42 in the game Barcelona (red) vs Real Betis (green) showing the different behavioural zones for this example. The smaller blue dots represent the real trajectory of the ball in the next frames.

6.1.2 Players' reachable area.

Just like the rest of the models that have been used during the thesis, this one will also have a physical basis for modelling the possible movements of the players we want to analyse. For doing so, the same model described in equations (5) - (8) will be used to compute the points on the pitch that the players can reach during the play. The only thing that we need to know, apart from the current player position and speed (which will determine the centre of the reachable area with equation (7)) is the time between the start and end frame (which gives the radius of the area, as in equation (8)).

Once this reachable area is computed for all the players we want to analyse, 200 points are randomly sampled from each of them, the corresponding function to maximise is computed for every point and we keep the one that gives the optimal result according to the role that each player has in the play.

So once we have defined everything we need, to sum up, the process that needs to be followed for simulating an attacking play is the next one:

- **Find the play** that is going to be studied and obtain the correspondent starting and ending frame numbers from the tracking data.
- **Select the players** whose movements we want to analyse and **assign their roles**. This process, as said before, should be done manually by watching the real video of the game.
- **Compute the reachable areas** for all the players that we want to simulate and sample 200 random points from each of them.
- **Apply the respective function to maximise** (depending on the player's role) to each of those points and select the one that maximises it.

7 Results

With these different but complementary models that cover three of the most important collective situations in football (passing, spatial dominance and generating danger) we have a strong basis to simulate and analyse different in-game situations. Firstly, we will make use of the pass success probabilities and pass impact models for studying pass decisions and disruptive runs for opening spaces during an attacking situation and, in the end, our PIC model to simulate the optimal movements of the players will be put into practise together with the pitch control model.

All the example situations that will be used in this section (as well as the other two previously used for illustrating the models) are real in-game situations for different games played by F.C. Barcelona and Hammarby IF Fotboll. The tracking data from Barcelona's matches has been provided to us by this football club itself, while the Swedish-based company *Signality* is responsible for acquiring the ones corresponding to Hammarby's games.

It is also important to mention that the analyses that will be performed here, even though based on numerical and probabilistic results from the different individual models, will be done in a qualitative way, since the final goal in the future of these implementations is to be a useful tool for coaches and players. So, having semi-open and interpretable results is paramount in them, in order to be adaptable to the many different ways of understanding football that each coach and club has.

7.1 Analysing pass decisions and disruptive runs.

7.1.1 Hammarby incorrect pass decision.

The first situation that will be analysed comes from the match played on the 2nd match day of 2019 Allsvenskan league between Hammarby and Kalmar FF (video of the play available in *this link*). Figure 7 shows a visualisation of the situation using the tracking data, where Hammarby and Kalmar players are represented in green and red respectively and the ball trajectory for the next frames is shown with smaller blue dots:

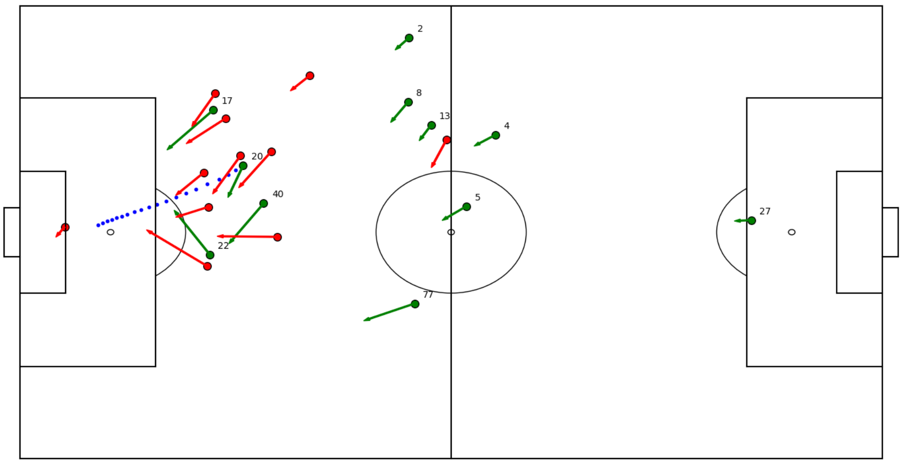

Figure 7: Tracking data view of the match situation in minute 11:44 of the first half in the game Hammarby vs Kalmar.

Hammarby is attacking to the left, with Alex Kacaniklic (number 20) carrying the ball and sending a pass forwards exactly at the plotted frame. It is easy to see just by observing the current players' positions and speeds that there are two possible dangerous pass options into the space behind the opponent's defensive line: one in front of Muamer Tankovic (number 22) and the other one ahead of Vidar Kjartansson (number 17). The sender of the pass chooses the first option, but the pass is not successful and ends up being intercepted by Kalmar's goalkeeper.

A great analysis of this rather simple-looking play can be performed with both the pass probabilities model that we implemented and the pass impact model, since they can give us the answer to questions like: Who made the mistake, the sender of the pass or the receiver, or none of them? Was the pass sent to the most appropriate player or should it have been sent to the other teammate? Even if the pass resulted being unsuccessful, was it worth to take that risk?

With the next figure 8 we will check the different possibilities that Kacaniklic had at the moment of the pass and a discussion about the first two questions will be done.

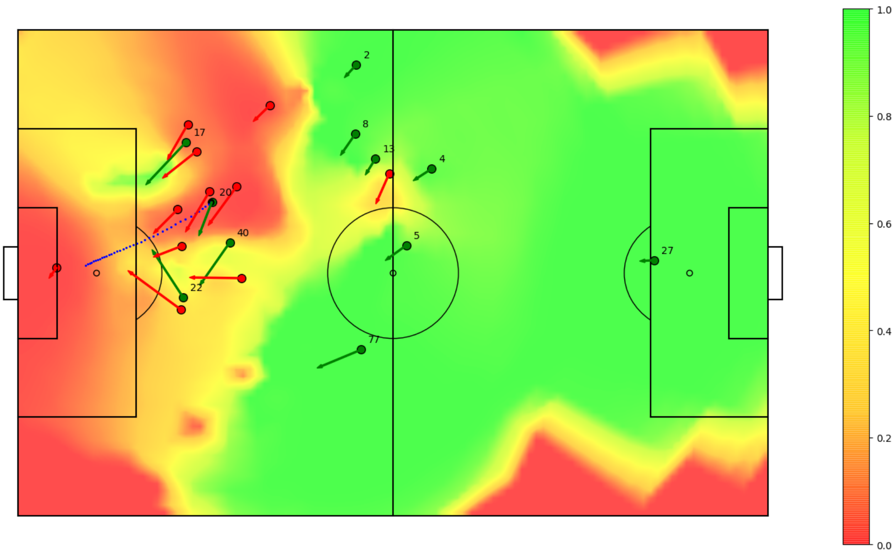

Figure 8: Pass success probabilities at the match situation in minute 11:44 of the first half in the game Hammarby vs Kalmar.

From this plot, the main thing that we can observe at first glance is that the sender of the pass chose the least safe option between the two open passing opportunities that he had. It can be seen that the probability of success for the pass that he made towards the space in front of Tankovic (number 22) is very low along the whole trajectory of the ball, with its maximum being only 22 %. However, the other pass option towards the open space created by Kjartansson's run (number 17) could have been the best one, reaching a probabilities of pass success around 90 % inside the opponent's penalty area.

Nevertheless, when you are in an attacking play, it is highly probable that you need to take risks when making decisions in order to create dangerous situations that could lead to scoring a goal. So, whether this was one of these cases when taking the risk was reasonable due to the possible outcome in the case of a teammate getting the ball or not can be determined with the mix of the already shown pass success probabilities model and the pass impact one.

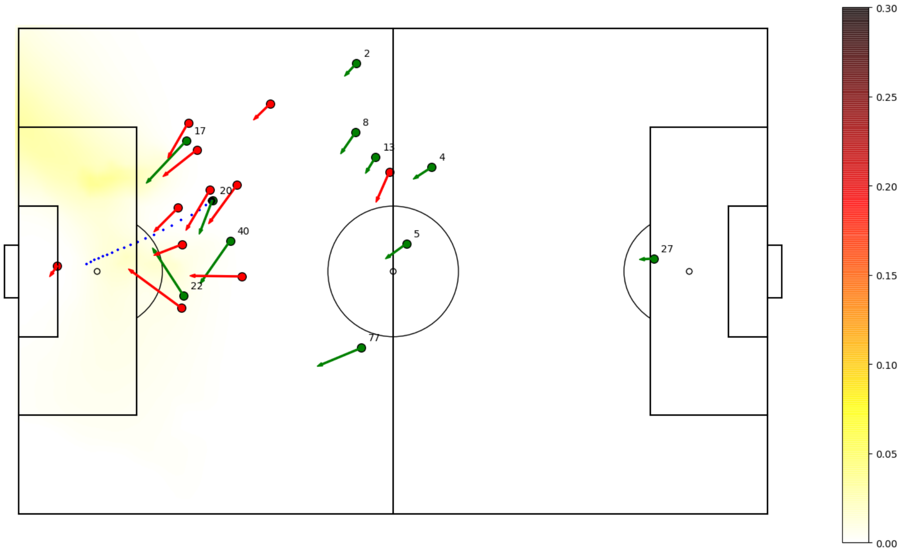

Figure 9: Goal scoring probabilities at the match situation in minute 11:44 of the first half in the game Hammarby vs Kalmar.

Figure 9 shows the multiplication of the probabilities of pass success and the pass impact value for this situation. The result of combining these two models could be considered to be a way of determining the total probability of scoring a goal as a function of the interception point of the pass, since we are taking into account the probability of successfully completing the pass and the probability of that pass ending up as a goal.

It can be seen also with that figure that the best option was sending the ball to Kjartansson instead of making the pass towards the penalty spot, because the probabilities of scoring a goal are also significantly higher for the alternative pass situation (4.5 % vs 1.8 %).

Finally, the last method of analysing the pass decision can be done computing pitch control as shown in figure 10 down below:

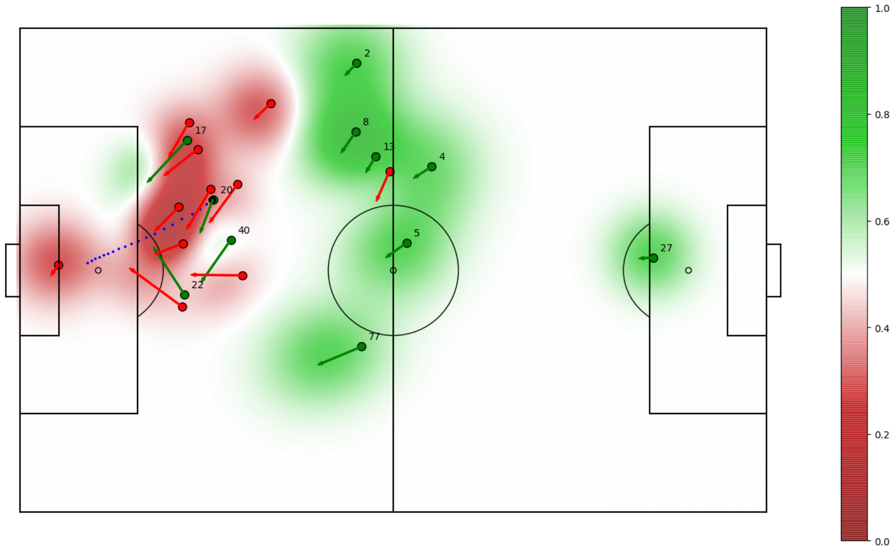

Figure 10: Pitch control at the match situation in minute 11:44 of the first half in the game Hammarby vs Kalmar.

Again, it can be clearly seen how player number 17 has created a spatial dominance in the area in front of him with his disruptive run, which means that he has a great chance of controlling the ball before any Kalmar's defender if a pass is sent there. On the other hand, if we follow the real trajectory of the ball during this situation in the match, we can observe again how the simulation tells us that successfully receiving the pass was fairly unprovable.

7.1.2 Hammarby disruptive run.

The second type of play that can be studied using the basic models are pure disruptive runs, where a certain player starts running forwards with the intention of receiving a deep pass behind the opponent's defensive line. By computing the pass possibilities at several frames during the run, it is possible to observe how open spaces are created behind the defenders and whether the passer sent the ball at the appropriate moment or not.

For this case of study we will analyse a play from the match Hammarby vs Djurgårdens match played on the 6th match day in Allsvenskan league. The situation takes place at minute 87:37, with Hammarby attacking to the right side of the pitch (watch the situation in *this video*). Figure 11 shows the situation represented using the tracking data.

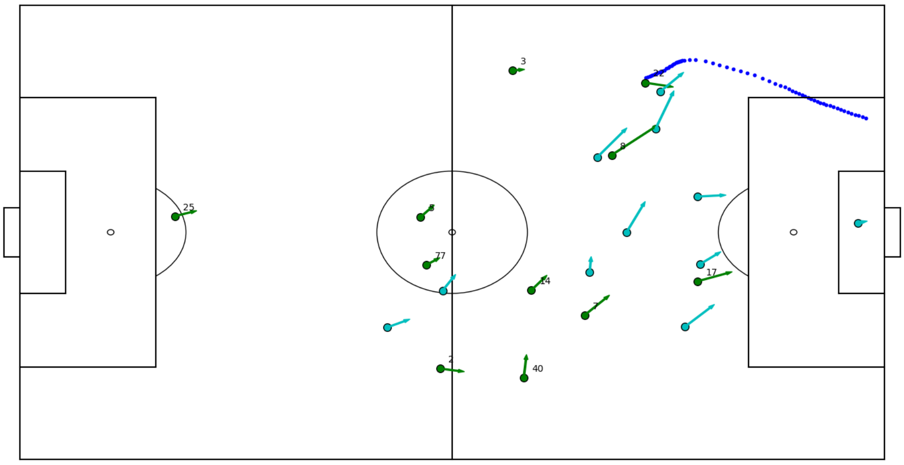

Figure 11: Tracking data view of the match situation in minute 87:37 in the game Hammarby (green) vs Djurgårdens (blue). The smaller blue dots represent the real trajectory of the ball in the next frames.

In set of figures 12 the evolution of the situation during the disruptive run that Jeppe Andersen (number 8) does can be observed at the top-right quadrant of the images. As can be seen on them, the run is perfectly done, opening more and more space in front of him as he moves through the defensive line. Furthermore, Tankovic (number 22) has enough patience to hold the ball until the space is fully open and a safe pass can be made, sending it with the appropriate direction which leads to a success probability of 75 %. A *.gif* file with the whole evolution of the pass success probabilities during the disruptive run is available through *this link*

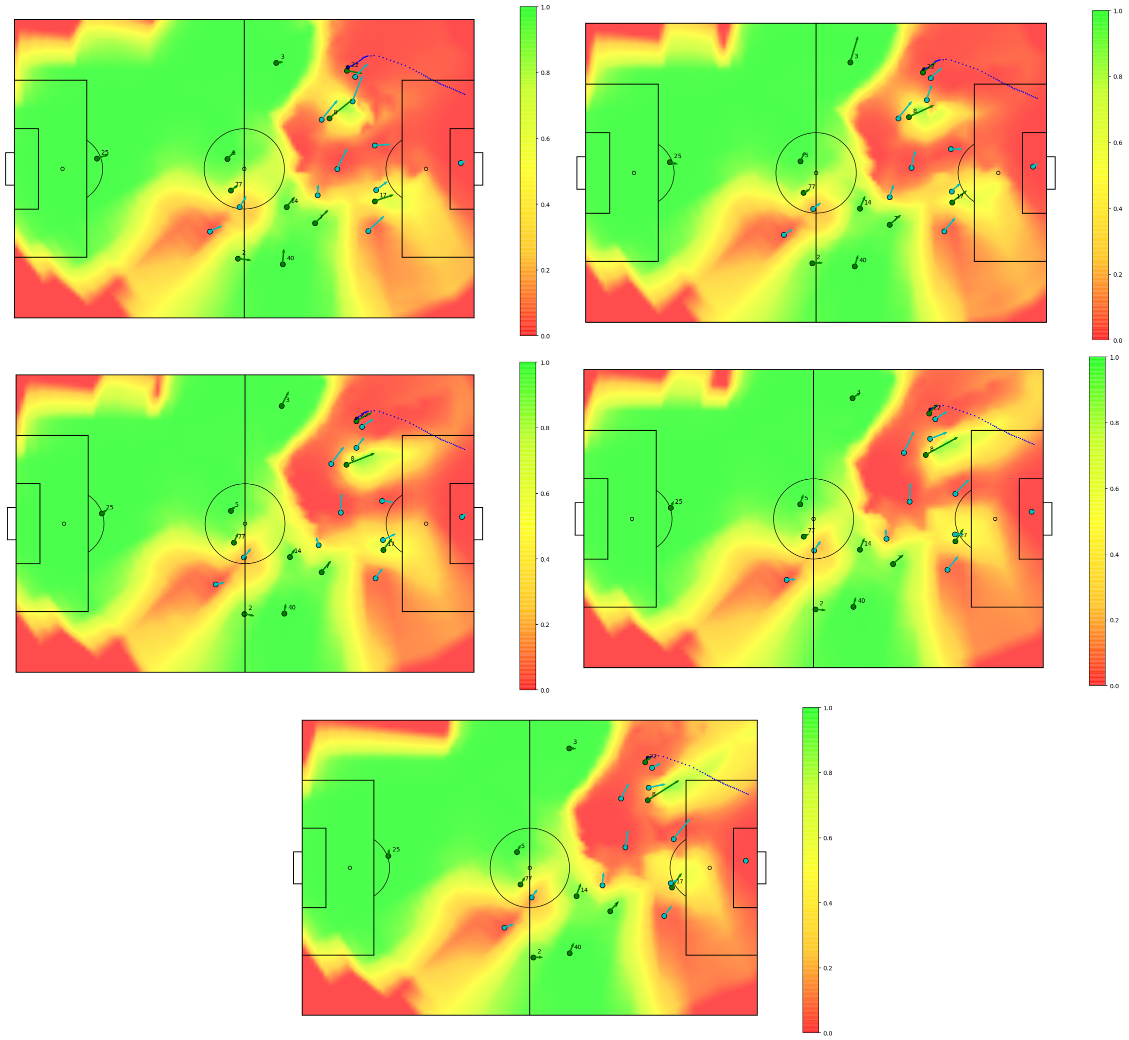

Figure 12: Evolution of the pass possibilities during the match situation in minute 87:37 in the game Hammarby (green) vs Djurgårdens (blue). Time advances from left to right and from top to bottom in steps of 10 frames (0.4 s).

7.2 Analyses on players positioning using the PIC model.

7.2.1 F.C. Barcelona vs Real Betis deep pass.

The first situation where our PIC model will be applied on is a play that belongs to the match F.C. Barcelona (red) vs Real Betis (green), the video clip of the attacking play can be accessed *here*. Figure 13 shows the positions of the players and their velocities at the moment of the start of the play, together with the roles that have been assigned to them.

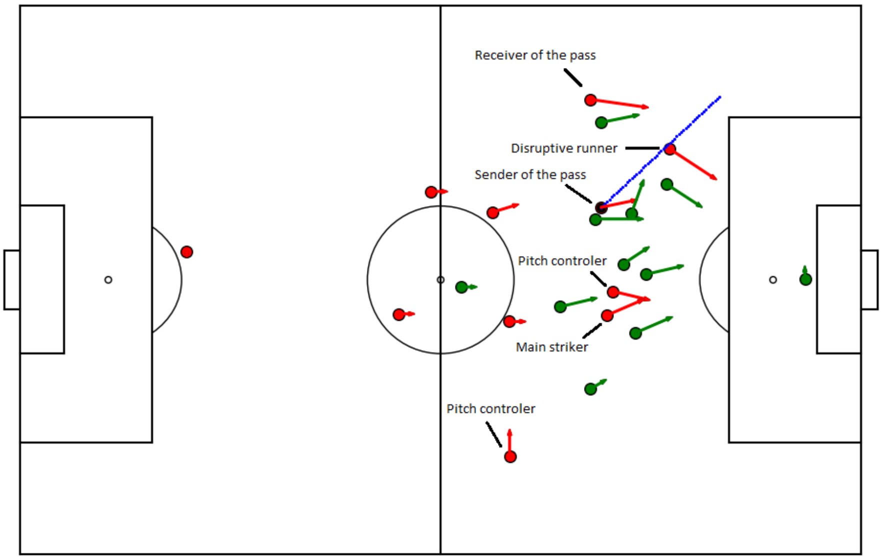

Figure 13: Tracking data view of the case of study during the match F.C. Barcelona (red) vs Real Betis (green).

This example play is one of the two main different kind of situations that can be studied with our optimal positions model, where a disruptive run takes place in order to open space behind the defensive line. However, this case has the particularity that the disruptive runner is not opening space for himself, but for the teammate that is running closer to the sideline of the pitch, by dragging one of the defenders with him because of his run. This space that has been opened by the disruptive runner is shown in figure 14, where the pass probabilities model is applied to the frame when the pass is made. It can be seen that the area around the point where the pass is received has a pass success probability between 65 and 70 %.

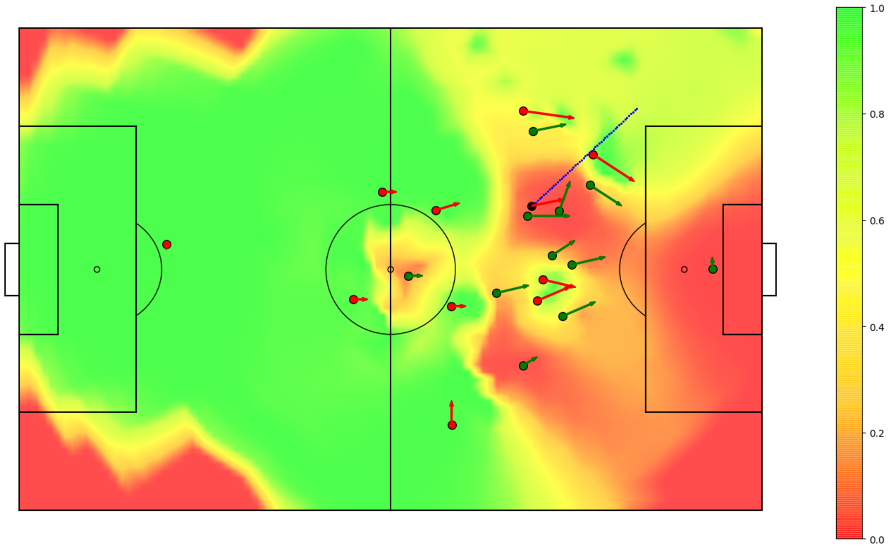

Figure 14: Pass success probabilities of the moment of the pass for the case of study at the match F.C. Barcelona vs Real Betis.

Once the pass is made, all the attacking players change their positions, preparing for the next situation and all of them trying to find the location that they find to be the best for their interests. The following figure 15 shows the pitch control for the moment when the pass is made and when the receiver intercepts the ball.

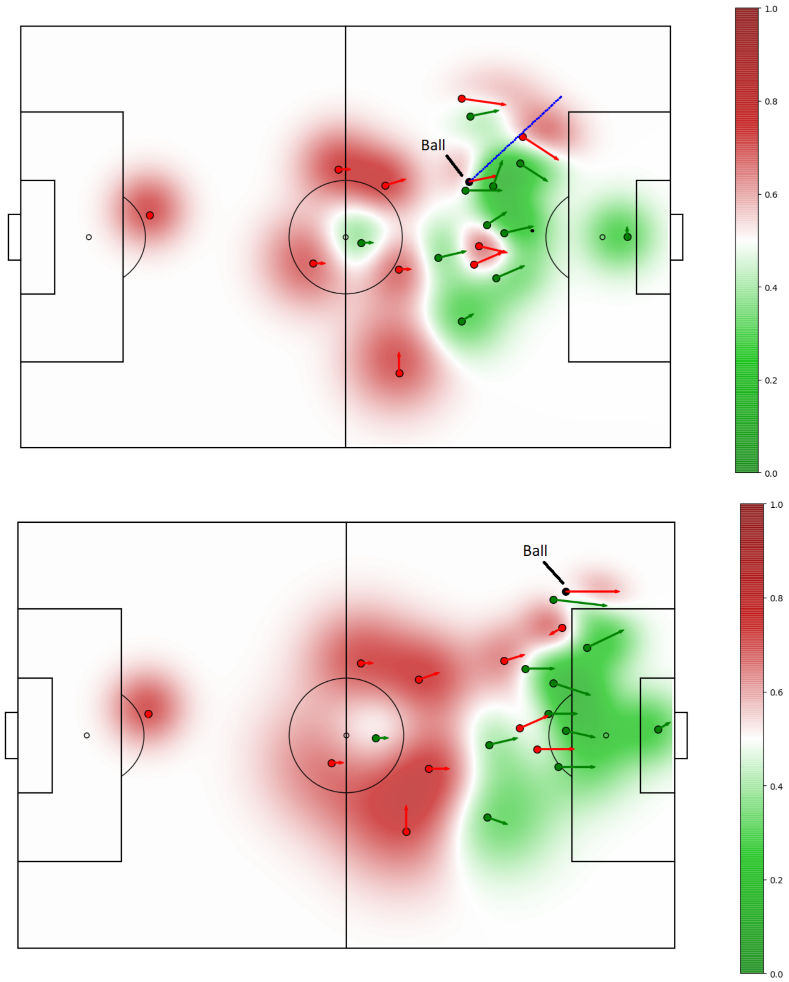

Figure 15: Pitch control for the example play at the moment of the pass (top plot) and at the moment of the interception (bottom figure).

At this point, knowing the real positions where the players moved once the pass is received by their teammate, we can apply our model to some of them to determine whether the movements that they did were close to the optimal ones according to our model or if they should have done something different in order to perform better in their role during the play.

We will start analysing a rather simple one, the right-back player, who was situated in what we called the cooperation zone, the area where players do not expect to receive the ball from the original pass. As we established, he should remain around the defensive zone, taking care of the opponent's strikers in case that there is a rebound or a quick counter-attack, i.e., his goal is to maximise pitch control around that area of the pitch. As can be seen in figure 16, where the final real and optimal positions of this player are shown, according to our model he should have moved a bit more forward, in order to "steal" some of the positional control that the opponent that is in front of him has.

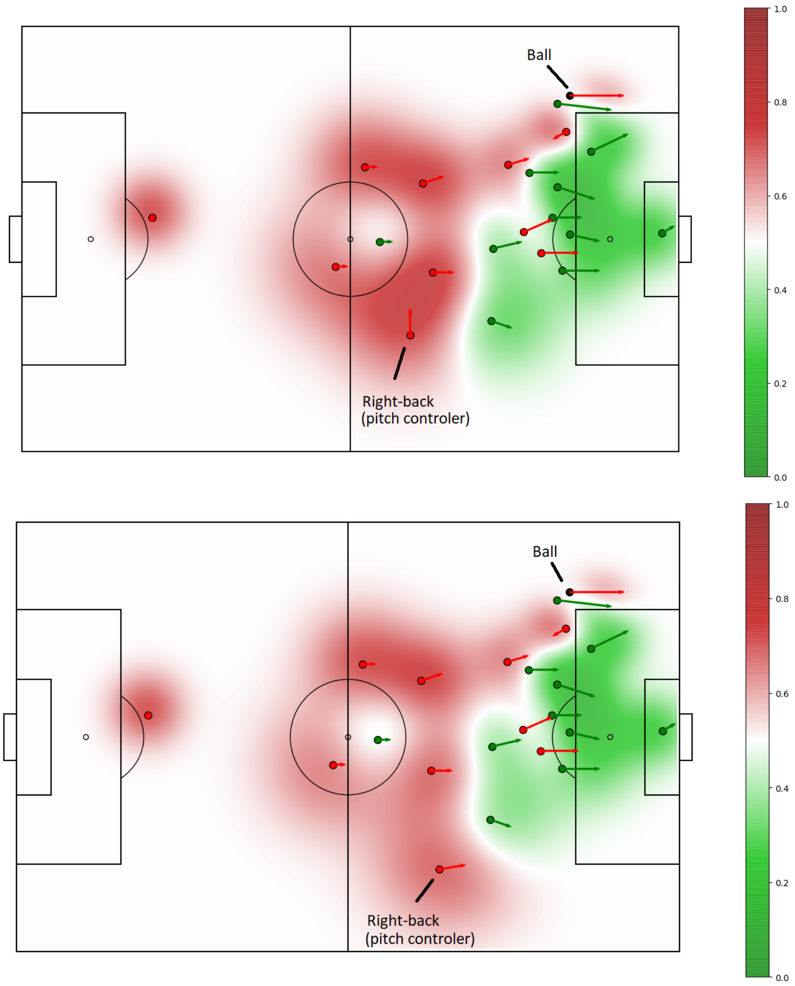

Figure 16: Real (top) vs optimal (bottom) position for the right-back player of F.C. Barcelona.

The next player that we focus on is the one that we decided that, considering his position and speed direction at the beginning of the play, should be the main striker, i.e., the player that should go towards a dangerous position with the maximum pass impact value. We can observe in figure 17 how his position gets significantly modified, having his optimal position in a more offensive point, being placed right at the border of the area and running in the direction of the opponent's first post. This position gives him, for sure, an advantage over the defenders if the ball is quickly sent into the area by the receiver or the pass, which is what we were looking for when we set his role in the play: being a menace for the defensive line.

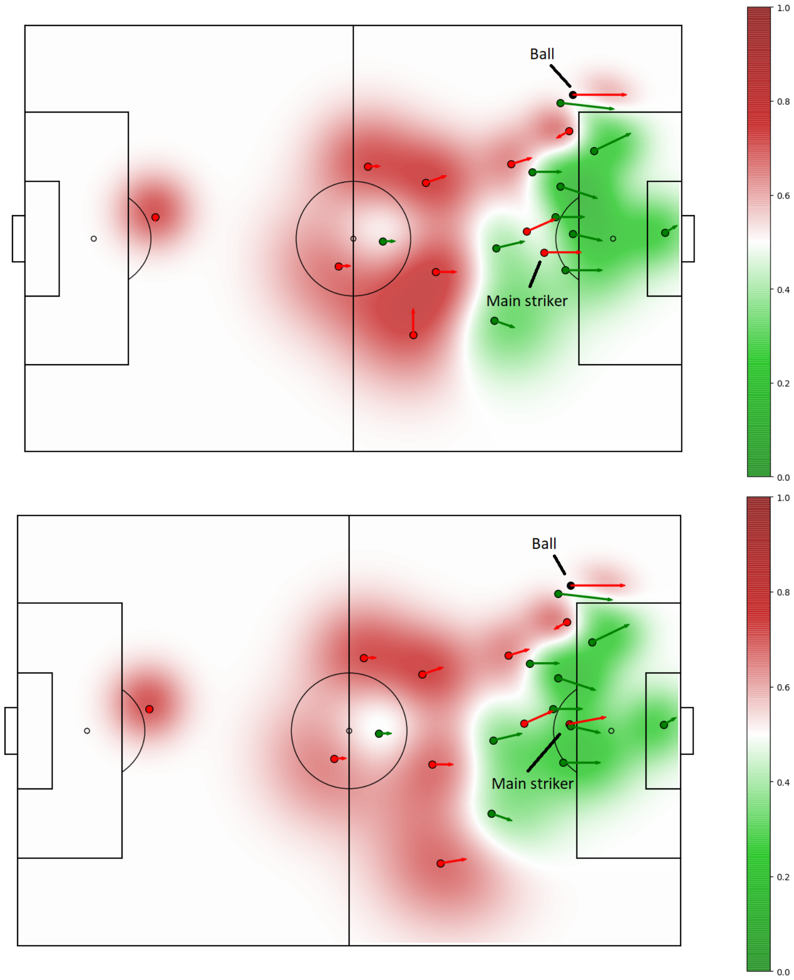

Figure 17: Real (top) vs optimal (bottom) position for the main striker of F.C. Barcelona.

When we get, as showed in figure 18, to analyse the position of the secondary striker, which is also set to gain as much pitch control as possible while having pass possibilities, we find a behaviour that differs from the previous two player that we have studied: in this case, the optimal position is situated some meters behind the real one. This allows the team to gain pitch control in a zone which is considered to be extremely valuable in football because a lot of rebounds end up there. Positioning this player in that area, results in the fact that the player of the defending team is not easily dominating it as he does in the real situation and, in the case of a rebound, there is an opportunity to keep the possession of the ball.

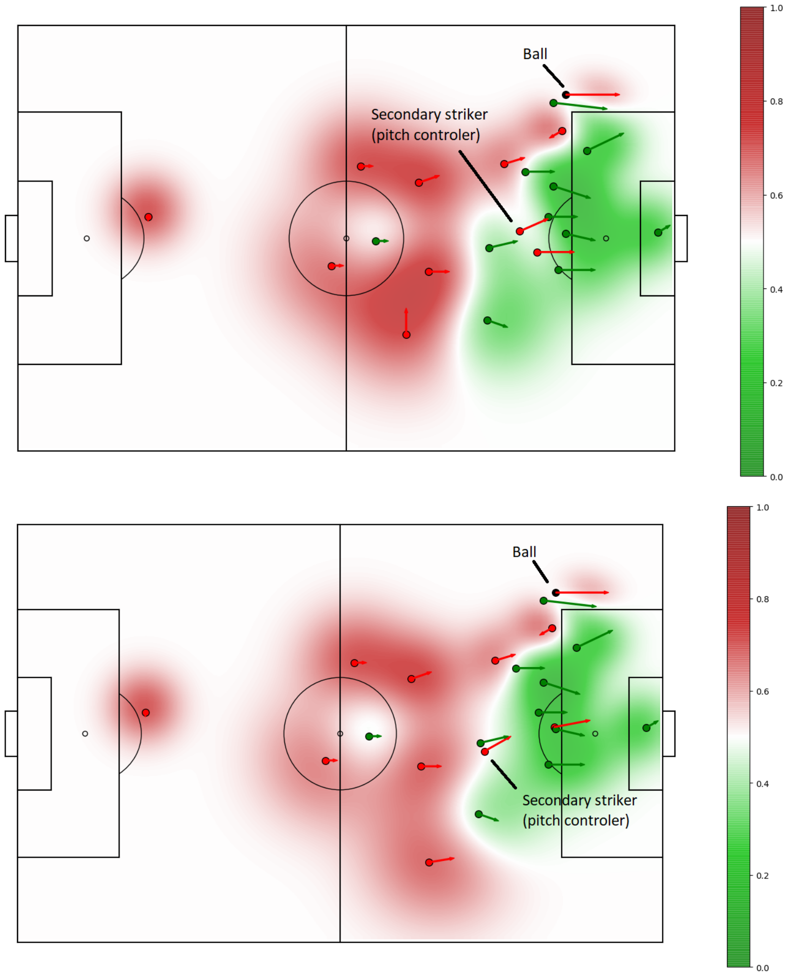

Figure 18: Real (top) vs optimal (bottom) position for the secondary striker of F.C. Barcelona.

However, among all of the players, the one that certainly catches our attention is the disruptive runner, as in the middle of the pass he stops his run and even moves slightly backwards to position himself right outside of the penalty area instead of keeping running towards a dangerous position inside the box (see the supporting video file). This might look strange at first if we think that he might want to receive a pass inside the box and try to score a goal. But the fact is that, as shown with our model, what he is doing is moving to the position with the highest possible value (as in our model he is set to go to the position that maximises the value of $PitchControl * PassProbabilities * PassImpact$). Figure 19 shows that his real position and the computed optimal one are almost the same, confirming that the player made a good decision in terms of having the best possible balance of controlling his pitch zone, having possibilities of receiving a pass and being in a dangerous position for the defenders.

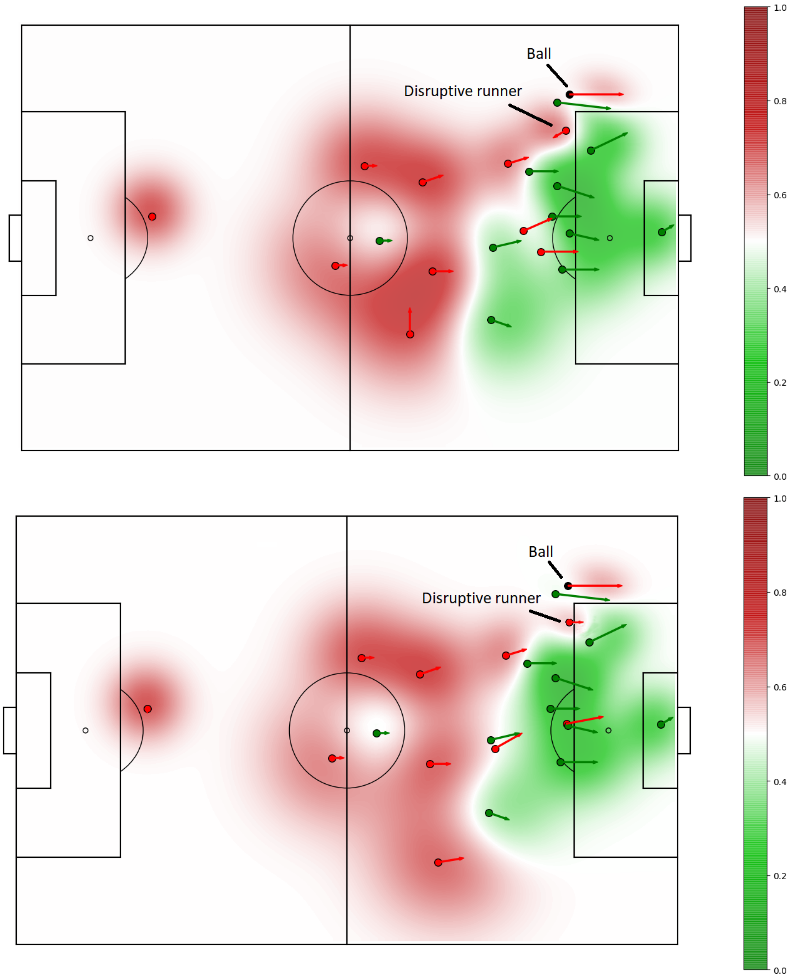

Figure 19: Real (top) vs optimal (bottom) position for the disruptive runner of F.C. Barcelona.

Finally, the following figure 20 shows the differences in pitch control between the real final situation and the optimal positions computed by our model. Note that in this case, the colormap does not show the value of the pitch control for each of the teams, but rather the areas where pitch control increases (in red) or decreases (in green) for F.C. Barcelona with respect to the original situation. There, it can be observed how there are important gains in positional control around the penalty spot due to the new dangerous position of the main striker and also in the zone situated at three fourths of the pitch due to the roles that have been assigned the right-back and the secondary striker, in order to have more dominance in that important area.

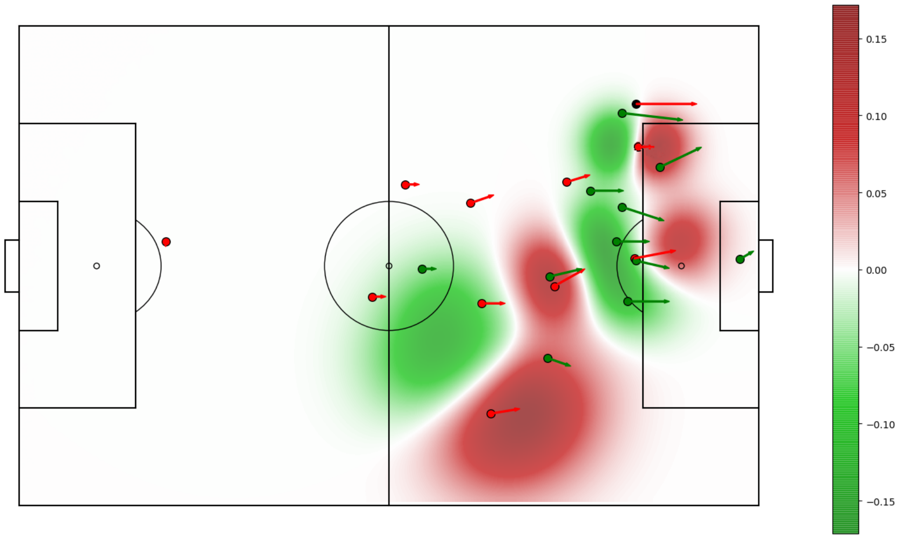

Figure 20: Differences in pitch control between the real and the optimal positions in the match F.C. Barcelona vs Real Betis. Red colours show the zones where Barcelona increases pitch control and green shows decreasing pitch control.

7.2.2 Hammarby vs Kalmar disruptive run and dribbling.

This second example play that we will analyse using our optimal positions model corresponds to the match played between Hammarby (green) and Kalmar (red) in the 2019 season of the Swedish league Allsvenskan. The situation takes place in the first half of the game, with Hammarby attacking to the left side of the pitch. Vidar Kjartansson (number 17) receives the ball at one of the corners of the penalty area and runs with it towards the end line of the pitch; in the meantime his teammates try to position themselves in the best possible spots for an imminent pass (see the *supporting video file*). Figure 21 shows the starting frame of the situation, with the respective roles assigned to the players.

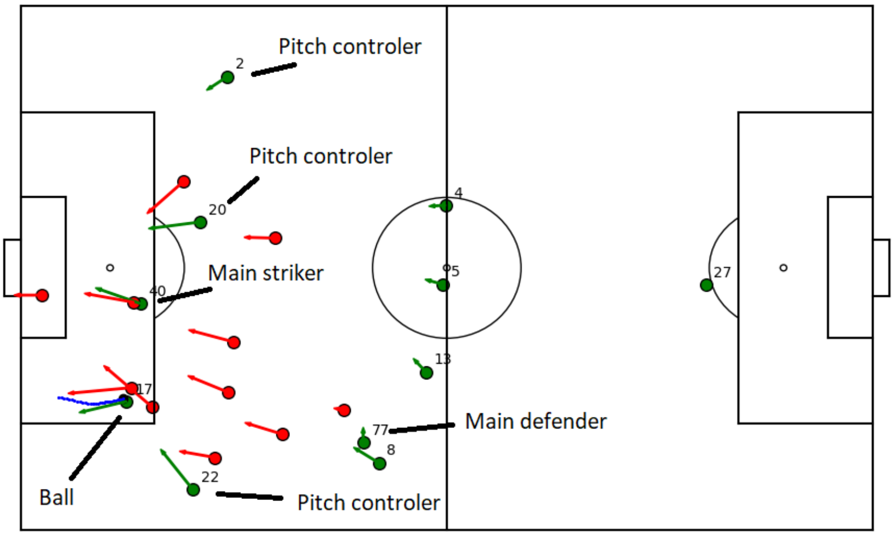

Figure 21: Tracking data view of the case of study during the match Hammarby IF (green) vs Kalmar FF (red).

The first player to simulate will be the main striker, Nikola Djurdjic. As stated in his role, he should have gone toward the maximum pass impact point inside his reachable area, and, as shown in figure 22, he did quite a good job ending up in a position which is really close to his optimal one.

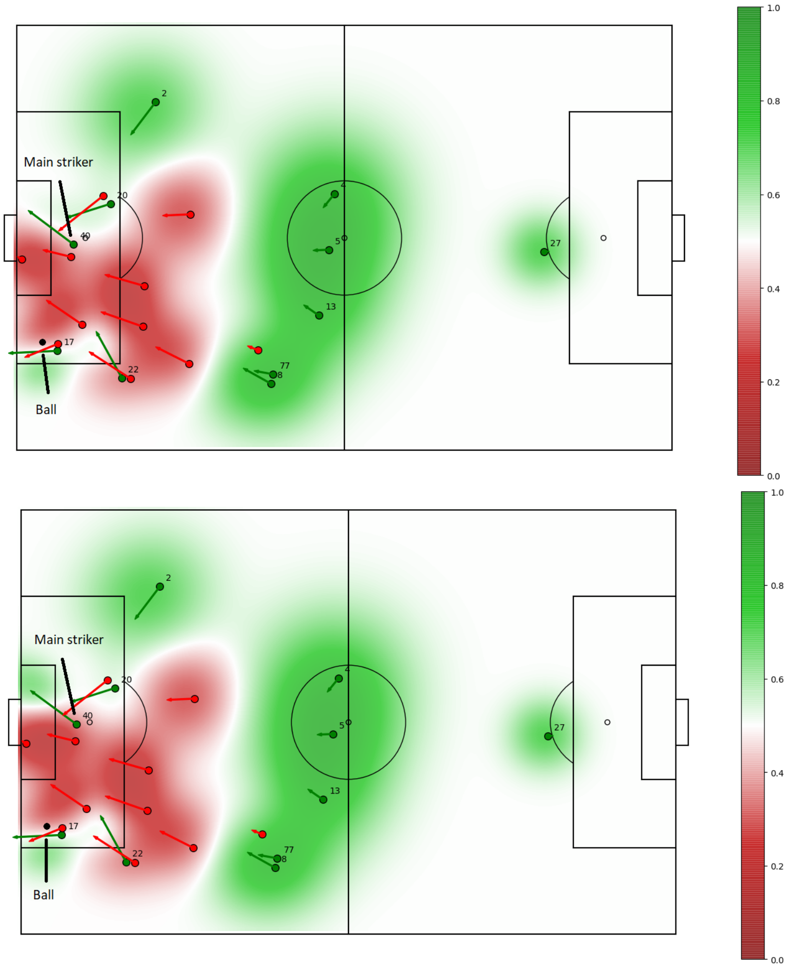

Figure 22: Real (top) vs optimal (bottom) position for the main striker of Hammarby.

Now we will focus at the same time on the three players that, even though are situated in different areas of the pitch far away from each other, have been assigned the same role: pitch controllers. Figure 23 shows their real positions and their optimal ones according to our model. It is interesting to see how, even having the same rule to follow, they move in a different way in this play. Starting with the right-winger, his optimal position is almost the same as the real one that he has at the end of the play, since he is holding a lot of positional control over the whole area around the top-left corner of the pitch. With respect to the secondary striker, he should have moved a bit more to the centre of the penalty area, so that in that way he would have had much more pitch control over the points around the penalty spot, which is a great and dangerous zone for receiving a pass and try a shot. Finally, the most interesting one of these three is the left-winger who in the real play stays too close to one of the defenders, having a useless position since he is barely controlling any part of the area around him and receiving a pass there is almost impossible. What the model says is that he should have stayed closer to the sideline, having in this way both much more control over the bottom-left corner in the case of a rebound and the possibility to receive a pass from the player with the possession of the ball in the case that he gets blocked by his defender.

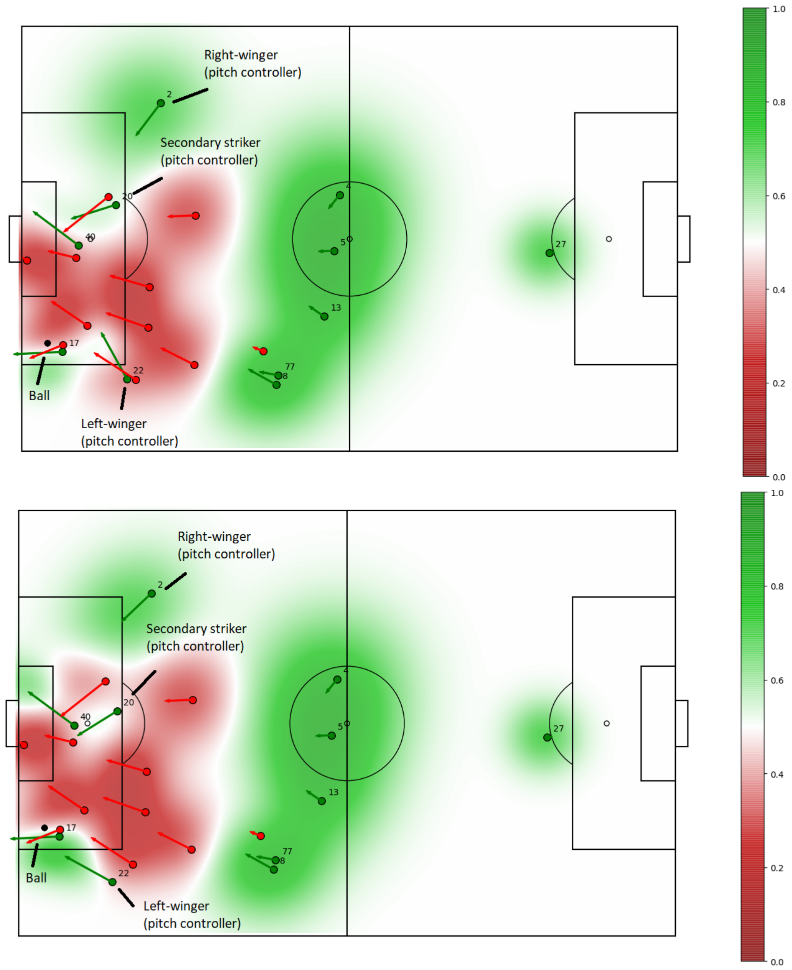

Figure 23: Real (top) vs optimal (bottom) position for the pitch controllers of Hammarby.

The last player that we simulate is the one with the "main defender" role, Mats Solheim (number 77), responsible of taking care of the most offensive player of the defending team. Even though in the real match situation he seems to be sufficiently close to the striker, the simulations tell that he should have been a bit closer, in order to gain some pitch control there, as shown in figure 24.

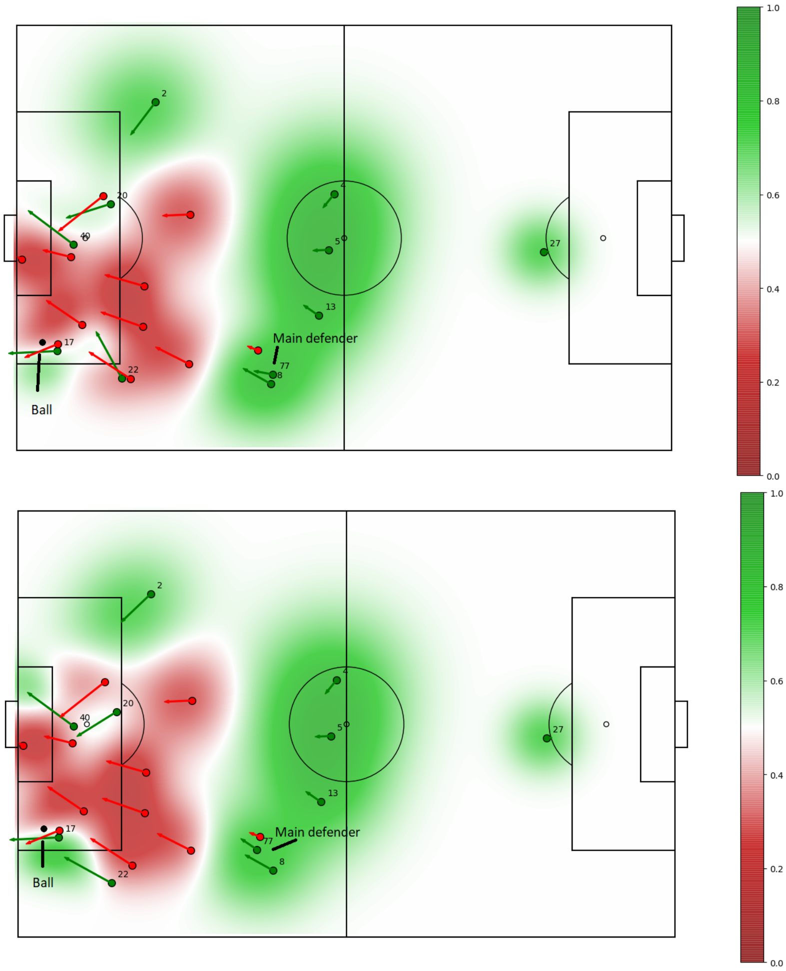

Figure 24: Real (top) vs optimal (bottom) position for the main defender of Hammarby.

Finally, as we did with the previous example, an image showing the differences between the pitch control in the real situation and the optimal one is can be seen in figure 25, where all the previously described changes in positional dominance are noticeable, specially in the highly valuable zone between the goal, the penalty spot and the penalty arc.

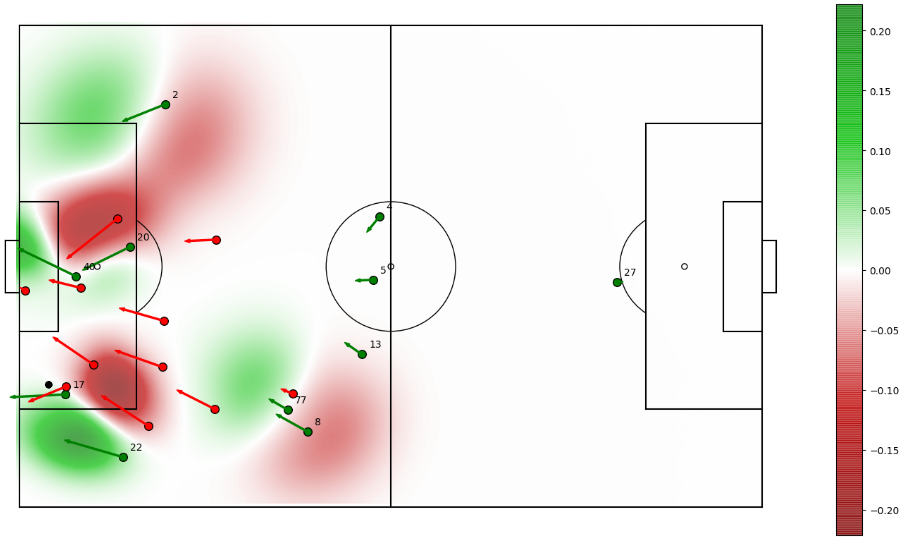

Figure 25: Changes in pitch control between the real and the optimal positions in the match Hammarby vs Kalmar. Green colours indicate gains and red shows losses in pitch control for Hammarby.

7.2.3 Hammarby vs Östersunds different choice of space occupation.

In this last situation we will show an example on how, even though the movements of the players during an attacking play could seem correct, there are always possibilities to change them and get a different ending situation that might result in higher chances of scoring. This case belongs to the game Hammarby vs Östersunds from the 2019 Allsvenskan; the local team is attacking to the right, with Kjartansson (number 17) dribbling with the ball and trying to send a pass, as shown in figure 26. We will simulate the positions for Tankovic, Djurdjic, Khalili and Andersen (numbers 22, 40, 7 and 8, respectively) to check what they could have done differently to perform better with their respective roles in the play (the video of this situation is available *here*).

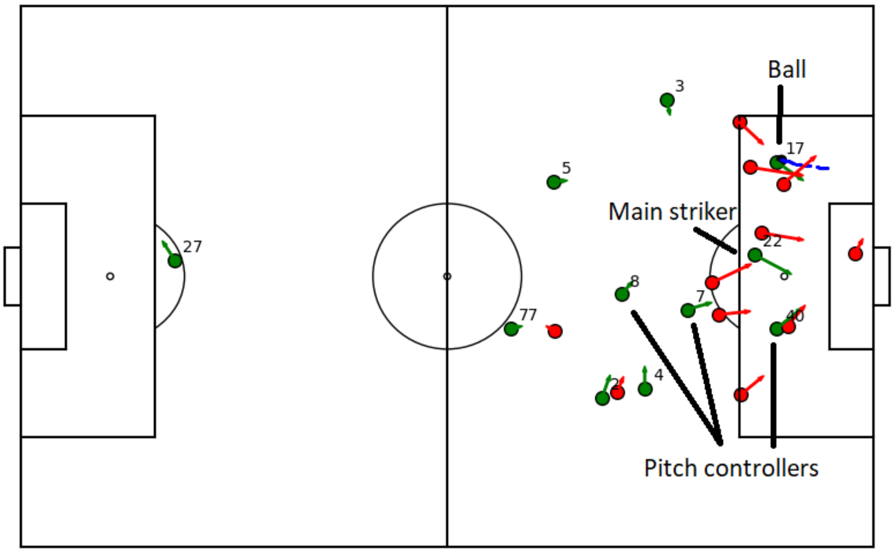

Figure 26: Tracking data view of the case of study during the match Hammarby IF (green) vs Östersunds (red).

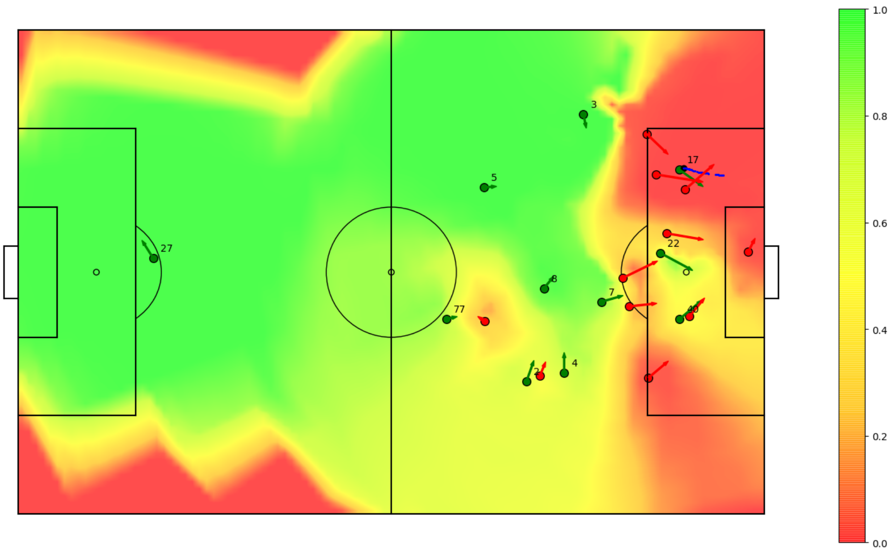

Figure 27: Pass success probabilities of the first frame of the play for the case of study at the match Hammarby vs Östersunds.

First of all, in figure 28, the initial and final real positions for the players are showed together with the pitch control. The first thing we observe that could have been a mistake is that none of the attackers position himself around the penalty spot, specially considering that, as showed in figure 27, there is a considerably high probability of success if a pass is sent there (around 65%)

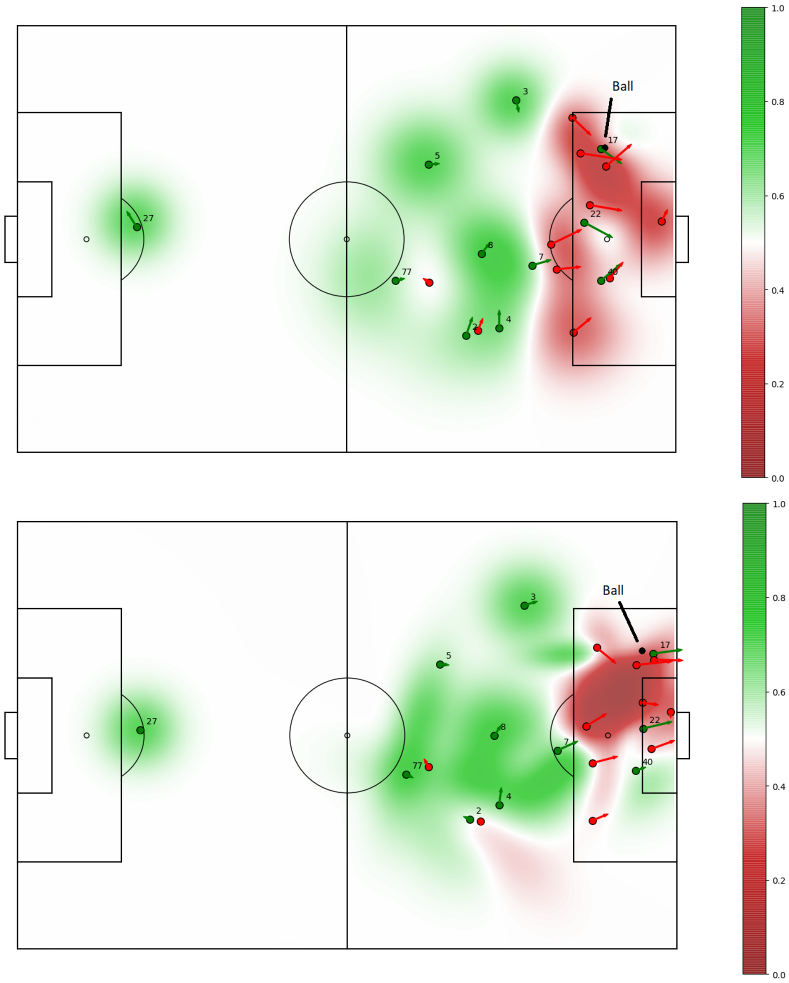

Figure 28: Pitch control for the example play at the first frame (top plot) and the last one (bottom figure).

We start computing the optimal position for Tankovic, who was selected to have the main striker role due to his position at the beginning of the play. In figure 29 we can see how he did almost exactly the same run as suggested in our model, going to the point of maximum pass impact.

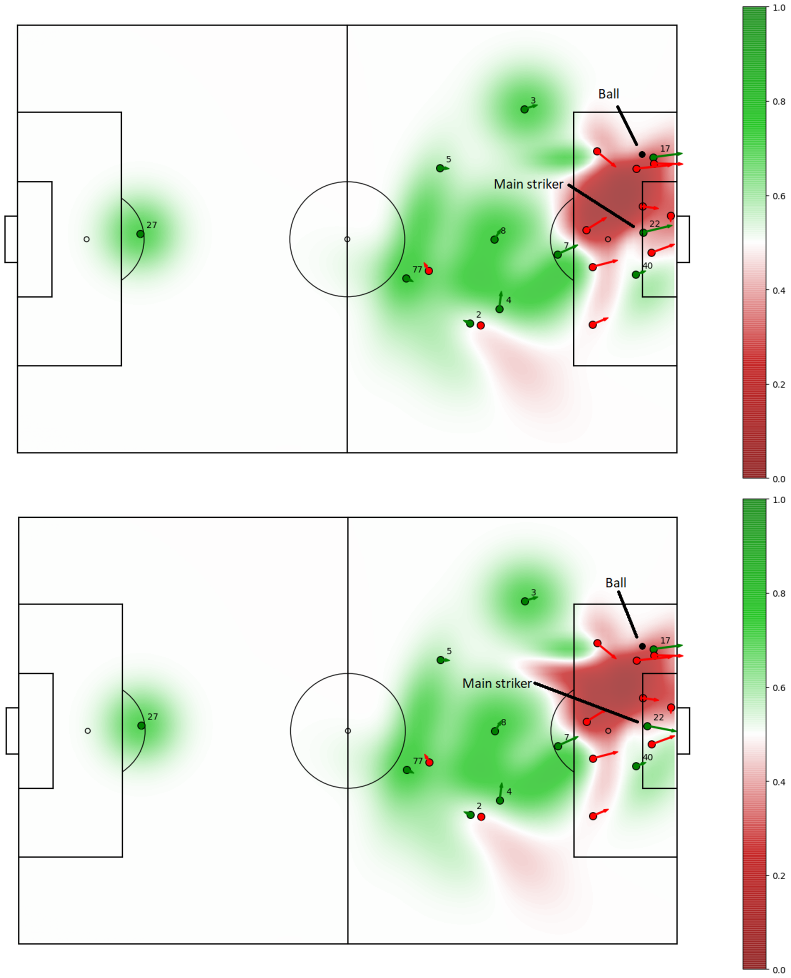

Figure 29: Real (top) vs optimal (bottom) position for the main striker of Hammarby.

Now we will compute the corrected positions for the other three players to simulate, as all of them are assigned to be pitch controllers and we will see in figure 30 how they complement each other's movements while going to better positions to receive a possible pass. The first player to focus on is Khalili (number 7), who should have positioned himself closer to the penalty spot inside the box. However, this position that the simulation suggests would have left a really valuable pitch area (the one right before the penalty arc) uncontrolled if no further adjusts are made, but this is solved when computing the optimal position for the player number 8 (Andersen), whose position is also moved forward to get control over that zone. Finally, Djurdjic's position (number 40) is also moved closer to the penalty spot, leaving some space between him and Khalili for having as much control of that area as possible.

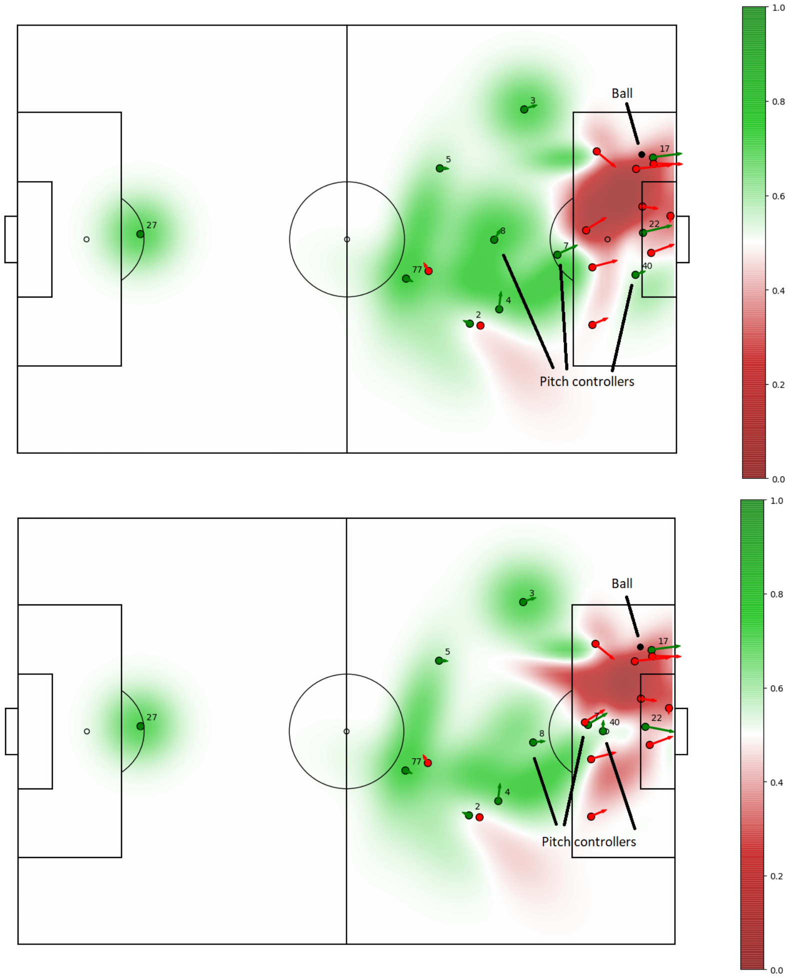

Figure 30: Real (top) vs optimal (bottom) position for the pitch controllers of Hammarby.

The last figure 31, shows the differences in pitch control between the real positions of the players and their optimal ones. As can be seen, pitch control is mainly increased in the central zone of the penalty box and arc, while it mainly decreases at the second post, due to the change in Djurdjic's position.

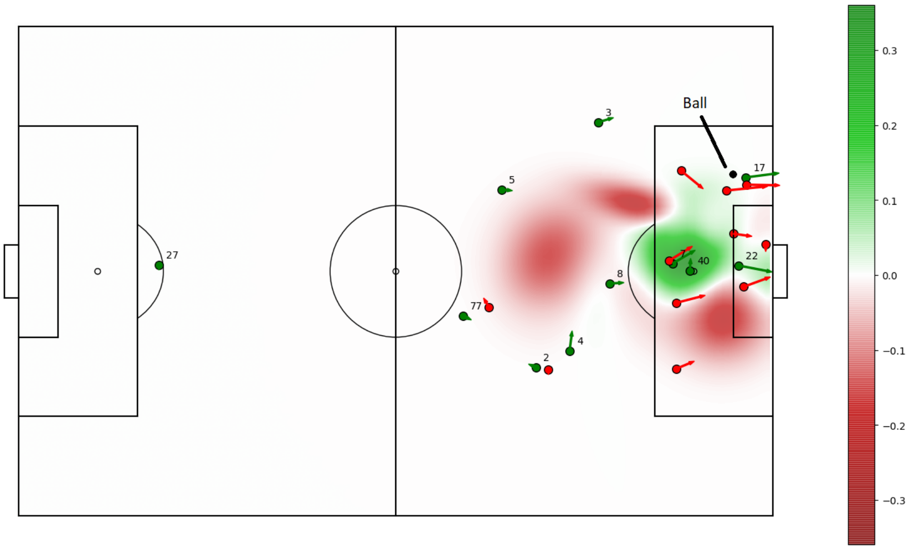

Figure 31: Differences in pitch control between the real and the optimal positions in the match Hammarby vs Östersunds. Green colours indicate gains and red shows losses in pitch control for Hammarby.

8 Conclusions

Throughout this project we have been able to successfully implement some models that allow us to quantify three paramount aspects in football: pass success probabilities, pitch control and pass impact. Furthermore, the three of them have been used as a basis to build up a completely new model designed that has allowed us to simulate and analyse the movements of football players during an attacking play, something that is new in the world of data analysis.

First of all, with respect to the pass probabilities model, we took the probabilistic approach and equations that Spearman et al. [18] proposed in their 2017 paper as a starting point for our implementations and made some changes and improvements to the equations of motion for both the players and the ball, resulting in a physics and probabilistic-based model that gave us a really solid foundation and ended up being the main pillar of the rest of our work. It allowed us to analyse and study a lot of different situations and made it possible to quantify in terms of probabilities what can be considered to be a good pass decision or not. Furthermore, combining it with *Twelve*'s pass impact algorithm gave us the possibility of studying these passes not only in terms of probability of success but also taking into account the danger that they can create, which immediately resulted in having a powerful tool to include in our optimal movements model to determine whether running towards a final position with relatively low pass success probabilities might be worth if it is threatening enough for the defenders.

Pitch control is another game aspect that we modelled and implemented for our simulations, with an attempt to use the algorithm from the extrapolation of the pass probabilities model from Spearman et al. [18], but finally adopting F.C. Barcelona's code [3] due to speed performance reasons. However, we think that Spearman's proposal for pitch control could have improved our simulations, since it would have made both the pass success probabilities and pitch control model to have the same type of approach and a strong physics basis. Making the implementation that we developed faster enough to be usable in practice is certainly one of the things that are left to be done in the future.

Finally, our PIC model was created and used as a tool for analysing the players' positioning during short attacking plays by simulating their optimal movements one or two seconds into the future. We decided to follow the procedure of establishing a few rather simple rules (even though based on complex models) that end up leading to interesting movements and decisions, just as is done in many researches about collective behaviour. These simulations that have been done in this project are a breakthrough in collective behaviour studies, since it is the first time that the approach followed by many studies in the animal world [2] [21] or in people crowds [12] has been applied to sports (where very few individuals are involved), allowing us to successfully predict and simulate the optimal future position for players during a one or two second attacking play.

We found out that the model worked fine for the kind of situations for which it was designed, being able to apply it systematically in many different cases with the only need to assign the roles to the players manually. The model has, in fact, turned out to have a great potential to become a really useful tool for managers and data analysts due to the possibility to easily change the rules that govern the movements of the players so that they adapt to the desired style of play of any team.

9 Future work

The main objectives of implementing several models to quantify different aspects in football and developing a good basis for a new model to simulate the movements of the players during an attacking play have been successfully achieved in the work that has been done. However, there are still many things that can be improved in order to make the model more efficient and even more trustworthy so that data analysts and coaches can include it as an interesting tool in their daily work.

The first main obstacle that we need to overcome is computational efficiency. If we want our algorithms to be used in real life situations, it is of utmost importance for them to be straightforward to use by anyone that is not familiarised with the code itself and quick enough to give out results within a few seconds (ideally not more than 30 or 40). Nevertheless, at the moment, the implementations that we have for the pass success probabilities and pitch control model from Spearman's paper can certainly be considered too slow to be used in practise, requiring running times of around 3 or 4 minutes for computing just one standard case of study, where the success probabilities of only one pass are computed. Also, there are some difficulties on finding the exact correspondence between the real in-game situation and its respective frames in the tracking data files, which significantly increases the required amount of time needed and makes the process of analysing situations in real time to be very difficult at this stage. Actually, an attempt to do these live analyses was performed during the Allsvenskan game between Hammarby and Östersunds (on the 9th match day of the 2019 season) and, even though we managed to analyse one of the plays during half time, there would not have been enough time left to show it to the players at the locker room if we would have wanted to, since we finished the computations when the second half was just about to start.

Some ideas about how to solve this issue have been considered during the development of the project; among those, the most doable ones are either converting the code into a GPU-parallelised programme or creating a neural network for obtaining the corresponding arrival times given the players' position and speed.

Finally, and with respect to the PIC model, there will always be aspects to improve in it, mainly regarding the roles that are assigned to the players, how they are assigned and how they should behave according to them. The most interesting path to follow in this case would be having meetings with professional coaches and players and compare their opinions on what they think that players should do in certain situations with the results obtained from the model and tune it according to this.

Bibliography

[1] Mustafa Cavcar. The international standard atmosphere (isa). *Anadolu University, Turkey*, 30:9, 2000.

[2] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.

[3] Javier Fernandez, F Barcelona, and Luke Bornn. Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *Sloan Sports Analytics Conference*, 2018.

[4] FIFA. Fifa quality programme for footballs. `https://football-technology.fifa.com/media/1133/2015_fifa-quality-programme-for-footballs_test_criteria.pdf`, 2015. [Online; accessed 16-May-2019].

[5] Akira Fujimura and Kokichi Sugihara. Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan*, 36(6):49–58, 2005.

[6] Floris R Goes, Matthias Kempe, Laurentius A Meerhoff, and Koen APM Lemmink. Not every pass can be an assist: A data-driven model to measure pass effectiveness in professional soccer matches. *Big data*, 2018.

[7] Joachim Gudmundsson and Thomas Wolle. Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47:16–27, 2014.

[8] Leroy F Henderson. On the fluid mechanics of human crowd motion. *Transportation research*, 8(6):509–515, 1974.

[9] Brian Hopkins, Andrew Churchill, Stefan Vogt, and Louise Rönnqvist. Braking reaching movements: A test of the constant tau-dot strategy under different viewing conditions. *Journal of motor behavior*, 36(1):3–12, 2004.

[10] Roger L Hughes. The flow of human crowds. *Annual review of fluid mechanics*, 35(1):169–182, 2003.

[11] Peter F Major and Lawrence M Dill. The three-dimensional structure of airborne bird flocks. *Behavioral Ecology and Sociobiology*, 4(2):111–122, 1978.

[12] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[13] Akira Okubo and HC Chiang. An analysis of the kinematics of swarming of anarete pritchardi kim (diptera: Cecidomyiidae). *Population Ecology*, 16(1):1–42, 1974.

[14] Julia K Parrish and Leah Edelstein-Keshet. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284(5411):99–101, 1999.

[15] Paul R Schrater, David C Knill, and Eero P Simoncelli. Mechanisms of visual motion detection. *Nature neuroscience*, 3(1):64, 2000.

[16] Jesse L Silverberg, Matthew Bierbaum, James P Sethna, and Itai Cohen. Collective motion of humans in mosh and circle pits at heavy metal concerts. *Physical Review Letters*, 110(22):228701, 2013.

[17] William Spearman. Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference*, pages 1–17, 2018.

[18] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. Physics—based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.

[19] Limonta Sport. Third generation artificial turf. `http://www.tadmur-trading.com/upload/file/1833051086_SOCCERPRO.pdf`. [Online; accessed 20-May-2019].

[20] Tsuyoshi Taki, Jun-ichi Hasegawa, and Teruo Fukumura. Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 815–818. IEEE, 1996.

[21] Andrew J Wood and Graeme J Ackland. Evolving the selfish herd: emergence of distinct aggregating strategies in an individual-based model. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1618):1637–1642, 2007.

A Appendix I. Solution to the players' equation of motion

The differential equation that we want to solve is:

$$m\frac{d}{dt}\vec{v} = \vec{F} - k\vec{v} \tag{15}$$

Which can be rewritten considering $\frac{d}{dt}\vec{v} = \vec{a}$ and, without any loss of generality, we can solve it in one dimension, having:

$$a = \frac{F - kv}{m} \tag{16}$$

$$\frac{dv}{dt} = \frac{F}{m} - \frac{k}{m}v \tag{17}$$

Equation that can be solved by separation of variables,

$$\int_{v=v_0}^{v} \frac{dv}{\frac{F}{m} - \frac{k}{m}v} = \int_{t=0}^{t} dt \tag{18}$$

Now we can make the change of variable $u = \frac{F}{m} - \frac{k}{m}v$, so $dv = -\frac{m}{k}du$ and we have

$$\int_{u=u_0}^{u} \frac{-\frac{m}{k}du}{u} = t \tag{19}$$

Which yields

$$-\frac{m}{k}\left[Ln\left|u\right|\right]_{u_0}^{u} = t \tag{20}$$

Applying the limits and going back to the original variable we have,

$$-\frac{k}{m}t = Ln\left|\frac{F}{m} - \frac{k}{m}v\right| - Ln\left|\frac{F}{m} - \frac{k}{m}v_0\right| \tag{21}$$

Where rearranging terms we finally obtain the solution for $v(t)$:

$$v(t) = \frac{F}{k} - \left(\frac{F}{k} - v_0 \right) e^{-\frac{k}{m}t} \tag{22}$$

Here we can substitute $V_{max} = \frac{F}{k}$ and $\alpha = \frac{k}{m}$ which represent the maximum reachable speed and the magnitude of the drag force, respectively:

$$v(t) = V_{max} - (V_{max} - v_0)\, e^{-\alpha t} \tag{23}$$

It is now possible to obtain the position as a function of time if we substitute $v = \frac{dx}{dt}$ and apply separation of variables again,

$$\int_{x=x_0}^{x} dx = \int_{t=0}^{t} \left(V_{max} - (V_{max} - v_0)\, e^{-\alpha t} \right) dt \tag{24}$$

Where we have two simple constant and exponential integrations, with solution:

$$x - x_0 = V_{max} t + V_{max} \left[\frac{e^{-\alpha t}}{\alpha} \right]_0^t - v_0 \left[\frac{e^{-\alpha t}}{\alpha} \right]_0^t \tag{25}$$

Substituting the limits and rearranging terms we finally get:

$$x - x_0 = V_{max} \left(t - \frac{1 - e^{-\alpha t}}{\alpha} \right) + \frac{1 - e^{-\alpha t}}{\alpha} v_0 \tag{26}$$

And we can finally recover the 3-dimensional nature of the problem and get the final expression:

$$\vec{x} - \vec{x_0} = V_{max} \left(t - \frac{1 - e^{-\alpha t}}{\alpha} \right) \vec{e} + \frac{1 - e^{-\alpha t}}{\alpha} \vec{v_0} \tag{27}$$