Physics-Based Modeling of Pass Probabilities in Soccer

Conference Paper · March 2017

5 authors, including:

William Spearman
Liverpool Football Club

Austin T. Basye
University of Illinois, Urbana-Champaign

Some of the authors of this publication are also working on these related projects:

Project   Soccer Analytics View project

Physics-Based Modeling of Pass Probabilities in Soccer

William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop

Hudl, Lincoln, NE 68508

Email: william.spearman@hudl.com

Abstract

In this paper, we present a model for ball control in soccer based on the concepts of how long it takes a player to reach the ball (time-to-intercept) and how long it takes a player to control the ball (time-to-control). We use this model to quantify the likelihood that a given pass will succeed we determine the free parameters of the model using tracking and event data from the 2015-2016 Premier League season. On a reserved test set, the model correctly predicts the receiving team with an accuracy of 81% and the specific receiving player with an accuracy of 68%. Though based on simple mathematical concepts, various phenomena are emergent such as the effect of pressure on receiving a pass. Using the pass probability model, we derive a number of innovative new metrics around passing that can be used to quantify the value of passes and the skill of receivers and defenders. Computed per-team over a 38-game dataset, these metrics are found to correlate strongly with league standing at the end of the season. We believe that this model and derived metrics will be useful for both post-match analysis and player scouting. Lastly, we apply the approach used to compute passing probabilities to calculate a pitch control function that can be used to quantify and visualize regions of the pitch controlled by each team.

1. Introduction

Various approaches have been discussed for how to model pass probabilities in soccer [1] [2]. These can be broadly categorized into two groups: time-to-intercept based and machine learning based. Time-to-intercept methods compute the time it would take a player to reach the ball from his/her starting position. This provides a metric for whether a specific player will be able to receive the pass. Machine learning methods can be much more flexible because they use real soccer data but the final product can be difficult to conceptualize. In this paper, we propose a tunable, physics-based approach. With this method, we compute the time to intercept as input to a statistical model with physically meaningful parameters. Our requirements for this model are as follows:

1) **Probabilistic:** The output of the model should be interpretable as a probability.
2) **Data Driven**: The model must be fit to real-world data so that the probabilities are statistically meaningful.
3) **Predictive:** We require that the pass probability model be a predictive one. In other words, the model must be able to predict the probability of the pass using only information from the moment the pass occurs.
4) **Smooth:** The probabilities should vary smoothly between edge cases. For example: if the time to intercept for two players is similar, they should have similar probabilities of receiving a pass.

Such a model forms the groundwork for four analyses: receiving and interception efficiency; pass value; spatial pitch control; and hypothetical passing. After describing the model, we examine each of these analyses.

2. Approach

We choose to treat each pass in soccer as a separate Bernoulli trial. This means that every pass must result in a *success* or a *failure*. Furthermore, we believe that there is fundamental uncertainty in passing so that if the same pass were executed under the same conditions, the outcome might be different. We define a *controlled touch* as any action by a player on the ball that indicates the player had a degree of control. A *receiver* is then defined as the next player to execute a controlled touch after a pass, excluding the passer. We call a pass *successful* if the receiver is on the same team as the passer. It follows, therefore, that to receive a pass, the receiver must first intercept the trajectory of the ball and then execute a controlled touch on it. Thus, our model is predicated on two notions: the receiving player must have enough time to *intercept* the ball before it moves past him and the player must be in the vicinity of the ball for sufficient time to *control* it.

2.1. Ball Trajectory

Because we want the model to be based on information available at the start of the pass we need to compute the trajectory of the ball rather than use the actual trajectory from tracking data. This is important so that the model can be used to predict the performance of hypothetical passes and thereby evaluate the decision making of players. Due to the granularity of the tracking data with respect to the ball, we average the first few frames of the pass to obtain a more resilient estimate of the ball's initial velocity. For our purposes, we found that a 10-frame average (corresponding to 0.4 seconds) works well for such an estimation. While in flight, the three fundamental forces that influence the trajectory of the soccer ball are the gravitational force, the Magnus force, and the drag force. A great deal of research in the field of fluid dynamics has been performed to understand and model the trajectory of the soccer ball [3] [4]. For our purposes however, we simplify the problem by assuming a constant coefficient of drag and we ignore the Magnus force. This leads to a simplified equation of motion for the soccer ball [5]:

$$\ddot{\vec{r}} = -g\hat{z} - \frac{1}{2m}\rho C_D A \dot{r} \dot{\vec{r}} \qquad (1)$$

g is the gravitational constant of 9.8 m/s^2, m is the mass of the ball assumed to be a constant 0.42 kg, ρ is the density of air and is set to 1.22 kg/m^3, C_D is the drag coefficient which is approximated to be constant at 0.25, and A is the cross-sectional area of the soccer ball and set to 0.038 m^2 [3] [4]. When solving this differential equation for the motion of the ball, the z-component of the velocity of the ball is set to 0 when the ball makes contact with the ground.

In totality, the simplifying assumptions we use have two primary non-physical effects: first, the drag force is still assumed to be aerodynamic even when the ball has made contact with the pitch. This correctly leads to a gradual reduction in velocity although we admit that a model which incorporates rolling friction would be more correct. Because such a model would be heavily reliant on playing conditions, we choose not to model it. Second, because we ignore the Magnus force, the ball will always travel in a straight line. This assumption is required because we have no information about the spin of the ball from tracking data. Our hope is that any uncertainty related to these simplifying assumptions will be encapsulated in the model parameters which are fit to data.

2.2. Time to Intercept

In its simplest form, we can compute the time it takes a player to reach every location along the ball's trajectory. For each of these locations, we can treat the ball as stationary and simply compute the expected time for a player to reach that location. This is the *time to intercept*.
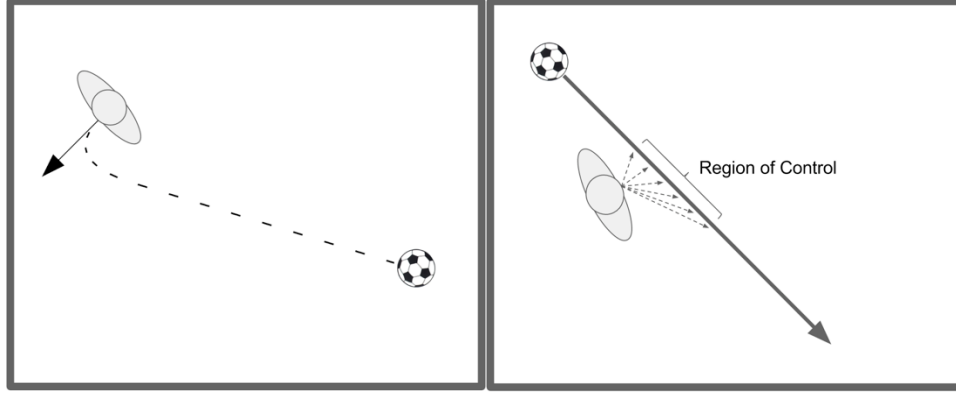

Figure 1. a) Visualization of the optimal interception trajectory for a stationary ball and a player moving with an initial velocity denoted by the black arrow. The dashes represent evenly spaces (in time) segments of the player's trajectory. b) Visualization of the region of control for a player along the trajectory of the ball. The dashed lines represent possible interception trajectories.

The time to intercept can be computed by solving the player's equation of motion:

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

with the following constraints: $|\dot{\vec{r}}| \leq v_{max}$ and $|\ddot{\vec{r}}| \leq a_{max}$. This equation is solved to minimize $t = t_{int}$ when $\vec{r}(t_{int})$ is set to the ball's location, $\vec{r}_{ball}$ for the flight time, T. Thus, the intercept time is a function of flight time $t_{int} = f_{int}(T)$. Conceptually if $\Delta t \equiv T - t_{int} \geq 0$ the player will be able to intercept the ball. We assume, however, that there is some temporal uncertainty, labeled σ, around Δt. The temporal uncertainty stems from a number of effects that we choose not to explicitly model such as differing speeds, reaction times, and effort levels of the players. We use the Logistic distribution to model this uncertainty. Thus, the probability that the player will intercept the ball at time T is given by the cumulative distribution function of the Logistic distribution[1]:

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

Figure 1a illustrates an example interception vector. The black arrow represents the initial velocity of the player; the player must stop running away from the ball before finally accelerating in the direction of the ball to intercept it.

2.3. Time to Control

To reflect the physical reality that players who remain in proximity to the ball for longer are more able to control the ball, we introduce a control rate parameter, λ. For a small region of time, Δt, the

---

[1] σ^2 represents the variance of the Logistic distribution.

probability that an actor can control the ball, assuming he can intercept the ball's trajectory, is given by: $\lambda \Delta t$. Integrating this over some temporal region during which the player is in physical proximity to the ball, the probability that the player is able to control the ball while in proximity within t seconds is given by the exponential distribution:

$$P(t) = 1 - e^{-\lambda t} \tag{4}$$

Due to the physical trajectory of the ball, a temporal region of control can be mapped to a spatial region of control. A visualization of the spatial region of control for a player along the trajectory of the ball is seen in Figure 1b.

2.4. Model

To help visualize the model, we show the cumulative distribution function of the two components of the model, the time to intercept and the time to control, in Figure 2. The time to intercept is described by the cumulative distribution function of the Logistic distribution and the time to control is described by the cumulative distribution function of the exponential distribution. The free parameter, σ, for the first distribution can be thought of as the uncertainty in time it takes a player to intercept the ball. Meanwhile the free parameter, λ^{-1}, can be thought of as the time it takes a player to achieve control of the ball.
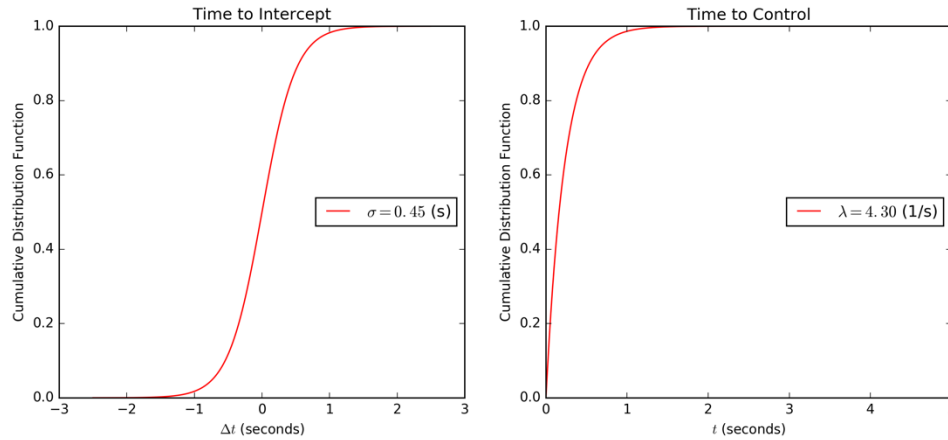

Figure 2. The cumulative distribution functions for the two components of the model. a) the time to intercept and b) the time to control. The displayed parameters of each are from the global fit described below.

Using the above components, we recursively construct the partial derivative of the probability that player j has received the pass by time, T as:

$$\frac{dP_j}{dT}(T) = \left(1 - \sum_k P_k(T)\right) P_{\text{int,j}}(T)\lambda \tag{5}$$

where $dP_j(T)$ represents the probability that player j will be able to control the ball during time T to $T + dT$. The probability that a player j will receive the pass within time t is given by:

$$P_j(t) = \int_0^t \left(1 - \sum_k P_k(T)\right) P_{\text{int}}(T)\lambda \, dT \qquad (6)$$

Computationally, we calculate this integral using a discrete simultaneous numerical integration for each player in the limit where $t \to \infty$. This gives us the total probability P_j that a specific player j will receive the pass. To determine the probability of a successful pass, we sum the probabilities for each player on the passing team, excluding the passer (call it set A): $P_A = \sum_{j \in A} P_j$.

2.4.1. Understanding the Model

To understand the model, we present a simple example in Figure 3 that highlights how the probabilities for different players to receive the pass evolve over time and space.
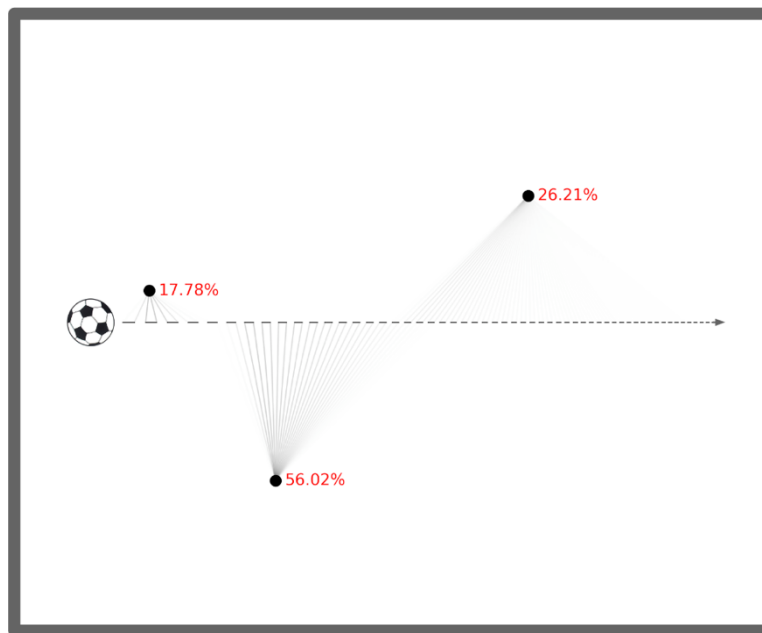

Figure 3. This figure shows the trajectory of the soccer ball and possible interception trajectories for three players. Each segment of the ball's trajectory corresponds to equally spaced temporal intervals. The darkness of the interception lines for each player represent the instantaneous probability that the player receives the pass where darker lines indicate a higher probability.

In this example, the ball is moving left to right with speed decreasing with drag proportional to v^2. The first of the players would need to move very quickly to intercept the fast moving soccer ball and accordingly, he only has a 17.8% chance of receiving the pass. The second player has multiple interception trajectories along a greater temporal and spatial region of control so his probability of receiving the pass is 56.0%. Finally, the last player has many possible interception trajectories, but since it is likely that the first two players have already intercepted the ball, his pass reception probability is only 26.2%. The overall probability distribution and cumulative distribution functions versus distance for this example are shown in Figure 4.
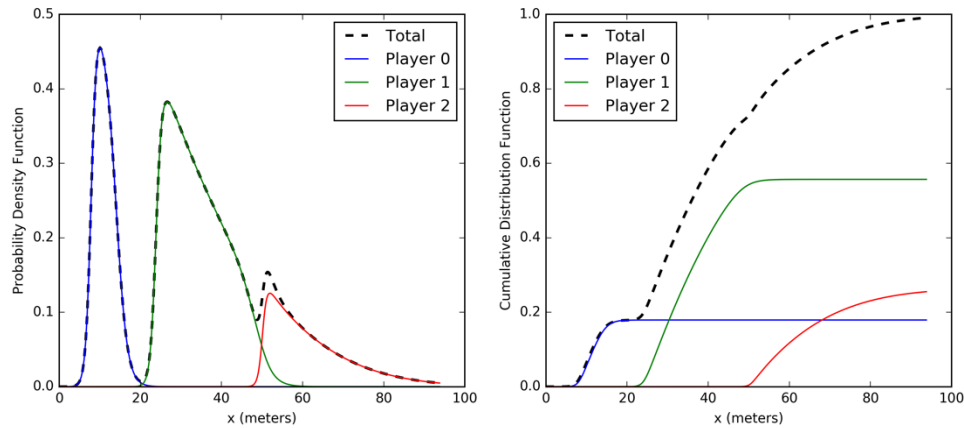
Figure 4. a) The probability density function for the example presented in Figure 3. b) The cumulative distribution function for the example presented in Figure 3. The PDF and CDF for each player is represented separately by the colored lines while the sum for all players is represented by the dashed black line.

3. Data

Our data set consists of 38 games played by Crystal Palace from the 2015-2016 season of the Premier League. Of these games, 11 were randomly selected to train the model, 12 were reserved to evaluate the performance of the fitted model while all 38 were used for applications and extensions. For each game, we have access to *tracking data* which gives the position of every player and the ball at 25 frames per second for the entire match. Additionally, we have access to *event data* which records events that occur during the match such as goals, passes, fouls, and many other soccer specific events. The event data incorporates information about when and where passes occur however it does not include information about the intended or actual recipient.

3.1. Processing and Selection
The first step in our data processing is to identify the actual recipient of all passes by looking forward in the event data to determine who received the ball. We use the following event types as indicative of a controlled touch and therefore, the receiver: pass, tackle, claim, ball recovery, and keeper pick-up. This means that events that do not indicate clear possession such as participation in an aerial duel are not considered when determining the eventual pass recipient.

The next step is to synchronize the event and tracking data so that events are precisely linked to the frame number at which they occur in the tracking data. This allows us to identify the start frame of every pass in the tracking data. From this we can know the location and velocity of all players and the ball at the start of the pass.

Once synchronized, we select passes that have a duration between 0.5 and 10 seconds and exhibit a track curvature of less than 5%. Additionally, passes where the ball goes out of bounds are removed. For the 23 games used to train and test the model, 10,875 passes remain after filtering, of which 5,404 passes are in the training set and 5,471 are in the testing set.

4. Fitting

Because we treat each pass as a Bernoulli trial, we require that our model follow a Bernoulli distribution for each pass. The probability mass function for the Bernoulli distribution is given by:

$$P(k|\sigma, \lambda, x) = \begin{cases} 1 - p & \text{for k=0} \\ p & \text{for k=1} \end{cases} \tag{7}$$

where k represents the outcome of the pass with success=1 and failure=0. The likelihood of a set of parameters, σ and λ, given outcome k and game state at the start of the pass, x, is given by:

$$\mathcal{L}(\sigma, \lambda|k, x) = P(k|\sigma, \lambda, x) \tag{8}$$

We wish to find the set of parameters that maximizes this likelihood for every pass in the training sample. This can be done by maximizing the product of the likelihood for each pass, which is equivalent to minimizing the sum of the negative log likelihood for every pass:

$$\min_{\sigma, \lambda \in \{\mathbb{R}, \mathbb{R}\}} \left\{ -\sum_{i \in P} \log[\mathcal{L}(\sigma, \lambda|k_i, x_i)] \right\} \tag{9}$$

where P represents the set of passes in the training samples with outcome k_i and inputs x_i. We perform the minimization by computing equation (9) for each value of σ and λ in our search space.
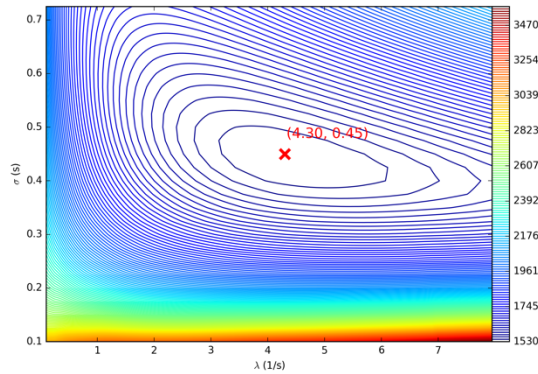

Figure 5. The negative log likelihood summed over all passes in the training set and varied over different values of, σ and λ. The contours represent intervals of 5 standard deviations in the likelihood. This indicates that the fit is very stable statistically. The systematic error, evaluated by fitting these parameters for specific games and finding the standard deviation among them. The best fit is found when $\sigma = 0.45 \pm 0.05\ s$ and $\lambda = 4.30 \pm 1.14\ s^{-1}$..

The best fit from the model is found for parameters: $\sigma = 0.45 \pm 0.01$ *(stat) ±0.04 (syst) s* and $\lambda = 4.30 \pm 0.28$ *(stat)* ± 1.1*(syst)* s^{-1}. As seen in Figure 2, these parameters give meaningful distributions for the time to intercept and time to control. It makes sense that players would have about a 95% likelihood to control the ball within the first second. A temporal uncertainty of half a second means that most interception times will be ± 1 second from the computed time to intercept.

4.1. Fit Results

When evaluated on the test sample of 12 games discussed in Section 3, the model is found to have an overall accuracy of 80.5% when predicting the success or failure of the pass. The accuracy when predicting the specific receiving player is 67.9%. Team strategy will often dictate that a player other than the intended receiver should receive the pass so there will be times when a player will not attempt to receive the pass thereby lowering the per-player accuracy number. A confusion matrix demonstrating this per-player performance on one game from the test set is shown in Figure 6a while the receiver operating characteristic (ROC) curve for the model is seen in Figure 6b.
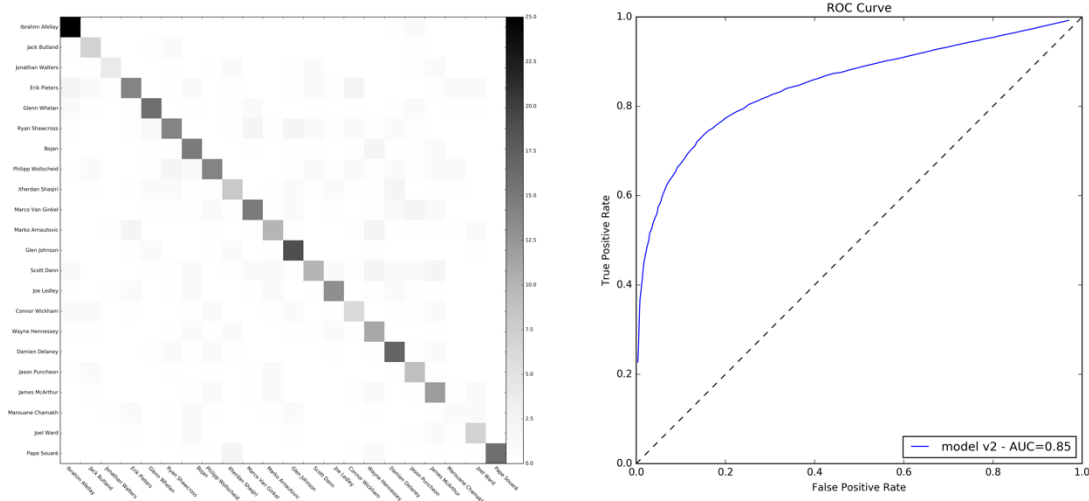

Figure 6. a) Confusion matrix of the for the Crystal Palace played on 19 December 2015. The y-axis is the actual receiver while the x-axis represents the receiver with the largest probability of receiving the pass. The total accuracy for pass receiver predictions computed by summing along the diagonal and dividing by the total number of passes is 73.1% b) The receiver operating characteristic of the model using predicted correct team as the accuracy metric.

Because more passes in soccer are successful than unsuccessful, the accuracy of the model can be increased to 81.9% by shifting the cutoff for successful pass from 0.5 to 0.27. Overall, in the test set, 78.9% of all passes are completed while the mean for the Poisson binomial distribution from the model is found to be 67.9%. This difference is likely due to features not accounted for by the model such as the Magnus force, player tendencies, inaccuracies in tracking data, or team passing strategy.

5. Applications

Although there is utility in using the pass probability model by itself, it is also possible to leverage it as the starting point for new metrics. In the subsequent sections, two new metrics are introduced: *receiving/interception efficiency* and *pass value*. These are shown to correlate with indicators of team success such as passes in the attacking third, shots, and even the Premier League standings at the end of the year.

5.1. Receiving/Interception Efficiency

Traditional metrics for measuring reception and interception skill are given by the percentage of passes received or the total number of interceptions made. Neither of these metrics takes into account the difficulty of receiving the pass. Using the pass probability model, we can predict how

many passes a player should receive over the course of a game. If the player receives more passes, it might indicate that they are more aggressively jockeying for the ball than average. Similarly, for interceptions, we can identify how many opportunities for an interception the player had and how many actual interceptions they achieved. Again, this provides a measure of how aggressive the player was. Oftentimes, team strategy affects player aggression so it is not always possible to assign credit or blame to players that over perform or underperform relative to the model's predictions. For example, a striker might have a chance to disturb a pass between opposing defenders, but team strategy dictates that he should maintain a defensive position. Mathematically, since each pass is treated as an independent Bernoulli trial, an aggregation of passes will follow the Poisson binomial distribution. The mean and variance of the Poisson binomial distribution are given in equation (10).

$$\mu = \sum_{i=1}^{n} p_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} (1 - p_i)p_i \tag{10}$$

Using this formulation, we can compute the expected number of passes that a player should receive and compare it to the actual number of passes they receive. We label this the *total receiving efficiency* and we split it into *receiving efficiency* for passes originating from a teammate and *interception efficiency* for passes originating from an opponent.

Position	Total			Receiving			Interception		
Defender	1.06	+/-	0.03	1.04	+/-	0.03	1.13	+/-	0.06
Midfielder	1.03	+/-	0.02	1.23	+/-	0.03	0.57	+/-	0.03
Forward	0.91	+/-	0.03	1.36	+/-	0.06	0.23	+/-	0.02
Goalkeeper	0.65	+/-	0.05	0.80	+/-	0.07	0.48	+/-	0.06

Table 1. This table shows the receiving efficiency for different position archetypes. Total receiving efficiency is broken into two categories: receiving (same team) and interception.

In Table 1, we see that defenders and midfielders have the highest total receiving efficiencies. When breaking the receiving efficiency into receiving and interception categories, the forwards are shown to have the highest efficiency when receiving passes from teammates while defenders have the highest efficiency when intercepting opponent passes.

5.2. Passing Value
To the trained analyst, it is possible to determine whether or not a specific pass was a good idea. Usually this takes into account three factors: the likelihood the pass will succeed, how much benefit would be gained by a successful pass, and how much the other team would benefit from an interception. A risky pass to a winger who is sprinting past defenders towards the goal is oftentimes a worthwhile one while a risky pass between defenders when an interception would grant the opposing striker a clear chance on goal is rarely justified.

Various heuristics have been proposed by analysts that quantify the value of a particular game state. Provided we have a function, $f(x)$, which quantifies the *value* of a specific situation, x, we can compute the value of a pass j using the following equation:

$$V_j = p_j f(x_{\text{suc}}) - (1 - p_j)f(x_{\text{fail}}) \tag{11}$$

where x_{suc} represents the game state in case of a successful pass, x_{fail} denotes the game state in case of a failed pass, and p_j represents the probability of the pass succeeding. We use a naïve model in which the value of a state, x, is given by the negative exponential of the distance between the possessor and the target goal. The functional form is shown below:

$$f_{\text{simple}}(x) = f_{\text{simple}}(\vec{r}_{ball}, \vec{r}_{target}) = ae^{-b\sqrt{(\vec{r}_{\text{target}} - \vec{r}_{ball})^2}} + c \tag{12}$$

for some constants of the model a, b, and c which are fit using event data[2]. With equation (11) and equation (12) we can compute the pass values for each pass. We approximate x_{suc} as the game state when the ball is received with possession retained by the passing team and x_{fail} as the game state at the beginning of the pass but with possession given to the intercepting team.

5.3. Correlations

Using data for the 38 games involving Crystal Palace during the 2015-2016 Premier League season, we compute the mean pass value and total reception efficiency for each opponent of Crystal Palace and Crystal Palace herself during these games. There are correlations between these new metrics and other statistics considered to be good indicators of a successful team. The correlation (Pearson product-moment correlation coefficient) between total reception efficiency and shots is 0.64 while the correlation with passes in the attacking third is 0.70. Similarly, for mean pass value, the correlation with shots is 0.63 and with passes in the attacking third is 0.83.

Team Standing	Shots (Mean)	Reception Efficiency (Total)	Pass Value (Mean)
1	10.0	0.92	0.0081
2	20.5	1.13	0.0126
3	23.0	1.06	0.0106
4	19.0	1.04	0.0110
5	10.5	1.09	0.0116
6	17.0	1.02	0.0097
7	21.0	1.05	0.0111
8	20.5	1.07	0.0114
9	19.0	1.05	0.0094
10	18.0	1.07	0.0120
11	14.0	1.03	0.0109
12	11.5	1.00	0.0086
13	12.5	0.97	0.0093
14	9.0	0.97	0.0079
15	12.4	0.96	0.0087
16	9.0	1.03	0.0097
17	14.5	1.05	0.0072
18	8.5	0.96	0.0065
19	15.0	1.06	0.0096
20	12.0	1.02	0.0100

[2] The values of a, b and c are set to 0.93, 0.14 and 0.02 respectively. These numbers were found by looking at the likelihood of a possession resulting in a score when a pass occurred at the specified location on the pitch.

Table 2. Comparison of league standing (at end of season) with score differential for 38 games involving Crystal Palace.

Given these correlations, we believe that these new metrics can provide a useful tool for evaluating whether a team's passing and receiving is having a positive impact on their performance.

6. Extensions

In this section, we introduce two extensions that build upon the pass probability model: *spatial pitch control* and *hypothetical passing*. Each provides a novel framework for interrogation of soccer tracking data.

6.1. Pitch Control

To compute how teams control regions on the soccer pitch, we compute the stationary pass probability for an imaginary ball placed at every point on the pitch. We call this real valued scalar field the *pitch control function* (PCF). The PCF can be evaluated for the individual player or the team. As with the pass probability model, the team PCF is merely the sum of the PCFs of the constituent players. A visualization of the team and individual PCF is shown in Figure 7.
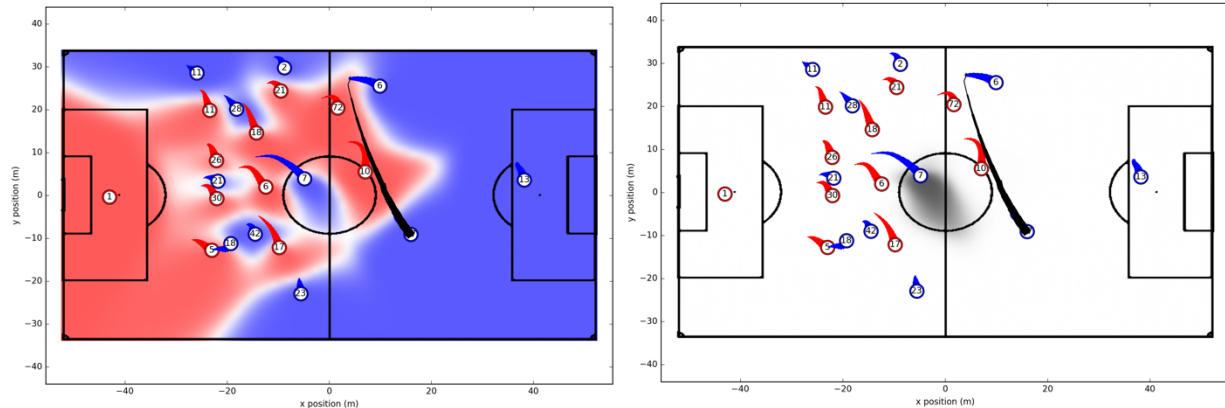

Figure 7. a) Computed pitch control function (PCF) for Crystal Palace. Blue regions are those controlled by Crystal Palace while red regions are those controlled by the opposition, regions in white are contested. b) Computed pitch control function (PCF) for Yohan Cabaye. The shaded region on the pitch is the region controlled by Cabaye at that instant. In both plots, the circles with numbers represent players (identified by their jersey number) and the black line represents the track of the ball. The trail behind each represents the position over the past 3 seconds.

In tactical terms, the PCF can be thought of as the probability that a player or team will be able to control the ball if it were at that location. Because it incorporates initial velocities and the time it takes to control the ball, the PCF can be used to extend previous studies [6] [7] [8] that make use of Voronoi diagrams in the context of sports analytics. It is possible to quantify how much space is controlled by each player or team during critical moments in the match such as corner kicks or transitions. In Table 3, we see the mean PCF for the region around the target goal during corner kicks in which a shot was taken. These are split into situations where a shot saved taken and those where a goal was scored.

Radius (m)	Mean PCF (Saved Shot)	Mean PCF (Goal)	Difference
5	0.82	0.79	0.04
10	0.75	0.74	0.01
15	0.68	0.67	0.01
20	0.62	0.61	0.01

| 25 | 0.53 | 0.53 | -0.01 |

Table 3. This table shows the mean PCF for the defending team within the specified radius of the goal for corner kick scenarios..

Conventional knowledge is that it is very important to control the region directly in front of the goal during a corner kick. In the table above, we see the mean value of the PCF near the goal for the defending team. This can be interpreted as the control the defending team exerts over the region within some radius of the goal. For situations when a shot is scored, the defending team exerts, on average 4% less control in the <5 m region than when a shot is saved. Unsurprisingly, this correlation confirms the conventional wisdom regarding the defense of corners.

6.2. Hypothetical Passing

Using the pass probability model, we can analyze hypothetical passes by inputting a hypothetical ball velocity. First, we find the ideal pass by choosing the ball speed and passing angles that maximize the to-player pass probability for the intended receiver (within some specified constraints for possible ball velocities[3]) using a simulated annealing process [9]. Next, we calculate the stability of this passing solution by tracking how the pass probability changes when perturbations $\delta \vec{v}$ are introduced to the ideal ball velocity vector $\vec{v}_{ideal}$. Treating these perturbed velocities as a set V, we can find the range of probabilities corresponding to these velocities, set P. The difficulty of the pass can be approximated by looking at the variance of probabilities in set P, σ_P^2. The mean plus the standard deviation of this set $\mu_P + \sigma_P$ is a heuristic for how probable the pass would be if well-kicked and the mean minus the variance $\mu_P - \sigma_P$ describes how probable the pass would be if poorly-kicked. If both values are high, it indicates a pass that is easy to execute while if both values are low, it indicates a pass that is nearly impossible. If the first value is high and the second value is low, the pass is feasible but it would require great skill to complete.
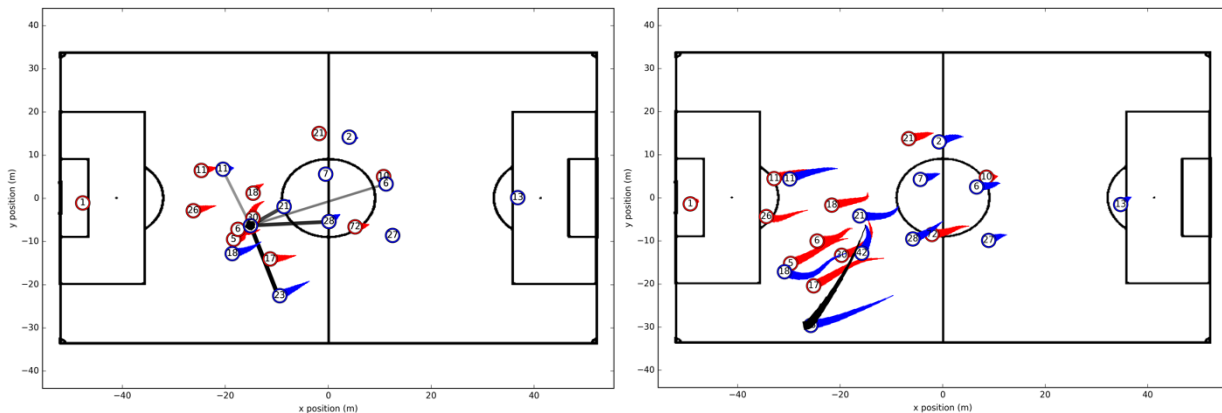

Figure 8. Plots represent a passing situation for Crystal Palace. a) Jason Puncheon (42) has the ball. His passing options are represented by the black lines. The color of the line represents the 1-sigma upper bound on the probability of passing to that player while the line width indicates the 1-sigma lower bound on the probability of passing to that player. The main target probability ranges are given as follows: Wilfried Zaha (11) 0.04 – 0.66, Connor Wickham (21) 0.49 – 0.88, Joe Ledley (28) 0.66 – 0.86, and Pape Souaré (23) $0.9\overline{9}$ – 1.00. b) The result is that Puncheon completes the safe pass to Souaré.

---

[3] Vast simplifications are made when deciding what is within the physical constraints of a player's kicking ability and an entire analysis dedicated to how a player can kick the ball given his facing and speed would be needed.

In Figure 8 we show an example passing situation. The width of the pass is proportional to $\mu_P - \sigma_P$ and the opacity of the pass line is given by $\mu_P + \sigma_P$. Until we replace simulated annealing with a less computationally expensive algorithm to determine the ideal passes, computational limitations prevent us from performing an analysis of hypothetical passes on a large scale. When such limitations are surpassed, we will be able to construct a metric to judge player decision making while passing.

7. Conclusions

In this paper, we have introduced a new passing model for soccer that is based on the physical concepts of interception and control time. The model uses a statistical framework to accurately predict the precise receiver in the majority of passes using only the game state at the point the ball is kicked. With the model we are able to introduce new metrics that correlate with winning at the team level and are also a good descriptor of a team's style of play. As these metrics are constructed at the per-player level, we can use them to evaluate specific players or teams.

Additionally, with this physics-based statistical model, we have constructed a new approach for interrogating tracking data. Using it, we can quantify the control a team exerts over any arbitrary region on the pitch. Furthermore, we can begin to evaluate passer decision-making using models that can determine the likely outcome of possible passing options.

Acknowledgements

References

[1] L. Stanojevic and R. Gyarmanti, "QPass: a Merit-based Evaluation of Soccer Passes," *2016 ACM KDD Workshop on Large-Scale Sports Analytics,* pp. 0-3, 2016.

[2] J. Wolle and T. Gudmundsson, "Football analysis using spatio-temporal tools.," *Computers, Environment and Urban Systems,* vol. 47, pp. 16-27, 2014.

[3] T. Asai, K. Seo, O. Kobayashi and R. Sakashita, "Fundamental aerodynamics of the soccer ball," *Sports Engineering,* vol. 10, no. 2, pp. 101-109, 2007.

[4] L. Saetran and L. Oggiano, "Aerodynamics of modern soccer balls," *Procedia Engineering,* vol. 2, no. 2, pp. 2473-2479, 2010.

[5] D. Halliday, R. Resnick and J. Walker, Fundamentals of Physics, 7th Edition, New York: John Wiley & Sons, Inc, 2005.

[6] S. Kim, "Voronoi Analysis of a Soccer Game," *Nonlinear Analysis: Modeling and Control,* vol. 9, no. 3, pp. 233-240, 2004.

[7] R. Maheswaran, Y. Chang, J. Su, S. Kwok, T. Levy, A. Wexler and N. Hollingsworth, "The Three Dimensions of Rebounding," in *8th Annual MIT Sloan Sports Analytics Conference*, Boston, 2014.

[8] S. Fonseca, J. Milho, B. Travassos and D. Araujo, "Spatial dynamics of team sports exposed by Voronoi diagrams," *Human Movement Science,* vol. 31, no. 6, pp. 1652-1659, 2012.

[9] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, Numerical Recipes: The Art of Scientific Computing 3rd Edition, New York: Cambridge University Press, 2007.