A New Metric for Pitch Control based on an Intuitive Motion Model

Lucas Wu and Tim B. Swartz

Abstract

With the availability of tracking data, the determination of pitch control (field ownership) is an increasingly important topic in sports analytics. This paper reviews various approaches for the determination of pitch control and introduces a new field ownership metric that takes into account associated sporting dynamics. The methods that are proposed utilize the movement of the ball and players. Specifically, physical characteristics such as current velocity, acceleration and maximum velocity are considered. The determination of pitch control is based on the time that it takes the ball and the players to reach a given location. The main result of our investigation concerns the validation of the resultant pitch control diagram. Based on a sample of 5887 passes, the team identified as having pitch control was the observed recipient of the pass with 91% accuracy. The approach is generally applicable to invasion sports and is illustrated in the context of soccer. Various parameters are introduced that allow a user to modify the methods to alternative sports and to introduce player-specific maximum velocities and player-specific accelerations.

Keywords : spatio-temporal modelling, tracking data.

1 INTRODUCTION

Pitch control (field ownership) has become an increasingly discussed topic in sports analytics ([9]). At a given point in time during a match, each location on the field may be "dominated" by a particular player. Maps that depict this domination are referred to as pitch control diagrams and the construction of these maps is the subject of this paper.

The determination of pitch control is an important component in the investigation of tactics. For example, related to pitch control is the quantification of scoring opportunities from off-the-ball positions in soccer ([24]). In turn, this provides an analysis of optimal player movement, player tendencies, the evaluation of player movement and attacking capability ([15]). [16] shows that pitch control models can be useful in applications such as expected goals and the assessment of player market value. With pitch control as a basic ingredient, [9] investigates a number of important questions of interest to the professional soccer community including on-ball performance and off-ball performance. Even in sports that are greatly different from soccer, pitch control is topic of increasing interest; see [19] concerning an application in the National Football League (NFL).

The availability of tracking data is fundamental to the analysis of pitch control. Tracking data in soccer consists of the (x, y) coordinates of the ball and the 22 players on the pitch recorded at regular and frequent time intervals. With tracking data, we know the locations of all of the players at all times during a match, and this facilitates the determination of pitch control. A feature of our research is that the determination of pitch control is easily automated whenever teams have access to tracking data. [13] provide a review paper on spatio-temporal analyses used in invasion sports (including soccer) where tracking data are available. [11] discuss a range of tactical problems in soccer based on the use of tracking data. For example, the visualization of team formations is a problem that has received particular attention ([29]). The analysis of tracking data has also been prominent in the sport of basketball; see for example, [18]. For a review of statistical contributions that have been made across major sports, see the text by [1].

In Section 2, we provide a literature review of prominent approaches that have been utilized in the construction of pitch control diagrams. Commentary is provided that discusses the relative strengths and weaknesses of the approaches. In Section 3, we introduce a new pitch control metric which is illustrated in the context of soccer but is adaptable to other invasion sports. A motivation of the approach is that it considers relevant soccer dynamics. That is, we consider the location, speed and direction of both the ball and the players in anticipation of their movement. We also consider realistic aspects of motion such as maximum player velocity and acceleration. The approach is conceptually straightforward compared to some of the approaches that have been introduced in the literature. This is an important as it allows users to vary parameters to suit their sporting application. Our approach defines three categorical regions of control (Team A, Team B and neither team). This has advantages over nonprobabilistic shaded regions that cannot be assessed in terms of accuracy. In Section 4, we implement the proposed pitch control metric in the context of an example. We evaluate the accuracy of the proposed pitch control metric, an analysis that is novel in the pitch control literature. We conclude with a short discussion in Section 5.

Since much of the code that has been developed in the pitch control literature is proprietary, we have provided an R function that determines pitch control (see Supplementary file). The code is realistic in terms of physical motion and can be adapted to various sports via intuitive parameters.

2 LITERATURE REVIEW

The simplest and a common approach for gaining appreciation of field ownership is based on methods that compute the centroid of player activity. Using tracking data, the (x, y) coordinates for a player are averaged during a period (e.g. perhaps the entire match), and this provides the centroid $(\bar{x}, \bar{y})$. The centroid gives an impression of where a player spends most time on the field. Of course, it says nothing about field ownership at a specific point in time and hence the centroid approach is mostly limited to analyzing average team formations. Websites such as whoscored.com provide positional maps for top-flight matches that illustrate the basic centroid approach. The centroid idea has been extended in many directions with various applications; see for example, [17].

The starting point for investigations of pitch control at instantaneous points in time is Voronoi tessellations ([27]). In the two-dimensional setting with n distinct points, the Voronoi diagram uniquely partitions the plane into n polygons such that each polygon contains exactly one point where every location within the polygon is closer to that point than to any other point. Figure 1 provides a Voronoi diagram on the unit square based on $n = 5$ randomly generated points; Voronoi diagrams are easily produced using the function *deldir* in R. From Figure 1, we observe that larger polygons correspond to points that are located further away from other points. In a sporting context, each of the points represents a player and the corresponding polygon represents the region of pitch control or dominance by the player. That is, if the ball were placed anywhere in a player's polygon, then it would be expected that the player would gain possession/control of the ball by virtue of the player being closer to the ball than any other player.

Pitch control as determined by Voronoi diagrams has been utilized by various investigators in soccer including [14]. However, there are some important limitations of the Voronoi diagrams as they do not consider relevant features of the corresponding sporting application. For example, a player could be moving in a particular direction. This is relevant because it is easier for the player to control regions in the direction of travel and it is more difficult to control regions that are opposite to the direction of travel. Clearly, player velocities impact pitch control and this is not addressed with Voronoi diagrams. [6] provides a critique of Voronoi diagrams in soccer and discusses properties that pitch control methods should possess.

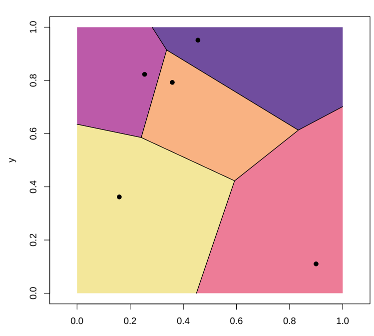

Figure 1: Voronoi diagram based on $n = 5$ points generated on the unit square.

In Figure 2, we provide a Voronoi diagram that is based on the instantaneous player positions in an actual soccer match. The coloring no longer refers to individual players, but rather the two teams. Consequently, we observe that the field is partitioned into two sets.

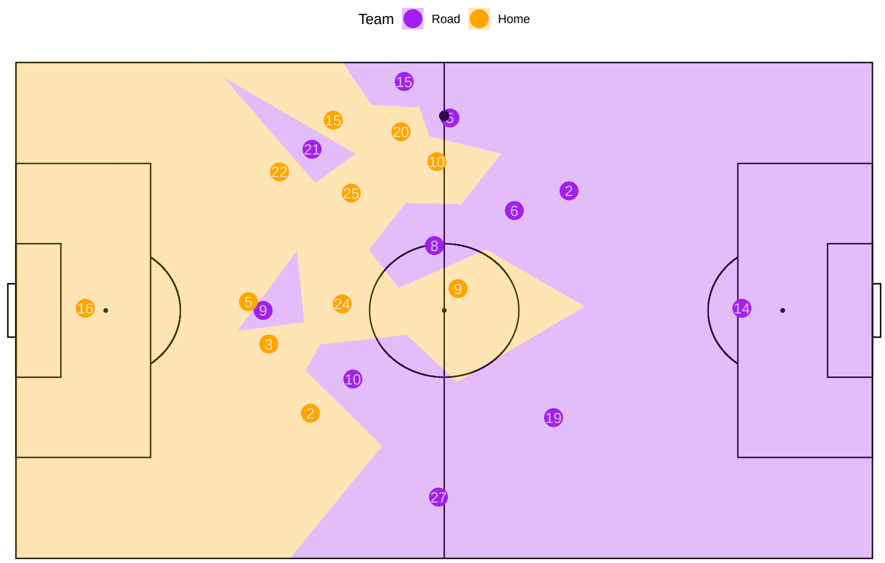

Figure 2: Voronoi diagram applied to a given snapshot of a soccer game based on the location of the 22 players on the pitch. The shaded orange and purple areas correspond the dominant regions for the home and away teams, respectively.

Prior to the advent of commercial tracking data, [25] developed a motion analysis system whereby cameras and software were utilized to extract features from a soccer match. In particular, they introduced the "dominant region" concept which addresses pitch control. Unlike Voronoi diagrams, the dominant regions of [25] are not polygons, but areas delineated by smooth curves. A dominant region for a player is defined as the set of points where the player can arrive earlier than all other players. A player's time of arrival at a given point takes into account the player's current location, current speed and potential acceleration. The acceleration model is not fully disclosed in [25] but it is based on patterns from an average player and permits reduced accelerations in the direction of movement. It also appears that the calculation of arrival time does not utilize a cap on player velocities which seems to be a practical limitation of the approach. [10] address the limitation of maximum velocity in their modified approach to pitch control. A notable difference between the dominant region approach and the metric proposed in Section 3 is that [25] does not take the dynamics of the ball into account. For example, there could be a location which player A can reach in time t_A and player B can reach in time t_B where $t_A < t_B$. However, it would be incorrect to assert that player A has dominance at this location if it takes time $t > t_B$ for the ball to reach the location. In this case, neither player has unique control over the location. [12] extend these ideas where individual player characteristics are estimated from data. [26] provide more detail on the dominant region approach and use the framework to evaluate teamwork in soccer.

[2] provide a probabilistic approach to the construction of pitch control diagrams based on machine learning methods. They refer to this area of research as movement models. The movement model for a given player is described by the conditional density function

$$P_{t_\Delta}(p \mid p_t, v_t) \tag{1}$$

where p is the location attained during time horizon t_Δ. The density (1) is conditional on the player's position p_t and velocity v_t at the current time t. As locations refer to positions on the field, p, p_t and v_t are two-dimensional variables. [2] suggest various algorithms for the estimation of the conditional density (1) and its corresponding discrete approximations. The algorithms are computationally intensive and use historical data to inform player movement. In contrast to the approach developed in Section 3, the conditional density (1) does not consider various aspects of the state of the game (e.g. ball position, ball movement, and current player acceleration). It is also important to note that not all historical movement data is pertinent to a given situation. For example, players move slowly in situations where they are not active, and these movements are not relevant to situations where they need to be active.

Having estimated the conditional density (1) for every player, [2] then define the "zone of control" for a given player as the locations on the field where the player's density (1) is greater than the density for any other player. The intuitive idea is that the zone is attained with higher probability than other players.

[7] also use statistical methods to determine pitch control where their movement models are based on parametric distributions. For the ith player, they define an *influence degree* $I_i(p, t)$ for location p at a future time t. The influence degree is a density ratio corresponding to a bivariate normal distribution that takes into account various soccer dynamics such as current player velocity. Field ownership by a team is then assessed through a kernel-based non-parametric point process which considers the cumulative influence degrees of all players on the field. An important feature of their approach is that the method elicits degrees of field ownership rather than a binary outcome. It is also important to note that the degree of field ownership is not a probability, and that the Gaussian distributions do not permit skewness. It would be interesting to calibrate the degrees of field ownership with empirical probabilities of field ownership. [16] provide additional critique on the approaches developed by [2] and [7].

William Spearman has developed pitch control models for the Liverpool Football Club where ideas are sketched out in a conference presentation ([23]) and also in the YouTube video https://www.youtube.com/watch?v=X9PrwPyolyU. The approach is also probabilistic and is based on the estimated time t_i that it takes the ith player to reach a given location. Players are labelled $l_i = 1(0)$ according to whether they play on the home(road) team. The probability that the home team ends up with possession at the given location is given by

$$\text{Prob(Home)} = \left(\frac{\sum_i l_i t_i^{-\beta}}{\sum_i t_i^{-\beta}} + 1 \right) /2 \tag{2}$$

where $\beta > 0$ is a specified tuning parameter. An important insight that is mentioned in the video (but is not fully disclosed) is that the motion of the ball is taken into account. Although there is some intuition surrounding (2), it is not clear that there is a physical process which suggests the proposed probability formulation. Also, it is not clear how the estimation of the times t_i impact the probabilities given by (2). The core idea in [22] is based on how long it takes a player to reach and control the ball. Our approach emphasizes how long it takes a player to reach a certain location. In a Friends of Tracking video ([21]), Laurie Shaw provides details on implementing a pitch control model based on some of the ideas developed by [23].

A nice feature of the probabilistic models is that they conveniently lead to expected possession value (EPV) calculations that assess the decision of passing the ball to a given location. At its most basic level, the calculation is essentially

$$\text{EPV} = (\text{Pitch Control Prob}) * (\text{Pass Impact}) \tag{3}$$

where pass impact is the probability of scoring a goal from the given location. Pass impact is determined by historical data and soccer considerations such as the location on the field, the relative position of defenders, etc. Of course, the simple formula given by (3) does not take into account the difficulty of executing the pass successfully. A more sophisticated way to build an EPV framework is discussed in [4] and [8].

All of the approaches reviewed in this section have a common element. The commonality is that there is an estimated minimal time for each player to reach a location. This is even true in the case of Voronoi diagrams where the minimal time is proportional to distance under the assumption that all players travel at a constant speed. Ownership of the location is then determined by assessing these minimal times by providing more weight to player ownership for players that can reach the location faster. In the following section, we take a philosophically different approach where it is asserted that if two players can reach a location by the time of ball arrival, neither player has a claim or advantage involving field ownership. This results in a metric that is relatively simple and provides regions of team dominance that do not form a partition of the field. Rather, there are regions of dominance and contested areas. This type of mapping may prove advantageous in some applications.

It is difficult to assert which of the approaches to pitch control are dominant in soccer. We know that many "big" teams are interested in pitch control and have their in-house programs. However, teams wish to maintain a competitive edge and are generally reluctant to communicate their technical work.

3 METHODS: A PITCH CONTROL METRIC

We stress that a distinguishing feature of the proposed approach is that there exist regions that are not dominated by either team. With some methods, all regions on the pitch are dominated by either one team or the other team. Alternatively, other methods provide color-coded regions that express degrees of dominance. An advantage of the proposed approach is that we can assess the accuracy of the metric using historical data, as is carried out in Section 4.2. Such an analysis is less straightforward with color-coded regions where the colors do not correspond to probabilities.

Another distinguishing feature is that our approach takes intent into account. When we consider the movement of a player to a particular location, the modelling assumes that the player moves to that location as quickly as possible to gain pitch control. We contrast this with the probabilistic approaches reviewed in Section 2. With the probabilistic approaches, a given trajectory is assessed probabilistically with respect to the population of potential trajectories.

The objective of this research is the development of a pitch control metric that is intuitive, realistic and adaptable. Accordingly, we let t_b denote the time that it takes the ball to reach a location of interest on the pitch. Similarly, we define t_h and t_r as the minimum times that it takes a player from the home team and the road team, respectively, to reach the location of interest. According to this setup, pitch control at the location of interest is determined according to the simple heuristics of Table 1.

3.1 Criteria for Pitch Control

In the first two lines of Table 1, we have the ball reaching the location of interest after the arrival of both players. Since both players are waiting for the ball, neither team can lay claim to pitch control. In the third line of Table 1, the player on the home team awaits the ball, and the ball arrives before the road team player can reach the location. In this case, the home team has pitch control, and by symmetry, the road team has pitch control according to the conditions in the fourth line of Table 1. In the fifth line of Table 1, the ball arrives first. In many instances, the ball can be passed slower than stipulated which modifies the inequality to $t_h < t_b < t_r$, and provides the home team with pitch control. By symmetry, the road team has control according to the sixth line of Table 1.

Note that we may introduce a tuning parameter $\epsilon > 0$ in Table 1 which corresponds to a time interval. For example, we might introduce a more stringent condition involving pitch control for the home team in the fifth line by modifying $t_b < t_h < t_r$ to the condition $t_b < t_h + \epsilon < t_r$. This states that the player on the home team must have a little bit of time to gain possession (relative to the arrival of the road player) in order for pitch control to be established. For example, we might set $\epsilon = 0.5$ seconds for professional soccer players. The consequence of such modifications is to partition the field with larger regions corresponding to "neither team". The introduction of free spaces that are controlled by neither team has also been investigated by [3].

Time Inequality	Pitch Control
$t_h < t_r < t_b$	neither team
$t_r < t_h < t_b$	neither team
$t_h < t_b < t_r$	home team
$t_r < t_b < t_h$	road team
$t_b < t_h < t_r$	home team
$t_b < t_r < t_h$	road team

Table 1: The determination of pitch control at a given location given time inequalities involving t_b, t_h and t_r.

In what follows, we use the Cartesian coordinate system where locations, velocities and accelerations are described by ordered pairs. Distances and times are measured in metres and seconds, respectively.

3.2 Timing of the Ball

According to the previous discussion, we wish to estimate the time t_b that it takes the ball to reach the location of interest. We define

$$\begin{aligned}(x_0, y_0) &\quad - \quad \text{current ball location} \\ (x_1, y_1) &\quad - \quad \text{location of interest} \\ s &\quad - \quad \text{speed that the ball is passed}\end{aligned}$$

[21] provide speeds of 15 metres/sec for balls that are passed between teammates. Maximum speed of kicked soccer balls has been estimated as high as 30 metres/sec (https://soccerballworld.com/soccer-ball-physics/). For this application, we have set the speed of the ball at $s = 18$ metres/sec based on an investigation of all completed passes during the 2019 season of the Chinese Super League (CSL).

This leads to the quantity of interest

$$t_b = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}/s \tag{4}$$

which is based on historical data using the approximation that a kicked ball travels at a constant speed.

3.3 Timing of Players

The time that it takes a player to reach the location of interest is a more complex calculation than the time t_b in (4) for the ball to reach the location of interest. Although there are $n = 22$ players on the soccer pitch, we use simplified notation where we suppress player subscripts. We consider a single player where we use t to denote the time that it takes the player to travel from their current location to the location of interest with intent. We further define

$$\begin{array}{rcl} (x_0, y_0) & - & \text{current player location} \\ (x_1, y_1) & - & \text{location of interest} \\ (x_t, y_t) & - & \text{location of player at time } t \\ (v_{x_0}, v_{y_0}) & - & \text{current velocity of player} \\ (v_{x_t}, v_{y_t}) & - & \text{velocity of player at time } t \\ (a_x, a_y) & - & \text{acceleration of player} \\ a & - & \text{non-directional acceleration of player} \\ s_{\max} & - & \text{maximum speed of player} \end{array}$$

Before proceeding with the calculation of the time t that it takes the player to reach the location of interest, we discuss some of the assumptions related to our motion model. We label these assumptions and the associated discussions (A) - (C).

Assumption A regarding $(\mathbf{v_{x_0}, v_{y_0}})$: We obtain the current velocity (v_{x_0}, v_{y_0}) directly from the tracking data. The current player velocity may be approximated from the tracking data by considering the change in location by the player during a small time increment surrounding the current time; e.g. $v_{x_0} = \Delta_{x_0}/\Delta_t$ where Δ_{x_0} is the distance travelled in the x-coordinate direction in a window of time length Δ_t surrounding the current time. For this application, we have set $\Delta_t = 0.4$ seconds. Some exploratory procedures for accurately estimating speed in tracking data are discussed by [28].

Assumption B regarding $\mathbf{s}_{\max}$: We impose a limit on maximum player speed. This is realistic and is player-dependent. The values for $s_{\max}$ can be estimated from the tracking data. For context, the Canadian international Alphonso Davies who plays for Bayern Munich has the top speed of 36.0 km/hour recorded in the 2020/2021 season of the German Bundesliga (see https://www.bundesliga.com/en/bundesliga/stats/players/top-speed). For contrast, the 356th player on the list is Jiri Pavlenka of Werder Bremen with a top speed of 31.0 km/hour. For illustration in the remainder of the paper, we set the common value $s_{\max} = 9.2$ metres/sec which corresponds to 33.1 km/hour. For comparison, we note that [7] uses $s = 13.0$ metres/sec and [2] use $s_{\max} = 8.0$ metres/sec for maximum speed. With the two-dimensional representation of velocity, we therefore introduce the constraint $v_{x_t}^2 + v_{y_t}^2 \leq s_{\max}^2$ for all times t.

Assumption C regarding $(\mathbf{a_x}, \mathbf{a_y})$:

Acceleration is the component of our motion model which allows players to change direction and velocity. Like player velocities, player accelerations are player-dependent, and can be informed from the tracking data. However, for illustration, we set the common value $a = 5.0$ metres/sec^2 for all players. For comparison, we note that [21] uses $a = 7.0$ metres/sec^2 and [2] use $a = 4.2$ metres/sec^2. The differences in the literature involving maximum speeds and accelerations is notable and is one reason for differences in various pitch control diagrams. With the two-dimensional representation of acceleration, we therefore introduce the constraint $a_x^2 + a_y^2 = a^2$ during periods of acceleration.

Our most limiting assumption concerning acceleration is that players accelerate with constant non-directional acceleration a until they reach maximum velocity. Once players reach their peak speed $s_{\max}$, their acceleration drops to zero. This movement assumption is related to intent, where a player attempts to reach the location of interest as soon as possible. There is support for the constancy of acceleration (at least for short time periods) as displayed in nearly linear velocity curves for sprinters ([5]).

We now return to the calculation of the time t that it takes the player to travel from the current location (x_0, y_0) to the location of interest (x_1, y_1). The first step is the determination of the time t_* that it takes the player to reach the prescribed maximum speed $s_{\max}$, where t_* is a function of the acceleration (a_x, a_y) profile. Therefore, consider an acceleration vector (a_x, a_y) which lies on the circle $a_x^2 + a_y^2 = a^2$. Given the acceleration, maximum player speed is achieved at time t_* when

$$(v_{x_0} + a_x t_*)^2 + (v_{y_0} + a_y t_*)^2 = s_{\max}^2 \; . \tag{5}$$

The quadratic equation (5) is solved via

$$t_* = t_*(a_x, a_y) = \frac{-(v_{x0}a_x + v_{y0}a_y) \pm \sqrt{(v_{x0}a_x + v_{y0}a_y)^2 - (a_x^2 + a_y^2)(v_{x0}^2 + v_{y0}^2 - s_{\max}^2)}}{a_x^2 + a_y^2} \tag{6}$$

where negative solutions in time are non-sensical. It can be proven that there is a unique solution $t_* > 0$.

Based on the stated motion assumptions, the location of the player at time t is therefore given by

$$\begin{aligned} x_t &= \begin{cases} x_0 + \int_0^t (v_{x_0} + a_x t)\ dt & t < t_* \\ x_0 + \int_0^{t_*} (v_{x_0} + a_x t)\ dt + \int_{t_*}^t (v_{x_0} + a_x t_*)\ dt & t > t_* \end{cases} \\ &= \begin{cases} x_0 + v_{x0}t + (1/2)a_x t^2 & t < t_* \\ x_0 + v_{x0}t_* + (1/2)a_x t_*^2 + (t - t_*)(v_{x_0} + a_x t_*) & t > t_* \end{cases} \end{aligned} \tag{7}$$

and similarly,

$$y_t = \begin{cases} y_0 + v_{y0}t + (1/2)a_y t^2 & t < t_* \\ y_0 + v_{y0}t_* + (1/2)a_y t_*^2 + (t - t_*)(v_{y_0} + a_y t_*) & t > t_* \end{cases} . \tag{8}$$

Equations (7) and (8) present an algorithm for computing t by setting the location of interest $(x_1, y_1) = (x_t, y_t)$. The algorithm proceeds by stepping through acceleration vectors (a_x, a_y) according to the constraint $a_x^2 + a_y^2 = a^2$. For a given (a_x, a_y), we determine $t_* = t_*(a_x, a_y)$ according to equation (6). Having solved for t_*, this simplifies equations (7) and (8). For equation (7), we have the solution $t^{(x)}$, and for equation (8), we have the solution $t^{(y)}$. If $t^{(x)} = t^{(y)}$, this means that for the acceleration vector (a_x, a_y), the player arrives at the coordinates x_1 and y_1 at the same time, and we have a solution $t = t^{(x)} = t^{(y)}$. If there are multiple solutions for different values of (a_x, a_y), then we select the minimum time according to the assumption that players go to the location of interest with intent.

To operationalize the algorithm, suppose that for a given (a_x, a_y), we determine the unique solution t_* according to equation (6). Then, if $t < t_*$ and $a_x \neq 0$, we solve a quadratic equation and obtain

$$t^{(x)} = \frac{-v_{x_0} \pm \sqrt{v_{x_0}^2 - 2a_x(x_0 - x_1)}}{a_x} \ . \tag{9}$$

If $t < t_*$ and $a_x = 0$, then

$$t^{(x)} = (x_1 - x_0)/v_{x_0} \ . \tag{10}$$

However, if $t > t_*$, we solve a linear equation and obtain

$$t^{(x)} = \frac{x_1 - x_0 + (1/2)a_x t_*^2}{v_{x_0} + a_x t_*} \ . \tag{11}$$

Analogous equations to (9), (10) and (11) are available for $t^{(y)}$. Non-sensical solutions $t^{(x)}$ and $t^{(y)}$ imply that it is not possible for the player to reach the location of interest (x_1, y_1) under the given acceleration vector (a_x, a_y).

4 RESULTS

In the development of our pitch control metric in Section 3, we emphasized that control of the pitch needs to be unambiguous. That is, a player on the team on possession must be able to reach a location before a player on the opposing team, and the ball must be delivered in a timely fashion. Consequently, our regions of dominance are typically smaller than alternative pitch control diagrams.

Using the tracking data, we computed the current velocity vector (v_{x_0}, v_{y_0}) for the same example as displayed in Figure 2. The arrows indicating velocity are depicted in Figure 3. We observe that some of the velocities are large and some are small; this together with the relative positioning of players determines the resultant pitch control diagram.

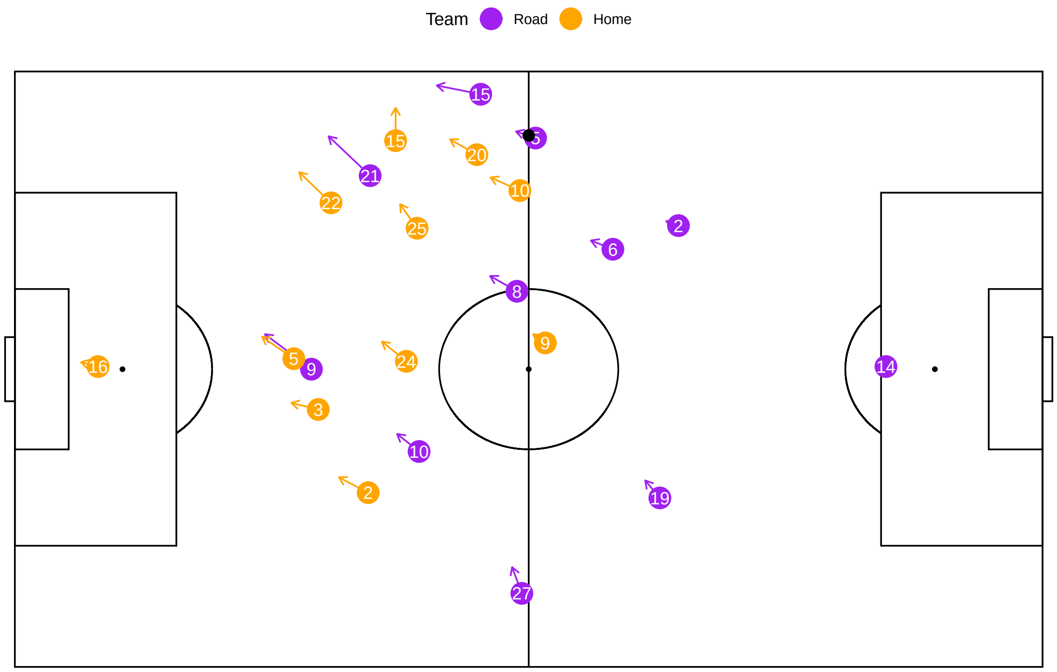

Figure 3: Current velocity vectors for the example depicted in Figure 2.

In Section 3.3, we presented a motion model that derived the time t that it takes a player to reach the location (x_1, y_1) on the pitch given the initial location (x_0, y_0) and given initial velocity (v_{x_0}, v_{y_0}). In Figure 4, the time is presented for both a stationary player (left plot) and a player with a northwest velocity (right plot). In the left plot, we observe colors that radiate in circles such that the player can reach any location of constant distance in the same amount of time. This corresponds to the Voronoi tessellations. We observe non-circular color contours in the right plot where the player can reach positions in the northwest quicker than in other directions of similar distance. The right plot introduces the reality of players having initial velocities that impact the time that it takes to reach various locations.

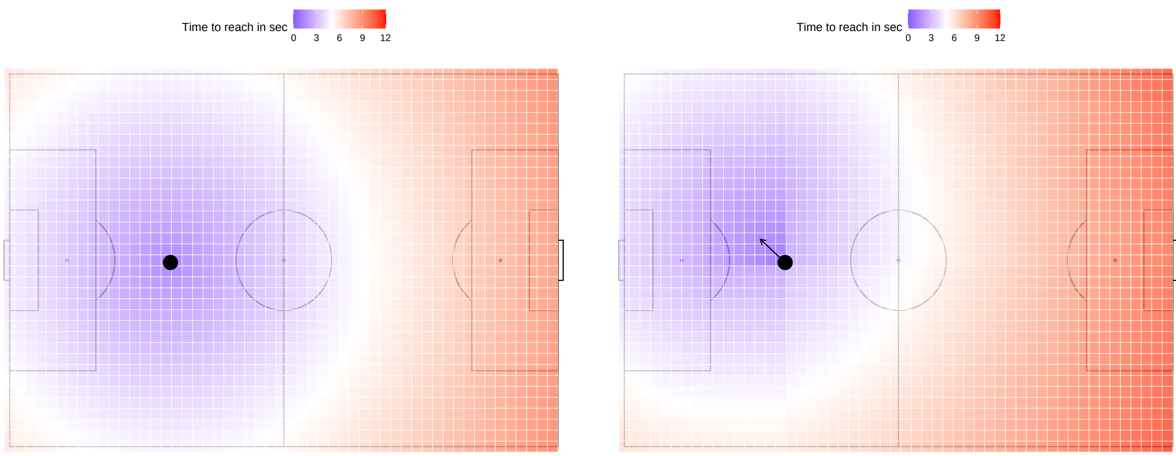

Figure 4: The left plot uses colors to depict the time that it takes a stationary player to reach field locations given the current location marked with a dot. The right plot does likewise but introduces an initial velocity (arrow) for the player.

To determine pitch control regions, we discretize the soccer field (105x68) metres into 1-by-1 metre grids and compute the time taken to reach the centre of each grid for each player. We have set the tuning parameter $\epsilon = 0.5$ seconds as described in Section 3.1 which requires that players arrive at least 1/2 second earlier than opponents to achieve pitch control. The time variables t_b, t_h and t_r are computed according to the methods described in Sections 3.2 and Sections 3.3, and Table 1 is used to determine the pitch control regions. Figure 5 provides the resultant pitch control map. When comparing the proposed approach illustrated in Figure 5 with other approaches, we emphasize that Figure 5 contains grey ambiguous areas where neither team has pitch control. This is sensible as there are locations where players on both teams can arrive before the ball, and therefore neither team can lay claim to the location. There are also grey locations which a player on one team cannot reach $\epsilon = 0.5$ seconds in advance of the opponent. For example, there are locations in the top-left regions of the field where player #22 (home) and player #21 (road) would arrive at roughly the same time.

Comparing Figure 5 with Figure 2 (Voronoi), we see that initial velocities play an important role in the determination of the pitch control diagram. For example, in Figure

5, there is a location of interest immediately southeast of player #8 (road), yet it is controlled by the home team. This is not the case in Figure 2. Here, player #9 (home) can reach this location quicker than player #8 (road) even though player #9 is further away. The reason is that player #9 is moving towards the location whereas player #8 is moving away and needs to reverse direction.

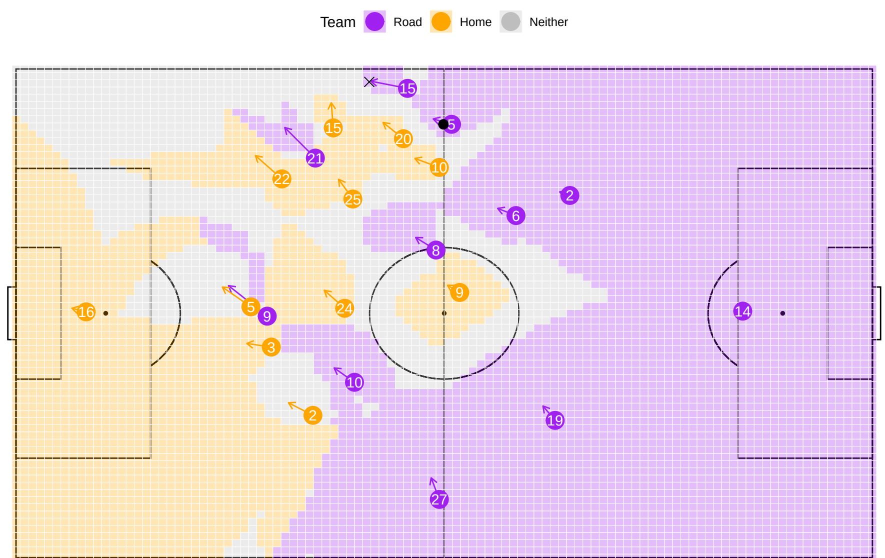

Figure 5: Pitch control diagram using the proposed methods for the example depicted in Figure 2.

4.1 Computation

It takes approximately 0.2 seconds on a laptop computer to evaluate a pitch control decision for a given location on the field. Whereas this seems reasonable, pitch control applications become computationally intensive as there are typically many locations of interest and many temporal-spatial snapshots of interest involving the initial locations and velocities of the 22 players. Fortunately, parallel processing may be implemented for the repeated tasks.

As an example of a realistic application, suppose that a team was interested in a striker's decision making based on all of his touches in a given match. For example, for each touch, an analyst may wish to investigate the location of the player's passes relative to pitch control dominance. For top-level soccer, [20] observe that strikers have in the vicinity of 30-60 touches. Therefore, with 50 touches, and 10 parallel machines, the evaluation of pitch control at 2500 1-by-1 metre grid locations (attacking third of the pitch) would require approximately 50(2500)(0.2 sec)/10 $\rightarrow$ 42 minutes of computation.

For a given target location on the field, we perform a grid search over 2000 combinations of a_x and a_y subject to the constraint $a_x^2 + a_y^2 = a^2$ to find the pair (a_x, a_y) that gives the shortest time t to reach the target location. For a pitch control diagram, the computational time is heavily affected by the number of grids specified for the soccer field. By halving the size of grids, we could reduce the computation time by a factor of four. More sophisticated optimization algorithms for the evaluation of t could be considered in future work. Such algorithms may avoid the consideration of unpromising acceleration pairs (a_x, a_y). It may also be possible to introduce computing efficiencies by eliminating some players in the determination of minimum time t to reach a location. Players who are distant from the location of interest cannot reach the location in minimum time.

4.2 Accuracy of the Metric

Whereas the literature has introduced various approaches to pitch control, the literature is sparse on validation. Validation is particularly difficult when color-codings are not probabilistic. In our approach, we are able to investigate validation as the field is segmented in three regions according to ownership by the two teams and neither team.

We sampled 10 games from the 2019 season of the CSL and obtained data on 7901 intended passes. We first classified the passes as either successful or intercepted. We then calculated our pitch control metric for each of these passes and further classified the destination location as either controlled by the intended team, the opponent or neither team. The results of the two-way classification are given in Table 2.

An initial observation from Table 2 is that there are naturally more successful passes (6826) than intercepted passes (1075). This corresponds to a successful pass rate of 86%. If we omit the "neither team" designation, there are 5887 passes of which $5275 + 55 =$

5330 are controlled by the predicted team according to the pitch control model. This is suggestive of a 91% accuracy rate in pitch control designation. However, it can be demonstrated that this is a conservative estimate. For example, there may be some passes that arrive at a player's feet and should be controlled. However, by some technical error on the part of the player, the opponent gains control. When we look at the passes whose pitch control designation is "neither team", we observe that this corresponds to 25% of the passes (i.e. $1470+544 = 2014$ passes out of 7901). These cases are truly more doubtful, as only 27% of them are received as intended (i.e. 1470 passes out of 2014). We emphasize that our model provides two tuning parameters that allow us to increase/decrease the number of passes that are classified as "neither team". By increasing the speed s of the ball (Section 3.2), the ball will rarely lag the players to the location of interest, and consequently, this will reduce the size of ambiguous regions according to Table 1. Also, by increasing the time ϵ that a player needs to arrive before the opponent (Section 3.2), this will increase the size of ambiguous regions.

	Successful Pass	Intercepted Pass
PC by Intended Team	5275	476
PC by Opposition Team	81	55
PC by Neither Team	1470	544

Table 2: The classification of 7901 intended passes according to whether pitch control (PC) was designated to the intended team, the opponent or neither team.

5 DISCUSSION

From the original work on pitch control established via Voronoi tessellations, there has been various attempts to define field ownership. In this manuscript, we have provided a motion model and heuristics (Table 1) that are straightforward but adhere (at least approximately) to the physics of running. A difference between our approach and most of the methods proposed in the literature is that we define regions corresponding to the home team, the road team and neither team. This allows us to validate the accuracy of pitch control diagrams (Section 4.2) whereas this is not possible with approaches that provide non-probabilistic color-codings. We have reported a 91% accuracy rate of the proposed pitch control region which is a conservative estimate. In addition, unlike some of the proprietary methods for pitch control, we have provided code (see Supplementary file) for determining pitch control regions. The code is realistic in terms of physical motion and can be adapted to various sports via intuitive parameters.

Our pitch control model offers opportunities to enhance/modify the approach. For example, it is possible to introduce player-specific maximum velocities and player-specific accelerations. For example, it is well-known that a fast player in a wide open space poses an attacking threat, and this may be better illustrated through the introduction of player-specific parameters. Such parameter settings may be obtained through historical data.

It is also possible to vary the tuning parameter ϵ which dictates the amount of time needed by players to gain control of the ball. This affords flexibility in the approach where, for example, youth soccer players require more time to achieve ball control than professional soccer players. The parameter ϵ may also be formally explored in sensitivity studies to see how the regions of ambiguous pitch control decrease with increasing ϵ. For example, one may vary ϵ to achieve the greatest accuracy rate according to the pitch control study of Section 4.2.

We have argued that the categorization of pitch control into three regions (Team A, Team B, neither team) is realistic. However, if we wish to reduce the categorization to the more common control regions (Team A, Team B), this is easily implemented by adjusting the tuning parameters. Specifically, one sets $\epsilon = 0.0$ seconds such that players require no time to control the ball and $s_{\text{max}} = \infty$ metres/sec such that the ball arrives instantaneously at the location of interest.

A limitation of our work is the setting of the ball speed at $s = 18$ metres per second. Clearly, a player in possession has some control over the speed of a pass. Also, long passes tend to be made at faster speeds than short passes. Therefore, $s = 18$ metres/sec ought to be viewed as an approximation to reality. In future work, it may be possible to improve on this by estimating the speed of the pass based on the distance of the pass. We note that faster ball speeds decrease the size of the ambiguous regions where neither team has pitch control.

6 REFERENCES

[1] Albert, J.A., Glickman, M.E., Swartz, T.B. and Koning, R.H., Editors (2017). *Handbook of Statistical Methods and Analyses in Sports*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Boca Raton.

[2] Brefeld, U., Lasek, J. and Mair, S. (2019). Probabilistic movement models and zones of control. *Machine Learning*, 108, 127-147.

[3] Caetano, F.G., Barbon Junior, S., Torres, R da S., Cunha, S.A., Ruffino, P.R.C., Martins, L.E.B. and Moura, F.A. (2021). Football player dominant region determined by a novel model based on instantaneous kinematics variables. *Scientific Reports*, 2(1), 1-10.

[4] Cervone, D., D'Amour, A., Bornn, L. and Goldsberry, K. (2016). A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514), 585-589.

[5] Chatzilazaridis, I., Panoutsakopoulos, V. and Papaiakovou, G.I. (2012). Stride characteristics progress in a 4-M sprinting test executed by male preadolescent, adolescent and adult athletes. *Biology of Exercise*, 8(2), 5-23.

[6], Efthimiou, C.J. (2021). The Voronoi diagram in soccer: A theoretical study to measure dominance space. https://arxiv.org/pdf/2107.05714.pdf

[7] Fernández, J. and Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. *MIT Sloan Analytics Conference*, retrieved February 24/21 at https://www.sloansportsconference.com/research-papers/wide-open-spaces-a-statistical-technique-for-measuring-space-creation-in-professional-soccer

[8] Fernández, J., Bornn, L. and Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, 110(6), 1389-1427.

[9] Fernández, J. (2023). A framework for the analytical and visual interpretation of complex spatiotemporal dynamics in soccer. *PhD thesis at Universitat Politècnica de Catalunya. Departament de Ciències de la Computació*, http://hdl.handle.net/10803/673529

[10] Fujimura, A. and Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan* 36(6), 49-58.

[11] Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Ridrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S. and Lemmink, K.A.P.M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sports Science*, 21(4), 481-496.

[12] Gudmundsson, J. and Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47, 16-27.

[13] Gudmundsson, J. and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), Article 22.

[14] Kim, S. (2004). Voronoi analysis of a soccer game. *Nonlinear Analysis: Modelling and Control*, 9(3), 233-240.

[15] Link, D., Lang, S. and Seidenschwarz, P. (2016). Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS ONE*, 11(12), 1-16.

[16] Martens, F., Dick, U. and Brefeld, U. (2021). Space and control in soccer. *Frontiers in Psychology*, Volume 3, Article 676179.

[17] Memmert, D., Lemmink, K.A.P.M. and Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1), 1-10.

[18] Miller, A., Bornn, L., Adams, R.P. and Goldsberry, K. (2014). Factorized point process intensities: A spatial analysis of professional basketball. In *Proceedings of the 31st International Conference on Machine Learning* - Volume 32, JMLR.org, Beijing, 235-243.

[19] Reyers, M. and Swartz, T.B. (2021). Quarterback evaluation in the National Football League using tracking data. *AStA Advances in Statistical Analysis*, DOI: 10.1007/s10182-021-00406-8

[20] Saunders, T. (2018). The average touch success ratio for strikers from the 'Big Six'. *GIVEMESPORT*, https://www.givemesport.com/1428519-the-average-touch-success-ratio-for-strikers-from-the-big-six/

[21] Shaw, L. (2020). Advanced football analytics: Building and applying a pitch control model in Python. *Friends of Tracking*, YouTube video accessed February 25/21 at https://www.youtube.com/watch?v=5X1cSehLg6s

[22] Spearman, W., Basye, A., Dick, G., Hotovy, R. and Pop, P. (2017). Physics-based modeling of pass probabilities in soccer. *MIT Sloan Sports Analytics Conference*, Accessed on December 14, 2020 at https://www.researchgate.net/publication/315166647_Physics-Based_Modeling_of_Pass_Probabilities_in_Soccer

[23] Spearman, W. (2016). Quantifying pitch control. *2016 OptaPro Analytics Forum*, DOI: 10.13140/RG.2.2.22551.93603

[24] Spearman, W. (2018). Beyond expected goals. *MIT Sloan Sports Analytics Conference*, Accessed on September 21, 2020 at http://www.sloansportsconference.com/wp-content/uploads/2018/02/2002.pdf

[25] Taki, T., Hasegawa, J. and Fukumura, T. (1996). Developmentof motion analysis system for quantitative evaluation of teamwork in soccer games. *Proceedings of 3rd IEEE International Conference on Image Processing*, Volume 3, 815-818.

[26] Taki, T. and Hasegawa, J. (2000). Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings of the International Conference on Computer Graphics*, 227-235.

[27] Voronoi, G. (1907). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Primiere Mémoire: Sur quelques prepriétés des formes quadratiques positives parfaites, *Journal für die reine und angewandte Mathematik*, 133, 97-108.

[28] Wu, L. and Swartz, T.B. (2022). The calculation of player speed from tracking data. *International Journal of Sports Science and Coaching*, https://doi.org/10.1177/17479541221124036

[29] Wu, Y., Xie, X., Wang, J., Deng, D., Liang, H., Zhang, H., Cheng, S. and Chen, W. (2019). ForVizor: Visualizing spatio-temporal team formations in soccer, *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 65-75.

Acknowledgments: L. Wu is a PhD candidate, and T. Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been partially supported by the Natural Sciences and Engineering Research Council of Canada. This work has also been supported by a CANSSI (Canadian Statistical Sciences Institute) Collaborative Research Team (CRT) in Sports Analytics. The authors thank two anonymous reviewers whose comments helped improve the manuscript.

Author Contribution Statement: L.W. carried out all of the computing associated with the contribution and T.S. wrote the paper. The development of ideas and methods were shared.

Additional Information: The authors declare no competing interests. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.