Wide Open Spaces: A statistical technique for measuring space creation in professional soccer

Javier Fernandez

F.C. Barcelona

javier.fernandezr@fcbarcelona.cat

Luke Bornn

Simon Fraser University, Sacramento Kings

lbornn@sfu.ca

1 Introduction

Soccer analytics has long focused on the outcomes of discrete, on-ball events; however, much of the sport's complexity resides in off-ball events. In the words of Johan Cruyff: "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." The creation and closure of spaces is a recurrent subject in observation-based tactical analysis, yet it remains highly unexplored from a quantitative perspective.

We present a method for quantifying spatial value occupation and generation during open play. Here direct space occupation refers to space created for oneself, while space generation refers to opening up space for teammates by attracting opponents out of position. We first build a novel parametric pitch control model that incorporates motion information, relative distance to the ball, and player position in order to provide a smooth surface of potential ball control. Through the mixture of all players' control surfaces we obtain a fuzzy degree of potential ball control at the team level in any given moment. We also construct a model for the relative value of any pitch position, based on the position of the ball and using feed forward neural networks. From all this (a player's invested pitch zones, a team's pitch control, and the relative value of each zone), we employ the full spatio-temporal dynamics of each player to construct two novel spatial value creation metrics, accounting for both occupation and generation of spaces.

Through the analysis of a first division Spanish league match, we show a handful of approaches to better understand a missing key factor for performance analysis in soccer: off-ball attacking dynamics. The quantification of space occupation gain and space generation allows us to observe Sergio Busquets' high relevance during positional attacks through his pivoting skills, the dragging power of Luis Suarez to generate spaces for his teammates, and the capacity of Lionel Messi to occupy spaces of value with smooth movements along the field, among many other characteristics.

The level of detail we can reach with automated quantitative analysis of space dynamics is beyond what can be reached through observational analysis. The capacity of evaluating space occupation and generation opens the door for new research on off-ball dynamics that can be applied in specific matches and situations, and directly integrated into coaches' analysis. This information can be used not only to better evaluate players' contributions to their teams, but also to improve their positioning and movement through coaching, providing a key competitive advantage in a complex and dynamic sport.

2 Occupation and Generation of Spaces

During the last two decades, successful elite soccer teams have been increasingly adapting possession-centered playing styles, especially within Spanish, English and German leagues. Beyond the basic idea of simply controlling the ball as much as possible, these strategies comprise of a large set of on-ball and off-ball actions to generate better scoring chances. Some of these are: creating superiorities (numerical, positional or qualitative), creating disorder on the opponents defense through movement and team-collaboration, building up plays from the goalkeeper, and executing passes with offensive intention, among many others. Above all these actions there is a main underlying concept: the generation and occupation of spaces. Pep Guardiola once said:

> We have to pass the ball, yes, but with clear intention. Pass it to drag players to one side and creating space in the opposite side. Then, move the ball there. That's our game.

Occupying space on the field is fundamentally about a player's act of continually positioning himself in an area of high value. The value of space can be defined in terms of the relative position of the ball, the closeness to the opponent's goal, and more specifically the level of ownership of space, regarding the density of opponents within the given area. Furthermore, we can cluster the types of occupation of space depending on the speed of the player. Specifically we identify two types: active occupation, when the player moves at running speed to earn the space, and passive occupation, when the player is below running speed (jogging or walking). For instance, if a player is closely marked and then runs towards a free space faster than the opponent, he will obtain a gain on owned space through active space occupation. As another example, if the player is walking towards a given area and nearby opponents move away from that area, the player will be gaining space through passive occupation.

A more complex concept is that of space generation. We define the generation of space as the action of dragging opponents out of certain areas to create new available space in previously covered areas. Specifically, we identify situations where a player drags an opponent away from another teammate whom the opponent was close to originally. The dragging concept is, at its simplest, creating space for a teammate by pulling their defender towards oneself. Notice that unoccupied space could also be generated when dragged players leave a clear area; however, we are not considering this case for this study. Similarly to the Space Occupation Gain (SOG) concept above, later we also explore the concept of Space Generation Gain (SGG). In this way, we separate out space created for oneself from space created for teammates, in both a passive and active manner.

Figure 1 presents an example of both space occupation and space generation during an official Spanish first division match. The three images show a process where Andrés Iniesta moves to clear up space away from the ball and then attacks a high value space inside the box. When he moves to this space he drags three defenders towards himeslf while also receiving a pass. The attraction of the three defenders leaves open space for Lionel Messi, who in this newfound space receives a pass free of a mark, and subsequently sends a lob pass onwards for Suarez, who meanwhile was running towards the goal line in search of space of value to score. A more detailed video example of occupation and generation of spaces can be seen at the following link, where players are highlighted when adding space for themselves or teammates: http://www.lukebornn.com/sloan/space_occupation_1.mp4

Before providing explicit details on how to calculate space occupation and generation, we first need a better notion of space ownership and value, as creating space for your own goalkeeper who is 80 meters
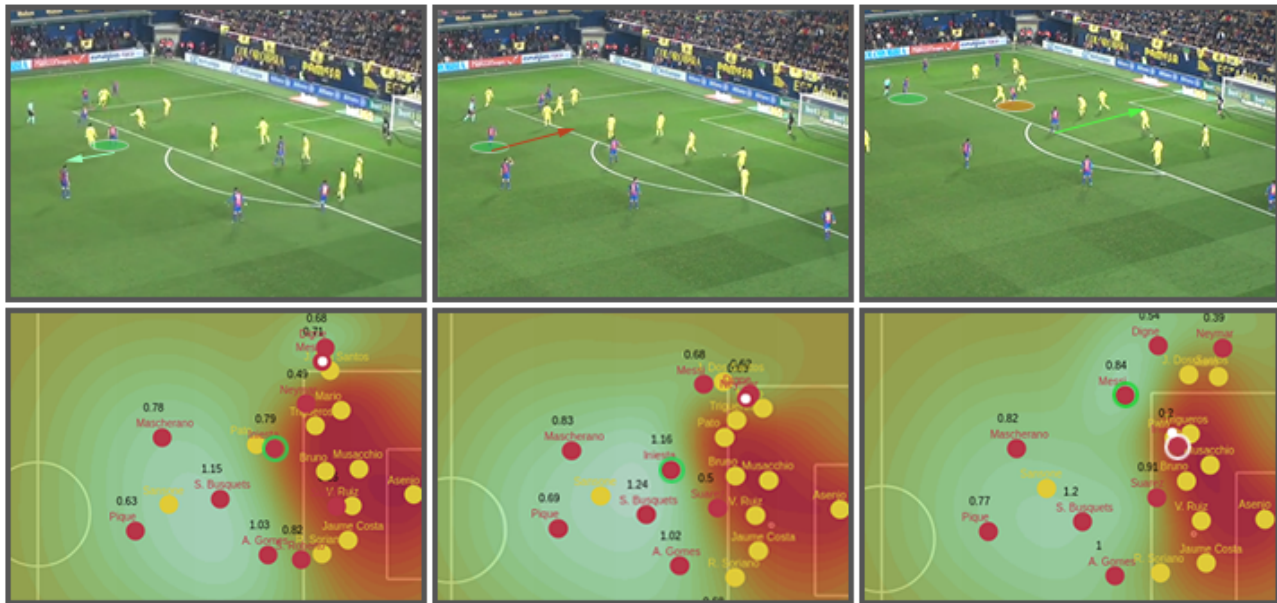
Figure 1: A game situation presenting both space occupation and generation. From left to right: in the first frame Iniesta moves back to occupy a space of value with higher control. In the second frame, Iniesta observes an open space to attack. He moves towards the space, dragging three defenders. In the third frame, the three dragged defenders leave an open space for Messi that can now receive the ball free of mark, while Suarez runs towards the goal line enabling him to receive a pass.

behind the location of play is worth much less than creating space in high-threat areas nearer the ball and goal. The next two sections present a novel pitch control model for evaluating space ownership, and a dynamical model for space value according to ball and player position.

3   Modelling Pitch Control: A Parametric Approach

Pitch control is a recurrent concept in the analysis of space dominance in team sports. It can be defined as the degree or probability of control that a given player (or team) has on any specific point in the available playing area. The emergence of player tracking data has given rise to different pitch control (or dominant region) models. A widely applied model is the Voronoi tesselation, which takes into account the position of all players on the field and calculates the closest player to each given spatial point, finding dominant cells for each player. This model has been used for quantifying the dominant area of attacking and defending teams in constrained playing areas [1], to evaluate the space dominance based on passing behaviour [2], to improve models for pass probabilities [3], and to evaluate positional value of players on rebounds in basketball [4], among many other applications. From the original model for team sports presented by Taki and Hasegawa [5] there have been several extensions for faster computation, as well as extensions to incorporate motion and weighted valuation of dominant space [3, 6]. Beyond its benefits, all the different Voronoi tesselation-based approaches start from the idea of finding regions that are exclusively dominated by a given player. This concept disregards the concept that ownership of space is continuous, not discrete, with uncertainty in who controls areas between players. Also, the distance between players and the ball is also believed to influence the relative positioning and degree of space control, especially for sports with wider playing spaces such as soccer; however, this is not taken into account by the mentioned approaches.

We propose a novel pitch control model that takes into account the location, velocity and distance to the ball of all the players, providing a smooth surface of control for each team. For any given location, the influence that every player has in that place is computed and summarized, resulting in a probability of control. An additional objective of this approach was to provide a model that could be applied in a given data frame, without requiring significant data for learning its parameters. This is particularly important for clubs in competitions such as the Spanish League where tracking data is not available for direct usage. Also, such a model would allow easier reproducibility.

3.1 Player Influence Area

distance to the ball --> area he could reach the ball in if it approached him

Depending on their location in time, players might have different levels of influence on nearby zones. When a player is far away from the ball his level of influence can be understood as a wider area, based on the reasoning that if the ball moves towards the player he would have a more time to reach the ball within a larger space. On the opposite, when closer to the ball, the player has less possibilities of reaching the ball if it moves from its current location. Also, the player's velocity plays an important role in defining the area of influence. A player at running speed might have more influence in the areas in the direction of speed compared to if they were walking or jogging. Further, the player may have higher levels of influence in close spaces than in farther spaces.

player velocity --> direction of control!

Based on this reasoning we propose defining the player influence area through a bivariate normal distribution, whose shape can be adjusted to account for the player's location, velocity, and relative distance to the ball. At any given location a degree of influence or control can be queried through the distribution's probability density function.

Specifically, the player's influence I at a given location p for a given player i at time t is defined by a bivariate normal distribution with mean $\mu_i(t)$ and covariance matrix $\Sigma_i(t)$, given the player's velocity $\vec{s}$ and angle θ. For a given location in space p at time t, the probability density function of player i influence area is defined by a standard multivariate normal distribution. The player's influence likelihood is then defined as the normalization of f at the given location p by the value of f at player's current location $p_i(t)$, as shown in Equation 1.

$$I_i(p, t) = \frac{f_i(p, t)}{f_i(p_i(t), t)} \tag{1}$$

This formulation provides an initial model for obtaining a degree of influence within a $[0, 1]$ range for any given location on the field. The mean and covariance matrix can be dynamically adjusted to provide a player dominance distribution that accounts for location and velocity. In Appendix A.1 we provide specific details for this equation.

Figure 2 presents the player influence area in two different situations regarding the player's distance to the ball and velocity. Here we can observe how depending on the distance to the ball the range of influence of the player varies. Also, the distribution of player influence is reshaped to be oriented according to the direction of movement and stretched in relation to the speed. If the player is in motion, the distribution is translated so the higher level of influence is near points where the player can reach faster, according to his speed. This model can easily be expanded to handle player-specific movement characteristics, such as acceleration and maximum speed.
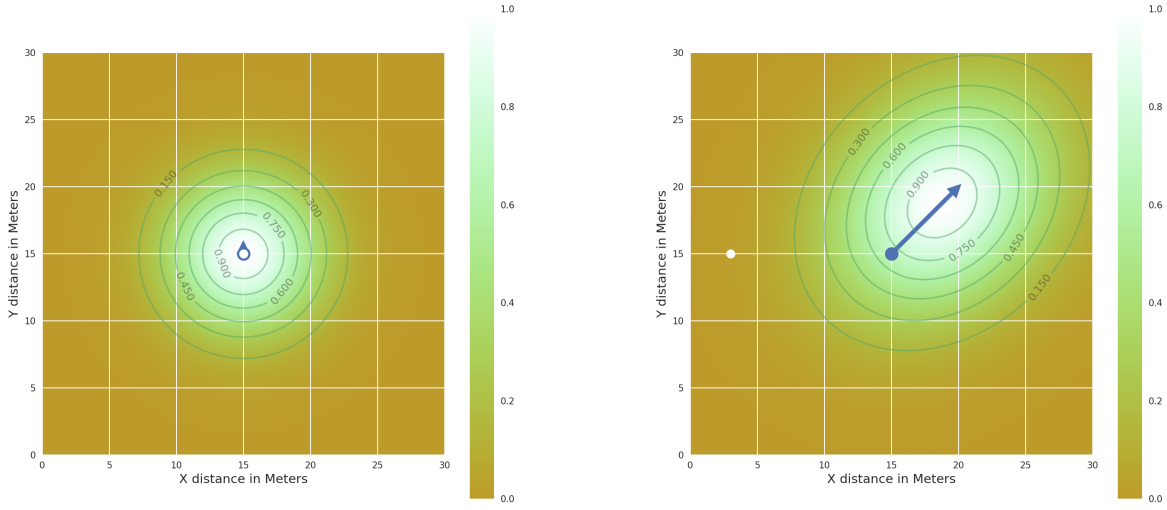
(a) Player influence area for player in possession of the ball and no speed (lower than walking speed)

(b) Player influence area for player 15 meters away from the ball, running at 6.36 m/s in a 45 degrees angle

Figure 2: Two situations representing the player influence area

3.2 Modelling Team Pitch Control

When defining a team's degree of control at any given location on the field, it is desirable to take into account the level of influence each individual player of both teams is having on that point. Since many players can have influence at a given location at a certain time, the model should be able to account for the aggregated influence of each player and provide a value of control within a continuous range, instead of strict areas such as the case of Voronoi Tesselation.

Based on this, we present a pitch control model that summarizes the level of influence of every player, and outputs a degree of control for any part of the pitch. Equation 2 presents the pitch control level at a location p at time t, where i and j refers to the index of the player in each opposing team. Here the logistic function transforms the substraction of the accumulated individual influence area of each team into a degree of control within the $[0, 1]$ range. Also, since we are defining a team-oriented pitch control model, a single player without any influence of any other player at its current location only controls $logistic(1) = 0.73$ of the space. This provides the need of higher density of players near a given area to provide higher level of control in that area. For the statistically-inclined, note that this formulation represents the probability of control of a given team, where each team's latent surface is captured by a kernel-based non-parametric point process,

$$PC(p,t) = \sigma(\sum_i I(p,t) - \sum_j I(p,t)) \tag{2}$$

where σ is the logistic function. Since the pitch control model follows the definition of player influence area in Figure 2, the model is taking into account the location of the ball, the players' velocities and the location of all the players on the field. Equation 2 is a simplified version of pitch control calculation based
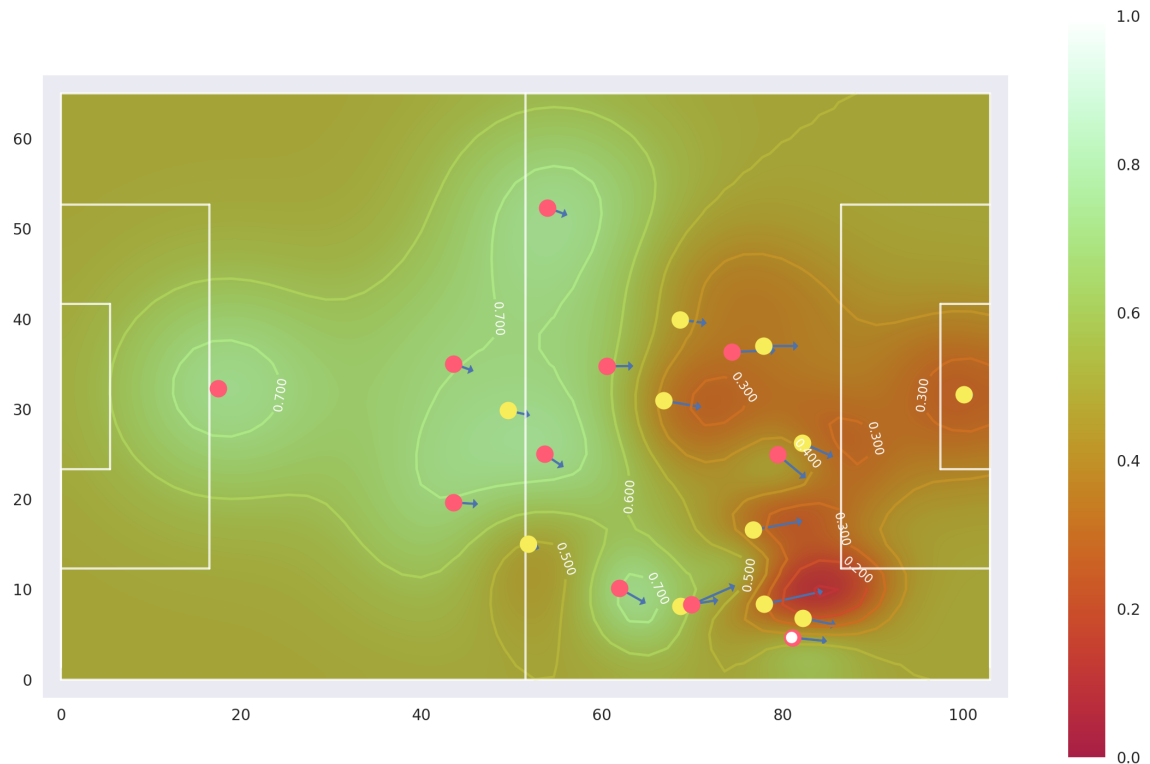
5

Figure 3: Pitch control surface indicating the degree of control for team in red. Arrows show players velocities, and contour lines allow to visualize the surface geometry. Numbers in white indicate the pitch control value at their drawing location. Axis dimensions are in meters.

on players' influence areas. Note we can include a constant within σ to add more flexibility, if desired. Figure 3 presents the pitch control surface in a given situation of the match. At location $(82, 8)$, near the ball, it can be observed clearly how the yellow team's high density provides lower level of control for the red team near the ball. Also the velocity of the player in possession of the ball (red team) provides the red team an advantage in the running direction. At location $(80, 25)$, the red player is creating a positional advantage. Meanwhile, at location $(50, 30)$ the yellow player has minimal control of space because of the high density provided by the three surrounding opponent players. For a single time frame, this pitch control model provides a synthesis of player locations, player velocities and ball-relative positioning in one variable. Also, by exploiting the dynamics of pitch control time, it becomes a versatile tool for evaluating multiple types of spatio-temporal characteristics of the game such as the creation of positional advantages, the influence of density and pressure speed in defending situations, and the creation and generation of spaces.

4   Quantifying Pitch and Space Value

Although the control of space is a fundamental element to identify occupation and generation of spaces, we still need an additional part of the equation: the value of space. The sole fact of moving for finding better passing options is an advantage itself. However, it can be easily argued that not every space has the same value. A trivial method for determinating the value of space is its distance to the opponent goal. Its well known that spaces near the goal have an increased value, given the advantage that would provide to dominate them. But exploring more deeply into the dynamics of soccer, and based on the opinion from F.C. Barcelona expert analysts, it can be also argued that the value of space changes dynamically depending on multiple positional factors, such as the position of the ball and the players. In order to quantify in a detailed way the value of the space generated or occupied we provide a novel model for finding the relative pitch value on every position of the field, depending on the location of the ball. The following link presents a video where the dynamic evaluation of pitch value depending on the ball position, as detailed below, can be observed: http://www.lukebornn.com/sloan/field_value.mp4.

Instead of defining a priori a model for space valuation we would like to extract a sense of space value from the spatio-temporal behaviour of players during multiple matches. For this we set the following hypothesis: considering a sufficiently high number of situations, the defending team distributes itself throughout the field in a manner which covers high value spaces. Although it is clear that at any given point defenders will deviate based on overloads, specific offensive player positioning, and other scenarios, in general, most players will remain close to high value areas. An extreme example of this will be the case where the attacking team places all players in the middle of field. It is arguable that, although this would impact the position of the defending team, they will most probably still keep players near the box and their own goal. Note that similar ideas are used when identifying defensive matchups based on defender locations in basketball [7].

Based on this, we propose learning the sum influence that a defensive team would have in a given location on the field, given the location of the ball. Let $V_{k,l}(t)$ be the value of location $p_{k,l}$ of the pitch at time t, and let $p_b(t)$ be the location of the ball at time t, we want to learn a function fn with parameters θ that values space as a function of of the ball,

$$V_{k,l}(t) = fn(p_b(t), p_{k,l}(t); \theta) \tag{3}$$

At a first glance it seems the relation between the position of the ball and the field location for predicting location value is complex and non-linear. While we have tested several linear models, we have observed significantly improved performance with non-linear alternatives, so stick to those here. In order to solve the proposed problem, we use a feed forward neural network with one hidden layer, that aims to learn the parameters θ of the mapping defined in Equation 3. For the learning process we build a dataset where the target value $V_{k,l}(t)$ is calculated by minimizing the loss 4, which corresponds to the sum of player influence for every defending player d in a given situation.
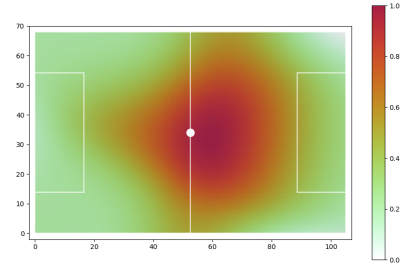
value of space is estimated by the usual degree of control by the defensive unit has

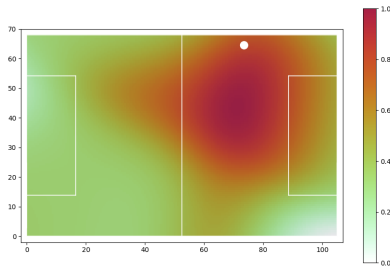$$D_{k,l}(t) = \sum_d I_d(p_b(t), p_{k,l}(t)) \tag{4}$$

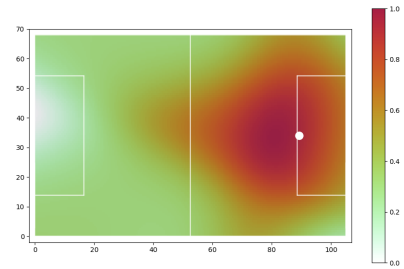$$\hat{V}_{k,l}(t) = \begin{cases} 1 & D_{k,l}(t) > 1 \\ D_{k,l}(t) & \text{otherwise} \end{cases}$$
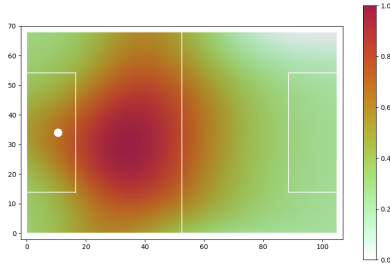
(a) Pitch value for ball vertically centered at the first quarter of the field

(b) Pitch value for ball at the center of the field

(c) Pitch value for ball at the third quarter of the field on top of the left lane
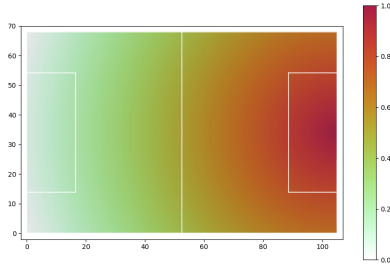

(d) Pitch value for ball vertically centered in the fourth quarter of the field

Figure 4: Predicted pitch value in a [0,1] range for given ball location (white circle)

Defensive situations are found by selecting game situations where the opponent has possession of the ball. Then, the sum of player influence is found for every location (k, l) within a 21 by 15 grid, for every defending player i. Situations are selected so they are separated in time by at least three seconds. Here we employ Metrica Sports tracking-data of 20 matches of first and second B Spanish division, consisting of 2.4 million examples. For learning the parameters we use a feed forward neural network with one hidden layer using the *adam* optimization algorithm [8]. Specifically, we aim to find the optimal parameters θ_* that minimize the loss function L as presented in Equation 5. We selected mean square error as loss function L and sigmoid function as the activation function f.

$$\mathcal{L}(\theta) = \arg\min_{\theta} \frac{1}{n} \sum_{e=1}^{n} L(y_e, f(x_e, \theta)) \tag{5}$$

We found the best model through a 10-fold cross-validation process. In order to obtain a valuation of a field location for a given ball location, we now query the learned model. Figure 4 shows three different ball position scenarios and the obtained field valuation. This model has learned that nearby locations to the ball have increasing value for a certain range, while understanding effectively how to translate this value depending on ball position. The model still lacks from the natural intuition that space generated at the higher valued locations of the first quarter of the field should not have an identical valuation than those of higher valued locations at the last quarter. In other words, the cumulative value of space is higher when further up the field, closer to the opponent's goal. In order to adapt to this intuitive thinking we
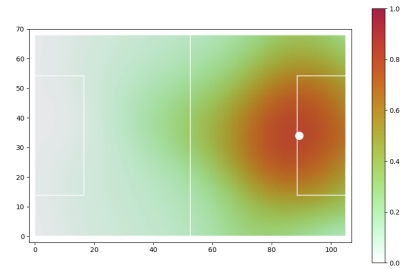
(a) Distance to goal pitch value normalization surface
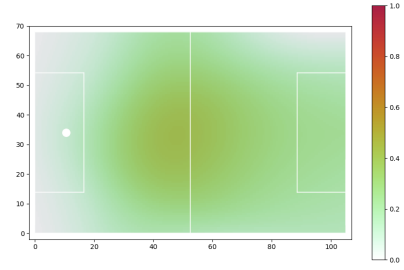
(b) Normalized pitch value for ball vertically centered at the first quarter of the field
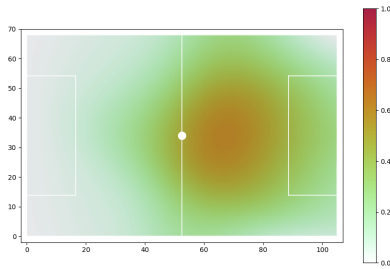
(c) Normalized pitch value for ball at the center of field

(d) Normalized pitch value for ball vertically centered the fourt quarter of the field

Figure 5: Predicted pitch value in a [0,1] range for given ball location (white circle) normalized by a distance to goal model

normalize the obtained pitch value by the distance to the goal of every location normalized on a $[0,1]$ range. Figure 5 presents the normalization surface and three different pitch value situations, where the results still adapt to ball location but show a more consistent valuation of the pitch which adjusts for the threat of the ball location, according to expert analysts. We see that when one's own goalkeeper has the ball, the overall value of space is limited, but when in the opponent's box, space is much more valuable alongside the looming threat of a shot on goal.

5 Occupation and Generation of Spaces

Previously, we approached the occupation and generation of spaces as actions focused on the improvement of the quality of team positioning, with the purpose of reaching better goal scoring chances. The quality of positioning of a given player is then related with having the best possible control of the space, and doing so for spaces with higher value. We could then express the quality of owned space Q as a function of the level of ownership (control) PC and the value of space V, as presented in Equation 6.

$$Q_i(t) = PC_i(t)V(t) \tag{6}$$

Based on the definitions of Sections 3 and 4 we can model $PC_i(t)$ through our team pitch control model, and $V(t)$ using the ball-relative field value model. Figure 6 presents the team pitch control, the pitch value

9

and the obtained quality of owned space, in a given match situation. We can now define with detail our two main proposed concepts: Space Occupation Gain and Space Generation Gain.

5.1 Space Occupation Gain

Now that we have the necessary tools to represent the value of space ownership in a given time, we can define a model for identifying gain in space occupation in time. As mentioned in section 2 we propose the Space Occupation Gain (SOG) concept as the relative amount of quality of owned space earnt during a time window. An opposite concept is that of Space Occupation Loss (SOL), which relates to a negative gain during the time window. We first define the concept of gain in time G as the mean difference of quality of space occupation Q during a time window $[t + 1, t + w + 1]$, for a given player i. This is expressed in Equation 7.

$$G_i(t) = \frac{\sum_{t'=t+1}^{t+w+1} Q_i(t')}{w} \tag{7}$$

Given the dynamic nature of football, players are involved in a continuous process of winning and losing space. A small gain of space can happen when the nearby defenders follow the ball when it moves away from the player, leaving the player a better control of space. However the same can happen in a high speed running situation between the attack and the defender, where the attacker is moving slightly faster. In another case, a medium or high gain of space can happen when the player moves towards a free space. Given this, it is necessary to define a level of space gain from which the earned space can be considered an actual occupational advantage and not a consequence of slower-moving contextual factors in a given situation. We set a constant ϵ as a threshold to account for space occupation gain only in the cases the gain is above that threshold. We can do the equivalent for space occupation loss. Both expressions are defined in Equations 8 and 9.

$$SOG_i(t) = \begin{cases} G_i(t) & \text{if } G_i(t) \geq \epsilon \\ 0 & otherwise \end{cases} \tag{8}$$

$$SOL_i(t) = \begin{cases} -G_i(t) & \text{if } G_i(t) \leq -\epsilon \\ 0 & otherwise \end{cases} \tag{9}$$

An additional concept for refining the idea of gaining space quality is the way that space is gained, specifically regarding a player's speed. We present two definitions: active and passive space occupation

(a) Pitch control surface (b) Pitch value based on ball position (c) Value of the owned space as product of pitch control and field value
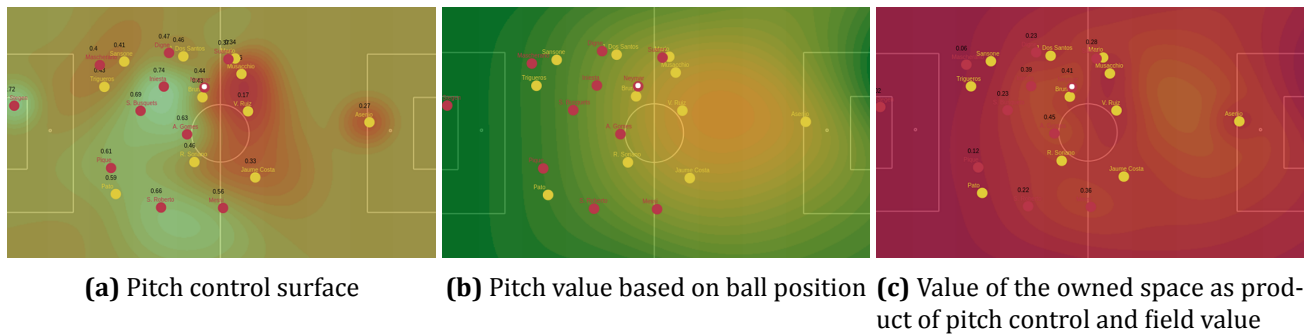
Figure 6: Pitch control, field value and value of owned space for attacking team in red, for attacking direction left to right

gain. A space is occupied actively when the player moves towards that space with a greater speed than a jogging pace (>1.5m/s [9]). Otherwise we consider that space to be occupied passively.

5.2 Space Generation Gain

The generation of space for teammates is a concept that involves two or more teammates during certain attacking situation. Two main types of actors are present: one generator and one or more receivers. The generator is a player that moves toward a certain space while dragging opponents during the process. This dragging behaviour causes the freeing up of space previously occupied by the dragged opponent. When that opponent was previously close to one or more other teammates we say those players are receiving space generated by the attracting player. In order to express this concept mathematically, we need to define a value for closeness. We will say that a player is close to another if the distance between them in a given time t is below a constant δ. Also, it is desirable to define another constant α for constraining the minimum attracting distance, which refers to the difference in distance between the starting and end position of the generator and the attracted opponent. This allows us to avoid innaccurate attractions when players are very close to each other initially. Given this, let $d_{i,j}(t)$ be the distance between players i and j, Equation 10 expresses the concept of space generation SG between any pair of teammates (i, i') and any opponent j, for a time window $[t, t + w]$.

$$SG_{i,i'}(t) = \exists_j (d_{i',j}(t) \leq \delta) \wedge (d_{i,j}(t+w) \leq \delta) \wedge (d_{i',j}(t+w) > \delta) \wedge (d_{i,j}(t+w) - d_{i,j}(t) < \alpha) \quad (10)$$

Once we can identify when a space generation behaviour is occurring, we would like to focus on the cases in which we actually have a gain in space due to the dragging effect. Analogously to the SOG definition, we express the Space Generation Gain (SGG) as space generation situations where the gain is above a threshold ϵ, as presented in Equation 11.

$$SGG_{i_j}(t) = \begin{cases} G_j(t) & \text{if } SG_{i,j}(t) \wedge G_j(t) \geq \epsilon \\ 0 & otherwise \end{cases} \quad (11)$$

Essentially, we are attributing space gain to a player when a defender leaves his mark and moves towards a teammate, subject to the conditions that the defender was close to the player and ended close to the teammate during a time window. It is important to clarify that while SOG and SGG represent two frequent and relevant cases of space gain within soccer, other types of situations and movements might contribute as well to the total space created by a player during a match. An additional possible concept is that of potential space, referring to a space that the player is more likely to reach, within his positioning, but not in his immediate influence area. We will now focus on analyzing SOG and SGG within a match context.

6 Match Analysis

The ability to create and occupy spaces are two commonly trained concepts in modern soccer. During training, coaches interrupt and reshape individual drills to teach players how to orient and move toward spaces and away from low value local zones on the field. When analyzing off-ball performance, coaches appeal to video analysis. Although elite soccer analysis staff typically have a great capacity to understand complex concepts through match visualization, the dynamics of space creation are so frequent and happen in such short time windows, that it becomes impractical for video analysts to grasp them all, even for

a single match. However, is important to note that there is no existance of ground truth data regarding the quantification of spaces in soccer. Hence we have performed an extensive validation of the developed concepts through video and studying individual situations within games, with the help of two expert soccer video analysts from F.C. Barcelona, in order to fine-tune our quantitative approach. The following videos are examples of the video-based validation tool we have used: http://www.lukebornn.com/sloan/space_occupation_1.mp4, http://www.lukebornn.com/sloan/space_occupation_2.mp4

Based on this, we provide a complete summary of off-ball movement statistics for a specific Spanish first division official match between F.C. Barcelona and Villareal F.C. in January 2017. Specifically, we provide an analysis focused on the concepts of space occupation and space generation, using Metrica Sports optical tracking data. This match ended with a 1-1 result, where the first goal was scored by Villareal F.C. at the 49th minute (second half), and the F.C. Barcelona equalizer came at the 90th minute by Lionel Messi. Situationally, this presents a game where F.C. Barcelona was in need of scoring during the final minutes, and were required to occupy and generate the most spaces possible to reach scoring chances. In order to identify space occupation and generation actions we calculate for the attacking situations of F.C. Barcelona all the instances where a player had controlled possession of the ball with his feet. From each of those situations, and alongside expert football analysts from F.C. Barcelona, we define a window w of three seconds after each of these cases, reaching a total of 845 different situations. The closeness factor δ is set to 5 meters, based on the minimum distance an opponent is on average to a player in possession of the ball. We also set the minimum attraction distance for space generation α to 3 meters.

Table 1 presents the space occupation statistics for F.C. Barcelona, sorted in descending order by the total amount of Space Occupation Gain (SOG). At first glance it can be seen that over 41% of gain of space occupation was performed by Iniesta, Sergio Busquets and Lionel Messi. Notably, these three players occupy different positions and have different roles within the team. Busquets is a pivot and has a specific role of helping to drive the ball with controlled possession during build-ups, and to accompany the game creation during positional attacks. Iniesta is an attacking midfielder with great control of the ball, and special skills in moving and finding spaces between lines. Messi is an attacker but not attached to a specific position, and is allowed to cover wide areas of the pitch to find space and request the ball. The three players share, however, a long-time tradition of possession-centered and off-ball movements quality during their career. Suarez and Neymar, two highly mobile players, appear with a lower count of situations where space was gained. This can be associated with the high level of strictly closed marking these players suffered during the match.

It is interesting to observe that for most players the active occupation of spaces is considerably more frequent than passive occupation. This is particularly noticeable on left and right backs Digne and Sergi Roberto, who need to cover wider spaces and show a high mean distance to ball for SOG, a characteristic shared by central defenders Pique and Mascherano. A remarkable case is that of Lionel Messi, whose passive SOG is considerably higher than the active one. The passive characteristic of SOG does not mean the player is not occupying the space intentionally, but rather that he is not moving at running speed, but slower. Much has been argued in recent years about several moments during matches where Messi walks through zones of the field. However, that walking behaviour is not a detachment from the match but a conscious action to move through empty spaces of value and claim the control of valuable space, and ultimately the ball. Messi does this very effectively, placing him near the top of players in terms of space gained during the whole match, despite the lack of active gain. A relevant characteristic of this is that 71% of the time the gain in space is done in front of the ball rather than behind. The in front and behind the ball

Name	# SOG	$\sum$ SOG	μ SOG	Active (%)	Passive (%)	FRT	BEH	MBD	Mins
Iniesta	96 (14.8%)	15.77	0.16	56.25	43.75	49	47	15.19	94.86
S. Busquets	90 (13.9%)	14.85	0.16	47.78	52.22	44	46	16.65	94.86
Messi	81 (12.5 %)	14.72	0.18	33.33	66.67	58	23	17.50	94.86
A. Gomes	74 (11.4%)	12.58	0.17	68.92	31.08	40	34	15.93	68.61
Suarez	70 (10.8%)	12.27	0.18	57.14	42.86	57	13	13.46	94.86
Neymar	61 (9.4%)	9.46	0.16	59.02	40.98	48	13	18.31	94.86
S. Roberto	51 (7.8%)	7.34	0.14	78.43	21.57	25	26	25.10	94.86
Pique	29 (4.4%)	4.92	0.17	48.28	51.72	6	23	21.05	94.86
Mascherano	29 (4.4 %)	4.54	0.16	41.38	58.62	2	27	22.03	94.86
D. Suarez	22 (3.4%)	4.07	0.18	77.27	22.73	13	9	17.47	26.25
A. Turan	17 (2.6%)	3.51	0.21	52.94	47.06	12	5	12.71	23.32
Digne	26 (3.2%)	3.48	0.13	80.77	19.23	13	13	16.23	71.54

Table 1: Statistics of space occupation for F.C. Barcelona in an official Spanish League match against Villareal F.C. Symbols #, $\sum$ and μ represent the total, sum and mean of their associated variable. SOG refers to Space Occupation Gain, while FRT and BEH indicate the amount of times SOG occurs in front or behind the ball. MBD represents the mean ball distance, and Active (%) and Passive (%) the player percentage of times the space was occupied through active or passive occupation.

statistics show a clear tendency for central defenders to gain space behind the ball, while attackers show a higher rate of space gain in front of the ball. Noticeably Busquests, Iniesta and the right and left backs (Digne and S. Roberto) have a balanced ratio of space gain behind and in front of the ball.

Table 2 presents the statistics for Space Occupation Loss (SOL) and Space Generation Gain (SGG). The SOL statistics show a clear tendency of higher space loss for players that are more often in possession of the ball such as Iniesta, Messi, Neymar and Suarez. The space loss can be directly associated with pressure by the opponent, who tends to increase density near to attacking players to reduce their range of action, especially for highly skilled players. Regarding the generation of space, we obtain a different picture from the space occupation skills. Here, Neymar and Suarez appear to be, alongside Messi, the players that most often drag opponents to create space. With a 4-3-3 system and high-quality players, a specific attacking strategy is that of spreading out attacking players to drag defenders out of position and provide wider spaces for attacking action. Busquets, a pivoting specialist, appears also at the top of the table showing his value in supporting space creation. Notably the left and right back, Digne and S. Roberto do not generate much space. Given that they move towards the border lines of the field, it is less likely that opponents are dragged by back defenders.

A more detailed perspective of space generators and receivers is presented in Figure 7. Here we can observe the amount of times generators are producing space for receivers, and discover some collaborative playing behaviour. First to observe is that Busquets receives space from most of the players at least once, possibly showing his ability to stay at the center of play. A renowned skill of F.C. Barcelona is the third-man pass, which consists of the following: if a player A wants to pass to player C, but is marked, he passes to player B, dragging the opponents toward him, enabling C to receive the ball in more space. This plot might show a third-man behaviour through Busquets. Notably, Suarez, Neymar and Messi generate space commonly for each other, especially Suarez who provides considerable space to both. A special connection

Name	# Generated	# Received	$\sum$ SGG	μ SGG	# SOL	$\sum$ SOL	μ SOL	Mins
Neymar	28 (18.9%)	6 (4.1%)	5.97	0.21	51	-8.53	-0.17	94.86
Suarez	25 (16.9%)	18 (12.3%)	5.60	0.22	52	-9.12	-0.18	94.86
Messi	22 (14.9%)	24 (16.4%)	4.32	0.20	68	-11.61	-0.17	94.86
S. Busquets	15 (10.1%)	24 (16.4%)	3.83	0.26	38	-6.16	-0.16	94.86
Pique	14 (9.5%)	9 (6.2%)	3.66	0.26	19	-2.77	-0.15	94.86
Iniesta	13 (8.9%)	21 (14.4%)	2.62	0.20	75	-11.79	-0.16	94.86
A. Turan	8 (5.4%)	7 (4.8%)	2.26	0.28	8	-1.29	-0.16	23.32
S. Roberto	7 (4.7%)	2 (1.4%)	1.55	0.22	31	-4.62	-0.15	94.86
A. Gomes	9 (6%)	18 (1.2%)	1.49	0.17	44	-6.25	-0.14	68.61
Mascherano	5 (3.4%)	9 (6.2%)	0.80	0.16	23	-3.39	-0.15	94.86
D. Suarez	2 (1.4%)	8 (5.5%)	0.46	0.23	16	-3.14	-0.20	26.25

Table 2: Statistics of space generation, and space occupation loss for F.C. Barcelona in an official Spanish League Match against Villareal F.C. Symbols #, $\sum$ and μ represent the total, sum and mean of their associated variable. # Generated and # Received indicate the total times a player generated or received generated space, accompanied by the team-relative percentage. SGG refers to Space Generation Gain and SOL refers to Space Occupation Loss.

between Suarez and Messi is also shown for this game, where both were able to generate a high amount of space for each other.

A further vision of space gain and generation can be grasped from Figure 9. Here we present the spatial heatmap for SOG and SGG situations. At first glance we observe the amount of space gained through occupation is considerably higher than through generation, a more complex process. Iniesta presents an interesting case where he can generate more space next to the left border line of the field, while he is better at gaining spaces for himself at the interior of the field. Also, he produces a notable amount of space near the box. Busquets shows an incredible collaborative behaviour by generating space almost anywhere around the field. He also presents wide areas of SOG, but more intensively near the midfield, his natural habitat. Suarez presents a notable ability to generate space within the box, where he concentrates most of his generating contribution. Here he arises as a specialist in dragging defenders either while making spaces for himself or while generating spaces for others. Messi also shows a great ability in generating spaces around the attacking zones of the field, while Neymar concentrates on the left wing, focused on high speed diagonal runs towards the box. Defenders, as expected, show very little generation of space.

7 Discussion

In a sport where the average possession of the ball by a player is 3 minutes in a 90 minute game, the analysis of team-collective dynamics through off-ball movements becomes a critical element for understanding performance. We have shown how through spatio-temporal data it is possible to extract meaningful information relating the occupation of spaces of value and the generation of spaces for teammates. Beyond the bigger picture that overall performance statistics of multiple matches can provide, the understanding of off-ball movements demands the need for a more specialized per-match or even per-situation analysis. Through the understanding of the frequencies, quality, position and effectiveness of space occupation and generation, a coach can provide specific guidance to players to help the team playing dynamics beyond what he can do with the ball.
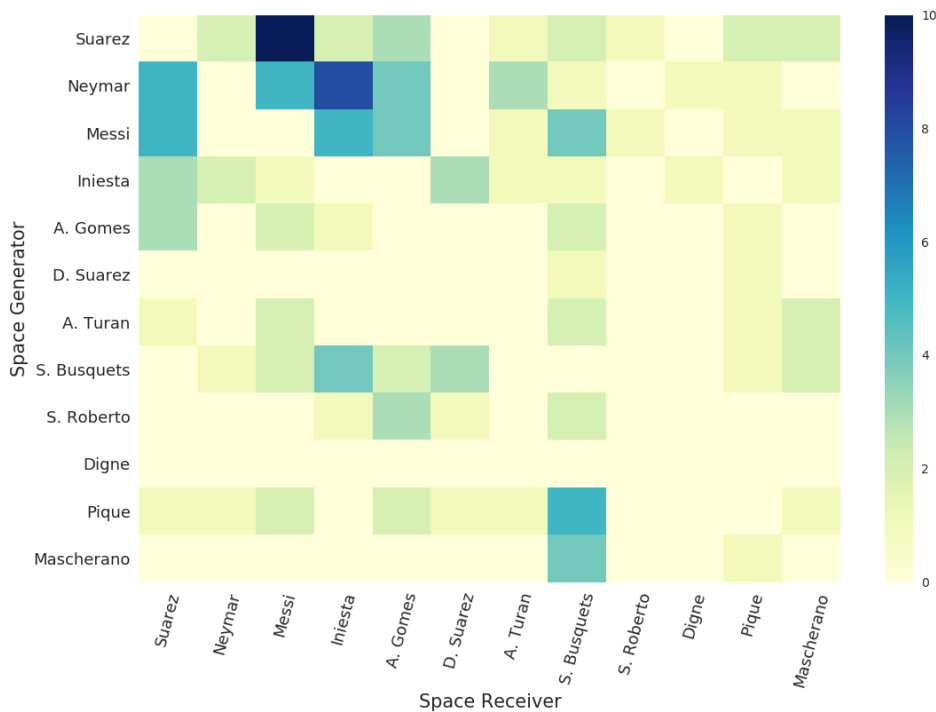
Figure 7: A heatmap showing the total times space was generated by generators (y-axis) for receivers (x-axis)

In order to provide a deeper understanding of space, we have presented two novel approaches for pitch control and pitch value modelling. Our pitch control model takes into account critical factors when understanding the dominance of space such as the velocity and position of the player. It also provides a key element that was missing in previous dominant region models: the idea of a soft surface of control where for a given location on the field, nearby players have a certain level of influence, instead of defining strict dominance margins such as in Voronoi-based models. On the other hand, the proposed pitch value model presents a way of quantifying the value of every location on the field in a dynamic way, relative to the location of the ball. This way, we can account for both the control of space a team has and the value of that space, to obtain a measure of spatial value controlled.

For future studies, the proposed pitch control and field models can be directly applied for reaching more comprehensive pass probability and reward models, and in general to incorporate a new perspective on dominant regions based approaches for understanding team sports. But more generally, this study sets a base for new research on off-ball behaviour in soccer. New perspectives are still to be studied, such as the effect of different pressure strategies, the concept of potential space and how it could be exploited, the overall dynamic balance of space control between the two teams and its association to performance, as well as many other research lines that address a critical question when training to succeed in soccer: what should I do when my teammate is in possession of the ball.
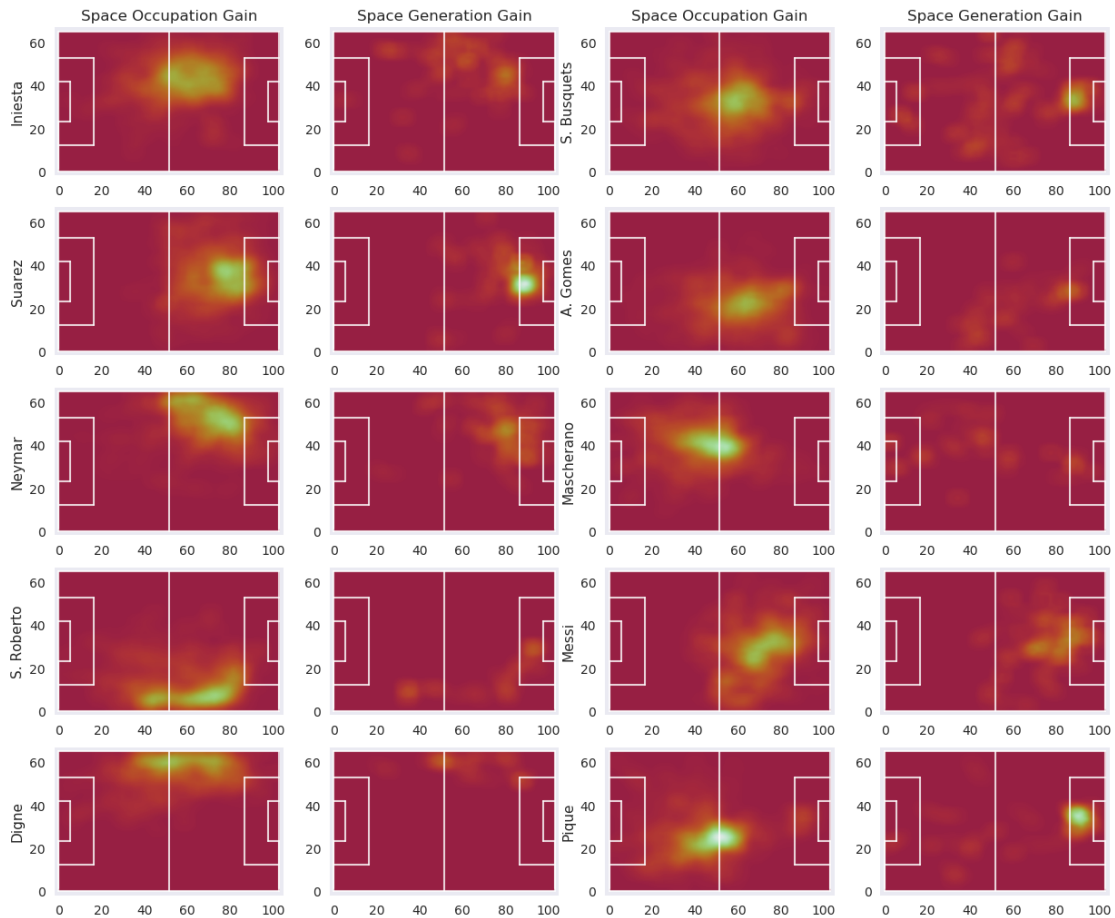
15

Figure 8: Space Occupation Gain and Space Generation heatmap for every field player playing over 60 minutes. The scaling factor is based on the maximum Space Occupation and maximum Space Generation among all the team, respectively.

References

[1] Pedro Silva, Paulo Aguiar, Ricardo Duarte, Keith Davids, Duarte Araújo, and Júlio Garganta. Effects of pitch size and skill level on tactical behaviours of association football players during small-sided and conditioned games. *International Journal of Sports Science & Coaching*, 9(5):993–1006, 2014.

[2] Robert Rein, Dominik Raabe, Jürgen Perl, and Daniel Memmert. Evaluation of changes in space control due to passing behavior in elite soccer using voronoi-cells. In *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, pages 179–183. Springer, 2016.

[3] Michael Horton, Joachim Gudmundsson, Sanjay Chawla, and Joël Estephan. Classification of passes in football matches using spatiotemporal data. *arXiv preprint arXiv:1407.5093*, 2014.

[4] R Masheswaran, Y Chang, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler, and Noel Hollingsworth. The three dimensions of rebounding. *MIT SSAC*, 2014.

[5] 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.

[6] Dan Cervone, Luke Bornn, and Kirk Goldsberry. NBA court realty. MIT Sloan Sports Analytics Conference, Boston, MA, USA, 2016.

[7] Alexander Franks, Andrew Miller, Luke Bornn, and Kirk Goldsberry. Counterpoints: Advanced defensive metrics for NBA basketball. In *9th Annual MIT Sloan Sports Analytics Conference, Boston, MA*, 2015.

[8] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] Angel Ric, Carlota Torrents, Bruno Gonçalves, Lorena Torres-Ronda, Jaime Sampaio, and Robert Hristovski. Dynamics of tactical behaviour in association football when manipulating players' space of interaction. *PloS one*, 12(7):e0180773, 2017.

[10] A geometric interpretation of the covariance matrix. http://www.visiondummy.com/2014/04/geometric-interpretation-covariance-matrix. Accessed: 2017-11-10.

A   Appendix

A.1   Player Influence Area Details

Section 3.1 presents a player influence model that accounts for the position, velocity and distance to ball of a given player. The influence degree I_i at Equation 13 is expressed in terms of the probability density function of a bivariate Gaussian distribution defined by Equation 12. In this section we detail the calculation of each of the elements involved in the equation.

$$f_i(p,t) = \frac{1}{\sqrt{(2\pi)^2 detCOV_i(t)}} exp(-\frac{1}{2}(p - \mu_i(\vec{s}_i(t)))^T COV_i(t)^{-1}(p - \mu_i(t))) \tag{12}$$

$$I_i(p,t) = \frac{f_i(p,t)}{f_i(p_i(t),t)} \tag{13}$$

The covariance matrix can be dynamically adjusted to provide a player dominance distribution that accounts for location and velocity. Using the singular value decomposition algorithm we can express the covariance matrix as a function of its eigenvectors and eigenvalues as expressed in Equation 14, where V is the matrix whose columns are the eigenvectors of Σ, and L is the diagonal matrix whose non-zero elements are the corresponding eigenvalues [10]. Let $R = V$ and $S = \sqrt{L}$, we can define R as a rotation matrix and S as a scaling matrix, allowing to express the covariance as in Equation 15. Based on this, the rotation matrix and scaling matrix can be defined as Equations 16 and 17, where θ is the rotation angle of the speed vector and, s_x and s_y are the scaling factors in the x and y direction.

$$\Sigma = VLV^{-1} \tag{14}$$

$$\Sigma = RSSR^{-1} \tag{15}$$

$$R = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \tag{16}$$

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \tag{17}$$

In order to find the scaling factors, we take into account both the player's magnitude of speed $S_i(t)$ (as meters per second), and the distance to the ball $D_i(t)$. Based on the opinion of expert soccer analysts we have defined the range $[4, 10]$ as the minimum and maximum distance in meters of player's pitch control surface radius $R_i(t)$, based on the distance to the ball, following the transformation function shown at Figure 9. Setting $13m/s$ as the maximum possible speed reachable, we calculate the ratio between players and the maximum speed, as shown in Equation 18. Then, the scaling matrix is expanded in x direction and contracted in y direction by this factor, as expressed in Equation 19. Given this, we can express a function COV for obtaining the covariance matrix as shown in Equation 21. Finally, the distribution mean value $\mu_i(t)$ is found by translating the players location at time t by half the magnitude of speed vector $\vec{s}$, following Equation 21.

$$Srat_i(s) = \frac{s^2}{13^2} \tag{18}$$

$$S_i(t) = \begin{bmatrix} \frac{R_i(t)+(R_i(t)Srat_i(\vec{s}_i(t)))}{2} & 0 \\ 0 & \frac{R_i(t)-(R_i(t)Srat_i(\vec{s}_i(t)))}{2} \end{bmatrix} \tag{19}$$

$$COV_i(t) = R(\theta,t)S_i(t)S_i(t)R(\theta_i(t),t)^{-1} \tag{20}$$

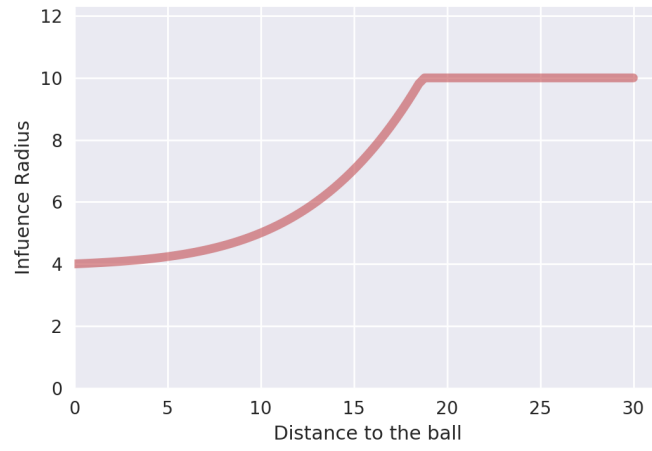$$\mu_i(t) = p_i(t) + \vec{\hat{s}}_i(t) \cdot 0.5 \tag{21}$$

Figure 9: Player influence radius relation with distance to the ball