The Geometry of Pressing: Spatial Analysis of Playing Out of the Back

By: Shane Hauck

Introduction

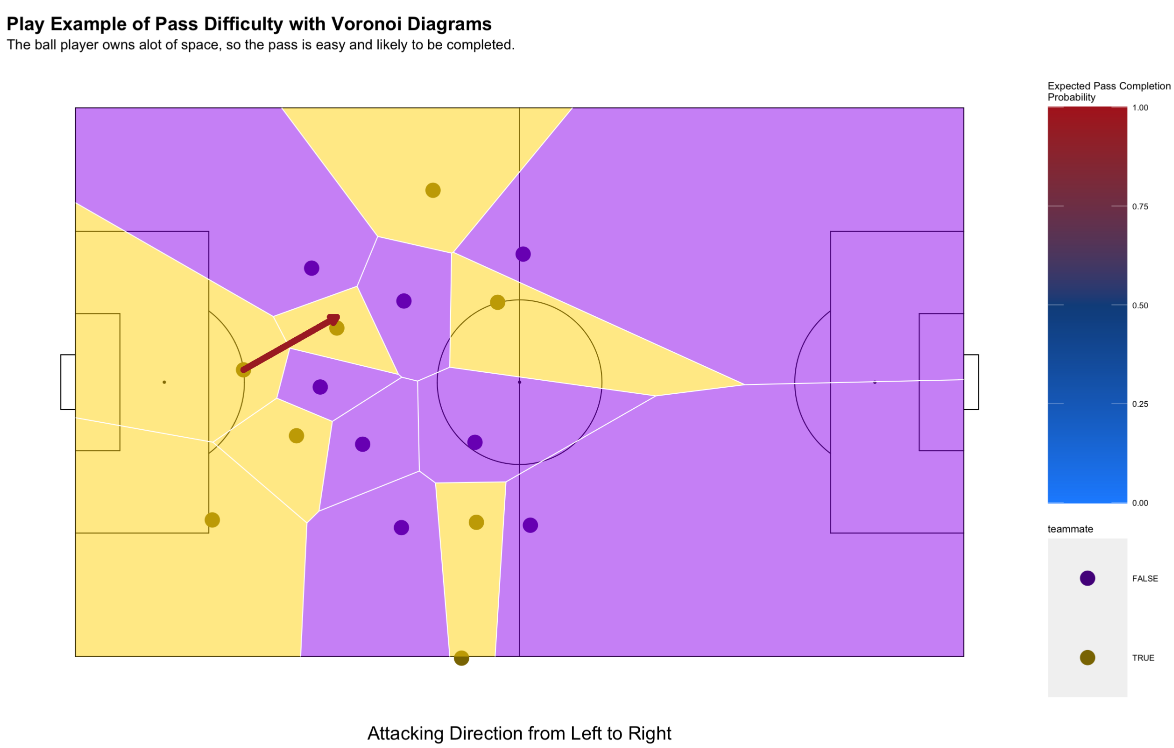
High pressing when the opposition are building out from the back is a tactical approach often used in modern soccer. This strategy involves applying intense pressure on the opposition high up the pitch, especially when they are in the initial stages of building an attack from their defensive area. The primary objective is to disrupt the opponent's build-up play, force errors, and regain possession quickly in advanced areas of the pitch, creating opportunities to attack the opposition's goal.

The more space and time a player has, the easier it is to make a considered and accurate pass. Space allows a player to better observe the field, assess options, and execute a pass without hurried decision-making. Time provides the opportunity to adjust body position and choose the best technique for the pass. When an opponent applies high-intensity pressure, the available space and time for the player in possession significantly decrease. This scenario increases the difficulty of making a successful pass. Under pressure, players have less time to make decisions and are more likely to be forced into errors or less optimal decisions.

The goal of this paper is to use StatsBomb 360 data of the 2022 FIFA World Cup to use a method of space ownership, the voronoi diagram, to quantify the amount of space a player has when playing out from the back and the effect that has on the pass difficulty.

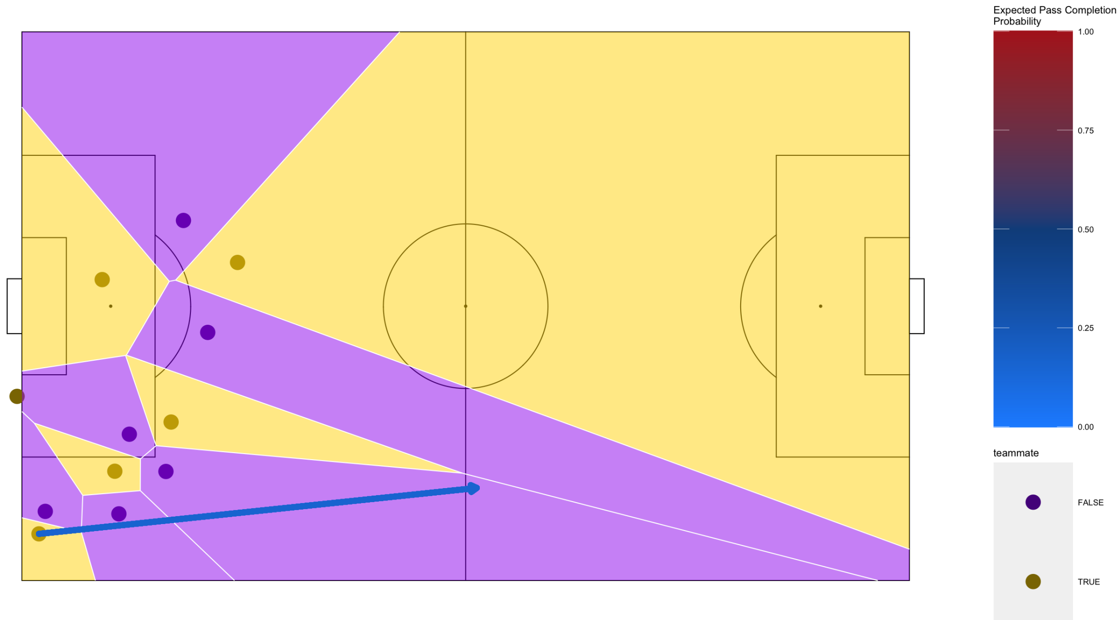
Methodology

The Voronoi Diagram, also known as Dirichlet Tessellation, is a method of spatial control that divides the pitch into regions based on the location of the players. If we consider each player on a pitch as a point, the Voronoi Diagram divides the pitch into regions where every point in a given region is closer to its corresponding player than to any other player. It quantifies the amount of space each player "controls" or can potentially influence, providing insights into the effectiveness of pressing situations in soccer. https://en.wikipedia.org/wiki/Voronoi_diagram



Play Example of Pass Difficulty with Voronoi Diagrams

The ball player owns a small amount of space, so the pass is difficult and unlikely to be completed.

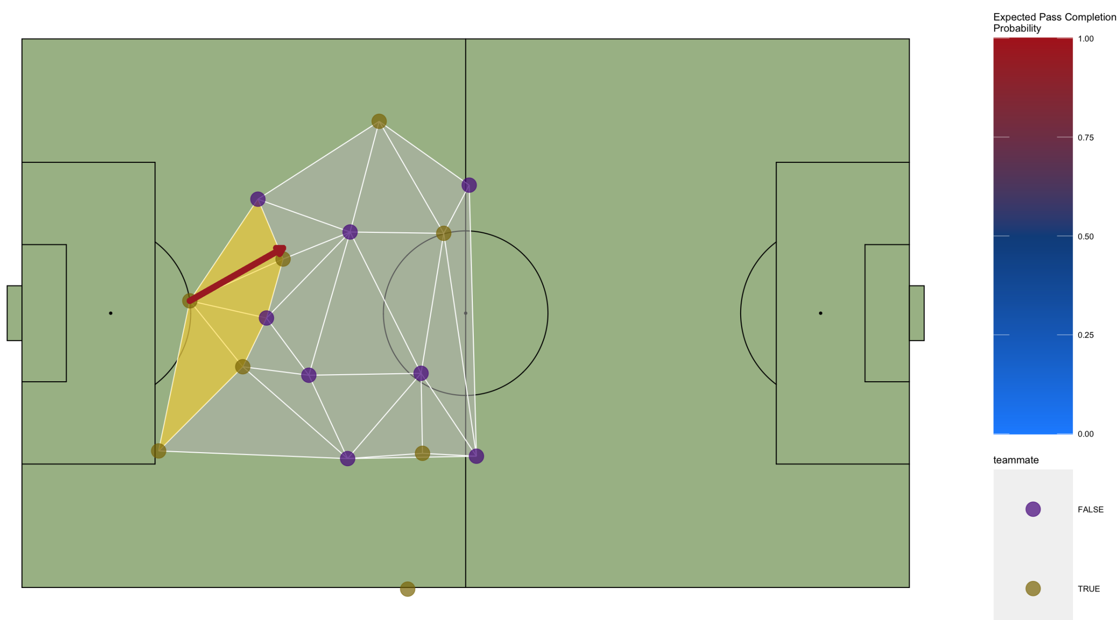


Attacking Direction from Left to Right

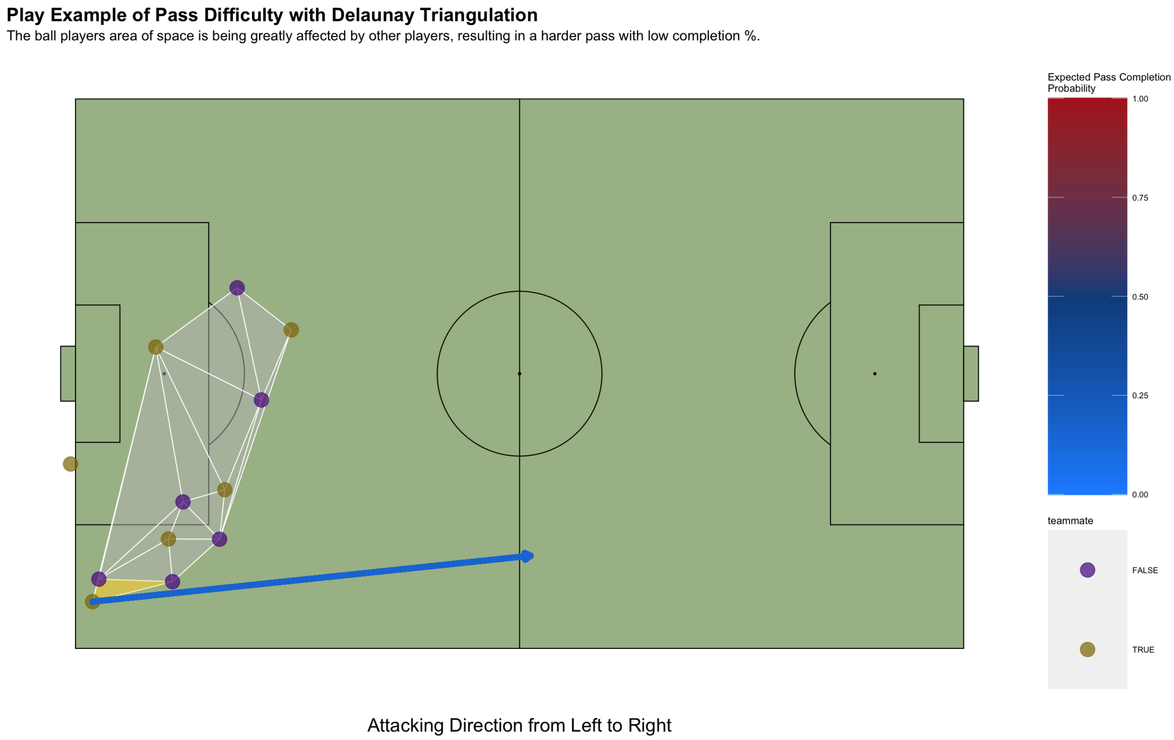
Additionally, Delaunay Triangulation is a method for connecting a set of points with triangles in such a way that no point lies inside the circumcircle, or the circle that passes through all three vertices, of any triangle. This technique aims to produce triangles that are as equiangular as possible, minimizing the occurrence of very acute angles resulting in a more uniform distribution of triangles, avoiding narrow, elongated triangles. Considering each player on a pitch as a point, this approach creates a network of triangles connecting players, ensuring that each triangle's vertices (players) are positioned so that the circle passing through them doesn't encompass any other player. By emphasizing the most efficient connections between them, it offers a geometric perspective on player positioning highlighting the direct spatial relationships and potential interactions between players. https://en.wikipedia.org/wiki/Delaunay_triangulation

Play Example of Pass Difficulty with Delaunay Triangulation

The ball players area of space is not being greatly affected by other players, resulting in an easier pass with high completion %.



Attacking Direction from Left to Right



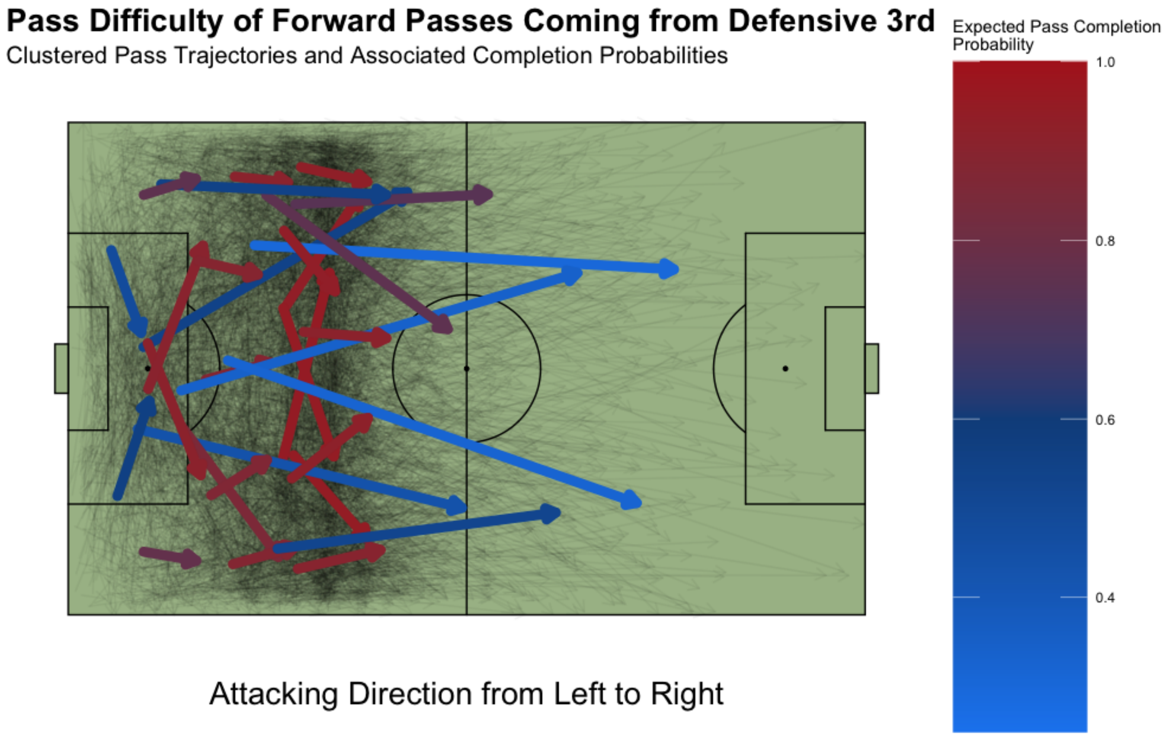
The combination of the Voronoi Diagram and Delaunay Triangulation techniques offers a multifaceted approach to quantifying player positions and spatial control on a soccer pitch. When evaluating the space around a player making a pass, various metrics derived from these techniques can be applied to quantify spatial relationships and control:

- **n.tri:** Represents the number of Delaunay triangles emanating from a player's point. This measure indicates the direct spatial connections a player has with teammates and opponents.
- **del.area:** Calculates one-third of the total area of all the Delaunay triangles emanating from the player's point, providing an estimate of the space influenced by the player through these connections.
- **del.wts:** This is the weight assigned to each player based on the del.area, normalized by the sum of areas for all players. It reflects the proportion of influence a player has in the network of triangles.
- **n.tside:** Counts the number of sides, within the rectangular playing field, of the Voronoi (Dirichlet) tile surrounding the player.
- **nbpt:** The number of points at which the Voronoi tile intersects the boundary of the rectangular field, indicating how a player's area of influence reaches the edge of the play area.
- **dir.area:** The area of the Voronoi tile surrounding the player, quantifying the actual space controlled by the player.
- **dir.wts:** Similar to del.wts, this variable provides a normalized weight based on the Voronoi area (dir.area), offering insight into the player's relative spatial dominance on the pitch.

These metrics measure the space around a player at the moment of passing, considering the positions of all other players in the frame. Utilizing these variables enables the development of a Pass Difficulty model that predicts the likelihood of a pass being successful. This model allows for a statistical evaluation of how significantly the metrics related to space ownership contribute to the odds of pass completion.

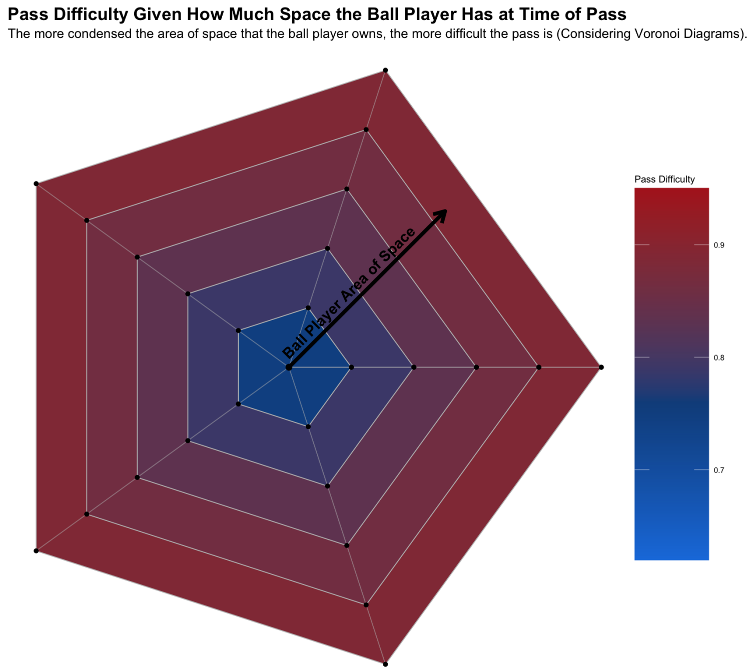
The Pass Difficulty model uses logistic regression to estimate the probability of a pass being completed while considering various spatial and temporal factors that can affect pass success. The model outputs a score between 0 and 1, where 1 indicates a guaranteed pass completion and 0 means the pass is not completed. In the regression, each of these factors is represented by coefficients that are statistically tested for their influence on the pass's outcome. The square roots of some of the metrics are used, to normalize the distribution of the data or to reduce skewness, which help attempt to satisfy the assumptions of logistic regression. The model was fit with the goal of minimizing the AIC to balance the complexity of the model against the fit of the model to the data.

Since the focus of this analysis is on high pressing and the difficulty of playing out from the back, only forward passes originating from the defensive third are considered. This approach ensures that the analysis is centered on the specific scenario of building out from the back under different levels of pressure, providing insights into the spatial control and pass difficulty in these situations. The Pass Difficulty model is trained on a subset of the data, and the model's performance is evaluated using a separate test set to ensure its generalization to new data.

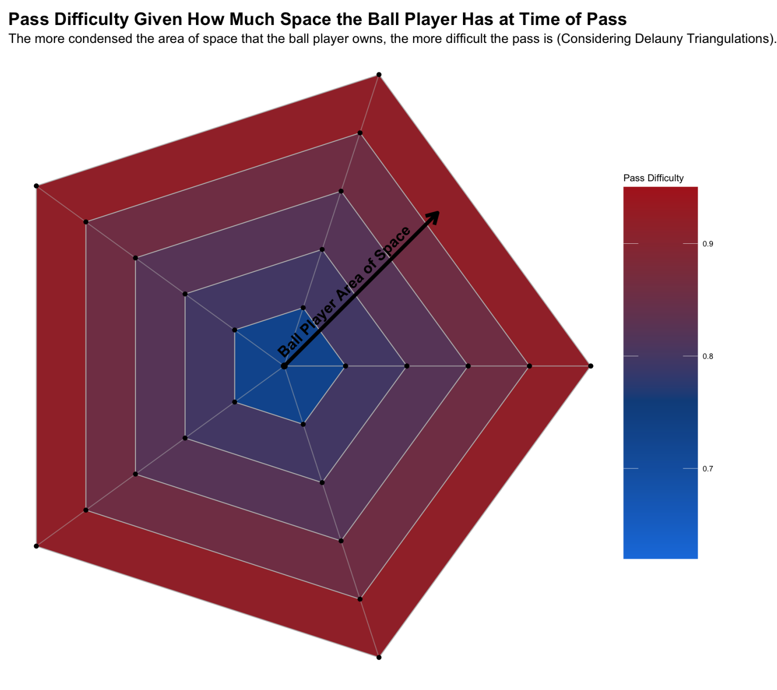


Results

The coefficients of the Pass Difficulty model provide insights into the significant influence of spatial metrics on the likelihood of a pass being completed. It was found that the metrics relating to the Delaunay Triangulation (n.tri, del.area, del.wts) were more significant than the metrics relating to the Voronoi Diagram (n.tside, nbpt, dir.area, dir.wts). However, the coefficients for the spatial metrics aligned with our hypothesis that the more space, the easier it is to make a considered and accurate pass. Interestingly there was a significant interaction between the Delaunay Triangulation Area (del.are) and the duration of time the passer had on the ball. This suggests that space and time are not independent of each other and that the more space a player controls, the more time they have to make a intelligent decision with regards to keeping possession of the ball.



Dir.Area Range	Mean Pass Difficulty
1.57 - 7.46	0.7483498
7.46 - 9.81	0.7904459
9.82 - 12.23	0.8335563
12.23 - 15.19	0.8627585
15.2 - 31.3	0.8871784



Del.Area Range	Mean Pass Difficulty
0.16 - 3.87	0.7285035
3.87 - 5.35	0.7974017
5.35 - 7.03	0.8218324
7.04 - 9.14	0.8580964
9.14 - 18.03	0.9164776

When looking at the expected pass probabilities in regards to space ownership of the ball player, it is clear that the more space a player controls, the higher the probability of a successful pass. Both Voronoi Diagrams and Delaunay Triangulations offer unique insights into the spatial control exerted by players. The Pass Difficulty model underscores the importance of these metrics in predicting the probability of successful pass completion. Notably, the analysis revealed that Delaunay Triangulations serve as a more effective predictor in this context.

The chi-squared tests from the ANOVA (Analysis of Variance) compare the full model against a null model, which lacks the spatial metrics. The significant chi-squared statistic ($p < 2.2e\text{-}16$) suggests that the full model with spatial metrics provides a significantly better fit than the null model. Additionally, the AIC of the full model is 3934.7, while the null model has an AIC of 4150.3. The difference in AIC between the two models is 215.6, indicating that the full model is a better fit while balancing the complexity of the model against the fit of the model to the data. Additionally, the calibration plot shows that the model's predicted probabilities align well with the actual pass completion rates, indicating that the model is well-calibrated and accurately predicts the likelihood of a pass being completed. (models and plots can be found in the appendix).

Discussion

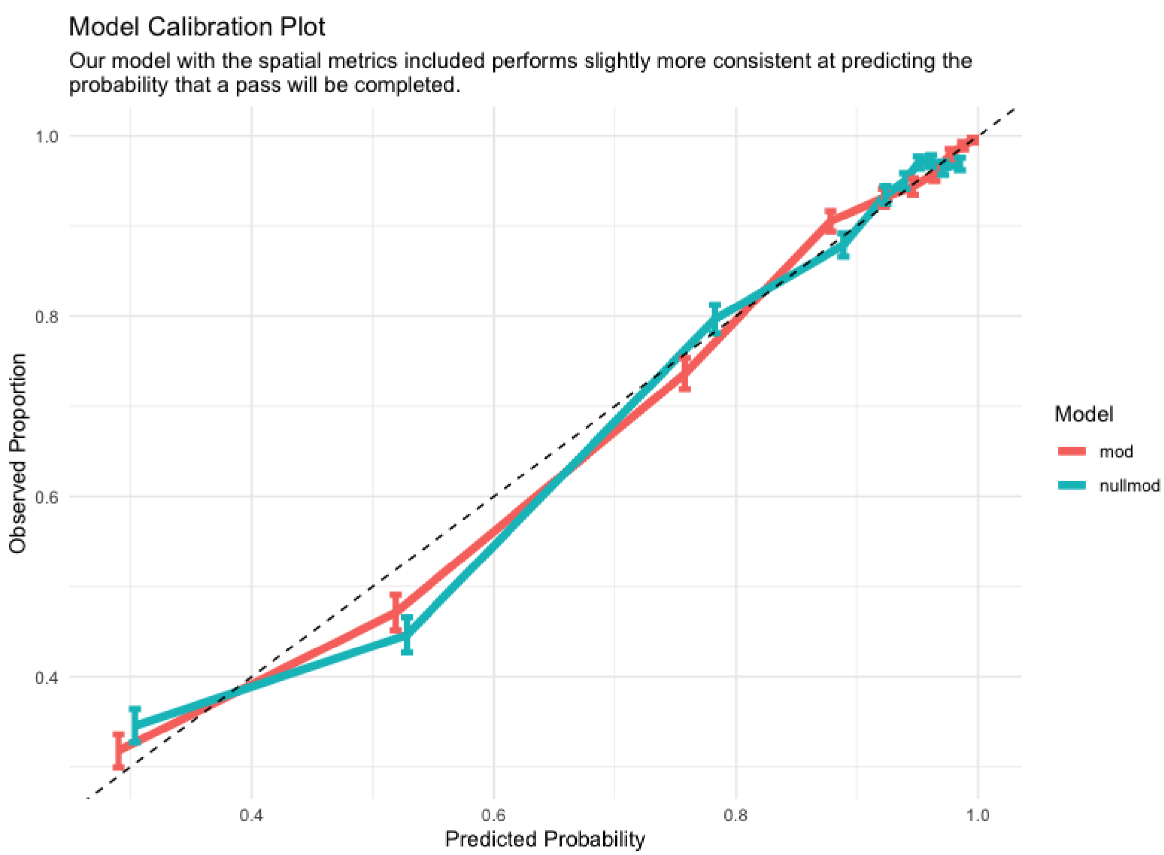
When considering the tactical applications of spatial analysis in soccer, particularly through the lens of high pressing, Voronoi diagrams and Delaunay triangulations offer an intricate view of how teams and individual players manage space. High pressing, as a tactic, is predicated on limiting the space and time that the opposition has to make decisions, thus forcing errors and creating opportunities for quick transitions. The spatial metrics derived from these geometrical constructs provide a quantitative measure of the effectiveness of such tactics.

The spatial control exerted by each player, as depicted by Voronoi diagrams, can be directly linked to their ability to influence the game both offensively and defensively. For instance, a player's capacity to find space or close down an opponent can be visualized and quantified, offering a more nuanced understanding of their contribution beyond traditional statistics. A team's collective shape, the compactness of their formation, and their ability to create overloads in certain areas of the pitch can also be discerned from these diagrams. Delaunay triangulations complement this by highlighting potential passing lanes and defensive coverages, revealing the most efficient connections between players.

Furthermore, the Pass Difficulty model, informed by these spatial analyses, provides an objective assessment of passing decisions a player makes considering their relative positioning. It brings to light the interplay between space and time — the more space a player controls, the more time they have to execute a decision.

Incorporating such spatial analysis into traditional soccer analytics enriches the dialogue around tactical effectiveness and player influence, moving beyond what can be seen with the naked eye. It allows analysts, coaches, and fans alike to appreciate the subtle nuances of spatial management in soccer, painting a clearer picture of the beautiful game's complex tactical landscape.

Appendix



```
Call:
glm(formula = pass.outcome.id ~ sqrt(del.area) + sqrt(del.wts) +
    sqrt(n.tri) + sqrt(n.tside) + sqrt(dir.area):sqrt(nbpt) +
    sqrt(dir.wts):sqrt(nbpt) + sqrt(TimeInPoss) + sqrt(duration) +
    sqrt(duration):sqrt(del.area) + location.x + location.x:pass.end_location.x +
    as.factor(pass.height.id) + as.factor(pass.body_part.id),
    family = binomial(link = "probit"), data = modeling_df)

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  -1.763e+00  3.111e-01  -5.665 1.47e-08 ***
sqrt(del.area)                3.854e-01  4.134e-02   9.323  < 2e-16 ***
sqrt(del.wts)                -1.766e+00  7.028e-01  -2.512 0.011995 *  
sqrt(n.tri)                  -6.605e-01  1.212e-01  -5.451 5.01e-08 ***
sqrt(n.tside)                 3.054e-01  1.275e-01   2.395 0.016622 *  
sqrt(TimeInPoss)             -1.636e-02  1.016e-02  -1.611 0.107243    
sqrt(duration)                1.659e+00  1.475e-01  11.247  < 2e-16 ***
location.x                    8.665e-02  5.514e-03  15.716  < 2e-16 ***
as.factor(pass.height.id)2   -1.004e+00  7.776e-02 -12.909  < 2e-16 ***
as.factor(pass.height.id)3   -1.341e+00  7.112e-02 -18.856  < 2e-16 ***
as.factor(pass.body_part.id)38  2.361e-01  1.601e-01   1.474 0.140475    
as.factor(pass.body_part.id)40  2.315e-01  1.580e-01   1.465 0.142969    
as.factor(pass.body_part.id)68 -9.337e-02  3.011e-01  -0.310 0.756504    
as.factor(pass.body_part.id)69  1.628e+00  4.558e-01   3.572 0.000355 ***
as.factor(pass.body_part.id)70  2.630e-01  3.942e-01   0.667 0.504625    
as.factor(pass.body_part.id)106  3.840e+00  5.527e+01   0.069 0.944614    
sqrt(dir.area):sqrt(nbpt)    -1.043e-02  1.117e-02  -0.934 0.350370    
sqrt(nbpt):sqrt(dir.wts)      1.006e+00  5.299e-01   1.898 0.057753 .  
sqrt(del.area):sqrt(duration) -1.814e-01  2.454e-02  -7.394 1.43e-13 ***
location.x:pass.end_location.x -1.082e-03  8.147e-05 -13.275  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6022.6  on 6448  degrees of freedom
Residual deviance: 3894.7  on 6429  degrees of freedom
AIC: 3934.7

Number of Fisher Scoring iterations: 12
```

```
Call:
glm(formula = pass.outcome.id ~ sqrt(TimeInPoss) + sqrt(duration) +
    location.x + location.x:pass.end_location.x + as.factor(pass.height.id) +
    as.factor(pass.body_part.id), family = binomial(link = "probit"),
    data = modeling_df)

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  -9.531e-01  2.053e-01  -4.643 3.43e-06 ***
sqrt(TimeInPoss)             -1.627e-02  9.451e-03  -1.722 0.085120 .  
sqrt(duration)                1.007e+00  8.586e-02  11.734  < 2e-16 ***
location.x                    8.952e-02  5.047e-03  17.738  < 2e-16 ***
as.factor(pass.height.id)2   -1.104e+00  7.583e-02 -14.560  < 2e-16 ***
as.factor(pass.height.id)3   -1.491e+00  6.913e-02 -21.562  < 2e-16 ***
as.factor(pass.body_part.id)38  2.620e-01  1.586e-01   1.652 0.098532 .  
as.factor(pass.body_part.id)40  2.802e-01  1.567e-01   1.788 0.073705 .  
as.factor(pass.body_part.id)68 -7.689e-02  2.969e-01  -0.259 0.795650    
as.factor(pass.body_part.id)69  1.555e+00  4.211e-01   3.693 0.000222 ***
as.factor(pass.body_part.id)70  1.830e-01  3.938e-01   0.465 0.642073    
as.factor(pass.body_part.id)106  3.612e+00  5.754e+01   0.063 0.949945    
location.x:pass.end_location.x -1.094e-03  7.678e-05 -14.248  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6022.6  on 6448  degrees of freedom
Residual deviance: 4124.3  on 6436  degrees of freedom
AIC: 4150.3

Number of Fisher Scoring iterations: 12
```

```
Analysis of Deviance Table

Model 1: pass.outcome.id ~ sqrt(del.area) + sqrt(del.wts) + sqrt(n.tri) +
    sqrt(n.tside) + sqrt(dir.area):sqrt(nbpt) + sqrt(dir.wts):sqrt(nbpt) +
    sqrt(TimeInPoss) + sqrt(duration) + sqrt(duration):sqrt(del.area) +
    location.x + location.x:pass.end_location.x + as.factor(pass.height.id) +
    as.factor(pass.body_part.id)
Model 2: pass.outcome.id ~ sqrt(TimeInPoss) + sqrt(duration) + location.x +
    location.x:pass.end_location.x + as.factor(pass.height.id) +
    as.factor(pass.body_part.id)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      6429     3894.7
2      6436     4124.3 -7  -229.63 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Analysis of Deviance Table

Model: binomial, link: probit

Response: pass.outcome.id

Terms added sequentially (first to last)


                               Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                            6448     6022.6
sqrt(del.area)                  1   193.08      6447     5829.5 < 2.2e-16 ***
sqrt(del.wts)                   1    48.10      6446     5781.4 4.046e-12 ***
sqrt(n.tri)                     1     8.24      6445     5773.2  0.004091 **
sqrt(n.tside)                   1   112.10      6444     5661.1 < 2.2e-16 ***
sqrt(TimeInPoss)                1     8.25      6443     5652.8  0.004067 **
sqrt(duration)                  1   289.18      6442     5363.6 < 2.2e-16 ***
location.x                      1   242.14      6441     5121.5 < 2.2e-16 ***
as.factor(pass.height.id)       2   886.38      6439     4235.1 < 2.2e-16 ***
as.factor(pass.body_part.id)    6    28.28      6433     4206.8 8.305e-05 ***
sqrt(dir.area):sqrt(nbpt)       1     1.95      6432     4204.9  0.162961
sqrt(nbpt):sqrt(dir.wts)        1     1.80      6431     4203.1  0.179191
sqrt(del.area):sqrt(duration)   1   116.32      6430     4086.7 < 2.2e-16 ***
location.x:pass.end_location.x  1   192.04      6429     3894.7 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```