Assessing Defensive Players in Football

Student Name : Harleen Gulati , Supervisors: Matthew Penn, Professor Christl Donnelly

August 25, 2023

1 Introduction

Oxford City is a football club which have invested in data of 2 year's worth of matches. Some analysis of this data has been done to deduce how good players are at attacking , passing etc. Whilst analysis has been done on defending, more could be done. Also, the current analysis done on defending is biased towards a team the player belongs to (for example, if a player belongs to a mostly attacking team, they will do little defending and so the current analysis will deem the player as a 'bad defender' though this may not necessarily be the case).

2 Intended Outcome

Our goal is to form defensive metrics for players which do not depend on which team the player belongs to and give further information on how good of a defender a player is. We desire to minimise the following objective function (where AIJ is the ability of player I in season J - the smaller this is, the less variation between defence abilities across different seasons and so the less dependent the ability is on a given season and so a given team).

$$\frac{\sum_i \sum_j \sum_k (A_i^j - A_i^k)^2}{\sum_i \sum_m \sum_j \sum_k (A_i^j - A_m^k)^2} = \frac{\text{(Variation from players across seasons)}}{\text{(Total variation)}}$$

3 Methodology

3.1 Method 1

The first defence metric for player I in season J is : total defences made by player I in season J / total time player I spent defending in season J.

Here defences are an umbrella term for interceptions, successful defensive duels and successful aerial duels (explained in the key-word section).

The intuitive sense behind this approach is we are now considering the total number of defences as a ratio of the time spent defending. For example, for a defence score of X if you spend longer defending, your score will go down and so you need to make more defences to reach X however if you spent little time spent defending, your score will be higher and so you need to make fewer defences to reach X.

The idea here is if you spent less time defending, each defence you make speaks more of your ability as a defender however if you spend longer defending you need to do more defences to prove yourself more because you had the chance to do so.

3.2 Method 2

We consider where a defence is made. If a defence is made closer to the goal the player is protecting, this defence can be thought of as more impactful than if it were made further away, because the defence made closer to the goal could have saved a potential shot and thus goal.

We split the pitch into 5 20x100 blocks (usually the pitch is 120x80 however the data we use considers the pitch to be 100x100 – we adjust coordinates in our final results accordingly) and for each player stored how many defences they made in each block. We gave each block a 'weight' where the weight reflects how impactful a defence in that block is in determining the defending ability of the player. For example, the

block closest to the goal the player is protecting had a weight of 0.05 whereas the block furthest away had a weight of 0.0005.

The final defence ability for each player in season X was then: sum of ((total defences made in each block * weight of each block) / time spent defending in season X)

We divide by the time the player spent defending to avoid team biases (same reasons as methodology step 1).

3.3 Method 3

We had data which for a pass, would state if the pass was intercepted or not and what the chance of intercepting the pass was.

If the pass was intercepted, we were also given who the interceptor was.

We considered if a pass was difficult to be intercepted, but was intercepted nonetheless then this spoke more of the ability of the player as a defender than if the pass was easy to be intercepted and indeed was intercepted (we'd expect this to be the case).

So, we split interceptions into two categories: easy interceptions and hard interceptions. If the chance of intercepting was high for a successful interception, we added this to the easy interceptions category otherwise we added this to the hard to intercept category.

For each player I in season J this defence metric then does : sum of ((interceptions made in each category by player I in season J * weight of each category) / total time player I spent defending in season J)

3.4 Method 4

Defensive duels have four possible outcomes. We either stop the progress of the ball (ball goes out of pitch say) or we recover possession of the ball (ball now belongs to the defending player) or we do both or neither.

Each outcome speaks of the defending ability of the player in different amounts (e.g. doing both suggests a defender is better than one who does neither).

Each outcome is then assigned a weight, which reflects how good of a defender each outcome implies and defence measure 4 for a player I in season J is sum of ((number of defensive duels in each category by player I in season J * weight of each category) / time player I spent defending in season J)

3.5 Method 5

We found the objective function for each defence measure which is shown in the table below. Since the first defence measure has a much higher objective function than the rest, we discarded it.

We then assigned a weight to each other defence measure which stated how much it impacted the final defence measure of a player.

Then the final defence measure for a player I in season J is : sum of (each defence measure of player I in season J * weight of this defence measure).

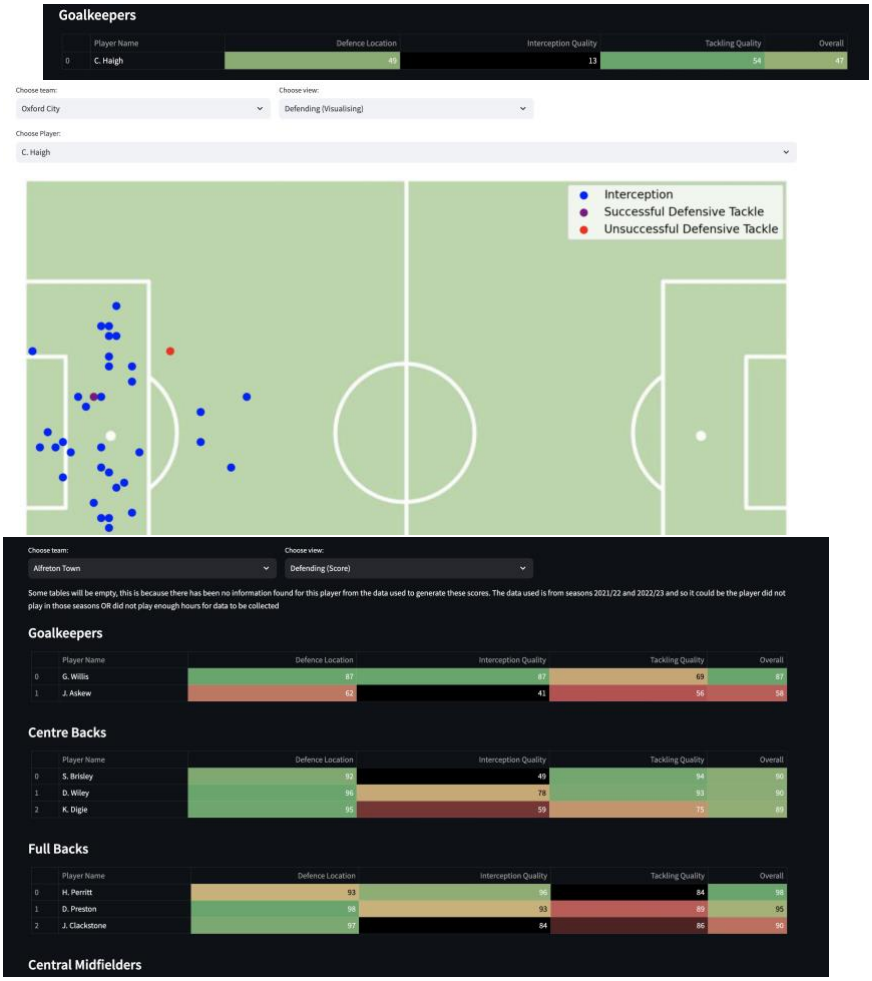
Note we do not need to divide by defending time of player I in season J since each individual defence measure does this.

Below is the objective function for each defence measure:

Defence Measure	Objective Function Value
Total defences / time spent defending	0.000183
Defence location / time spent defending	0.000136
Interception quality / time spent defending	0.000148
Defending duel quality / time spent defending	0.000154
Combining all the above / time spent defending	0.000140

4 Results

For each player, we found the percentiles of their scores with respect to other players in the same category. For example, all the goalkeeeepers scores were collected and percentiles found, same for centre backs etc. This was to follow the format done of previous data analysis and this makes the results easier to analyse. Below is the results for the goalkeeper Chris Haigh and the results for more players. We see from the visualisation Chris Haigh makes many interceptions close to his goal, which is reflected in the table as his defence location score is high. We observe his interception quality seems weak, however the visualisation does not reflect this as it does not show easy or hard interceptions. The tackling quality is average which is reflected in the visualisation as two tackles were made one successful and one unsuccessful.



5 Further Research/Next Steps

5.1 Tweaking Weights

We tweaked the weights for the defensive duel quality metric and for the interception quality metric using SciPy's minimisation function. However this minimisation function had issues with the defence location metric and combination metric. A next step could be writing your own minimisation functions to tweak the weights for these metrics so they are optimal in minimising the objective function.

5.2 Improving Current Approaches

Our defence location approach splits the pitch into 5 segments. A next step could be looking at splitting the pitch into further smaller segments. This approach is slightly biased since a goalkeeper will mostly be in the block closest to his goal and so be scored higher than say an attacker. A next step could be to tackle this bias, by perhaps considering what position a player plays (e.g. defender) and adjusting the score to reflect biases the position may introduce.

5.3 Further Metrics

As we have for successful and unsuccessful defensive duels, perhaps we can add a metric for unsuccessful interceptions. For example, for a pass made which was not intercepted we could see which player of the defending team was closest to this pass. We can then score them down for not taking the opportunity to intercept (here considering to score them down greater if the chance of interception was high). We can add further metrics like these to provide more information on players defending abilities.

6 Conclusion

Our metrics seem feasible and reflect the players defence ability fairly. Dividing by the time spent defending reduces biases on the teams ability as a whole in defending or attacking.

However, adding more metrics and tweaking the current metrics could aid in further player defence metrics.

7 Acknowledgments

Thank you to Matthew Penn for the data, support and constant guidance throughout this internship. Thank you to Professor Christl Donnelly for her support and kindness throughout.

8 Keywords

Interception – a player intercepts a pass made between two attacking players.

Defensive duel – trying to get the ball off an attacking player usually in a one-to-one tackle (which either result in stopping the progress of the ball e.g., ball goes off pitch or recovering the possession of the ball (player defending gets ball of player attacking) or both or neither.

Aerial duels – players try to get ball that is in the air, whoever touches the ball first wins the duel