Athens Journal of Sports 2023, volume 10: 1-21
https://doi.org/10.30958/aspo.X-Y-Z

A Mathematical Analysis of Team Impact and Individual Player Contribution in Football

By Jeffrey Leela[*], Donna M. G. Comissiong[±] & Karim Rahaman[°]

In this paper, we present an important application of the Hungarian Method - a well-known combinatorial optimization tool for solving assignment problems. For our purposes, we consider the assignment of players to specific roles in a football team. It involves the broad classification of team players as defensive, midfield or attacking, while assigning the main roles associated with each of these positions. This provides insight on specific role of each individual player, thereby facilitating an optimal team selection. To illustrate this method, we utilize the average player statistics per game for two teams from the 2016/2017 Premier League Season. In addition, a team rating index is created by identifying six sub-indices. The first is called team contributions - which includes set piece goals, percentage tackles won, percentage take-ons won, percentage aerial duels won, number of interceptions, number of blocked shots, number of clearances, number of red and yellow cards. To visualize the method, a multiple correlation is carried out on team data for the 2016/2017 Premier League season to generate a correlation coefficient for each contribution. The resulting team index can be a useful tool for measuring the overall strengths of competing teams in a football league.

Keywords: *Hungarian method, football, team rating index, multiple correlations, team comparisons*

Introduction

It is a well-known fact that the most successful teams are the ones that are best balanced, not necessarily those comprised of the best collection of available players. Nevertheless, each player's individual contribution is vital for the overall team performance, and coaches are continually seeking the most effective techniques for identifying the most outstanding players. Modern-day football scouts can make use of data-driven analysis techniques to assess any player's potential based on the available performance metrics. After recruiting the players with the best ratings, it is then up to the coaching staff to conduct appropriate training sessions to get the players to work together, harnessing each individual player's strengths to optimize overall team performance.

[*]Senior Lecturer, Department of Mathematics, The College of Science, Technology and Applied Arts of Trinidad and Tobago, Trinidad, West Indies.
[±]Senior Lecturer, Department of Mathematics and Statistics, The University of The West Indies, St. Augustine Campus, Trinidad, West Indies.
[°]Senior Lecturer, Department of Mathematics and Statistics, The University of The West Indies, St. Augustine Campus, Trinidad, West Indies.

One of the most important tasks of a football coach is team selection, according to the match being played, and after careful consideration of the opposing team's strengths and weaknesses. While is not necessarily true that a collection of players with the best individual performance ratings would be the optimal choice for the selected team, once these players have trained together and have a well-defined game plan, it is not unreasonable to expect a favourable match outcome. The difficulty lies with the process of team selection when the available players have similar attributes. In such cases, it would be beneficial for coaches to have a scientific method to distinguish between closely matched players with similar abilities. Given the large sums of money on offer for winning professional football leagues and national team titles, the ability to select the most suitable team players has become an indispensable skill (Qadar et al. 2017).

The Hungarian method can be employed for the effective solution of assignment problems, where a set of tasks must be assigned to workers who each have a different level of ability. The problem is solved by creating the cost matrix associated with each worker-task pair, and consequently finding the optimal assignment of workers to tasks through a series of iterative steps. The objective is to minimize the total cost or to maximize the total benefit associated with the completion of the assigned jobs. As the Hungarian method guarantees an assignment solution that is both feasible and optimal, it can conceivably be employed to determine the optimal team selection for any team sport. The method was successfully applied by Britz and Maltitz for the optimal selection of baseball players for the most effective team (Britz and Maltitz 2010). After assessing a group of novice baseball players to determine their abilities in key practical aspects of the game, they successfully employed the Hungarian method to determine the optimal team, according to these metrics. This same approach could conceivably be adapted for team selection in football, utilizing the available player performance metrics freely available on online football data repositories.

In this paper, we utilize these player performance metrics – available for free download from Whoscored.com – to create an efficiency matrix for key players in a football team. We do this by classifying players according to their specific role on the team, and extracting the relevant statistical data associated with the jobs typically allocated to defenders, midfielders and strikers. Subsequently, we apply the Hungarian algorithm to the efficiency matrix to determine the maximal defensive, midfield and striking scores for the team. This facilitates an unbiased comparison of competing teams in a professional football league using summarised player statistics obtained from a recently completed season. Given that large sums of money are spent each year on recruiting new players, it would be very helpful to have another scientific tool to analyse immediate past team results to effectively identify problem areas where player recruitment might be opportune.

Next, we present a general method for the overall rating of teams in a professional football league, using the Premier League to illustrate this objective. To formulate our team rating system, we establish a set of criteria which characterizes all-round team play. Inspired by the player ranking methodology presented by McHale et al., we introduce an appropriate number of sub-indices, the first being called "team contributions" which is sub-divided into set piece goals, duels won percentage, defensive actions and discipline (McHale et al. 2012). The required data for each registered team for the 2016/2017 Premier League season was sourced from Squawka.com (Squawka 2017). A multiple correlation analysis is performed with points achieved by each team as the reference variable, with the other variables being set piece goals, tackles, take-on, aerial duels won, interceptions, blocked shots, clearances, red and yellow cards.

As football fans around the world will attest, the final result of a match does not often represent the actual performance of a football team. Our proposed team index is a single score that effectively rates the collective contributions of all team players. While there are several predictive tools that are available for use in team football, our analysis will provide an avenue for evaluating the overall team performance after the season has ended. A quick comparison with the overall team standings at the end of the playing season can easily demonstrate the effectiveness of the method, lending credibility to its usefulness for coaching staff when planning for future seasons.

Literature Review

The mathematical foundation for the Hungarian algorithm was established by the Hungarian mathematicians Konig (1913) and Egevary (1931). Harold Kuhn later devised a computational algorithm that efficiently employs the Hungarian method for the solution of an assignment problem (Kuhn 1955). The algorithm was studied independently by James Munkres in 1957, and for that reason, it is sometimes referred to as the Kuhn–Munkres algorithm or the Munkres assignment algorithm (Munkres 1957). The method reduces the associated cost matrix in such a manner that at least one zero in each row and column will be obtained. The positions of these zeros in the matrix are representative of the optimal assignment solution, thus facilitating the calculation of the minimal opportunity cost.

Britz and Maltitz utilized the Hungarian algorithm for team selection in baseball, by assigning the most effective player to respective positions on the field (Britz and Maltitz 2010). They considered different weighted combinations of player roles on a baseball field, while considering the overall balance that must be achieved between offensive and defensive plays. They then tested their proposal on a group of novice baseball players by conducting skill tests to determine the relevant ratings for the associated efficiency matrix. They then employed the Hungarian algorithm to identify the optimal team. To the best of our knowledge, this methodology has not yet been adopted for team selection purposes in football.

As explained by McHale et al., "performance assessment is a fundamental tool for quantitative analysts and operational researchers" (McHale et al. 2012). Rating systems are often utilized to measure team or player performance, and there are well-established rating systems for ranking opposing teams in competitive sports competitions. In individual sports such as tennis, it is relatively straightforward to analyse recent results of player competitions to generate an ordered list of the top ranked players. As these official rankings are often used to seed players in a tournament, this can also affect the overall outcome of the tournament, as top seeded players are effectively guaranteed an easier route to the final rounds of matches. It is true however that there are limitations to any ranking system, and absolute trust cannot be placed on rating systems that rely only on past player performances. Even the official rankings provided by the well-established Association of Tennis Professionals (ATP) might prove somewhat deceptive for sports enthusiasts placing bets on the top ranked tennis players (McHale and Morton 2011).

Tennis is not the only sport to have used officials' rankings to predict future performance. Forrest and McHale found that for men's professional golf, increased forecasting power can be achieved by incorporating up-to-date results with an established forecasting model which utilizes world rankings as a predictor (Forrest and McHale 2007). In a similar study for football, McHale and Davies determined that recent match results of international teams can add much value to the forecasting model (McHale and Davies 2007). Thus, the evidence from tennis, golf and football suggests that although official rankings of players and teams are useful as predictors, they do not determine match outcomes with absolute certainty. Reliable team ratings are required for the calculation of betting odds, and substantial funds are generated when sports fans place bets on their preferred teams. The availability of methods for the evaluation of team performance is therefore of great interest not only to players and coaching staff, but also to the wider community of sports enthusiasts.

Methodology

Individual Player Contribution

In assignment problems, the main objective is the allocation of jobs to an equal number of persons at a minimum cost for maximal profit. Let us suppose that there are 'n' jobs to be performed and 'n' persons available to take these jobs. We assume that each person can complete an assigned job in a specified time with a varying level of efficiency. Let c_{ij} be the cost associated with the i^{th} person being assigned to the j^{th} job. Our goal is to determine the optimal job assignment such that the total cost for performing all the jobs is minimized. Typical examples of assignment problems include the allocation of machines to jobs, classes to classrooms, players to a team, etc.

Basic Mathematical Formulation

Cost matrix: $c_{ij} = \begin{matrix} c_{11} & c_{12} & \ldots\ldots & c_{1n} \\ c_{21} & c_{21} & \ldots\ldots & c_{2n} \\ \ldots & \ldots & \ldots\ldots & \\ c_{n1} & c_{n2} & \ldots\ldots & c_{nn} \end{matrix}$

We wish to minimize cost: $z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij}$ $\quad i = 1, 2,\ldots, n \;;\; j = 1, 2,\ldots, n.$

subject to the conditions

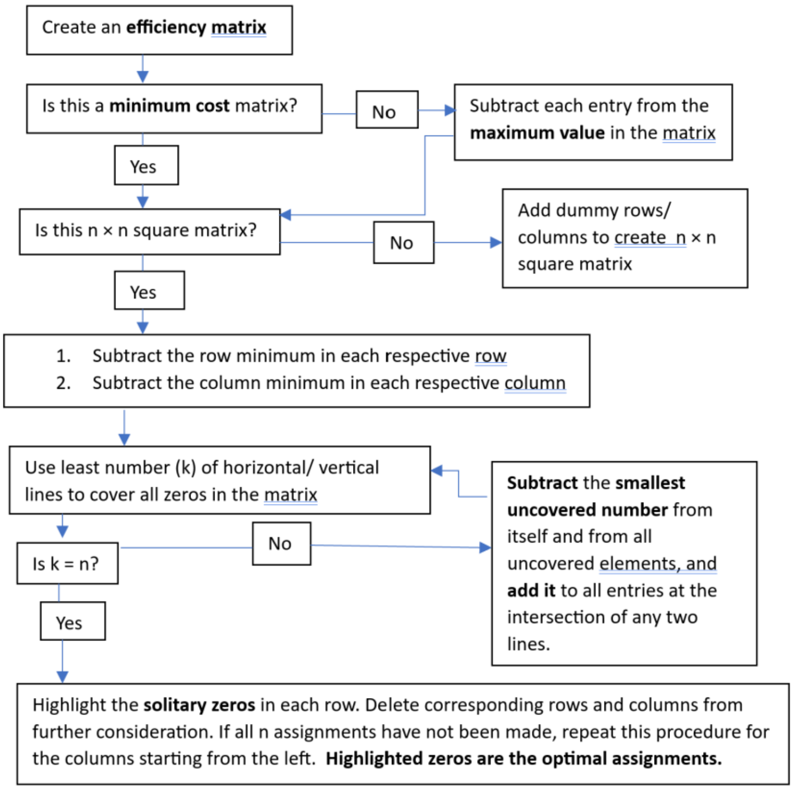
$$x_{ij} = \begin{cases} 1 \text{ if } i^{th} \text{ person is assigned to } j^{th} \text{ job} \\ 0 \text{ otherwise} \end{cases}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \text{ (one job is done by the } i^{th} \text{ person } i = 1, 2, \ldots, n)$$

$$\sum_{i=1}^{n} x_{ij} = 1 \text{ (only one person should be assigned the } j^{th} \text{ job } i = 1, 2, \ldots, n)$$

where x_{ij} denotes the j^{th} job to be assigned to the i^{th} person.

The Hungarian Algorithm (Britz and Maltitz 2010)



The position that a player occupies on the field defines the role and responsibility of that particular player. There are three main positions for outfield players in a football team: defender, midfielder or striker. Players may be asked to perform multiple tasks/jobs in accordance with the team formation/tactical directives provided by the coaching staff. These jobs include passing, tackling, blocking, intercepting, clearing, shooting, assisting, and dribbling. Most football players are better at mastering one or two of these jobs, although there are a few exceptional players who exhibit extraordinary levels of talent and can therefore perform multiple functions with equally high levels of competence. In general, regardless of the position that they occupy, players must be able to perform all these jobs effectively - since football is a team sport, and successful teams are comprised of players who can adapt quickly to changing situations on the pitch.

To illustrate the method, we utilize the average player statistics per game for two teams from the 2016/2017 Premier League Season (Whoscored 2017). We select the team that placed first that year: Chelsea, and the team that placed sixth: Manchester United. Our main objective is to investigate the roles performed by the players from each team, and in so doing, to provide reasons for the gulf in class between these two teams. This type of critical analysis can help the coaching staff to identify what is working well for their team, and what needs to be improved.

We begin by classifying the players on each team according to their main roles – as defender, midfielder or striker. Defenders are given five major jobs while midfielders are given seven and strikers four. As midfielders must perform both defensive and offensive duties, there is some overlap in the tasks to be performed by defenders and midfielders as well as by midfielders and strikers. We use the available data to assign the players to jobs, noting that whenever there are more players than jobs, the resulting matrix is not square. As the Hungarian Algorithm requires a square matrix, in such cases, "dummy jobs" must be created to facilitate the analysis. Although not all the players will be given a legitimate job as a result, the analysis will still allow us to identify the most efficient combination of players on the team to perform all the tasks outlined. Our objective is to maximize each team's defensive and offensive statistics, based on the available data. This will allow important comparisons to be made between the two teams. Our results will provide reasonable justifications for the gap in points scored between the teams and for the overall performance of each team as a unit.

We will illustrate the method by considering the defensive statistics for Manchester United. The nine defenders used for the majority of the 2016/2017 Premier League season by Manchester United are listed in Table 1 with their associated averages for the five defensive jobs considered crucial for their position. Note that each number indicates the average for that particular job per game, and that passing data is based on quoted pass percentages. For example, Smalling has an 89% successful passing rate.

Table 1. *Defenders for Manchester United*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	0.7	2.0	2.4	1.4	1.5	2.4	1.1	2.4	2.1
Clearing	6.9	4.7	2.1	6.8	2.1	5.0	2.5	4.1	7.6
Blocking	0.5	0.4	0.3	0.3	0.2	0.9	0.1	0.1	0.8
Intercepting	0.7	1.9	1.5	1.6	1.3	2.5	1.1	2.3	1.6
Passing	0.89	0.86	0.86	0.86	0.83	0.86	0.86	0.81	0.89

Our problem is to maximize the defensive statistics for the team - by identifying the combination of five selected players that results in the maximum overall defensive score for the team with respect to the five tasks identified: talking, clearing, blocking, intercepting, and passing. Now, to turn this into a maximization type problem for the Hungarian algorithm we must first develop the effective matrix. To do this, we must first subtract the largest entry (7.6) from each other entry of the matrix. The resulting matrix is shown in Table 2.

Table 2. *Subtract the Smallest Entry from Each Row from all other Entries in that Row*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	6.9	5.6	5.2	6.2	6.1	5.2	6.5	5.2	5.5
Clearing	0.7	2.9	5.5	0.8	5.5	2.6	5.1	3.5	0
Blocking	7.1	7.2	7.3	7.3	7.4	6.7	7.5	7.5	6.8
Intercepting	6.9	5.7	6.1	6.0	6.3	5.1	6.5	5.3	6.0
Passing	6.71	6.74	6.74	6.74	6.77	6.74	6.74	6.79	6.71

Next, we add dummy jobs to make the number of rows to equal the number of columns. The dummy jobs are denoted as F, G, H and I, as illustrated in Table 3.

Table 3. *Effective Matrix – After Addition of Dummy Jobs F, G, H and I*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	6.9	5.6	5.2	6.2	6.1	5.2	6.5	5.2	5.5
Clearing	0.7	2.9	5.5	0.8	5.5	2.6	5.1	3.5	0
Blocking	7.1	7.2	7.3	7.3	7.4	6.7	7.5	7.5	6.8
Intercepting	6.9	5.7	6.1	6.0	6.3	5.1	6.5	5.3	6.0
Passing	6.71	6.74	6.74	6.74	6.77	6.74	6.74	6.79	6.71
F	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0

We can now proceed with the steps listed in the Hungarian algorithm. Subtract the minimum element in each row from each element in that row. As there are zeros in every column, there is no need to subtract the minimum element from each column from all elements in that column. The result is shown in Table 4.

Table 4. *Modified Matrix – After Subtraction of Minimum Element from all Rows*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	1.7	0.4	0	1	0.9	0	1.3	0	0.3
Clearing	0.7	2.9	5.5	0.8	5.5	2.6	5.1	3.5	0
Blocking	0.4	0.5	0.6	0.6	0.7	0	0.8	0.8	0.1
Intercepting	1.8	0.6	1	0.9	1.2	0	1.4	0.2	0.9
Passing	0	0.03	0.03	0.03	0.06	0.03	0.03	0.08	0
F	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0

As there are zeros in every column, there is no need to subtract the minimum element from each column. We must now cover all the zeros with the minimum number of horizontal and vertical lines. This yields eight lines, as shown in Table 5.

Table 5. *Cover all Zeros with Minimum Number (8) of Horizontal/Vertical Lines*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	1.7	0.4	0	1.0	0.9	0	1.3	0	0.3
Clearing	0.7	2.9	5.5	0.8	5.5	2.6	5.1	3.5	0
Blocking	0.4	0.5	0.6	0.6	0.7	0	0.8	0.8	0.1
Intercepting	1.8	0.6	1	0.9	1.2	0	1.4	0.2	0.9
Passing	0	0.03	0.03	0.03	0.06	0.03	0.03	0.08	0
F	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0

As the order of the matrix is nine, the optimal assignment cannot be made. We proceed by subtracting the minimum uncovered element from all uncovered elements and add this minimum uncovered element to the covered elements at the line intersections only. From Table 5, we see that the minimum uncovered element is 0.03. To cover all the zeros with the minimum number of horizontal and vertical lines in the resulting matrix, we will again require eight lines (as shown in Table 6), so once again, the optimal assignment cannot be made.

Table 6. *Zeros Covered with the Minimum Number (8) of Horizontal/Vertical Lines*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	1.73	0.4	0	1.0	0.9	0.03	1.3	0	0.33
Clearing	0.7	2.87	5.47	0.77	5.47	2.6	5.07	3.47	0
Blocking	0.4	0.47	0.57	0.57	0.67	0	0.77	0.77	0.1
Intercepting	1.8	0.57	0.97	0.87	1.17	0	1.37	0.17	0.9
Passing	0	0	0	0	0.03	0.03	0	0.05	0
F	0.03	0	0	0	0	0.03	0	0	0.03
G	0.03	0	0	0	0	0.03	0	0	0.03
H	0.03	0	0	0	0	0.03	0	0	0.03
I	0.03	0	0	0	0	0.03	0	0	0.03

We must repeat the steps of the Hungarian algorithm. The smallest uncovered number is 0.17, so we subtract 0.17 from all uncovered numbers, and we add 0.17 to the covered numbers that are located in any position where two lines intersect. Nine lines can be used to cover all zeros in the resulting matrix, as shown in Table 7. The optimal assignment can now be determined.

Table 7. *Zeros Covered with the Minimum Number (9) of Horizontal/ Vertical Lines*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling	1.73	0.4	0	1.0	0.9	0.2	1.3	0	0.5
Clearing	0.53	2.7	5.3	0.6	5.3	2.6	4.9	3.3	0
Blocking	0.23	0.3	0.4	0.4	0.5	0	0.6	0.6	0.1
Intercepting	1.63	0.4	0.8	0.7	1	0	1.2	0	0.9
Passing	0	0	0	0	0.03	0.2	0	0.05	0.17
F	0.03	0	0	0	0	0.2	0	0	0.2
G	0.03	0	0	0	0	0.2	0	0	0.2
H	0.03	0	0	0	0	0.2	0	0	0.2
I	0.03	0	0	0	0	0.2	0	0	0.2

Zeros are then eliminated to leave one zero in each row and column, thus ensuring that each player is assigned to one job. The resulting matrix is presented in Table 8, where the symbol ⊗ indicates an eliminated zero while (0) indicates the assigned player in the respective column with the corresponding job in the respective row.

Table 8. *Matrix Displaying the Optimal Assignment Solution*

	Smalling	Blind	Valencia	Rojo	Young	Bailly	Shaw	Darmian	Jones
Tackling			(0)					⊗	
Clearing									(0)
Blocking						(0)			
Intercepting						⊗		(0)	
Passing	(0)	⊗	⊗	⊗			⊗		
F		(0)	⊗	⊗	⊗		⊗	⊗	
G		⊗	⊗	(0)	⊗		⊗	⊗	
H		⊗	⊗	⊗	(0)		⊗	⊗	
I		⊗	⊗	⊗	⊗		(0)	⊗	

For verification purposes, we utilized the MATLAB Hungarian Algorithm for linear assignment problems (V2.3) developed by Yi Cao (Yi 2023). Deployment of the program "munkres.m" with the effective matrix (see Table 3) yielded the same result displayed in Table 8, which effectively confirms our manual calculations. The optimal team defensive score is subsequently calculated by adding the original performance scores by the players selected for the optimal solution (refer to Table 1). This type of analysis facilitates a comparison of defensive systems employed by different teams in the Premier League, with the highest overall defensive team score expected to correspond with the team with the most effective defensive records.

A similar analysis can be carried out for midfield players and strikers, for their respective jobs. For brevity, our calculations for the two football teams under consideration will be summarized in the results section.

Analysis of Team Impact

In formulating a system for rating, we must first establish a number of criteria which constitutes the all-round play of each registered team in the league under consideration. The necessary data for each team for the 2016/2017 Premier League Season was collected from Squawka (Squawka 2017). A multiple correlation analysis was then carried out with 'number of points obtained' as the reference variable with the other variables being 'number of set piece goals', 'tackles percentage won', 'take-ons percentage won', 'aerial duels percentage won', 'number of interceptions', 'number of blocked shots', 'number of clearances', 'number of red cards' and 'number of yellow cards'. The results are presented in Table 9.

Table 9. *Multiple Correlation Analysis Results*

Multiple Correlation Analysis	Correlation Coefficients
# Set Piece Goals & # Points	0.4844
Tackles % Won & # Points	0.4239
Take-ons % Won & # Points	0.1644
Aerial Duels % Won & # Points	0.0333
# Interceptions & # Points	-0.4651
# Blocked Shots & # Points	-0.7879
# Clearances & # Points	-0.6562
# Red Cards & # Points	-0.2820
# Yellow Cards & # Points	-0.2411

As positive correlation coefficients are indicative of a relationship between two variables that tends to move in the same direction, our results indicate that the more set piece goals scored by a team, the more points will be obtained. This is also the case for tackles percentage won. We note that take-ons percentage won and points are positively correlated but only distantly so, while aerial duels percentage won and points have almost no correlation. This indicates that there may be other underlying factors that we have not considered for those two variables.

As expected, defensive actions and points are all negatively correlated, i.e., more of these actions is indicative of fewer points attained by the team. We note that teams with a larger number of interceptions, blocked shots and clearances are constantly under attack. As a result, those teams will be defending frantically to stay in the game, with much less focus on offensive play. Discipline also impacts on the team, so it is no surprise that red/ yellow cards and points are negatively correlated.

We are now ready to establish the first sub-index of the rating system by multiplying the number of actions of each team by the correlation coefficient obtained for that action and then summing these products. As the number of set piece goals, tackles won, take-ons won and aerial duels won will be significantly less than the number of defensive actions (namely blocked shots, clearances and interceptions), we expect that the associated index will be negative. In general, team data for the number of set piece goals and the percentages of tackles, take-ons and aerial duels won is in the tens, while the data corresponding to defensive actions is markedly higher, often measuring in the hundreds or thousands. To compensate for this imbalance, we multiply set piece goals by one hundred - since the number of goals scored decides the match outcome and the number of points awarded to the team. The percentages associated with tackles, take-ons and aerial duels won will also be multiplied by one hundred. This will allow for a better balance in terms of the tabulated offensive and defensive actions of each team.

Sub-Index 1

Team Contributions Index

$$I_1 = 100\ [0.4844(x_1) + 0 \cdot 4239(x_2) + 0 \cdot 1644(x_3) + 0 \cdot 0333(x_4)] - 04651(x_5) \quad - 0 \cdot 7879(x_6) - 0 \cdot 6562(x_7) - 0 \cdot 282(x_8) - 0 \cdot 2411(x_9)$$

where x_1 = number of set piece goals, x_2 = tackles % won, x_3 = take-ons % won, x_4 = aerial duels % won, x_5 = number of interceptions, x_6 = number of blocked shots, x_7 = number of clearances, x_8 = number of red cards, x_9 = number of yellow cards.

Sub-Index 2

Goal Difference Index

This sub-index awards points to a team based on net goals. The specific number of points awarded has been calculated by converting goals into points. Over the 2016/2017 Premier League Season, there was a total of 1064 goals scored, and 1056 points won. Therefore, we can estimate how many points one goal is worth as $\frac{1056}{1064} = 0 \cdot 9925$ points for each goal. This means that on this index, a team receives 0.9925 points for each goal the team scores. The points awarded to a team for goal difference is simply points per goal multiplied by a team's goal difference scaled by a factor of ten - to keep in line with the weight of the first sub-index, as well as not to outweigh it.

$$I_2 = 10 \times \text{goal difference}_i \times 0 \cdot 9925$$

where $i = 1,2,3,\cdots,20$ denotes the each of the 20 teams in the Premier League Season 2016/2017.

Sub-Index 3

Assists Index

An assist is defined as a pass which leads to a goal. Therefore, from our previous estimate a goal is worth $0 \cdot 9925$ points. We can place an assist on this same scale. Hence, each assist by a team is worth $0 \cdot 9925$ points. The points awarded for the assist for each team is simply the number of assists multiplied by the points for each assist. As for the previous sub-index, we scale by a factor of ten to get:

$$I_3 = 10 \times \text{assists}_i \times 0.9925$$

where $i = 1,2,3,\cdots,20$ denotes the each of the 20 teams in the Premier League Season 2016/2017.

Sub-Index 4

Key Pass Index

A key pass is defined as a pass that creates a goal scoring opportunity. At times, a key pass leads to an assist. The total chances created by each team is a combination of the key passes and assists of each team. From the 2016/2017 Premier League Season Data there were a total of 7067 chances created. 717 of these chances created were assists, therefore:

$$\frac{717}{7067} \times 100 = 10 \cdot 146\%$$

From this analysis we can conclude that approximately 10% of the chances created resulted in goals. This leads to approximately 90% of the total chances created to be classified as key passes. Therefore, a chance created is nine times more likely to not result in a goal as to result in a goal. The points awarded per assist are $0 \cdot 9925$. As a result, the points awarded per key pass should be $\frac{0 \cdot 9925}{9}$ which is close to one ninth of the value of an assist, i.e., $0 \cdot 1103$. As before, we scale by a factor of ten to obtain

$$I_4 = 10 \times \text{key passes}_i \times 0 \cdot 1103$$

where $i = 1,2,3,\cdots,20$ denotes the each of the 20 teams in the Premier League Season 2016/2017.

Sub Index 5

Work Rate Index

The seasonal points obtained per team based on distance covered which again is scaled by a factor of ten (10).

$$\text{Work rate: } I_5 = \frac{\text{distance covered}_i \times \text{points}_i \times 10}{\sum_{i=1}^{20} \text{distance}_i}$$

where $i = 1,2,3,\cdots,20$ represents the each of the 20 teams in the Premier League Season 2016/2017.

Work rate is a measure that contributes significantly less than the other sub-indices. This is mainly because players in general tend to run more and cover more distance when they are not in possession of the soccer ball. This could translate to being under pressure from opposing teams. Hence, absorbing such pressure takes a high level of concentration and should be merited. In terms of team rating this would not place a team at the summit by any means. However, it could separate teams with fine margins in ratings.

The final index is calculated by taking the sum of the five sub-indices calculated previously:

$$\text{The Final Index } = I_1 + I_2 + I_3 + I_4 + I_5 .$$

Note that some of the ideas in creating this index were utilised and modified from (McHale et al., 2012).

Results

a. Hungarian Method Results:

The optimal defensive assignment (jobs → player) of Manchester United in the 2016/2017 Premier League season was as follows:

Passing → Smalling; **Tackling** → Valencia; **Blocking** → Bailly;
Intercepting →Darmian; **Clearing** → Jones.

To determine the maximum defensive assignment score for Manchester United, we combine the initial average data for the specific job that is assigned to each of these five defenders as follows:

$$0.89 + 2.4 + 0.9 + 2.3 + 7.6 = 14.09$$

The same analysis can be carried out for the midfielders and strikers from Manchester United, resulting in the following assignments.
Midfield assignment for Manchester United:

Passing → Lingard; **Shots per Gam**e → Pogba; **Through Balls** →Mkhitarian; **Key Passes** → Mata; **Tackling** →Fellaini; **Assists** → Herrera; **Intercepting** → Carrick.

The associated maximum midfield assignment for Manchester United is calculated to give

$$0.88 + 3.1 + 0.2 + 1.8 + 2 + 6 + 1.9 = 15.88$$

Striker assignment for Manchester United:

Assists → Martial; **Successful Dribbles** → Rashford;
Shots per Game → Ibrahimovic; **Fouled per Game**→ Rooney.

The maximum assignment for the strikers of Manchester United tallies to:

$$6 + 1.3 + 4.1 + 0.5 = 11.9$$

The same analysis on the Chelsea team for defense is as follows:

Intercepting → Azpilicueta; **Blocking** → Cahill; **Clearing** → Luiz;
Passing → Terry; **Tackling** → Aké.

The maximum assignment for defence in the Chelsea team is therefore:

$$1.9 + 0.5 + 5.3 + 0.92 + 2 = 10.62$$

We now apply the analysis to the Chelsea midfielders. The optimal assignment is as follows:

Shots per Game → Hazard; **Assists** → Fabregas; **Through Balls** → Willian; **Intercepting** → Matic; **Key Passes** → Oscar; **Tackling** → Kanté; **Passing** → Loftus-Cheek.

The maximum midfield assignment score for Chelsea is therefore:

$$2.1 + 12 + 0.1 + 1.4 + 1.7 + 3.6 + 0.84 = 21.74$$

The optimal assignment for the strikers of Chelsea produces the following:

Assists → Pedro; **Successful Dribbles** → Hazard;
Shots per Game → Costa; **Fouled per Game**→ Moses.

The maximum striker assignment for Chelsea calculates is therefore given as:

$$3.2 + 8 + 3.9 + 0.9 = 16.$$

b. Performance Rating Results:

Figure 1 presents a scatter diagram displaying the multiple correlations for each variable. These are used in the calculation of the first sub-index 'team contributions'. The four remaining sub-indices are 'goal difference', 'assist', 'key pass', 'work rate'. The five calculated sub-indices are summarized in Tables 10 and 11.

Figure 1. *Multiple Correlations with Reference Variable 'Points'*

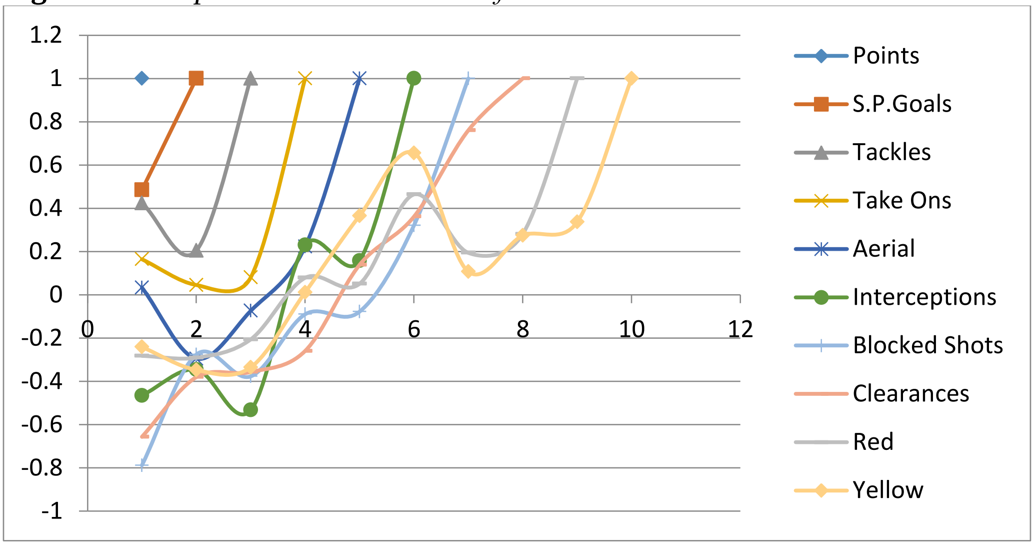


Table 10. *Sub-indices 1 & 2 - Team Contributions & Goal Difference*

Team	**Score (Team Contributions Index)**	**Score (Goal Difference Index)**
Chelsea	3258·15	516·10
West Bromwich Albion	2694·62	−79·40
Tottenham Hotspur	2899·47	595·50
Swansea City	2665.90	−248·13
Liverpool	2944·11	357·30
West Ham United	2577·96	−168·73
Bournemouth	2629·69	−119·10

Manchester City	2735·41	406·93
Burnley	2214·87	−158·80
Hull City	2431·96	−426·78
Arsenal	2741·94	327·53
Crystal Palace	2478·15	−129·03
Everton	2518·69	178·65
Watford	2196·54	−227·90
Stoke City	2424·60	−148·88
Leicester City	2123·46	−148·88
Manchester United	2058·51	248·13
Southampton	2404·40	−69·48
Sunderland	2073·07	−397·00
Middlesbrough	2099·02	−258·05

Table 11. *Sub-Indices 3, 4, 5 - Assist, Key Pass and Work Rate*

Team	**Score (Assist Index)**	**Score (Key Pass Index)**	**Score (Work Rate Index)**
Chelsea	555·80	430·17	47·35
West Bromwich Albion	337·45	276·85	23·18
Tottenham Hotspur	585·58	490·84	44·18
Swansea City	317·60	283·47	20·65
Liverpool	545·88	486·42	39·996
West Ham United	287·83	355·17	22·18
Bournemouth	357·30	330·90	23·48
Manchester City	516·10	474·29	40·38
Burnley	228·28	262·51	20·47
Hull City	535·95	269·13	17·08
Arsenal	496·25	415·83	37·88
Crystal Palace	337·45	302·22	19·79
Everton	456·55	380·54	29·99
Watford	258·05	295·60	19·38
Stoke City	228·28	296·71	21·53
Leicester City	327·53	285·68	21·51
Manchester United	406·93	447·82	32·96
Southampton	258·05	403·70	22·57
Sunderland	119·10	269·13	11·58
Middlesbrough	208·43	247·07	14·29

The data on average distance (km) covered per game by each team and the total distance covered by each team for the season was obtained from the Express UK online (Express UK 2017). We found that the work rate index is substantially smaller than the assist and key pass indices. Also, the goal difference index can be either positive or negative. It can therefore add to or subtract from a team's rating. It is quite clear that the team contributions index carries the greatest weighting in the overall rating index. Table 12 presents the overall rating calculated for the twenty Premier League teams for the 2016/ 2017 season, arranged in order from the highest to the lowest rating.

Table 12. *Final Team Ratings*

Position	Team	Team Rating
1.	Chelsea	4807
2.	Tottenham Hotspur	4616
3.	Liverpool	4374
4.	Manchester City	4173
5.	Arsenal	4059
6.	Everton	3445
7.	West Bromwich Albion	3253
8.	Bournemouth	3222
9.	Manchester United	3194
10.	West Ham United	3074
11.	Swansea City	3039
12.	Southampton	3019
13.	Crystal Palace	3009
14.	Stoke City	2822
15.	Hull City	2788
16.	Leicester City	2609
17.	Burnley	2567
18.	Watford	2542
19.	Middlesbrough	2311
20.	Sunderland	2076

Discussion

Recapping the optimal assignments for both teams, we observe that Chelsea's entire round total was higher that Manchester United's. However, Manchester United's average defensive assignment was higher that Chelsea's. As a result, Manchester United conceded fewer goals than Chelsea, twenty-nine as opposed to thirty-three. Note that Chelsea scored significantly more goals than Manchester United i.e., eighty-five to fifty-four. Chelsea also had a goal difference which was more than twice that of Manchester United. Chelsea's midfield and attack scored approximately five and four more assignment points per game respectively. This tells the story of how effective the link between midfield and attack worked for Chelsea. The midfield also helped martial the defence by creating a formidable barrier in front of the defence. We can see that the combined average assignment per game in defence and midfield for Chelsea was 32.36 as opposed to 29.97 for Manchester United. This shows how superior the midfield of Chelsea was in supporting the defence and linking up the attacks.

This type of analysis can inform the manager and coaching as to the best players for various roles. It will in fact aid in the selection of the team from the available players - depending on the team formation adopted for a particular game. For example, suppose that Chelsea was playing the (3-4-3) formation with three defenders, having identified that the three most important roles to counteract the opposition's weaknesses were to intercept, clear and pass optimally to neutralize the opposition's attack. In such a case, Azpilicueta, Luiz and Terry would have been the three best available options in that particular season.

It is interesting to observe that the top two and bottom two teams in our rating index placed exactly the same as the final league table for 2016/2017. The top five teams were also the same as the final league table, but with Manchester City and Liverpool switching positions. We note that in our final team rating index, Everton, West Bromwich Albion and Bournemouth all finished above Manchester United, however this did not happen in the final league table for 2016/2017. This was because offensively, their contributions on our ratings index were higher than that of Manchester United.

Conclusion

The selection of individual football players to function as a cohesive unit can be a very daunting task for coaches. Getting the right balance of strikers, midfielders and defenders is critical to the team's all-round performance. By using data from previous games on how players perform the various roles, coaches can explore the best combinations to use for upcoming matches. We have demonstrated how this can be achieved via the application of the Hungarian Algorithm. Web sources provide data on football statistics such as blocking, clearing, tackling, intercepting, dribbling, shooting, assisting, passing, etc. We are able to divide these attributes into defensive and offensive together with a combination of both to pick the best defence, midfield and offence to perform optimally as a unit.

We have also described how to use player statistics to create a ranking system for all registered teams in a football league. This can be achieved through the creation of a team index by way of a combination of five sub-indices. The first sub-index is called team contributions, and it accounts for the number of set piece goals, shots, blocks, tackles won, aerial duels won, clearances, red and yellow cards obtained by the players. For each team, the total number for each component is multiplied by an estimated correlation coefficient and the resulting values are added to determine the overall score that is representative of these contributions. The four remaining sub-indices are called goal difference, assist, key pass and work rate. Each of these sub-indices contributes a score to the overall team index, based on the overall numbers that the team amasses in each respective aspect of team play. The score for the five sub-indices is then totalled to produce the team index score, and the teams are ranked from highest to lowest based on the final index score.

We have used the 2016/2017 Premier League data to demonstrate the similarities between our team ranking index and the eventual position of each team in the league table at the end of the season. This suggests that our proposed team index can be used as a league predictor for future seasons and to set up betting odds for teams. Further analysis could be carried out to determine what proportion each sub index contributes to the all-round team index. This would allow conclusions to be drawn on the effectiveness of the various sub-indices and their relative importance in predicting the outcome of the league.

The ratings index that we have presented in this paper provides an additional tool for the comparison of teams. It allows us to analyse the overall performance, and subsequently to determine the best and worst teams in the league. Some of the ideas in creating this index were utilised and modified from (McHale et al. 2012). The team index is a single score used to rate the collective player contributions that directly influence overall team success. It provides a quantitative way to measure the differences between teams.

References

Britz SS, Maltitz MJ (2010) *Application of the Hungarian algorithm in baseball team selection and assignment.* Available at: https://www.ufs.ac.za/docs/librariesprovider22/mathematical-statistics-and-actuarial-science-documents/technical-reports-documents/teg419-2070-eng.pdf?sfvrsn=383cf921_0.

Express UK (2017) *Premier League distance covered.* Available at: https://www.express.co. uk/sport/football/794607/Premier-League-distance-covered-sportgalleries.

Forrest D, Mc Hale I (2007) Anyone for tennis (betting)? *European Journal of Finance* 13(8): 751–768.

Kuhn WH (1955) The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2(1–2): 83–97.

McHale I, Davies S (2007) Statistical analysis of the effectiveness of the FIFA world rankings. In *Statistical Thinking in Sports*, 77–90. 1st Edition. Chapman and Hall/CRC.

McHale I, Morton A (2011) A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting* 27(2): 619–630.

McHale I, Scarf P, Folker D (2012) On the development of a soccer player performance rating system for the English Premier League. *Interface Science* 42(4): 339–351.

Munkres J (1957) Algorithms for the assignment and transportation problems. *SIAM Journal on Applied Mathematics* 5(1): 32–38.

Qadar MA, Zaidan BB, Zaidan AA, Ali SK, Kamaluddin MA, Radzi WB (2017) A methodology for football players selection problem based on multi-measurements criteria analysis. *Measurement* 111: 38–50.

Squawka (2017) https://www.squawka.com/en/comparison-matrix/#avg.

Whoscored (2017) *Premier League 2016/2017*. Available at: https://www.whoscored.com/statistics.

Yi C (2023) *Hungarian algorithm for linear assignment problems (V2.3)*. Available at: https://www.mathworks.com/matlabcentral/fileexchange/20652-hungarian-algorithm-for-linear-assignment-problems-v2-3.