Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 45 (2024), No. 1, pp. 1–23
DOI : 10.47974/JIOS-1049

Stochastic optimization model to simulate football games as logistics networks

Kenan Mengüç *
Department of Industrial Engineering
Istanbul Technical University
Beşiktaş
Istanbul
Turkey

Nezir Aydin †
Department of Industrial Engineering
Yıldız Technical University
Beşiktaş
Istanbul
Turkey

Abstract

Football industry grows and improves day by day. All successes in universal tournaments results in huge revenue. Since the individual performances of the players determine the performance of the team, players` performances are measured by commercial companies with a higher reliability. Significant football models have been produced from several produced data. The proposed model, in this study, develops an attack routing model based on deviations in players` performances because the performances of players fluctuate from game to game. These fluctuations are handled by incorporating stochasticity. Thus, the developed mathematical model is stochastic and compares attack routes according to variations in performances of players. As a case study, one of the best team, which ranked in top three in its league for last two years, is analyzed and results show the adequacy of the model very clearly.

Subject Classification: *Primary 65K05, Secondary 91B70.*

Keywords: *Transshipment problem, Optimization, Spatio-temporal data, Football, Soccer, Stochastic modelling, Chance constraint.*

* *E-mail:* **menguck@itu.edu.tr (Corresponding Author)**
orcid.org/0000-0002-7536-2124
† *E-mail:* **nzraydin@yildiz.edu.tr; orcid.org/0000-0003-3621-0619**

1. Introduction

The increasing brand values of football clubs and watching rates of football worldwide have brought not only economic interest but also deep sociological studies. Many technical and tactical research have been studied out for football industry where pitch success contributes the most. The data sets used for these studies are quite varied. Thus, several commercial companies perform many measurements which range from the physical dynamism of the players to the team performance. A stochastic mathematical model being suitable for the nature of football was produced by using these measurements and the results were interpreted clearly. The model, which aims to determine an effective offensive course, offers different routes for all possible pressure of opponent or poor performance of team.

There have been a great deal of publications on optimization in sport over the last 50 years (Wright, 2009). The rules of the game in team sports are also effective in speeding up the game (Takeuchi, Ramadan & Iida, 2014). Football is a dynamic game rather than static game as structure of its nature (Rahnamai Barghi, 2015). The importance and amount of big data has been increased with the growing of football industry. Both the training and competition performances of the athletes are measured in detail by the Optasport and the reliability of these data are also concluded to be high (Liu, Hopkins, Gómez & Molinuevo, 2013). Optasport data are used to determine a strategy for team sport or individual sport (Stein et al., 2017). In another study, the players` data used to select the most ideal squad by Tavana et. al., (2013). Furthermore, Rao and Shrivastava (2017) used artificial intelligence to estimate the players' positions on the pitch.

The media generally interprets statistical data, but it is important to analyze the performance of the player on the pitch. Therefore, spatio-temporal data give more realistic results and Gudmundsson and Horton (2017) used spatio-temporal data rather than classical percentage data in their study. In an important study, which is conducted for precious leagues in Europe, indicated that passes between players takes a very valuable place in football because the teams with higher pass amount take higher places in the scoring (Cintia et. al., 2015). Accordingly, Pena and Touchette (2012) mentioned that the pass structures have brought about important results in football game (Pena and Touchette, 2012). Besides, the passes differ according to the style of ball's transshipment (Malqui, 2017).

Football and logistics problems have some similarities according to make routes. The transshipment model of Krishnan and Rao's model is modified to implement in this study (Krishnan & Rao, 1965). Considering to effective offensive routes in the football, the positions of important players were analyzed to identify goal attempt motifs in offensive routes during the match for deeper analysis (Bekkers, 2017).

Mclean et. al., (2019) analyzed matches by categorizing their results as win, loss and draw. They show that the network of passing is changed according to the result of the match Furthermore, they concluded that even though the score affects the network structure, there is no significant difference between individual performance and team performance (Mclean, et. al., 2019).

Yu et al. (2018) has constructed the unidirectional transshipment game model by using model TOPSIS (Yu et. al., 2018). Similarly, Wei Xu et. al., estimated the risk using the TOPSIS method to make decisions under uncertainty. In our study, the risk has been determined by statistical distributions (Xu, et. al., 2017). Stochastic modeling is developed for situations where individual situations are not clear (Kolbin, 1977). In such cases stochastic programming is used (Aydin, 2020). Furthermore, it is preferred for problems where probability and statistics are on the agenda (Pr´ekopa, 1978). Stochastic models generally are solved by transforming them into deterministic version. Commonly, three version of such problems are handled and are developed by Taha including transforming them into deterministic forms: Maximum Expected Value Model (E -Model) (Taha, 2007), Minimum Variance Model (V -Model) (Taha, 2007), Maximum Probability Model (P-Model) (Taha, 2007; Díaz-García and Garay-Tápia, 2007).

Shapiro brings a different dimension to above studies (Shapiro, 2009). The uncertainty in data is used to model by random coefficients with known probability distribution (Sengupta et. al., 1963; Atalay and Apaydin, 2011). Normal distribution was mostly used in stochastic models, but Atalay and Apaydin presented a different approach and used chi-square distribution (Atalay and Apaydın, 2011). In football, stochastic modeling was used to determine the most ideal transfer policy for the next three seasons for football management (Pantuso, 2017). The goal of this study is to propose an offensive routing model that predicts changes in players' performances. The results are also compared with a played game to show the effectiveness of the proposed work.

2. Model Development and Case Study

In this study we first presented the data collected for the case study and then the development of transshipment problem model is shown. In case study, Basaksehir F.K. is selected as the team to be analyzed. The data of players are gathered from Optasport (Liu, et. al., 2013). Optasport data were open sourced in Turkey during the 2014/2015 football season.

The measures of the pitch are calculated in meters and then separated into coordinates. Next, target directions are determined. Passes are scored and classified based on their performance in bringing ball closer to penalty zone. Overall score for a player is created considering his performance in passing through all season. For each direction (right, left, back, forward) a score is calculated. Further, the position of each player is determined based on his location on the pitch through the season. For each direction at a position the most successful two players are determined. To create an effective attack the best combination of players and passes are determined by the model (Mengüç, 2019). The deterministic model is run for each problem type and results are obtained determined (Deterministic Model Result-DMR). The results (objective function of deterministic model) are used in the stochastic model as constraints. For the played match, the stochastic model is solved for three different positions and for different ratio of confidence interval by using Performance of Players (POP). The flowchart of the steps is summarized in Figure 1.

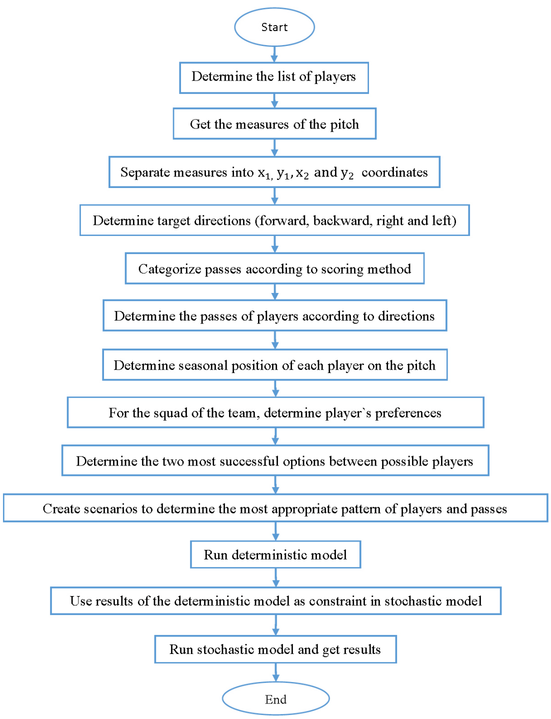

Figure 1

Stochastic model formation flowchart to determine effective route in football.

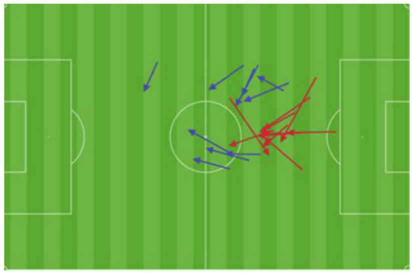

Figure 2

Passes of a player in a match.

Passes are analyzed according to x_1, y_1, x_2 and y_2 coordinates as shown in Figure 2. Considering left (l), right (r), back (b) and forward (f) as directions, the pass of each player is classified according to equations in (1- 4).

$$x_2 - x_1 > 0 \Rightarrow f = x_1 - x_2 \tag{1}$$

$$x_2 - x_1 < 0 \Rightarrow b = x_1 - x_2 \tag{2}$$

$$y_2 - y_1 > 0 \Rightarrow r = x_1 - x_2 \tag{3}$$

$$y_2 - y_1 < 0 \Rightarrow l = x_1 - x_2 \tag{4}$$

$$P_\lambda = \frac{\left|S_\lambda^1\right|}{\left|S_\lambda^1 \cup S_\lambda^0\right|} \tag{5}$$

Where,

P_λ **:** successful pass rate for each λ. λ **= {r, l, f, b}**

S_λ^1 : successful pass set for λ.

S_λ^0 : unsuccessful pass set for λ.

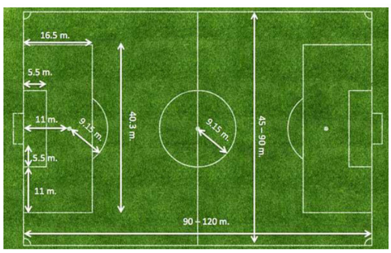

Figure 3

General pitch measurements (Menguc, 2019).

The size of the most stadiums in Turkey are designed as 68 meters wide to 105 meters long as in Figure 3. The data used in this study is processed considering above mentioned sizes. Further, the equation (6) and equation (7) are used for right/left and forward/back direction passes, respectively.

$$a_{ijv} = \frac{\left|\frac{e}{2} - y_2\right|}{\left(\frac{e}{2}\right)} * \left|y_2 - y_1\right| \tag{6}$$

$$a_{ijh} = \frac{x_2}{b} * \left|x_2 - x_1\right| \tag{7}$$

i: Start node of pass.

j: End node of pass.

a_{ijv} ***:*** Success rate of vertical (v) passes between nodes i, j.

a_{ijh} ***:*** Success rate of horizontal (h) passes between nodes i, j.

e: length of the short side of the pitch.

b: length of the long side of the pitch.

Equation (8) and (9) value the passes made to the center and to the opponent's goal, respectively.

$$a_{12v} = \frac{\left|34 - y_2\right|}{34} * \left|y_2 - y_1\right| \tag{8}$$

$$a_{12h} = \frac{x_2}{105} * \left|x_2 - x_1\right| \tag{9}$$

Where, 34 shows the half of the 68, which is the length of the short side of the pitch and 105 shows the length of the long side of the pitch.

Successful and failed passes are shown with blue red and arrows, respectively, in Figure 2. Then, all passes are classified according to their directions as a_{ijl}, a_{ijr}, a_{ijb} and a_{ijf}.

$$\mu_\lambda = \frac{\sum_{i=1}^{k}(d_{i\lambda}^1 + d_{i\lambda}^0)}{n_\lambda} \qquad (\forall\lambda) \tag{10}$$

Where,

μ_λ **:** Average score of the passes for each direction *λ= {r, l, f, b}.*

$d^1_{i\lambda}$ **:** success score of pass ***i*** in direction *λ*, and $i \ge 0, i = N, i \in (S^1_\lambda \vee S^0_\lambda)$

$d^0_{i\lambda}$ **:** failure score of pass ***i*** in direction *λ*.

n_λ : Sum of the passes done toward direction *λ*, and $n_\lambda \ge 0,\ n_\lambda = s(S^1_\lambda \vee S^0_\lambda)$

Linear approach:

The average performance score of each player and the average length of pass they made for a season is determined according to equation (5) and (10). Then, distance-based score statistics are created using the equations provided below:

$$P^{\acute{y}}_{\lambda k} = \frac{\mu_{\lambda k} * P_{\lambda k}}{d_{ik}} \tag{11}$$

if $$P^{\acute{y}}_{\lambda k} \le 100,\ c_{\lambda k} = P^{\acute{y}}_{\lambda k} / 100 \tag{12}$$

else $$P^{\acute{y}}_{\lambda k} > 100,\ c_{\lambda k} = 1 \tag{13}$$

In these equations,

$P^{\acute{y}}_{\lambda k}$ **:** Success pass score of player ***k*** toward direction *λ* for specified distance.

$P_{\lambda k}$ **:** Success pass score of player ***k*** toward direction *λ* for average distance through the season.

$\mu_{\lambda k}$ **:** Average passing distance of player ***k*** for the direction *λ***.**

d_{ik} **:** Desired distance for ***i***$^{\text{th}}$ pass by player ***k*****.**

$c_{\lambda k}$ **:** Passing score toward direction *λ* by player ***k*** for specified distance**.**

c_{ijk} **:** Passing score of player ***k*** between node ***i*** and ***j*** for specified distance**.**

The distances of the passes are transformed to the scores using equations provided in (8 – 9).

a_{ijh} and a_{ijv} are scores in equations (13).

a_{ijh} : Back or forward direction passes are is converted to score by equations (8 - 9).

a_{ijv} **:** Left or right direction passes are converted to score by equations (8 - 9).

The proposed model is the revised version of the transportation model. This study examines how the possible fluctuations in player performances change routes of the ball. In addition, the chance constraint is used for this situation. c_{ijk} shows the performance of player $\boldsymbol{k}$ from zone $\boldsymbol{i}$ to zone $\boldsymbol{j}$. Solutions were produced for α=0.99, 0.95 and 0.90 confidence levels. A result was obtained according to the e-model. The objective function was written as a constraint, and this result was at the right side of the equation (16). The results of different confidence levels determined by calculating the standard deviation of the whole team were taken for the right-side value ($\boldsymbol{b_i}$) of the new constraint.

The Deterministic Model Result (DMR) was converted to stochastic model. The objective of the deterministic model is $\max z : \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{K} C_{ijk}{}' x_{ijk}$

The stochastic chance constraint version of the above model is constructed as follows:

$$\text{Max}\, z : f \tag{14}$$

Subject to:

$$P\left(\sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{k=1}^{K} c_{ijk}{}' x_{ijk} \leq f \right) \geq (1-\alpha), \tag{15}$$

$$\sum_{k=1}^{2} E\{c_{ijk}\} x_{ijk} + K_{c_{ijk} y_i} \leq b_i + K_{c_{ij}} \sqrt{Var_{b_i}}, \quad \forall i,j,k \tag{16}$$

$$\sum_{k=1}^{2} E\{c_{ijk}\} x_{ijk}{}^2 - y_i^2 = 0, \quad \forall i,j,k \tag{17}$$

$$x_{ijk} \neq x_{jzk}, \qquad \forall i,j,z,k \tag{18}$$

$$x_{ij1} + x_{ij2} = 1, \qquad \forall i,j \tag{19}$$

$$\sum_{j=1, j\neq i}^{m} x_{ijk} - \sum_{j=1, j\neq i}^{m} x_{jik} = 0, \quad \forall i = 1,2,\ldots.m, \forall k \text{ and } i \neq \gamma \tag{20}$$

$$\sum_{j=1, j\neq \gamma}^{m} x_{\gamma jk} - \sum_{j=1, j\neq \gamma}^{m} x_{j\gamma k} = -1, \qquad \forall \gamma, k \tag{21}$$

$$\sum_{i=\theta}^{t}\sum_{j=1,j\neq i}^{t} x_{ijk} - \sum_{i=\theta}^{t}\sum_{j=1,j\neq i}^{t} x_{jik} = 1, \qquad \forall k \tag{22}$$

$$if \begin{cases} x_{ijk} = 1, & c_{ijk} = c_{ijk} \\ x_{ijk} = 0, & c_{ijk} = 1 \end{cases}, \qquad \forall i,j,k \tag{23}$$

$$x_{ijk} \in \{0,1\}, \qquad \forall i,j,k \tag{24}$$

Equation (14): The objective function maximizes the total scores of players. This constraint refers to the deviation in the confidence interval of the deterministic performance (1 – α). f is the highest performance to be achieved stochastically.

Equation (15): The performance constraint was determined for different confidence intervals.

M: *m* shows the number of zones on a pitch, which is 60 in this case. Each zone is approximately 1 square meter.

x_{ijk} : =1 if player k passes the ball from zone *i* to *j*, otherwise =0.

c_{ijk} :pass performance of player ***k*** from zone ***i*** to ***j***. c_{ijk} is a continuous number between 0 and 1.

S: Set of players who are in the squad. ($k \in S$)

$K_{c_{ij}}$ ***:*** Confidence interval for DMR.

Var_{b_i} : Variance of football team performance.

$K_{c_{ij}y_i}$ ***:*** Confidence interval for players taken place in route ***i.***

Equation (16): Stochastic version of deterministic objective function.

$E\{c_{ijk}\}$: The average performance of player k from zone i to region j for the deterministic model such as (E -Model) (Taha, 2007).

Equation (17): The problem was converted to a separate program as in Taha’s study (Taha, 2007).

Equation (18): A player cannot throw pass to himself.

Equation (19): Each player can pass from only one zone.

Equation (20): It is the flow conservation constraint.

γ: Position of the ball where it stolen from the opponent and the attack organization begins. (γ= 1, 2....*m*)

F: The arc which starts in region θ and ends in region *t*. $(\theta, t \in F): \{50,51,52,53,56,57,58,59\} \in F$

Equation (21): Guarantee that the ball can enter the pitch only from certain zones in where it is stolen from opponent.

Equation (22): The attack may only end in certain zones.

Equation (23): Our model aims to maximize attack performance. This constraint is placed in order to prevent the ball from moving around the pitch too much in an attack organization. Full performance is determined for every situation where there is no transfer of the ball. Thus, while the ball goes to the goal, it is prevented that the ball goes around the field too much.

θ*:* First region where the attack ends $(\theta \in F)$

t: Last region where the attack ends $(t \in F)$

The effective region was determined as areas within the competing penalty zones that are 50, 51, 52, 53, 56, 57, 58 and 59th zones for this problem.

6.Region	5.Region *(1)	4.Region *(1)	3.Region *(1)	2.Region *(1)	1.Region *(11)
12.Region *(80)	11.Region *(1) *(80)	10.Region *(1)	9.Region *(1)	8.Region *(15)	7.Region *(11)
18.Region *(80) *(15)	17.Region *(15) *(11) *14 *(14) *(80)	16.Region *(11) *(14) *(15)	15.Region *(11) *(14) *(15)	14.Region *(11) *(14) *(15)	13.Region *(11)
24.Region *(7) *(8) *(15) *(80)	23.Region *(14) *(11)*(5) *(8) *(15)*(80)	22.Region *(14) *(8) *(5) *(23) *(15)*(11)	21.Region *(8) *(23) *(14) *(15) *(5)	20.Region *(8) *(23) *(15) *(5) *(14) *(11)	19.Region *(8) *(17) *(11) *(23) *(5)
30.Region *(8) *(7) *(17) *(15) *(80)	29.Region *(8) *(14) *(15) *(80) *(5)	28.Region *(8) *(14) * (15) *(23) *(5)	27.Region *(8) *(14) *(23) *(5) *(15)	26.Region *(11) *(23) *(8) *(14) *(15) *(5)	25.Region *(11) *(23) *(5) *(17) *(8)
36.Region *(7) *(8) *(17) *(80) *(15)	35.Region *(7) *(15) *(8) *(17) *(23) *(80) *(26) *(5) *(14)	34.Region *(15) *(8) *(17) *(23) *(26) *(5) *(14)	33.Region *15 *(8) *(17) *(23) *(26) *(5)	32.Region *(15) *(8) *(17) *(23) *(26) *(5) *(11)	31.Region *(15) *(8)*(17) *(23) *(26) *(5) *(11)
42.Region *(7) *(8) *(17) *(80) *(15)	41.Region *(7) *(15) *(8) *(17) *(23) *(80) *(26)*(5)	40.Region *(15) *(8) *(17) *(23) *(26) *(5)	39.Region *(15) *(8) *(23) *(26) *(5)	38.Region *(15) *(8) *(17) *(23)*(26) *(5)	37.Region *(8) *(17) *(26) *(5) *(11)
48.Region *(7) *(8) *(17) *(80) *(15)	47.Region *(7) *(8) *(17) *(80) *(26)	46.Region *(26)	45.Region *(5) *(26)	44.Region *(5) *(26) *(8) *(17)	43.Region *(8) *(17) *(26) *(11)
54.Region *(7) *(17) *(80)	53.Region *(7) *(80) *(26)	52.Region	51.Region	50.Region *(26)	49.Region *(17)
60.Region *(7)	59.Region	58.Region	57.Region	56.Region	55.Region

Figure 4

Regional assignment (Mengüç, 2019)

As a case study, the match made between Besiktaş J.K and Basaksehir F.K in the season 2014/15 was analyzed. The team of Başakşehir F.K., was captured from real data and the players positions are shown in Figure 4. The match between the two teams was modelled as a deterministic model in Mengüç's study and players line-up as shown in Figure 4 (Mengüç, 2019). However, the model proposed in this study is a more advanced because it predicts the variability of player performance via stochastic programming.

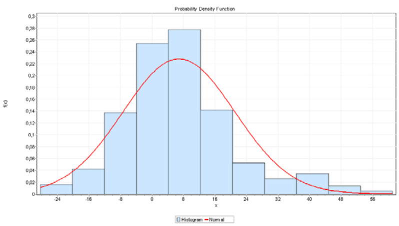

Figure 5

Performance distribution of player # 1,

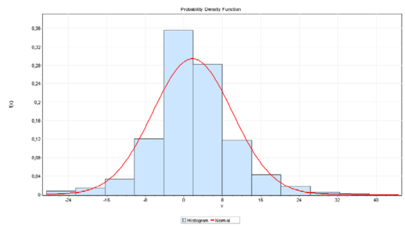

Figure 6

Performance distribution of player # 5.

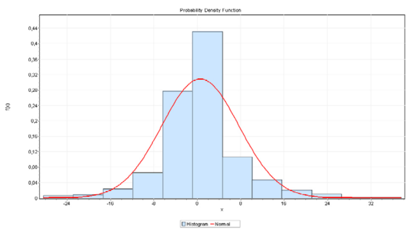

Figure 7

Performance distribution of player # 7.

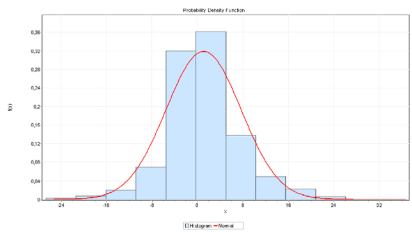

Figure 8

Performance distribution of player # 8.

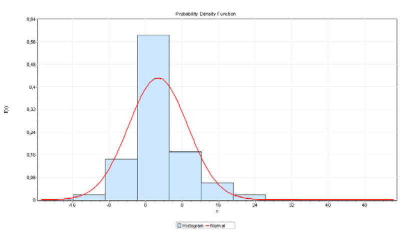

Figure 9

Performance distribution of player # 11.

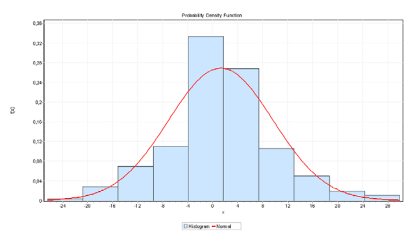

Figure 10

Performance distribution of player # 14.

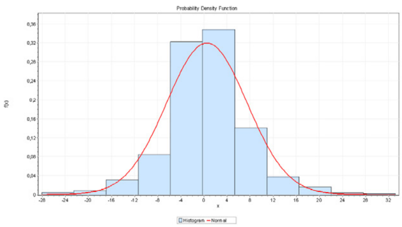

Figure 11

Performance distribution of player # 15.

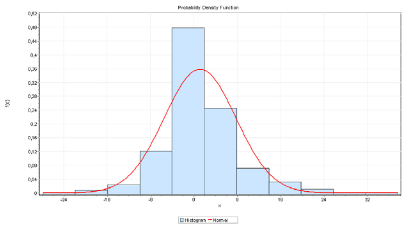

Figure 12

Performance distribution of player # 17.

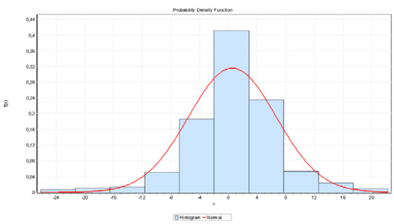

Figure 13

Performance distribution of player # 23.

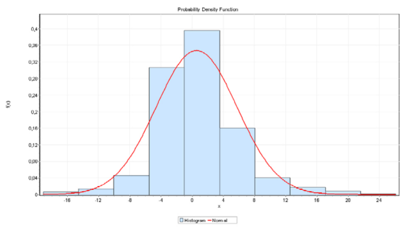

Figure 14

Performance distribution of player # 26.

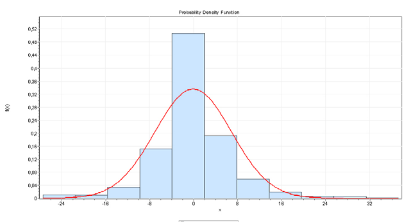

Figure 15

Performance distribution of player # 80.

The data show the number of passes made by each player in season 2016/2107. The normal distributions plotted in Figures 5-15 show the performance of players and are determined based on 500 to 2500 data points per player. According to the central limit theory, when the number of data is near infinity, it is thought that the distribution of the above player distributions will approach normal distribution (Bernstein, S. N, 1945). Therefore, the following standard deviations and variances for each player in Table 1 are calculated to use in the mathematical model.

Table 1

Statistics of players and team.

#	Player	Std. Dev.	Variance
1	Volkan	14,215	202,0662
5	Emre	8,282	68,59152
8	Mussoro	6,6029	43,59829
7	Visca	7,1122	50,58339
11	Eren Albayrak	6,4833	42,03318
14	Bekir İrtegül	8,3483	69,69411
15	Attamah	6,9477	48,27054
17	Cengiz	6,6227	43,86016
23	Holmen	6,1286	37,55974
26	Adebayor	5,1888	26,92365
80	Çaiçara	6,9793	48,71063
All	Team of BFK	7,5373	61,99012

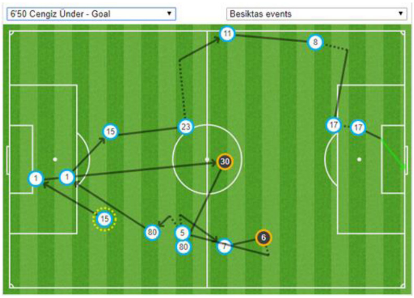

Figure 16

Attack at 6th minute.

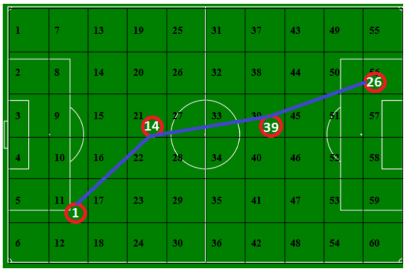

Figure 17

Simulation of attack at 6th minute (POP90%, DMR10%)

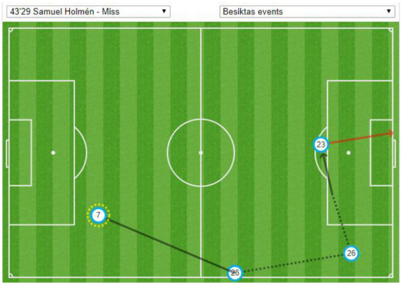

Figure 18
Attack at 43th minute.

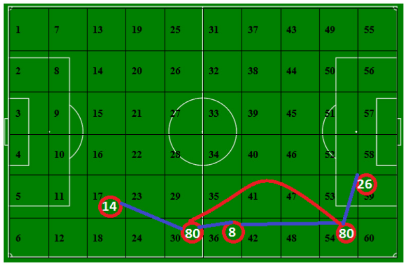

Figure 19
Simulation of attack at 43th minute (POP 90%, DMR95%).

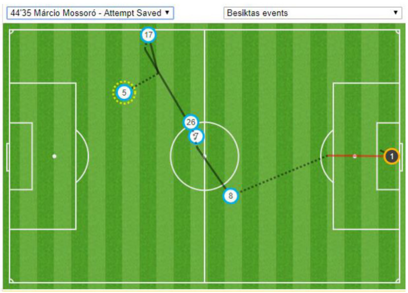

Figure 20
Attack at 44th minute

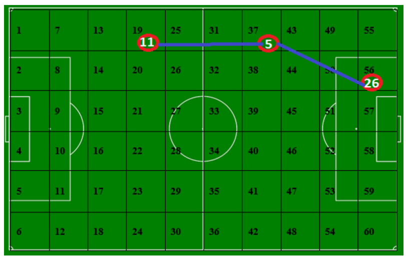

Figure 21
Attack at 44th minute for each POP and DMR.

Table 2
Attack's routes at 6th minutes for each confidence level.

	Routes						POP	DMR	Optimal score	Type of model
Player	1	15	23	11	8	17	Attack at 6th minutes			
Region	10	15	27	31	49	50				
Player	1	14	5	14	80	26	0	0	451,738582	deterministic
Region	11	17	23	35	54	59				
Player	1	14	5	26			99%	99%	465,1639897	stochastic
Region	10	21	39	59						

Contd...

Player	1	14	5	26			99%	95%	451,3996403	stochastic
Region	11	21	39	59						
Player	15	14	5	26			99%	90%	451,384092	stochastic
Region	8	26	38	56						
Player	15	14	5	26			99%	10%	451,3457338	stochastic
Region	8	21	39	54						
Player	No feasible solution						99%	0		stochastic
Region										
Player	1	14	5	26			95%	99%	451,6578025	stochastic
Region	10	22	40	56						
Player	1	14	5	26			95%	95%	451,524262	stochastic
Region	10	21	39	59						
Player	1	14	5	26			95%	90%	451,3996403	stochastic
Region	11	21	39	59						
Player	15	14	5	26			95%	10%	451,384092	stochastic
Region	8	26	38	56						
Player	No feasible solution						95%	0		stochastic
Region										
Player	1	14	5	80	26		90%	99%	451,6638308	stochastic
Region	9	27	34	54	59					
Player	1	14	5	26			90%	95%	451,5863836	stochastic
Region	9	27	39	59						
Player	1	14	5	26			90%	90%	451,524262	stochastic
Region	10	21	39	59						
Player	1	14	5	26			90%	10%	451,3996403	stochastic
Region	11	21	39	56						
Player	No feasible solution						90%	0		stochastic
Region										

Table 3

Attack's routes at 43th minute for each confidence level.

						POP	DMR	Optimal score	Type of model
Player	7	26	26	23		Attack route at 43rd minute			
Region	17	36	54	51					
Player	14	5	14	80	26	0%	0%	451,77781	deterministic
Region	17	23	35	54	59				
Player	14	8	80	26		95%	10%	451,4769	stochastic
Region	17	24	42	59					
Player	15	14	5	26		95%	90%	451,4648	stochastic
Region	17	22	40	56					
Player	15	14	5	26		95%	95%	451,4648	stochastic
Region	17	22	40	56					
Player	80	14	5	80	26	95%	99%	451,7158	stochastic
Region	17	23	34	54	59				
Player	No feasible solution					95%	0%	0	stochastic
Region									
Player	No feasible solution					90%	0%	0	stochastic
Region									
Player	No feasible solution					90%	10%	0	stochastic
Region									
Player	15	14	5	26		90%	90%	451,4648	stochastic
Region	17	22	40	56					
Player	14	80	8	80	26	90%	95%	451,7027	stochastic
Region	17	30	36	54	59				
Player	11	14	5	26		90%	99%	451,5016	stochastic
Region	17	29	41	59					
Player	No feasible solution					99%	0%	0	stochastic
Region									
Player	No feasible solution					99%	10%	0	stochastic
Region									
Player	15	14	5	26		99%	90%	451,429	stochastic
Region	17	30	42	59					

Contd...

Player	11	14	5	26		99%	95%	451,5015	stochastic
Region	17	29	41	59					
Player	14	8	80	26		99%	99%	451,4648	stochastic
Region	17	22	40	56					

Table 4

Attack' routes at 44th minute for each confidence level.

						POP	DMR	Optimal score	Type of model
Player	17	26	7	8	8	Attack at 44th minutes			
Region	17	27	27	40	51/52				
Player	11	5	26			0	0	452,1569	deterministic
Region	19	37	56					452,156972	
Player	11	5	26			99%	99%	452,1569	stochastic
Region	19	37	56						
Region	19	37	56			99%	90%	452,1569	stochastic
Region	19	37	56			99%	10%	452,1569	stochastic
Region	19	37	56			95%	99%	452,1569	stochastic
Region	19	37	56			95%	95%	452,1569	stochastic
Region	19	37	56			95%	90%	452,1569	stochastic
Region	19	37	56			95%	10%	452,1569	stochastic
Region	19	37	56			90%	99%	452,1569	stochastic
Region	19	37	56			90%	95%	452,1569	stochastic
Region	19	37	56			90%	90%	452,1569	stochastic
Region	19	37	56			90%	10%	452,1569	stochastic

Different POP and DMR values were examined with different confidence levels which provides different routes for the three different attacks as shown in Table 2, Table 3, and Table 4. The players have a seasonal average performance, but they perform with a different performances level in each competition. When POP was taken 90% and DMR was taken 10% a similar route was obtained, i.e., route at 6th minute in the match. Figure 16 shows the position at the 6-minute mark of the match, while Figure 17 shows the route taken when 90% of POP and 10% of DMR were used. This ratio provides meaningful information for the performance of the team to decision makers. The confidence level in Figure 17 creates a safer route by the time in Figure 18. This indicates that the team is not under great pressure in this position and therefore prefers the direct attack path despite the high risk of losing the ball. Figure 20 shows an attack that used different ratios of performance and DMRs. Interestingly, the model suggests the same route for this attack as shown in Figure 19. To be successful in this this attack, the team should follow a route with a lower number of alternatives. In addition, Figure 21 illustrates the attack route for POP and DMR at the 44th minute.

3. Conclusions

In real life, surprisingly, players show different performances when their teams are ahead or behind. This study provides strategies to decision makers for choose more or less risky routes by changing the right-hand side of the constraint given in the stochastic model. If the team wants a higher risky route, it should decrease the confidence level and vice versa. On the other hand, the results show that the teams under pressure prefer more risky routes when they are in an attack position.

This model will help the coach to make technical decisions based on the condition of the team during the match or the remaining time of the match. Every success achieved on the pitch leads to great gains in the football industry. Increased football player value, increased team value and increased revenue moves the team to a higher class as brand value.

As future studies, a more comprehensive model can be designed that includes certain characteristics of the competitor and the fiction of reality can be increased. Furthermore, stochastic model can be developed as a multi-stage stochastic model. Thus, different decision trees can be created according to the score status of the team during a live match. The strategies produced according to win, lose or draw situation will allow the strategy changes during the match, as well.

4. Discussion area and originality of the study

The aim of this study is to determine effective offensive routes according to the characters of the players in a team. The mathematical model proposed, in the study, gave consistent results with the effective routes of the team. While the method used in the study overlaps with some studies, it contradicts with others. Tahchiro and his friends suggested a clustering analysis for the pass behavior of a team in a match, but clustering is generally expected to vary across different player groups (Tahchiro et. al., 2014). This clustering hasn't provided a prediction for future match organizations (Kawasaki et. al., 2019). Since the developed model gives different routes for each player selection the usage area of the model increases. In some competitions, statistical superiority is strategically left to the opponent, since especially teams with more modest staff adopt speed offensive or counterattack football. Gamble et. al., (2019) conducted a study based on the statistical data of the passes, where the multiplicity of the number of passes was interpreted. However, the effectiveness of the passes are pondered rather than the number of passes in recent football matches. The passing data in the study was graded according to the region where pass was done. Meaningfully, passes that move the ball to the attack zone on the pitch were considered more valuable. In another study using statistical data, it was stated that the corner kicks in the last period of the match were more likely to result in goals (Casal et. al., 2015). This result can vary according to many variables such as squad structure, pitch, goalkeeper performance and climate. Therefore, the data used in this study is defined over each player`s performance since the variability of the team data will be high. The relationship between players developing and their birth dates have been also examined in the football leagues of England. In the study, Premier league, which has employed more players from different countries, has different results than the lower leagues in England (Fleming and Fleming, 2012). Since there is a wide range of player profiles in football leagues, it will be difficult to comment using general statistics. To overcome this issue spatial data was used in this study. Passing on the pitch was modeled by Mengüç (2019) as linear model. One of the most important differences between Mengüç (2019) and this study is that the fluctuations in players' performances are considered. By this contribution the model has become more realistic and applicable to the real world because it considers stochasticity.

In his study, Chassy and his friends mentioned that a mathematical model can be used for organizations in football (Chassy et. al., 2018). This study supports their prediction in using a mathematical model. Another study reported that the short pass sequences result in goal with a higher rate (Garratt et. al., 2017). The results of our study are consistent with the character of the pass routes that were expressed by Garratt et. al., (2017). The data set used in our study also has the reliability and consistency mentioned in Liu's study. Because the data set used obtained from the same source that Liu and friends used (Liu et. al., 2013). In their study Castellano and friends mentioned that teams behave differently against strong or weak teams (Castellano et. al., 2013). The confidence intervals used in our study were taken from different figures and then the model was re-solved. The purpose of this approach is to integrate possible changes in player performances into the model. Our reflection on these results has concluded in different results for different performance ranges. Cartesian coordinate was used for the separation of passes in Clemente's study (Clemente et. al., 2014). A similar approach is also applied in our study.

To summarize, it can be said that this study proposes a model that allows players to demonstrate effective offensive performance using their habits. A study which focuses on possible changes in player habits and makes out a routing according to these changes has not been encountered in the literature before.

5. Development proposal

In Taylor's study, team data were examined considering two consecutive seasons and strong consistency was observed between them (Taylor et al 2010). Pansure's work proposed a stochastic mathematical model on transfer policies (Pantuso, 2017). In our study, it is also suitable to apply in transfer strategies and can be used to estimate how much the possible transfer player will increase the reliability of the attack route of the team. In addition to this, a cost benefit analysis can also be performed. Moreover, a more complex model can be created that includes the competitor players` habits. The suggested model could reflect reality better.

Acknowledgements

The data of the study were used in the 2019 study conducted by Kenan Mengüç, one of the authors of this study. The author used the dataset, which he obtained as open source, in this study with his own permission.

References

[1] Atalay, K., & Apaydın, A. : Determınıstıc Equıvalents Of Chance Constraıned Stochastıc Programmıng Problems. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi-B Teorik Bilimler, 1(1), 1-18 (2011).

[2] Atalay, K. D., & Apaydin, A. : Gamma distribution approach in chance constrained stochastic programming model". *Journal of Inequalities and Applications,* 108, 1–13 (2011).

[3] Aydin, N. : Designing reverse logistics network of end-of-life-buildings as preparedness to disasters under uncertainty. *Journal of Cleaner Production*, 256, 120341 (2020).

[4] Bekkers, J. J. W. : Network Theory in Association Football: Defining Play Styles Using Flow Motifs (2017).

[5] Bernstein, S. N. : On the work of PL Chebyshev in Probability Theory. Russian.) Nauchnoe Nasledie PL Chebysheva. [The Scientific Legacy of PL Chebyshev.] Vol, 1, 43-48 (1945).

[6] Casal, C. A., Maneiro, R., Ardá, T., Losada, J. L., & Rial, A. : Analysis of corner kick success in elite football. *International Journal of Performance Analysis in Sport*, 15(2), 430-451 (2015).

[7] Castellano, J., Álvarez, D., Figueira, B., Coutinho, D., & Sampaio, J. : Identifying the effects from the quality of opposition in a Football team positioning strategy. *International Journal of Performance Analysis in Sport*, 13(3), 822-832 (2013).

[8] Chassy, P., Malone, J. J., & Clark, D. P. : A mathematical model of self-organisation in football. *International Journal of Performance Analysis in Sport*, 18(2), 217-228 (2018).

[9] Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D., & Malvaldi, M. : The harsh rule of the goals: Data-driven performance indicators for football teams. Paper presented at the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (2015).

[10] Clemente, M. F., Martins, F. M., Couceiro, S. M., Mendes, S. R., & Figueiredo, A. J. : Inspecting teammates' coverage during attacking plays in a football game: A case study. *International Journal of Performance Analysis in Sport*, 14(2), 384-400 (2014).

[11] Díaz-García, J. A., & Garay-Tápia, M. M. : Optimum allocation in stratified surveys: Stochastic programming. *Computational Statistics & Data Analysis*, 51(6), 3016-3026 (2007).

[12] Fleming, J., & Fleming, S. : Relative age effect amongst footballers in the English Premier League and English Football League, 2010-2011. *International Journal of Performance Analysis in Sport*, 12(2), 361-372 (2012).

[13] Gamble, D., Bradley, J., McCarren, A., & Moyna, N. M. : Team performance indicators which differentiate between winning and losing in elite Gaelic football. *International Journal of Performance Analysis in Sport*, 1-13 (2019).

[14] Garratt, K., Murphy, A., & Bower, R. : Passing and goal scoring characteristics in Australian A-League football. *International Journal of Performance Analysis in Sport*, 17(1-2), 77-85 (2017).

[15] Gudmundsson, J., & Horton, M. J. A. C. S. : Spatio-temporal analysis of team sports. 50(2), 22 (2017).

[16] Kawasaki, T., Sakaue, K., Matsubara, R., & Ishizaki, S. : Football pass network based on the measurement of player position by using network theory and clustering. *International Journal of Performance Analysis in Sport*, 19(3), 381-392 (2019).

[17] Krishnan, K., & Rao, V. J. J. O. I. E. : Inventory control in N warehouses. In, Vol. 16, pp. 212-& (1965).

[18] Kolbin, V. V., & Kolbin, V. V. : Stochastic programming (No. 14). Springer Science & Business Media (1977).

[19] Liu, H., Hopkins, W., Gómez, A. M., & Molinuevo, S. J. : Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport*, 13(3), 803-821 (2013).

[20] Malqui, J. L. S. : A visual analytics approach for passing strateggies analysis in soccer using geometric features (2017).

[21] Mengüç, K. : A Solution of a Mathematical Model Which Simulates Football Game as a Logistics Network (2019).

[22] Mclean, S., Salmon, P. M., Gorman, A. D., Dodd, K., & Solomon, C. : Integrating communication and passing networks in football using social network analysis. *Science and Medicine in Football*, 3(1), 29-35 (2019).

[23] Pena, J. L., & Touchette, H. J. A. P. A. : A network theory analysis of football strategies (2012).

[24] Pantuso, G. : The football team composition problem: a stochastic programming approach. *Journal of Quantitative Analysis in Sports*, 13(3), 113-129 (2017).

[25] Pr´ekopa A. : The use of stochastic programing for the solution of the some ploblems in statistics and probability. Technical Summary report 1983. University of Wisconsin-Madison (1978).

[26] Rahnamai Barghi, A. : Analyzing Dynamic Football Passing Network. Université d'Ottawa/University of Ottawa (2015).

[27] Rao, V., & Shrivastava, A. : Team strategizing using a machine learning approach. In 2017 International Conference on Inventive Computing and Informatics (ICICI), pp. 1032-1035. IEEE (2017, November).

[28] Sengupta, J. K., Tintner, G., & Millham, C. : "On Some Theorems of Stochastic Linear Programming with Applications". *Management Science,* 10(1), 143–159 (1963).

[29] Shapiro, A., Dentcheva, D., & Ruszczyński, A. : Lectures on stochastic programming: modeling and theory. *Society for Industrial and Applied Mathematics* (2009).

[30] Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., Grossniklaus, M. J. D. : How to make sense of team sport data: From acquisition to data modeling and research aspects. 2(1), 2 (2017).

[31] Taha, H. A. : Operations research: An introduction (2007).

[32] Takeuchi, J., Ramadan, R., & Iida, H. J. I. P. S. O. J. : Game refinement theory and its application to Volleyball. 2014, 1-6 (2014).

[33] Tavana, M., Azizi, F., Azizi, F., & Behzadian, M. J. S. M. R. : A fuzzy inference system with application to player selection and team formation in multi-player sports. 16(1), 97-110 (2013).

[34] Taylor, B. J., Mellalieu, D. S., James, N., & Barter, P. : Situation variable effects and tactical performance in professional association football. *International Journal of Performance Analysis in Sport,* 10(3), 255-269 (2010).

[35] Yu, L. E. : Effect of unidirectional transshipment on carbon emissions in supply chain from the perspective of non-cooperative game. *Journal of Interdisciplinary Mathematics,* 21(5), 1145-1150 (2018).

[36] Wright, M. J. J. O. T. O. R. S. : 50 years of OR in sport. 60(sup1), S161-S168 (2009).

[37] Xu, W., Liu, L., Zhang, Q., & Wang, X. : A multi-object decision-making method for location model of manufacturing industry under uncertain environment. *Journal of Interdisciplinary Mathematics,* 20(4), 1019-1028 (2017).

Recieved June, 2021