Football scout analysis models (based in the 2013/2014 Champions League)

http://dx.doi.org/10.11606/1807-5509201700010025

Rafael SONCIN*
Juliana PENNONE*
João Pedro PINHO*
Maraiza Campos DINIZ**
João Gustavo CLAUDINO*
Alberto Carlos AMADIO*
Júlio Cerca SERRÃO*
Bruno MEZÊNCIO*

*Escola de Educação Física e Esporte, Universidade de São Paulo, São Paulo. SP, Brasil.
**Faculdade Pitágoras, Betim, MG, Brasil.

Abstract

First places in major football competitions are no longer a few teams' achievement. A better understanding of the game becomes crucial with this highly competitive background. Our purpose was to determine scout variables that best describe the technical and tactical characteristics of the 2013/2014 Champions League's participating teams. All UEFA Champions League's qualifying phases and play-offs matches (32 teams in 126 matches) were analyzed using scout public domain data. The analyzed variables were goals scored, goals difference, total attempts, attempts on target, disarms, corners, offside, passes completed, passes uncompleted, fouls committed and ball possession. A best subsets multiple regression analysis reveal that total attempts (r^2 = 0.815), ball possession (r^2 = 0.748), passes completed (r^2 = 0.742) and scored goals (r^2 = 0.699) are the most relevant variables to describe the match. The present study was able to show how scout information can be reduced to the coach easily application.

Keywords: Statistics; Soccer; UEFA; Competitions.

Introduction

Football is one of the most practiced and acclaimed sports in the world[1]. The financial investments by football teams reach impressive amounts making national and international competitions major sources of income. Such conditions lead to a spectacle of grace and human performance raising the sports' visibility, so that there is an expectation of perfection in each event[2].

Athletes' top performance conditions demanded by investors and sponsors make teams' technical and tactical preparation indispensable[2]. First places in the major national and international competitions are no longer a few teams' achievement. Such a highly competitive scenario determines the end of this sport's powers' supremacy.

One of the most employed preparation strategies for an upcoming event is the observation of the opponents' matches. Such method allows identifying opponents' actions, as well as their own, producing information that helps planning the training process[3].

Scout analysis is one of the main instruments utilized to understand this modality. Used as well in other sports, it is a tool that earns increasing technological improvements due to its importance and applicability in numerous sports[4]. Scout analysis can provide several information such as total number of attempts on goal, goal difference, suffered and committed fouls, passes completed and uncompleted; among other technical elements that are assessed according to the team's technical committee needs. Ramos Filho and Oliveira Júnior[5] point out that even though extremely complex and with numerous random variables football can present

some regularity. Such regularity, in turn, allows the coach to setup a type of game.

Football game diversity of strategies and high number of unpredictable actions leads to a situation where the most important variables to be assessed remain unidentified. Being able to recognize such variables would allow the coaches understanding the reasons leading to success or failure in a match[6]. Previous studies tried to model the football game in order to predict less complex final outcomes using different dimensionality reduction techniques such as principal component analysis[7], logistic regression[8-9], discriminant analysis[10-12], means comparison tests[13]. An assumption in these approaches is that scout variables can determine winning or loosing. Given the game complexity, it seems a very optimistic assumption, since it disregards the influence of the scout variables on each other.

Another option to reduce the dimension of the football data is to perform a multiple regression analysis. This method allows establishing quantitative relationships between variables. In this way, it is possible to describe technical and tactical characteristics with a reduced number of scout variables and determine how these variables affects each other without the requirement of winning or losing association.

Thus, this study's aim is to determine the scout variables that best describe characteristics of the teams participated in the 2013/2014 Champions League.

Method

The present study is a descriptive analysis of all the 2013/2014 European Champions League's scout quantitative data. Eleven variables of the 32 teams in 126 matches (qualifying phase and play-off round) were analyzed (TABLE 1):

TABLE 1 - Scout variables description.

Goals Scored (GS)	Total goals scored;
Goals Difference (GD)	Difference between goals scored and suffered by the team;
Total Attempts (TA)	Total attempts to score a goal, either on or off target;
Attempts on Target (AT)	Total attempts to goal direction;
Disarms (DI)	Number of actions of gaining possession from an opponent player who is in possession of the ball;
Corners (CO)	Number of corner kicks performed;
Offside (OF)	Number of times a team has been caught in an offside position resulting in a free kick by the opposing team;
Passes Completed (PC)	Number of successful and intentional played balls from one player to another;
Passes Uncompleted (PU)	Number of intentional but uncompleted played balls from one player to another;
Fouls Committed (FC)	Number of infringementspenalized as foul by a referee;
Ball Possession (BP):	Duration when a team takes over the ball from the opposing team without any clear interruption, expressed as a proportion of total duration when the ball was in play.

Public domain data were obtained at the UEFA Champions League official website[14] and FootStats[15]. Best subsets multiple regression analysis was employed to verify the relationship between all scout variables. All variables were therefore, one by one, used as a dependent variable while all others as independent variables, for the n independent associated by the one dependent variable were calculated $n^*(n\text{-}1)!$ partial models with order 1 to n. The best partial model for each order was used to compound the output analysis with n partial models. The best partial model for each variable was accepted as the one that presented the larger number of variables in which all coefficients were found to have a statistical significance. According to this criteria, only variables that were influenced by more than three other variables were reported. A 5% probability significance level was adopted. All statistical procedures were performed in SigmaStat (version 3.5).

Results

Total attempts (TA), ball possession (BP), passes completed (PC) and goals scored (GS) showed significant regression models ($p < 0.05$). In TABLE 2 are shown the regression analysis results.

TABLE 2 - Best subset multiple regression analysis results.

Variable		Coefficient	Standard error	t	p
Total attempts ($r^2 = 0.815$)	Constant	1.087	1.043	1.043	0.298
	Goals scored	-0.344	0.136	-2.529	0.012
	Attempts on target	1.214	0.060	20.205	< 0.001
	Corners	0.265	0.062	4.267	< 0.001
	Fouls committed	-0.090	0.037	-2.420	0.016
	Passes completed	-0.006	0.002	-2.844	0.005
	Ball possession	0.108	0.030	3.586	< 0.001
Ball possession ($r^2 = 0.748$)	Constant	14.297	2.900	4.931	< 0.001
	Goals scored	-0.772	0.375	-2.056	0.041
	Goals difference	0.644	0.227	2.834	0.005
	Total attempts	0.380	0.063	5.993	< 0.001
	Disarm	-0.113	0.054	-2.102	0.037
	Fouls committed	0.216	0.074	2.931	0.004
	Passes completed	0.052	0.003	19.946	< 0.001
	Passes uncompleted	0.057	0.014	4.036	< 0.001
Passes completed ($r^2 = 0.742$)	Constant	78.513	40.588	1.934	0.054
	Total attempts	-5.291	1.763	-3.002	0.003
	Attempts on target	7.239	2.616	2.768	0.006
	Fouls committed	-5.720	1.061	-5.393	< 0.001
	Passes uncompleted	-0.930	0.205	-4.540	< 0.001
	Ball possession	11.689	0.575	20.326	< 0.001
Goals scored ($r^2 = 0.699$)	Constant	2.182	0.367	5.949	< 0.001
	Goals difference	0.413	0.025	16.605	< 0.001
	Total attempts	-0.053	0.020	-2.658	0.008
	Attempts on target	0.158	0.030	5.341	< 0.001
	Corners	-0.052	0.020	-2.596	0.010
	Fouls committed	-0.023	0.011	-1.987	0.048
	Passes uncompleted	-0.005	0.002	-2.092	0.037

Coefficient: Values of variables coefficients;
Error: 95% error associated to the variable coefficient;
t: T-value of t-student test;
p: probability value.

Total attempts regression equation (Eq. 1) was significantly influenced by goals scored (GS), attempts on target (AT), corners(CO), fouls committed (FC), passes completed (PC) and ball possession (BP); with a high determination coefficient ($r^2 = 0.815$).

$$TA = 1.087 - 0.344GS + 1.214AT + 0.265CO - 0.090FC - 0.006PC + 0.108BP \quad \text{Eq.1}$$

Ball possession regression equation (Eq. 2) was significantly influenced by goals scored (GS), goals difference (GD), total attempts (TA), disarm (DI), fouls committed (FC), passes completed (PC) and passes uncompleted (PU); with a high determination coefficient ($r^2 = 0.748$).

$$BP = 14,297 - 0,772GS + 0,644GD + 0,380TA - 0,113DI + 0,216FC + 0,052PC + 0,057PU \quad \text{Eq.2}$$

Passes completed regression equation (Eq. 3) was significantly influenced by total attempts (TA), attempts on target (AT), fouls committed (FC), passes uncompleted (PU) and ball possession (BP); with a high determination coefficient ($r^2 = 0.742$).

$$PC = 78,513 - 5,291TA + 7,239AT - 5,720FC - 0,930PU + 11,689BP \quad \text{Eq.3}$$

Goals scored regression equation (Eq. 4) was significantly influenced by goals difference (GD), total attempts (TA), attempts on target (AT), corners (CO), fouls committed (FC) and passes uncompleted (PU); with a high determination coefficient ($r^2 = 0.699$).

$$GS = 2,182 + 0,413GD - 0,053TA + 0,158AT - 0,052CO - 0,023FC - 0,005PU \quad \text{Eq.4}$$

Discussion

Our study aimed to identify scout variables that best describe the technique and tactical characteristics of the 2013/2014 Champions League teams. Total attempts, ball possession, passes completed and goals scored were pointed as the key variables.

Total attempts, sum of the attempts on target and attempts off target, was found to have a direct relationship with attempts on target, ball possession and corners; and to be inversely related with passes completed, fouls committed and goals scored. Greater ball possession allows more offensive actions enabling a higher number of attempts. Corners and total attempts directly proportional relationship seems to be due to the fact that in a dead ball situation, such as in a corner kick, there are many players next to goal, increasing the possibilities of an attempt.

Hughes and Franks[16] found the most part of total attempts happened after few passes, in agreement with the inverse relationship between these variables in the model. In this work, when the number of passes before the shot are normalized by the frequency of occurrence of different lengths of passing sequences, the more longer passing sequences produces more shots, in this way, retain the ball possession produces more attempts on target, in agreement with the relationship between this variables in the model.

Perin[17] showed that the team with greater number of total attempts won 51.61% of the times. Similarly, teams with greater attempts on target won 45.16% of the times. Accordingly to Tempone and Silva[18], based in 2010 FIFA World Cup, the victorious teams are those who have more total attempts. Armatas et al.[19] also found in the Greek Soccer First League that top teams made more attempts than bottom teams, and Szwarc[20], after examined 2002 World Cup, showed similar results and concluded that finalist teams made more attempts than unsuccessful teams.

Lago-Peñas et al.[12] analyzed the UEFA Champions League of the season 2007-2008, 2008-2009 and 2009-2010, the results showed that winning teams had significantly higher average values of total attempts, attempts on target, effectiveness, total passes, completed passes, and ball possession. Losing teams had significantly higher values in the variable yellow cards and red cards.

Spanish Men's Professional League 2008-2009 season showed that winning teams had averages higher for the total attempts, attempts on goal, effectiveness, assists, offsides committed and crosses against, while the losing teams had higher averages in the variable crosses, offsides received and red cards[11].

Castellano et al.[10] identified the total shots, shots on target and ball possession as the variables that distinguish the winner and losers teams.

Ball possession was found to have a positive relationship with goals difference, total attempts, fouls committed, passes completed and passes uncompleted; and a negative relationship with disarm and goals scored. When a team is winning, seek another goal is a more carefully planned action which leads to a higher ball possession. In addition, passes performed effectively also increases ball possession. Hughes and Franks[16] noted that teams with fast offensive actions, i.e. in which ball possession is not the main objective, had more attempts on target. Liu et al.[21] identified that there was a positive effect of Shot from Counter Attack on the probability of winning. These results are in agreement with previous findings on Norwegian football league[8-9] and Spanish Professional Football League[22], which suggested that counter-attacks were more effective in generating goals. This idea could explain the inverse relationship between ball possession and scored goal found in the present study. Jones et al.[23] found that the first places in the 2001/2002 English Premier League were held

by teams with higher ball possession. There is no evidence in relate literature to explain the relation of total attempts, disarms, fouls committed and passes uncompleted with ball possession.

Passes completed was found to have a positive relationship with attempts on target and ball possession; and a negative relationship with total attempts, fouls committed and passes uncompleted. Winning teams presents better values of passes completed, attempts on target and ball possession than losing teams[11, 19, 21]. Passes completed is affected by attempts on target most likely due to the teams' high accuracy. Similarly, low accuracy would raise total attempts number, explaining his inversely relationship with this variable. According to Leitão[24], pass represents 80% of all actions in a match and quality pass ensures ball possession. Likewise, Silva Junior[3] suggests that the pass is essential for a positive outcome in a match since a good pass could allow more offensive plays and better tactical options. In an inferior technical ability team the number of uncompleted passes is superior which as, as expected, a negative influence in the number of completed passes. Indeed, Silva Junior[3] states that a deficiency in this technique affects the team's attack because decreases the offensive chances due to ball possession loss.

Goals scored was found to have a positive relationship with goals difference and attempts on target; and a negative relationship with total attempts, corners, passes uncompleted and fouls committed. Total attempts showed an inverse relationship with goals scored whereas attempts on target showed a positive one. In other words, an accurate attempt is more important than the total number of attempts which in turn is related with higher goals balance, as described also by Yue et al.[13].

Equation 3 showed that fewer uncompleted passes leads to more completed passes. Therefore, higher pass efficiency increases the possibility of a higher number of attempts on target which could increase goals scored. Corroborating with these results, Silva Junior[3] showed that completed passes that resulted in goal are crucial to increase goals balance; and that the total number of attempts was not a goal score determinant.

Equation 1 showed that a play starting with a corner kick allows increasing the number of total attempts, however, equation 4 showed an inverse relationship between corners and scored goals. Therefore, the increased number of total attempts originated by a corner kick did not resulted in goal. This inefficiency can be explained by the dead ball time that allows the defensive repositioning, reducing the goal score's chance. Indeed, Ramos Filho and Oliveira Júnior[5] showed that 31% of goals scored in 2004 Euro Cup was a result of a dead ball situation (only 10% of these goals were as a result of a corner kick).

Our study also showed a negative relationship between fouls committed and scored goals. Hughes and Franks[25] found that in the 1990 World Cup 41% of the goals were scored in a dead ball situation. Thus, committing fouls seems to be a counterproductive strategy that does not allow scoring, increasing the vulnerability to take a goal.

A limitation of this study could be the data source (i.e. public domain), nevertheless it is available information source for Coaches. Thus, the possible data bias study is the same bias found by coaches using this information.

In conclusion, total attempts, ball possession, passes completed and scored goals are the variables that best describe the relationships between the different 2013/2014 Champions League scout variables. Moreover, attempts on target seem to be an even more success determinant than the total number of attempts. It was possible to show that plays originated in a corner kick were not a good strategy to score. Finally, we were able to show that pass accuracy does not allow the opponents' ball possession. Our study showed that scout information can be reduced to the coach easily application.

Resumo

Modelos de análise de scout no futebol (baseado na temporada 2013/2014 da Liga dos Campeões)

Conquistar as primeiras posições em competições de futebol é apenas para alguns times. Assim, um melhor entendimento do jogo se torna crucial mediante essa competitividade. O objetivo do presente estudo foi determinar variáveis de scout que melhor descrevem as características técnicas e táticas dos times que participaram de liga dos campeões na temporada 2013/2014. Todos os jogos acessíveis das fases de classificação e das finais da liga dos campeões da UEFA (32 times e 126 partidas) foram analisados usando o scout dos dados disponíveis para domínio publico. As variáveis analisadas foram gols marcados, saldo de gols, finalizações totais, finalizações a gol, desarmes, escanteios, tiros de meta, passes certos, passes errados, número de faltas cometidas e posse de bola. As variáveis mais relevantes para descrever uma partida, apontadas pela analise de regressão múltipla foram as finalizações totais (r^2=0.815), posse de bola (r^2=0.748), passes certos (r^2=0.742) e gols marcados (r^2=0.699). O presente estudo foi capaz de mostrar como as informações obtidas pelo scout podem ser agrupadas para fácil aplicação pelos treinadores.

Palavras-chave: Estatística; Futebol; UEFA; Competições.

References

1. Vendite C, Vendite LL, Moraes AC. Scout no futebol: uma ferramenta para a imprensa esportiva. XXVIII Congresso Brasileiro de Ciências da Comunicação. Rio de Janeiro: UERJ; 2005. p.1.
2. Bottaro LEV. Análise de scout em partidas de futebol: finalizações da equipe do Cruzeiro Esporte Clube nos jogos da fase de grupos da Taça Libertadores da América do ano de 2009 [monografia]. Belo Horizonte (MG): Universidade Federal de Minas Gerais, Escola de Educação Física, Fisioterapia e Terapia Ocupacional; 2009.
3. Silva Júnior MSL. Estudo descritivo sobre o desempenho do passe da Seleção Brasileira de futebol na Copa das Confederações de futebol 2009 [monografia]. Belo Horizonte (MG): Universidade Federal de Minas Gerais, Escola de Educação Física, Fisioterapia e Terapia Ocupacional; 2009.
4. França Neto CP. Análise do scout em partidas de futebol: fundamentos técnicos individuais da equipe profissional de futebol do Cruzeiro esporte clube no primeiro turno do campeonato brasileiro de 2010 [monografia]. Belo Horizonte (MG): Universidade Federal de Minas Gerais, Escola de Educação Física, Fisioterapia e Terapia Ocupacional; 2009.
5. Ramos Filho LAO, Oliveira Junior MH. Futebol: classificação e análise dos gols da Eurocopa 2004. Rev Bras Futebol. 2008;1:42-8.
6. Marques AT. Bases para a estruturação de um modelo de detecção de talentos desportivos em Portugal. Rev Espaço. 1993;1:47-55.
7. Moura FA, Martins LEB, Cunha AS. Analysis of football game-related statistics using multivariate techniques. J Sports Sci. 2014;32:1881-7.
8. Tenga A, Holme I, Ronglan, L, Bahr, R. Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. J Sports Sci. 2010;28:245-55.
9. Tenga A, Ronglan L, Bahr R. Measuring the effectiveness of offensive match-play in professional soccer. Euro J Sport Sci. 2010;10:269-77.
10. Castellano J, Casamichana D, Lago C. The use of match statistics that discriminate between successful and unsuccessful soccer teams. J Hum Kinet. 2012;31:139-47.
11. Lago-Peñas C, Lago-Ballesteros J, Dellal A, Gómez M. Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. J Sports Sci Med. 2010:9:288-93.
12. Lago-Peñas C, Lago-Ballesteros J, Rey E. Differences in performance indicators between winning and losing teams in the UEFA Champions League. J Hum Kinet. 2011;27:135-46.
13. Yue Z, Broich H, Mester J. Statistical analysis for the soccer matches of the first Bundesliga. Int J Sports Sci Coach. 2014;9:553-60.
14. UEFA Champions League [internet homepage]. Season 2014 [cited 2015 Sept. 28]. Available from: http://www.uefa.

com/uefachampionsleague/season=2014.
15. Footstat [internet homepage]. UEFA Champions League 2014 [cited 2015 Sept. 28]. Available from: www.footstats.net.
16. Hughes M, Franks IM. Analysis of passing sequences, shots and goals in soccer. J Sports Sci. 2005;23:509-14.
17. Perin DEB. Análises das finalizações e posse de bola em relação ao resultado do jogo de futebol [monografia]. Porto Alegre (RS): Universidade Federal do Rio Grande do Sul, Escola de Educação Física; 2012.
18. Tempone GMT, Silva CD. Análise de indicadores quantitativos de vitórias e derrotas na Copa do Mundo FIFA 2010. Rev Bras Futebol. 2012;5:42-6.
19. Armatas V, Yannakos A, Zaggelidis G, Skoufas D, Papadopoulou S, Fragkos N. Differences in offensive actions between top and last teams in Greek first soccer division. J Phys Educ Sport. 2009;23:1-5.
20. Szwarc A. Effectiveness of Brazilian and German teams and the teams defeated by them during the 17th Fifa World Cup. Kinesiology, 2004;36:83-9.
21. Liu H, Gomez MA, Lago-Peñas C, Sampaio J. Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. J Sports Sci . 2015;33:1205-13.
22. Lago-Ballesteros J, Lago-Peñas C, Rey E. The effect of playing tactics and situational variables on achieving scorebox possessions in a professional soccer team. J Sports Sci. 2012;30:1455-61.
23. Jones PD, James N, Mellalieu SD. Possesion as a performance indicator in soccer. Int J Perf Anal Sport. 2004;4:98-102.
24. Leitão RAA. Futebol: análise qualitativas e quantitativas para verificação e modulação de padrões e sistemas complexos de jogo [dissertação]. Universidade Estadual de Campinas, Escola de Educação Física; 2004.
25. Hughes M, Franks IM. Notational analysis of sport. 2nd edition. New York: Routledge; 2004.

ADDRESS
Juliana Pennone
Laboratório de Biomecânica
Escola de Educação Física e Esporte - USP
Av. Prof. Mello Moraes, 65
05508-030 - São Paulo - SP - BRASIL
e-mail: juliana.pennone@usp.br

Submitted: 15/04/2015
1st. Review: 18/06/2015
2nd. Review: 30/09/2015
Accepted: 09/10/2015