



The Dirty League: English Premier League Provides Higher Incentives for Fouling as Compared to other European Soccer Leagues

by

Ashwin A. Phatak¹, Robert Rein¹, Daniel Memmert¹

Fouling in soccer has been studied from an ethical standpoint as a measure of aggression. However, there is limited research related to fouling for performance. The present study investigated fouling as a factor influencing performance in European soccer leagues. Out of possession fouls (FPGNorm), yellow cards (YCFNorm), and their ratio (YCPFPG) were used as predictors of points (Pts) and goals conceded (GA) at the end of the season using three separate linear regression models. Furthermore, 5-fold cross-validation was used to measure out sample reliability. All the models significantly predicted GA and Pts ($p < 0.001$). Models predicting GA showed higher reliability than models predicting points. Cross validation (CV) results suggested that FPGNorm and YCPFPG models showed a small standard deviation (SD) in the R^2 results whereas the results from YCFNorm were not reliable to high SD in the 5-fold CV results. In summary, FPGNorm and YCPFPG seem to predict success (low GA and high Pts) across European soccer leagues, with EPL showing the maximum effect. The findings of the current study and the methodology can be applied to an actual game analysis by coaches in multiple invasion sports. Normalizing for out of possession time is a crucial step for the time spent in particular phases of play, which has not been done in previous research while analyzing 'key performance indices' (KPIs). Normalization can successfully introduce domain-specific knowledge into predictors, which can be used in complex algorithms improving predictions and investigation of underlying mechanisms.

Key words: fouls, normalization, performance analysis, soccer, English Premier League, aggressive behavior.

Introduction

'Ball or the man' is a term commonly associated with rough tackling teams in contemporary soccer. The phrase suggests that any strategy (legal or illegal), which stops an attacker is justified to ensure game success. Fouls and cards (yellow and red) are a measure of evaluating the fairness of play in soccer. They have been examined from an ethical standpoint as methods for assessing aggression in soccer players (Gümüřdağ et al., 2011; Sapp et al., 2018; Tanamati Soares, 2016). A study on 130 unsportsmanlike fouls in men's and women's basketball games from Olympics, European and World championships showed an overall advantage in the outcome of the game to the team with more fouls (Gómez et al., 2016). Few studies have investigated the influence

of fouls on game success in soccer. The majority of studies focus on correlating attacking KPIs with performance. A study that analyzed 177 FIFA world cup games from 2002 to 2010 is one of the few studies which examined defensive KPIs. The results suggested a significant influence of the number of fouls committed, yellow cards, and red cards on the teams' success (Castellano et al., 2012). There seems to be few studies focusing on defensive KPIs (Sarmiento et al., 2014). Hence there is a need to analyze defensive KPIs within game specific-scenarios.

Five full seasons in the Bundesliga were analyzed, and it was found that the probability of a win increased with increasing numbers of fouls (Anders and Rotthoff, 2011). In contrast, yellow

¹ - German Sports University, Institute of Exercise Training and Sport Informatics.

cards and red cards lowered the winning probability. Another study based on 17 matches in the Turkish Super League concluded that only 1.2 percent of all fouls were unintentional, indicating that players intentionally commit the vast majority of the fouls (Gümüřdađ et al., 2011). Based on a limited sample size, this finding warrants further investigation of fouling behavior as a tactical strategy and its relationship with game success in professional soccer leagues. Furthermore, research analyzed when a 'professional foul' (tactical foul) should be committed based on hypothetical game scenarios (M. Wright and Hirotsu, 2003). That study extends previous findings on red cards' effect by accounting for team strengths based on the end-of-season standings (Ridder et al., 1994). The study also combines the impact on the optimal timing of tactical decisions using the Markov chain prediction model (Hirotsu and Wright, 2002). Furthermore, the authors recommend rule changes in soccer based on the fouling team's advantage as a function of the foul's location. The same study also accounted for the strength of the opposition team, the foul's timing, the punishment awarded to the fouling player, and its implications for the fouling team affecting the change in goal probability. The authors conclude that stronger teams seem to benefit more from professional fouling. The penalty for a professional foul is not always fair when accounted for the probability of scoring a goal. Considering the above studies, there seems to be a gap for analyzing fouls as a key performance indicator (KPI) for successful gameplay in high-performance soccer.

Fouls and yellow cards indicate aggressive behavior in European soccer leagues (Sapp et al., 2018). The results suggested that the English Premier League (EPL) referees called the least number of fouls compared to the other four leagues. The number of tackles and yellow cards received had strong correlations with the final table position across all leagues. They further concluded that, although European soccer has become less aggressive over the years, the EPL still seems to be the most aggressive of the investigated leagues, especially away from home (Thomas et al., 2006). Variance in referee strictness was observed across different stages of the game in the German league. The same study elaborated the referee's calibration process in calling a foul or giving a card (Unkelbach and Memmert, 2008).

Assuming that this phenomenon may exist in different leagues to a varying degree is justified. Hence, it is crucial to examine and compare the effect of fouling behavior and its impact on another leagues' success.

Possession in soccer has been extensively studied as an indicator of performance and team strength. Retention of possession for a prolonged period has been linked to game success in multiple studies (Camerino et al., 2012; Casal et al., 2017; Jones et al., 2004; Lago-Ballesteros and Lago-Peñas, 2010). A systematic review suggested that the best-ranked teams maintained higher-level possession, home teams had more control than away teams on average, and differences in team strength translated to differences in ball possession with better teams having more possession (Sarmiento et al., 2014). An exception to this trend was seen in the case where teams were shown to have more possession while trailing (losing) as compared to when they were leading. Most studies treat possession as an offensive KPI, but the amount of possession for a team also depicts how much of the playing time was spent in attacking versus in defending phases (Casal et al., 2017; Sarmiento et al., 2014).

It also predicts the strength of the teams (Lago and Martín, 2007). It is a crucial factor that needs to be considered while evaluating a team's attacking and defensive performance. In the present study, we propose that possession should be neither an attacking nor a defensive KPI. It should be treated as a category of its own, while defensive and attacking KPIs should be investigated after factoring in the time duration in ball possession. There is a need for normalizing technical and tactical data for accurately accounting for the frequency of game specific actions (Hughes and Bartlett, 2002). Hence, the present study proposes a normalization method to investigate and accurately report teams' fouling behavior without ball possession. It better depicts defensive and offensive KPIs and their relationship to success in soccer.

There seems to be few studies that account for average in possession and out of the possession time as a factor when investigating both defensive and offensive KPIs. Possession seems to be a KPI that is descriptive of the phase of play and comparative team strength (Pratas et al., 2018; Sarmiento et al., 2014). Thus, the present study

addresses several methodological issues presented by multiple studies (Bergkamp et al., 2019; James, 2006; Mackenzie and Cushion, 2013;).

There seems to be limited research focused on fouling as a tactical strategy for success in soccer. Intending to fill this gap, the present study investigates fouls out of possession (FPGNorm), Yellow cards out of possession YCFNorm, and the average amount of fouls required to get a yellow card (YCPFPG) and their relationship to success in European soccer leagues. Success is defined as the end of season points (Pts) and total end of season goals conceded (GA). Furthermore, the study compares how this effect potentially varies across the 5 European top-flight leagues vs. English Premier League (EPL), Spanish LA Liga (LaLiga), German Bundesliga (BL), Serie A (SA), and the French league One (FL). Following is the list of hypotheses that the study shall investigate:

- (H1) The current study predicts a positive correlation of both FPGNorm and YCFNorm with the end of season points (Pts) and a negative correlation with the number of goals conceded (GA).
- (H2) The effects of fouling on performance will be significantly different in the EPL as compared to the other leagues.
- (H3) YCPFPG will show a negative correlation with the end of season points and positive correlation with GA, with the EPL showing significantly different effects as compared to the other leagues.

Methods

Participants

End of the season statistics from the EPL, SA, BL, LaLiga, and FL were obtained for the seasons from 2009-10 to 2018-19. Notational statistics for fouls per game (FPG), fouls which received yellow cards per game (YCF), average possession for the whole season (Percent Possession), and final table performance stats were obtained from (Whoscored.com, 2019) and defined in accordance with the definitions of the data source. Permission was obtained from whoscored.com for publishing the results of the present study under the creative commons open-source license.

Measures: Normalization

In soccer, fouls are mostly conducted by the defensive team, which is out of possession as

possession rates vary widely between teams and games (Sarmiento et al., 2014). FPG and YCF should be normalized according to possession to account for time spent out of possession. Thus, the following normalization equations were used, yielding a normalized number of fouls per game (FPGNorm) and a normalized number of yellow cards per game (YCFNorm):

$$FPGNorm = FPG / (1 - PercentPossession/100) \quad (1)$$

$$YCFNorm = YCF(\text{for fouls}) / (1 - PercentPossession/100) \quad (2)$$

Furthermore, the ratio YCPFPG between YCF and FPG was used to assess how many fouls per game on average were required to receive a yellow card:

$$YCPFPG = YCF / FPG \quad (3)$$

Design and Procedure

To model season and league effects, post hoc regression analysis was conducted and both variables were included in the study to account for mixed results. As FPGNorm and YCFNorm were found to be positively correlated with each other, two separate models were formed. Thus, YCPFPG was used as a performance indicator which could potentially indicate how fouls and yellow cards influence performance in combination.

Statistical Analysis: Multiple Linear Regression

Multiple linear regression was conducted separately for both Points and GA as dependent variables to indicate performance over the season. The independent variables were FPGNorm, YCFNorm, and YCPFPG, season, and league. The variations in baseline levels and the interaction effect of variables on each other were accounted for using three separate models for FPGNorm, YCFNorm, and YCPFPG. In each of the models, an interaction effect of the investigated independent variables with the league was introduced to explore league wise differences. Hence, the models used for investigation were as follows:

$$Pts \text{ or } GA = Season + (FPGNorm * League) \quad (4)$$

$$Pts \text{ or } GA = Season + (YCFNorm * League) \quad (5)$$

$$Pts \text{ or } GA = Season + (YCPFPG * League) \quad (6)$$

The above models were further tested for out of sample validity using K-Fold cross-validation (K = 5). All the analysis was conducted using Excel 2016 and custom routines written in Python 3.7.1 and R 3.6.2. The generalized linear

model routines from the H2O library were used to build the models and conduct k-fold cross-validation.

Results

The initial post hoc regression analysis results showed a change in the amount of FPGNorm and YCFNorm across leagues as the seasons progressed. The YCFNorm increased, and the FPGNorm seemed to decrease with each consecutive season. A correlation between YCFNorm and FPGNorm was also observed. Although there was a difference in the base rates (intercepts) of each league's fouling behavior, the correlation of YCFNorm and FPGNorm was consistently positive. It increased with each consecutive season across all leagues.

Regression Results

Table 1a below shows the regression results predicting the end of season points using FPGNorm, season, and league as independent variables. The results show significance while indicating points ($p < 0.001$, $Adj. R^2 = 0.072$). Furthermore, EPL teams show significantly lower average points total as compared to all the other leagues ($p < 0.001$). A unit rise in the FPGNorm showed an increase of 0.63 points across all leagues ($p = 0.055$), and for every unit increase in the season, there was a 0.83 rise in the average points total ($p < 0.001$) across leagues. FPGNorm shows a significant effect for predicting points in the EPL compared to the other leagues with a rise of 2.10 points per unit increase in FPGNorm ($p < 0.001$).

Table 1b below shows the regression results to predict the total end of season goals conceded (GA) using FPGNorm, season, and league as independent variables. The model significantly predicts GA ($p < 0.001$, $Adj. R^2 = 0.069$). A decrease of 0.81 units in GA was seen per unit reduction of FPGNorm ($p = 0.001$), and a decrease of 0.42 GA was observed for each consecutive season across leagues ($p = 0.007$). FPGNorm shows a significant effect for predicting GA in the EPL compared to the other leagues with a reduction of 0.99 GA per unit increase in FPGNorm ($p = 0.009$).

Figure 1a below shows the simple correlation between FPGNorm and Pts for all 5 leagues, while Figure 1b shows the relationship between FPGNorm and GA visualizing the results shown in Table 1,

Table 2a below shows the regression results predicting the end-of-season points with YCFNorm, season, and league as independent variables. The model significantly predicts points ($p < 0.001$, $Adj. R^2 = 0.032$). YCFNorm showed a significant effect for predicting points in the EPL ($p < 0.001$) compared to the other leagues, with an increase of 0.26 points per unit YCFNorm.

Table 2b shows the results of a regression performed to predict GA using YCFNorm, season, and league as independent variables. The model significantly predicts GA ($p < 0.001$, $Adj. R^2 = 0.030$). The EPL shows significantly higher GA as compared to all the other leagues ($p = 0.009$). YCFNorm shows a significant effect for predicting GA in the EPL compared to the other leagues with a reduction of 0.15 points per unit YCFNorm ($p = 0.015$).

Figure 2a represents the relationship between YCFNorm and Pts for all 5 leagues, while Figure 2b shows the relationship between YCFNorm and GA visualizing the differences in the leagues presented in Table 2.

Table 3a shows the regression results predicting the end of season points with YCPFPG, season, and league as independent variables. The model significantly predicts points ($p < 0.001$, $Adj. R^2 = 0.075$). A unit rise in YCPFPG showed a significant decrease of 6.07 points ($p < 0.001$), and a significant increase of 0.86 points was observed every consecutive season ($p < 0.001$). The BL shows the strongest effect of YCPFPG ($p = 0.045$) while predicting points, with a decrease of 5.21 points per unit YCPFPG.

Table 3b shows the results of regression predicting GA using YCPFPG, season, and league as independent variables. The model significantly predicts GA ($p < 0.001$, $Adj. R^2 = 0.075$). GA shows a significant rise of 5.13 units per unit YCPFPG in all leagues ($p = 0.003$), and in every consecutive season there was a fall of 0.46 GA across leagues (p

Table 1a*Pts vs. FPGNorm*

Names	Estimate	Std. Error	t-value	p
(Intercept)	21.691	11.168	1.942	0.0524
Season	0.836	0.211	3.956	< 0.001
FPGNorm	0.634	0.330	1.918	0.055
LeagueEPL	-42.541	13.124	-3.241	< 0.001
LeagueLaLiga	-9.272	14.091	-0.658	0.510
LeagueBL	-13.885	13.143	-1.056	0.291
LeagueFL	-19.605	12.628	-1.552	0.121
FPGNorm: LeagueEPL	2.096	0.506	4.14	<0.001
FPGNorm: LeagueLaLiga	0.372	0.481	0.776	0.438
FPGNorm: LeagueBL	0.283	0.434	0.652	0.515
FPGNorm: LeagueFL	0.713	0.429	1.665	0.096

Residual standard error: 15.95 on 969 degrees of freedom

Multiple R-squared: 0.08163, Adjusted R-squared: 0.07216

F-statistic: 8.613 on 10 and 969 DF, p: 1.379e-13

Table 1b*GA vs. FPGNorm*

Names	Estimate	Std. Error	t-value	p
(Intercept)	80.495	8.395	9.589	< 0.001
Season	-0.427	0.159	-2.689	0.007
FPGNorm	-0.809	0.249	-3.252	0.001
LeagueEPL	18.230	9.865	1.848	0.065
LeagueLaLiga	-14.523	10.591	-1.371	0.171
LeagueBL	-3.545	9.879	-0.359	0.720
LeagueFL	2.974	9.492	0.313	0.754
FPGNorm: LeagueEPL	-0.989	0.381	-2.599	0.0094
FPGNorm: LeagueLaLiga	0.527	0.361	1.460	0.145
FPGNorm: LeagueBL	0.091	0.326	0.280	0.779
FPGNorm: LeagueFL	-0.242	0.322	-0.751	0.453

Residual standard error: 11.99 on 969 degrees of freedom

Multiple R-squared: 0.07897, Adjusted R-squared: 0.06947

F-statistic: 8.309 on 10 and 969 DF, p: 4.922e-13

Note 1: GA = Total goals conceded at the end of the season, Pts = End of season points total, FPGNorm = Fouls Per game normalized for out of possession (See normalization sections in Methods), BL = German Bundesliga, FL = French League One, LaLiga = Spanish Soccer league, EPL = English Premier League.

Table 2a

<i>Pts vs. YCFNorm</i>				
Names	Estimate	Std. Error	t-value	<i>p</i>
(Intercept)	64.382	7.255	8.874	< 0.001
Season	0.082	0.183	0.450	0.653
YCFNorm	-0.102	0.052	-1.949	0.052
LeagueEPL	-28.327	9.588	-2.954	0.003
LeagueLaLiga	9.214	10.326	0.892	0.372
LeagueBL	4.417	10.603	0.417	0.677
League	-18.241	10.104	-1.805	0.071
YCFNorm: LeagueEPL	0.262	0.086	3.059	0.002
YCFNorm: LeagueLaLiga	-0.053	0.074	-0.719	0.472
YCFNorm: LeagueBL	-0.136	0.096	-1.415	0.157
YCFNorm: LeagueFL	0.144	0.085	1.697	0.090

Residual standard error: 16.29 on 969 degrees of freedom,
Multiple R-squared: 0.0425, Adjusted R-squared: 0.03262,
F-statistic: 4.301 on 10 and 969 DF, *p*: 6.745e-06

Table 2b

<i>Pts vs. YCFNorm</i>				
Names	Estimate	Std. Error	T-value	<i>p</i>
(Intercept)	40.952	5.451	7.513	<0.001
Season	0.133	0.137	0.967	0.334
YCFNorm	0.059	0.039	1.500	0.134
LeagueEPL	18.695	7.203	2.595	0.009
LeagueLaLiga	-9.884	7.757	-1.274	0.203
LeagueBL	-2.791	7.966	-0.350	0.726
LeagueFL	1.228	7.591	0.162	0.872
YCFNorm: LeagueEPL	-0.156	0.064	-2.428	0.015
YCFNorm: LeagueLaLiga	0.076	0.055	1.366	0.172
YCFNorm: LeagueBL	0.040	0.072	0.550	0.583
YCFNorm: LeagueFL	-0.023	0.064	-0.362	0.717

Residual standard error: 12.24 on 969 degrees of freedom
Multiple R-squared: 0.04067, Adjusted R-squared: 0.03077
F-statistic: 4.108 on 10 and 969 DF, *p*: 1.439e-05

Note 2: GA = Total goals conceded at the end of the season, Pts = End of season points total, YCFNorm = Fouls which received a yellow card per game, normalized (See normalization sections in Methods), BL = German Bundesliga, FL = French League One, LaLiga = Spanish Soccer league, EPL = English Premier League.

Table 3a

<i>Pts vs. YCPFPG</i>				
Names	Estimate	Std. Error	T-value	p
(Intercept)	67.464	6.749	9.996	< 0.001
Season	0.866	0.206	4.207	<0.001
YCPFPG	-6.068	1.556	-3.900	<0.001
LeagueEPL	-8.642	9.615	-0.899	0.369
LeagueLaLiga	17.710	10.614	1.669	0.096
LeagueBL	4.290	9.744	0.440	0.660
LeagueFL	-4.397	9.418	-0.467	0.641
YCPFPG: LeagueEPL	1.715	2.186	0.784	0.433
YCPFPG: LeagueLaLiga	-2.737	2.192	-1.249	0.212
YCPFPG: LeagueBL	-5.208	2.600	-2.003	0.045
YCPFPG: LeagueFL	0.025	2.257	0.011	0.991

Residual standard error: 15.93 on 969 degrees of freedom
 Multiple R-squared: 0.0845, Adjusted R-squared: 0.07505
 F-statistic: 8.944 on 10 and 969 DF, p: 3.466e-14

Table 3b

<i>GA vs. YCPFPG</i>				
Names	Estimate	Std. Error	T-value	p
(Intercept)	33.818	5.063	6.680	<0.001
Season	-0.465	0.154	-3.011	0.003
YCPFPG	5.135	1.167	4.399	<0.001
LeagueEPL	11.639	7.213	1.614	0.107
LeagueLaLiga	-3.846	7.963	-0.483	0.629
LeagueBL	0.052	7.310	0.007	0.994
LeagueFL	-3.732	7.065	-0.528	0.597
YCPFPG: LeagueEPL	-2.017	1.640	-1.229	0.219
YCPFPG: LeagueLaLiga	0.513	1.644	0.312	0.755
YCPFPG: LeagueBL	1.616	1.951	0.828	0.408
YCPFPG: LeagueFL	1.168	1.693	0.690	0.491

Residual standard error: 11.95 on 969 degrees of freedom
 Multiple R-squared: 0.08537, Adjusted R-squared: 0.07593
 F-statistic: 9.044 on 10 and 969 DF, p: 2.28e-14

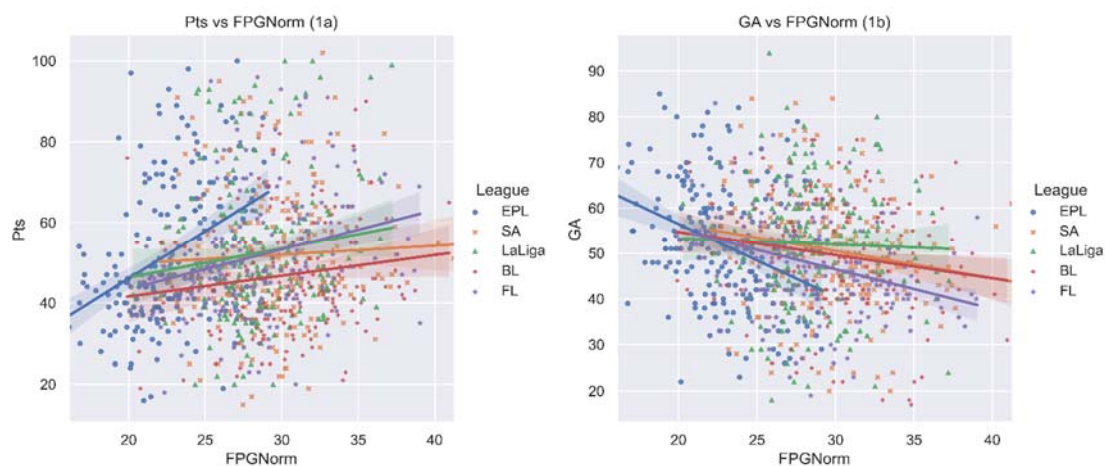
Note 3: GA = Total goals conceded at the end of the season, Pts = End of season points total,
 YCPFPG = Total number yellow card fouls per game divided by the total number of fouls per game
 (See normalization sections in Methods) BL = German Bundesliga, FL = French League One,
 LaLiga = Spanish Soccer league, EPL = English Premier League

Table 4

Variance results of 5-fold cross validation for models predicting Pts and GA

Model	R^2 Training	R^2 CV \pm SD
Pts = Season + FPGNorm*League	0.082	0.054 \pm 0.033
GA = Season + FPGNorm*League	0.078	0.056 \pm 0.011
Pts = Season + YCFNorm*League	0.042	0.012 \pm 0.013
GA = Season + YCFNorm*League	0.041	0.011 \pm 0.015
Pts = Season + YCPFPG*League	0.084	0.061 \pm 0.041
GA = Season + YCPFPG *League	0.085	0.062 \pm 0.043

Note 7: Pts = End of season points total, GA = Total goals conceded at the end of the season, YCPFPG = Total number yellow card fouls per game divided by the total number of fouls per game, YCFNorm = Fouls which received a yellow card per game, FPGNorm = Fouls per game normalized for out of possession (See normalization sections in Methods), R^2 Training = Variance in the training data, R^2 CV \pm SD = Variance in Cross Validation data with standard deviation.

**Figure 1**

Correlation between Out of possession fouls (FPGNorm) and End of Season Performance in terms of total goals conceded and total points accumulated for top 5 European soccer leagues

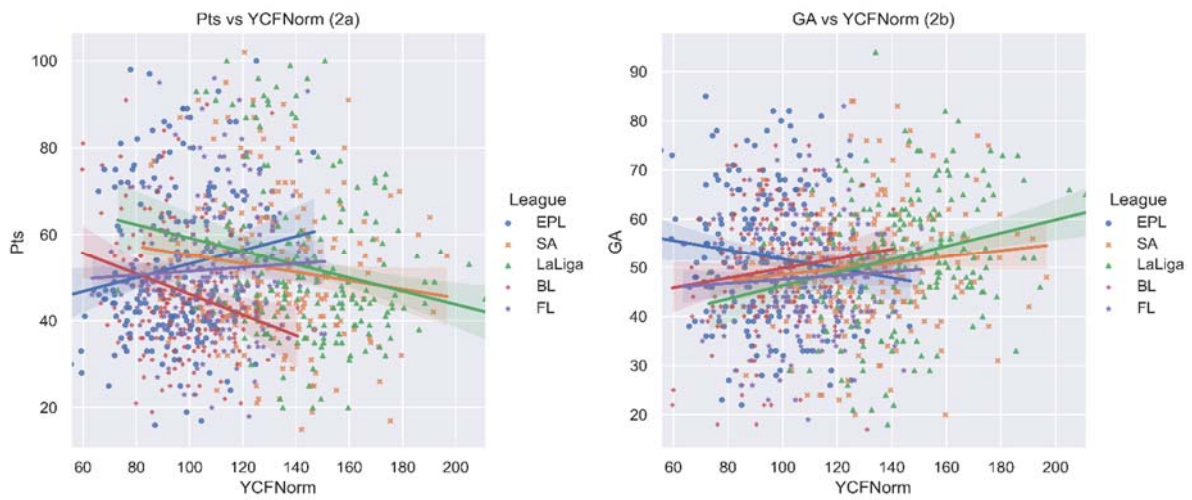


Figure 2
 Correlation between Normalized Yellow cards (YCFNorm) and End of Season Performance in terms of total goals conceded and total points accumulated for top 5 European soccer leagues

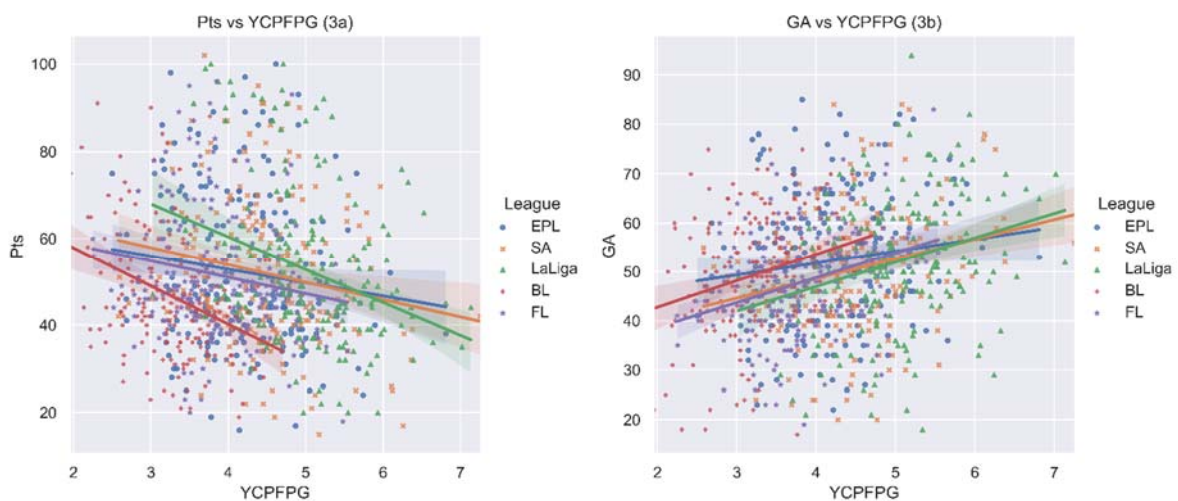


Figure 3
 Correlation between the ratio of yellow cards to fouls (YCPFPG) and End of Season Performance in terms of total goals conceded and total points accumulated for top 5 European soccer leagues

Appendix

Results for raw FPGNorm (FPG) data as an example

Table A and Table B show the results of Pts and GA predicted by the following equation:

$$\text{Pts or GA} = \text{Season} + (\text{FPG} * \text{League})$$

It is clear that the results in the coefficient column vastly differ when compared to Normalized results (FPGNorm) as seen in Table 2.

Table A: Pts vs. FPG

Names	Estimate	Std. Error	t value	p
(Intercept)	136.846203	9.73922	14.051	< 0.001 ***
Season	-1.230303	0.201237	-6.114	<0.001 ***
FPG	-4.58191	0.572257	-8.007	<0.001 ***
LeagueEPL	-12.077144	12.637097	-0.956	0.33947
LeagueLaLiga	-2.458947	11.997726	-0.205	0.83765
LeagueBL	-21.972394	11.519512	-1.907	0.05676
LeagueFL	-31.718143	11.873391	-2.671	0.00768 **
FPG:LeagueEPL	-0.38963	1.010742	-0.385	0.69996
FPG:LeagueLaLiga	-0.006783	0.817817	-0.008	0.99338
FPG:LeagueBL	1.166712	0.761644	1.532	0.12589
FPG:LeagueFL	1.985442	0.811966	2.445	0.01465 *

Residual standard error: 15.26 on 969 degrees of freedom

Multiple R-squared: 0.1599, Adjusted R-squared: 0.1512

F-statistic: 18.44 on 10 and 969 DF, p : < 2.2e-16

Table B: GA vs. FPG

Names	Estimate	Std. Error	t value	p
(Intercept)	9.21317	7.6367	1.206	0.2279
Season	0.77886	0.15779	4.936	<0.001 ***
FPG	2.06951	0.44872	4.612	<0.001 ***
LeagueEPL	4.76155	9.90898	0.481	0.631
LeagueLaLiga	0.63312	9.40764	0.067	0.9464
LeagueBL	-1.54081	9.03266	-0.171	0.8646
LeagueFL	16.58516	9.31015	1.781	0.0752
FPG:LeagueEPL	0.40009	0.79254	0.505	0.6138
FPG:LeagueLaLiga	0.16565	0.64127	0.258	0.7962
FPG:LeagueBL	0.01057	0.59722	0.018	0.9859
FPG:LeagueFL	-1.2575	0.63668	-1.975	0.0485 *

Residual standard error: 15.26 on 969 degrees of freedom

Multiple R-squared: 0.1599, Adjusted R-squared: 0.1512

F-statistic: 18.44 on 10 and 969 DF, p : < 2.2e-16

Figure 3a represents the correlation between YCFNorm and Pts for all 5 leagues while Figure 3b shows the correlation between YCFNorm and GA visualizing the results in Table 3.

Cross validation

Table 4 shows the cross-validation results for all six models. In all the cases below, the cross-validation R^2 scores are lower than that of the training data. In the case of YCFNorm predicting both Pts and GA accounting in the standard deviations, R^2 goes below zero. In the case of FPGNorm and YCPFPG, the standard deviations, although high, are still within the positive range.

Discussion

The present study investigates fouls as performance indicators in the European top-five soccer leagues. FPGNorm showed a significant positive correlation with points and a significant negative correlation with GA across all leagues. The EPL showed the maximum effect for both Pts and GA as compared to other leagues. YCFNorm showed a significant negative correlation with the end of season points and a significant negative correlation with GA on average for all leagues. The EPL seems to be the exception to the rule where the correlations for YCFNorm above are flipped with significance, confirming the study's initial hypotheses. YCPFPG showed a significant negative correlation with the end of season points and a significant positive correlation while predicting GA. In the BL's case, there was a significant negative correlation with YCPFPG while predicting points.

In all the above models, the cross-validation scores for R^2 were lower than that of the training set, which depicts lower out of sample reliability. Considering FPGNorm and YCPFPG, the standard deviations for CV results were high, but still showed a non-zero effect on the end of season Pts and GA. This was not the case with YCFNorm, which lost reliability due to the low value and high standard deviation of R^2 in the CV results while predicting both Pts and GA. Thus, we can conclude the initial hypotheses were confirmed except for YCFNorm due to low out of sample reliability.

Overall, across leagues, the total points increased every season, and GA decreased. The EPL consistently showed a significantly lower

average points total and a higher number of goals conceded with every consecutive season compared to the other leagues. Although initially not hypothesized, all the leagues seemed to commit lesser YCFNorm, while YCPFPG appeared to increase every consecutive season. These observations are in line with previous results confirming that the leagues may be getting less aggressive each season, with the EPL being most competitive and aggressive (Sapp et al., 2018). The present study shows that fouls, when accounted for possession, positively predict success, which differs from previous findings, which indicate a negative relationship between success and fouls (Castellano et al., 2012; Sapp et al., 2018). This may be since previous studies did not normalize the fouls for out of possession time.

The above results may suggest that fouling while out of possession is a good strategy for success in every league, but it seems to be the most effective in the EPL. Furthermore, FPGNorm seems to predict defensive performance more accurately than the overall end of season performance. This is evident from the cross-validation scores in Table 4, which show a lower standard deviation for R^2 scores while predicting GA as compared to the Pts. A large number of fouls committed at the elite level are intentional (Gümüřdađ et al., 2011; Wright and Hirotsu, 2003). Hence, further research on tactical fouls, the foul's position, and change in the expected goal (XG) after the foul is needed.

The results of YCFNorm may suggest that while committing fouls worthy of a yellow card may be a bad strategy for success overall, in the EPL committing yellow card worthy fouls seems to be beneficial. It may also suggest that most of the yellow cards received partly indicate tactical fouls as it is harder to get a yellow card in the EPL due to lower strictness levels of referees while calling fouls and giving yellow cards in this particular league (Gümüřdađ et al., 2011; Unkelbach and Memmert, 2008). It should be noted that the CV results indicate that the standard deviation on the R^2 values is high, which leads to low reliability of the given model and the effects of YCFNorm suggest a need for further investigation on fouling behavior in the EPL with a more extensive and more detailed sample including the location of foul, the position of players, etc.

The results of YCPFPG may suggest that fouling without getting a yellow card seems to be a good strategy for success in general, but neither of the individual leagues show this trend. Hence, this seems to be a novel case of Simpson's paradox, where the trend for an entire dataset is different compared to the component leagues (Simpson, 1951). Season and league differences in fouling behavior observed in the present study suggest that fouling behavior in each league may show different results in accordance with previous findings (Sapp et al., 2018). Hence, it is necessary to separate leagues while concluding for the purpose of practical implications.

Possession has been extensively studied as a performance indicator in various game situations, accounting for team strength, home-away factors, the zone of possession, etc. Previous studies suggest that possession seems to successfully predict success when accounted for performance in the whole season (Casal et al., 2017; Göral, 2015; Jones et al., 2004; Mayr et al., 2019; Mitrotasios et al., 2019). Due to the nature of the game of soccer, 'which team has possession of the ball' intrinsically defines the phase of play, i.e., whether the team is attacking or defending. The FPG and YCF in the present study would show different results if not normalized (see Appendix). This knowledge is crucial for modeling any KPIs for predicting performance, as it may lead to unreliable or inaccurate results due to missing normalization. Hence, the present study proposes both individuals and team KPIs should be normalized for in and out of possession time of the team in concern. Depending on what phase of play (defensive or offensive) the particular KPI is relevant for. Furthermore, the study of KPIs post normalization would have intrinsic meaning in sport-specific knowledge from a perspective of performance analysis.

Significant data and modern machine learning approaches are already a part of performance analysis, forecasting, and prediction (Couceiro et al., 2016; Dutt-Mazumder et al., 2011; Memmert et al., 2017; Rein and Memmert, 2016). The machine learning algorithms are excellent at improving the model's prediction and forecasting capability, but they still are a 'black-box' approach. Hence, building-in domain-specific knowledge in such algorithms' input features becomes crucial for interpreting the results along with data-based

objectivity, both of which are essential factors for performance analysis, talent identification, and recruiting (Bergkamp et al., 2019; Honegger, 2018; Sarmiento et al., 2018). The normalization technique in the present study, with the theoretical background, is an excellent 'used case' for building-in domain-specific knowledge. It provides a strong base for the application of cumulative research in notational analysis for invasion sports.

Conclusions

The present study confirms that out of possession fouls and the ratio of yellow cards to fouls seem to significantly predict success, i.e., low GA and high Pts. All models in the present study predict defensive performance better than overall performance (Table 4). Furthermore, all leagues seem to be getting stricter each year. In summary, 'dirty play' seems to work the best in the EPL compared to other European leagues in the current study. That may be due to the teams adopting a strategy to take advantage of the lower referee strictness. The normalization procedure suggested in the present study for in and out of possession actions can be applied in other analogous invasion sports.

With the rise of large databases and new machine learning algorithms, the suggested structure in the present study can further improve the efficiency of player recruitment, improve betting odds, and help improve the sports forecasting industry. Teams and coaches can apply the study's findings in formation on effective strategy with and without the ball. They can further use phase-specific notational statistics (post normalization) to select the right personnel based on self and the opposition team's data.

With big data, access to computational power, and sport-specific knowledge combined with Auto ML algorithms, there is a potential to explore new KPI's and seamlessly apply them in the industry (Balaji and Allen, 2018). The current study provides the first step towards fouling behavior and its effect on success in soccer and the application of normalization techniques for KPI's. Further detailed research with larger datasets and more sophisticated algorithms is needed to understand the current findings' full impact.

References

- Anders A, Rotthoff KW. Yellow cards: Do they matter?. *J Quant Anal Sport*, 2011; 7(1)
- Bergkamp TLG, Niessen ASM, den Hartigh RJR, Frencken WGP, Meijer RR. Methodological Issues in Soccer Talent Identification Research. *Sport Med*, 2019; 49(9): 1317–1335
- Camerino OF, Chaverri J, Anguera MT, Jonsson GK. Dynamics of the game in soccer: Detection of T-patterns. *Eur J Sport Sci*, 2012;
- Casal CA, Maneiro R, Ardá T, Mari FJ, Losada JL. Possession zone as a performance indicator in football. The game of the best teams. *Front Psychol*, 2017; 8(JUL): 1–11
- Castellano J, Casamichana D, Lago C. The use of match statistics that discriminate between successful and unsuccessful soccer teams. *J Hum Kinet*, 2012; 31(1): 139–147
- Couceiro MS, Dias G, Araújo D, Davids K. The ARCANÉ Project: How an Ecological Dynamics Framework Can Enhance Performance Assessment and Prediction in Football. *Sport Med*, 2016; 46(12): 1781–1786
- Dutt-Mazumder A, Button C, Robins A, Bartlett R. Neural network modelling and dynamical system theory: Are they relevant to study the governing dynamics of association football players?. *Sport Med*, 2011
- Gómez MÁ, Toro EO, Furley P. The influence of unsportsmanlike fouls on Basketball teams' performance according to context-related variables. *Int J Sports Physiol Perform*, 2016; 11(5): 664–670
- Göral K. One Of The Main Determinants Of Performance In Soccer: Ball Possession (Analysis Of Fifa U-20 World Cup). *Int Ref J Humanit Acad Sci*, 2015
- Gümüşdağ H, Yıldırım İ, Yamaner F, Kartal A. Aggression and fouls in professional football. *Biomed Hum Kinet*, 2011; 3(1): 67–71
- Hirotsu N, Wright M. Using a markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *J Oper Res Soc*, 2002; 53(1): 88–96
- Hughes MD, Bartlett RM. The use of performance indicators in performance analysis. *J Sports Sci*, 2002; 20(10): 739–754
- James N. Notational analysis in soccer: past, present and future. *Int J Perform Anal Sport*, 2006; 6(2): 67–81
- Jones PD, James N, Mellalieu SD. Possession as a performance indicator in soccer. *Int J Perform Anal Sport*, 2004; 4(1): 98–102
- Lago-Ballesteros J, Lago-Peñas C. Performance in team sports: Identifying the keys to success in soccer. *J Hum Kinet*, 2010; 25(1): 85–91
- Lago C, Martín R. Determinants of possession of the ball in soccer. *J Sports Sci*, 2007; 25(9): 969–974
- Mackenzie R, Cushion C. Performance analysis in football: A critical review and implications for future research. *J Sports Sci*, 2013; 31(6): 639–676
- Memmert D, Lemmink KAPM, Sampaio J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sport Med*, 2017; 47(1): 1–10
- Mitrotasios M, Gonzalez-Rodenas J, Armatas V, Aranda R. The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. *Int J Perform Anal Sport*, 2019; 19(3):
- Pratas JM, Volossovitch A, Carita AI. Goal scoring in elite male football: A systematic review. *J Hum Sport Exerc*, 2018; 13(1): 218–230
- Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *Springerplus*, 2016; 5(1): 1–13
- Ridder G, Cramer JS, Hopstaken P. Down to ten: Estimating the effect of a red card in soccer. *J Am Stat Assoc*, 1994;
- Sapp RM, Spangenburg EE, Hagberg JM. Trends in aggressive play and refereeing among the top five European soccer leagues. *J Sports Sci*, 2018; 36(12): 1346–1354
- Sarmiento H, Marcelino R, Anguera MT, Campaniço J, Matos N, Leitão JC. Match analysis in football: a systematic review. *J Sports Sci*, 2014a; 32(20): 1831–1843
- Sarmiento H, Marcelino R, Anguera MT, Campaniço J, Matos N, Leitão JC. Match analysis in football: a systematic review. *J Sports Sci*, 2014b; 32(20): 1831–1843
- Simpson EH. The Interpretation of Interaction in Contingency Tables. *J R Stat Soc Ser B*, 1951;

- Tanamati Soares J. Quantitative Analysis of Penalty Kicks and Yellow Card Referee Decisions in Soccer. *Am J Sport Sci*, 2016; 4(5): 84–89
- Thomas S, Reeves C, Smith A. English soccer teams' aggressive behavior when playing away from home. *Percept Mot Skills*, 2006; 102(2): 317–320
- Unkelbach C, Memmert D. Game management, context effects, and calibration: The case of yellow cards in soccer. *J Sport Exerc Psychol*, 2008; 30(1): 95–109
- Whoscored.com. Whoscored.com., 2019
- Wright C, Atkins S, Jones B. An analysis of elite coaches' engagement with performance analysis services (match , notational analysis and technique analysis). *Int J Perform Anal Sport*, 2019; 9(1): 304
- Wright M, Hirotsu N. The professional foul in football: Tactics and deterrents. *J Oper Res Soc*, 2003; 54(3): 213–221

Corresponding author:**Ashwin Phatak**

Alternative Email: Ashwin.phatak@gmail.com

Orchid Id: <https://orcid.org/0000-0003-1772-8949>

Email: a.phatak@dshs-koeln.de