Evaluating Passing Ability in Association Football

Else Marie Håland[a], Astrid Salte Wiig[a], Magnus Stålhane[a], and Lars Magnus Hvattum[b]

[a] *Department of Industrial Economics and Technology Management, NTNU, Alfred Getz vei 3, NO-7491 Trondheim, Norway*
[b] *Molde University College, NO-6402 Molde, Norway*

Abstract

In this paper, the passing ability of football players is determined by building three generalised additive mixed models that each explains a different aspect of a pass' success: difficulty, risk and potential. The models are built on data from the 2014-2016 seasons of the Norwegian top division Eliteserien, and their predictive power is tested on the 2017 season. The results provide insight into the factors affecting the success of a pass in Eliteserien. These include the location of the pass, the relationship to previous passes and to situations such as throw-ins, corners, free kicks, or tackles, as well as conditions specific to the Eliteserien, such as the time of season and the ground surface type. Finally, the key pass makers in the league are identified.

Keywords: soccer, passing, player rating, generalised additive mixed model, regression

1 Introduction

Association football, referred to as football in this paper, attracts more than 4 billion followers worldwide and is the most popular sport measured by participation, media coverage and key economic figures (Total Sportek, 2007). Enormous amounts of money are circulating in the world of football and the three teams generating the highest revenue earned a combined sum of €2 billion in the 2016/2017 season (Boor et al., 2018). Considering these numbers, there is no doubt that success in football is valuable and football teams are therefore constantly seeking new ways of improving their performance, with the aim of winning more titles and thereby receiving prize money and increased attention.

Decisions in sport have traditionally been made qualitatively by humans, being based on gut feelings or adherence to previous choices (Steinberg, 2015). Recently, the use of statistics in sports has gained popularity and this emerging field of research is known as sport analytics. By making use of data material to assist in decision-making, players' and teams' strengths and weaknesses can be evaluated and accordingly, changes can be made to training sessions with the aim of increasing performance.

This paper is focused on performance analysis of football players' passing ability. Passing is the most frequent event happening during a football match, and by successfully passing the ball forward on the pitch, the chance of creating goal-scoring opportunities increases. Although the passing ability of players has been studied in the past, the focus has mainly been on whether a pass successfully reaches a teammate. In this paper, three generalised additive mixed models (GAMMs) are developed and each model looks into a different aspect of each pass: its difficulty, risk and potential. The difficulty of a pass is measured through the probability that it successfully reaches a teammate. The risk associated with a pass is measured through the probability that the pass both reaches a teammate and that the next event after the pass is successful. That is, the risk is associated to the probability of losing possession directly after the pass. Ideally, a pass should help a team to perform well, which involves scoring goals. Goals are very rare events in football, and the potential of a pass is therefore measured as the probability that the pass is part of a sequence that leads to a shot by the passing team.

Several studies have been conducted to analyse passing behaviour in football. Szczepański and McHale (2016) evaluated football players' passing ability by building a GAMM to estimate the probability of a pass' success. Proxies were developed to account for the overall difficulty of the pass, and random effects of players, the players' team and opponent teams were included to investigate individual passing skill and teams' abilities to facilitate and hamper passes. A similar GAMM was developed by Tovar et al. (2017), and the model's ability to predict the performance of any player transferring from the Colombian to the Spanish league was demonstrated. McHale and Relton (2018) elaborated on the work by Szczepański and McHale (2016) and developed a GAMM using player tracking data to estimate the likelihood of a pass' success with the purpose of identifying the key passers in a team through a network analysis.

Whereas Szczepański and McHale (2016) considered only the difficulty of a pass, Power et al. (2017) covered both the difficulty and the potential of passes by estimating their risk and reward through the use of a supervised learning approach. The potential of passes was also explored by Brooks et al. (2016), who used machine learning techniques to measure the importance of a pass by examining the relationship between pass location in a ball possession and shot opportunities. Other studies related to pass potential include Rein et al. (2017), Gyarmati and Stanojevic (2016) and Mackay (2017). In the latter paper, ridge logistic regression with a sliding window approach is used to find the probability of different actions on the field resulting in a goal.

To summarise past research on passes, Szczepański and McHale (2016) showed that predicting players' completion rates over a season can be done significantly better by using a GAMM than by just using the completion rates of past seasons. Tovar et al. (2017) found that a slightly simplified GAMM can be used to predict how players will perform, with respect to passing, when moving from one league to another. Rein et al. (2017) demonstrated that scoring goals and winning matches are related both to passes creating majority situations and to controlling space. They found that passes from mid-field into the attacking area are most effective. Brooks et al. (2016) concluded that ranking players based entirely on their passing yields results that are consistent with general perceptions of offensive ability. Gyarmati and Stanojevic (2016) discovered that passes can be made that lose possession yet simultaneously increase the chances of winning the game. Mackay (2017) and McHale and Relton (2018) focused on identifying the best players and the key passers.

This paper builds on the study by Szczepański and McHale (2016) in three ways. First, a different data set is used, so that their findings may be independently confirmed. Second, additional binary variables and smooth terms are introduced, which may allow a more precise evaluation of each pass, using additional sources of data. Third, the evaluation of passes is done using three different GAMMs. One of these (Model 1) assesses the difficulty of passes, as in the work by Szczepański and McHale (2016). A second model (Model 2) is seemingly unique in the way the dependent variable is defined, and handles the tactical aspect of a pass by examining the probability that a pass can be followed up, i.e. the play can easily be continued without losing possession. The final model investigates pass potential (the probability that the pass is part of a sequence leading to a shot taken), which has been considered by many researchers in the past. However, the regression model produced is far more extensive.

The objective of this study has two main parts. On one hand, the attributes of passes in the Norwegian top division, Eliteserien, have not previously been studied. The study aims to reveal the characteristics of passes that are likely to reach a teammate, that are likely to maintain possession, and that are likely to result in shots taken. On the other hand, this may in turn allow the identification of good passers in the league, and coaches can get an indication of a player's ability to transfer performances in training sessions to match situations. In addition, the analysis may provide other insights that can be generalised and utilised by coaches and players to elevate their team's performance.

The remainder of the paper is structured as follows. In Section 2, the data and the methodology used are introduced. Then, the results are presented and discussed in Section 3, before the concluding remarks are given in Section 4.

2 Experimental Set-up

Data from three different sources are obtained to build detailed regression models. The models are developed to rate football players' passing abilities.

2.1 Data

For the analysis, event-data is obtained from Opta Sports's Opta24Feed (Opta Sports, 2018). The event-data covers all matches in the 2014-2017 seasons of Eliteserien, with all events happening in a match being logged with their characteristics and time stamp. As the data delivered by Opta Sports for Eliteserien only follows the movement of the ball, the whereabouts of the players not in possession of the ball are unknown. This is a major limitation considering the types of models which are examined in this paper, as the positioning of the opponents would be of importance.

Opta Sports is a credible provider of sports data; they have three analysts collecting data in each match (Greig, 2017). In the data processing however, some errors were identified, and as far as possible, these were dealt with in order to be able to use the observations that were faulty. The measures taken involve calculating the angle of the ball movement or length of a pass when it is wrongly set to zero, switching the order of shot events and out-of-play events when they occur in the wrong order, manually inserting player positions in cases where they are wrong or incomplete, and limiting the ball possession of a single player to be less than 30 seconds.

Supplementing the event-data, Elo ratings (ClubElo, 2018), measuring the strength of teams at the time of play, were collected together with information about the ground surface types on football pitches used in Eliteserien (Eliteserien, 2018) during the given seasons. In total, 960 matches and 749,859 passes performed by 831 different players are included in the data set. Passes include passes from open play, headed passes, long passes and crosses and exclude free kicks, corners and throw-ins.

2.2 Methodology

A total of three GAMMs looking into the three different aspects of a pass' success defined earlier are estimated through binary logistic regression to analyse football players' passing ability. GAMMs originate from Lin and Zhang (1999), and are advantageous as random effects can be incorporated and continuous variables can be modelled as smooth terms in addition to other fixed-effect variables. A GAMM can be written as (Wood, 2006):

$$\eta_i = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{\alpha} + f_1(x_1) + \cdots + f_j(x_j), \tag{2.1}$$

$$\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \Sigma_\sigma), \tag{2.2}$$

where $\boldsymbol{X_i}$ is a vector of fixed effects, $\boldsymbol{\beta}$ is a vector of fixed-effect coefficients, $\boldsymbol{Z_i}$ is a vector of random effects, $\boldsymbol{\alpha}$ is a vector of random-effect coefficients, and f_j are smooth functions of variables x_j. The random effects are assumed to be normally distributed with Σ denoting the covariance matrix, parameterised by the coefficient vector σ.

2.2.1 Variables

The dependent variables vary across the models. For Model 1 (difficulty) the dependent variable, Y_1, indicates whether the pass successfully reached a player on the same team, while the dependent variable in Model 2 (risk), Y_2, tells whether the next event after a pass was successful or not. A player shows an ability to keep an overview of the game if he is able to make passes that are more easily followed up. Considering Model 3 (potential), the dependent variable Y_3 takes the value of one if a pass was part of a passing sequence leading to a shot. If the passing sequence is interrupted by events that do not initiate a new sequence, success will also be indicated if a shot is made by the same team within 15 seconds of the last pass in the sequence. Own goals made by the opponent team are seen as shots by the team awarded with the goal.

The explanatory variables are presented in Table 1, and they are initially the same for all models, chosen based on their presumed influence on the dependent variable of Model 1. Some of the variables are based on the model produced by Szczepański and McHale (2016), especially the chosen smooth terms, while several new fixed-effect variables are introduced.

A preliminary test, based on the AIC criterion (Hosmer Jr et al., 2013), was performed to assess whether smooth terms or categorical fixed effects are the most suitable form for the variables. For example, using a 1-D smooth turned out to be preferred over categorical variables to represent the goal difference, and a 4-D smooth preferred over a discretization into zones to represent starting and ending positions of passes.

For many situations in which the ball is recovered in open play, opponents of the passer are nearby. In the case of tackles, aerial duels and interceptions, binary variables X_2, X_3 and X_4 are respectively created to indicate what type of ball recovery that happened in the event prior to a pass. An extra dummy variable, $X_{5.i}$, is added for each of the three situations to indicate whether the player attempting the pass also was involved in the specific prior event. If the same player is involved, it is reasonable to assume for certain that opponents are close, especially for the cases of tackles or aerial duels. All other ball recoveries in the data are assumed to be loose balls, and X_6 is an indicator of whether this occurred in the previous event. There is no indication of opponent pressure in the case of loose balls. These explanatory variables are adjustments to the model developed by Szczepański and McHale (2016), where interceptions and loose balls were not considered and only one variable was used to determine the involvement of a player in both the previous tackle and aerial duel.

Binary variables are included in the model to identify whether the previous pass (X_7) or the current pass (X_8) is a header. The rationale is that passes are likely harder to execute in such circumstances. Moreover, the smooth functions f_7 and f_8 are proxies for the pressure from opponents to expedite passing and the strength of the opponent team, respectively.

During set plays, players cluster together, and opponents surround the team in possession of the ball. Three variables are suggested to function as proxies of opponent pressure. Binary variable X_{10} is used to indicate whether there was a free kick in the previous event, and binary variable X_{11} is used to indicate whether the previous event was a throw-in. The variable X_{12} indicates whether or not a corner was taken within the previous five events. This variable is given a longer time frame due to the effect of a corner usually lasting longer. A new variable, $X_{13.i}$, indicates whether the team that made the set play action also attempted the pass.

A categorical variable, X_1, is included to investigate whether the initiating pass in a sequence

Table 1: Summary of explanatory variables initially used in the models. Fixed-effect variables are denoted by X, random-effects variables by Z, and smooth terms by $f(\cdot)$. The types of variables of fixed and random effects are continuous (C), categorical/factor (F), binary (B) and interaction (I).

Variable	*Description*	*Type*
$X_{1.s}$	Pass number category s in the current sequence of passes ($s = 1, 2, 3, 4$)	F
X_2	Tackle in the previous event	B
X_3	Aerial duel in the previous event	B
X_4	Interception in the previous event	B
$X_{5.i}$	The same player who took part in a tackle ($i = 1$)/aerial duel ($i = 2$)/interception ($i = 3$) also made the pass	B
X_6	Ball recovery due to a loose ball in the previous event	B
X_7	Previous pass was a header	B
X_8	Current pass is a header	B
X_9	Player performing the pass plays for the home team	B
X_{10}	Previous event was a free kick	B
X_{11}	Previous event was a throw-in	B
X_{12}	Corner taken within the past five events	B
$X_{13.i}$	The same team executing a corner ($i = 1$)/ free kick ($i = 2$)/ throw-in ($i = 3$) attempted the pass	B
X_{14}	The match is played on artificial grass	B
$X_{1.2}$*X_{10}	Pass sequence number 2 interacting with free kick in previous event	I
$X_{1.2}$*X_{11}	Pass sequence number 2 interacting with throw-in in previous event	I
$Z_{1.k}$	Player k passing the ball ($k = 1, \ldots, 689$)	F
$Z_{2.t}$	Team t the player is representing ($t = 1, \ldots, 19$)	F
$Z_{3.o}$	Opponent team o to the player passing the ball ($o = 1, \ldots, 19$)	F
$f_1(x_0, y_0, x_1, y_1)$	4-D smooth handling the start (x_0, y_0) and end coordinates (x_1, y_1) of a pass	C
$f_2(\bar{x_2}, \bar{y_2})$	2-D smooth representing the average position of the player given by coordinates ($\bar{x_2}$, $\bar{y_2}$)	C
$f_3(x_3)$	1-D smooth representing game time, x_3, in minutes	C
$f_4(x_4)$	1-D smooth handling time played, x_4, by player passing the ball	C
$f_5(x_5)$	1-D smooth representing the goal difference, x_5	C
$f_6(x_3, x_5)$	2-D smooth representing the interaction between game time and goal difference	I
$f_7(x_7)$	1-D smooth handling the time passed, x_7, since last occurred event	C
$f_8(x_8)$	1-D smooth representing the absolute Elo rating, x_8, of the opponent team	C
$f_9(x_9)$*X_{14}	1-D smooth functions representing month of play, x_9, interacting with type of grass	I

is more difficult to make due to opponent pressure. The variable can take on four possible values depending upon whether the pass is the: 1) first, 2) second, 3) third, fourth or fifth or 4) of a higher sequence number in the sequence. The first category is chosen to be the reference. Both open play ball recoveries and set play actions are allowed to be starting points of sequences. To investigate the impact of the first pass after a set play, two interaction terms are added to separate this pass from other passes with a sequence number of two: the interaction between $X_{1.2}$ and the binary variable telling whether the previous event was a free kick (X_{10}), and the interaction between $X_{1.2}$ and the indicator telling whether the previous event was a throw-in (X_{11}).

The position of the players passing and receiving the ball should have an impact on the success of the pass as the whereabouts could imply something about the pressure from opponents on both players. To account for this, the 4-D smooth function f_1 handles the start and ending coordinates of the pass.

The 2-D smooth function f_2 is a function of the average position of the player in the current match. For each match, the average position occupied by a player is calculated by averaging the x- and y-coordinates of all events a player is involved with. This is different from (Szczepański and McHale, 2016), where the anticipated player position based on the previously played matches with exponential weighting is considered to avoid endogeneity. As players change their player positions both within matches and between matches, and due to the same player position being occupied differently depending upon the team's strategy, the approach used in this paper seems to be more appropriate. As in (Szczepański and McHale, 2016), the y-coordinates are calculated as absolute distances from the centre of the axis to avoid cancellation of terms for players playing on both sides of the pitch. That is, a player appearing as a left winger in the first half and as a right winger in the second half will be identified as a winger, and not as a central player.

Several situational variables are added to the models. The game time in minutes is a stressing factor for the players and is treated as a smooth function, f_3, while the actual time a player has spent on the pitch when the player passes the ball is handled by f_4. Players' time spent on the pitch accounts for the exhaustion effect of players as not all players play the entire match. Whether game time turns out to be a stressing factor tends to depend upon the scores at the moment of play. The goal difference, relative to the team in possession of the ball, is therefore modelled through smooth term f_5, and to test the interaction with game time, the 2-D smooth term f_6 is added to the models.

Other variables describing the external circumstances of a football match which may influence the outcome of a pass include home team advantage, X_9 (Pollard and Pollard, 2005), ground surface type, X_{14} (Hvattum, 2015), and month of play, f_9 (Mohr et al., 2003). The ground surface type is either natural or artificial grass, and the condition on the fields may vary a lot throughout a season. Hence, the interaction between type of grass and month of play is included. As the month of play is treated as a smooth function, this interaction is directly included in the smooth term, and two separate smooths are used in the regression, one for each type of grass. The month of play represents a seasonal effect, affecting both ground conditions and the shape of the players. There should be a fatigue effect when the season is close to an end, while the performance of players is expected to increase in the beginning of the season as their match shape is improved.

The random effects $Z_{1.k}$, $Z_{2.t}$ and $Z_{3.o}$ are incorporated to assess players' individual passing ability and teams' ability to facilitate and hinder passes. It is expected that players getting high estimated coefficients are among those players with best passing skill in the league. Similarly, teams achieving high scores of facilitation and hampering are perceived to be good at these specific abilities. By assuming normally distributed random effects, noise in the models resulting from players exhibiting exceptionally good or bad passing skills can be reduced as the coefficients will shrink towards the mean.

2.2.2 Model Selection and Validation

The fixed-effect variables to be included in each final model are separately determined through Wald tests (Hosmer Jr et al., 2013). Although stepwise regression may cause some problems in identifying the best fitted model (Goodenough et al., 2012; Judd et al., 2011), it is used due to its effectiveness and low computational burden. Additionally, with a data set consisting of a high number of observations, the problem of having highly correlated variables is believed to be minimal.

Validation of the models is performed through methods used for binary classifier systems. The area under the receiver operating characteristic curve (AUC-ROC) and the area under the precision-recall curve (AUC-PR) (Fawcett, 2006) are considered.

The logistic regressions were run using the *bam* function from the *mgcv* package in RStudio (Wood, 2011), an open-source integrated development environment for the statistical programming language R (R Core Team, 2017). For all 1-D smooths, the cubic regression spline is chosen as basis function. This is due to the computational gain implied by this choice rather than using the default thin plate regression splines. Tensor product smooths are chosen for all the multidimensional smooths.

2.2.3 Predictions and Player Ratings

The three models are estimated based on data covering the 2014-2016 seasons of Eliteserien, and their predictive capability is tested for the 2017 season. To test the predictive power of the models, full predictions are utilised as in (Szczepański and McHale, 2016). Hence, all random effects are considered. Players and teams that are new to the 2017 season are given a corresponding coefficient of zero, the equivalent of being average. The predicted success rate for each player is compared to the actually observed rate for each model. This gives an indication for how well the models are able to predict the players' performance.

As the random coefficients obtained for players' passing ability are given for a total of three seasons, players are rated differently in this paper compared to Szczepański and McHale (2016), and the notion of average players is used to rate players per season instead. Consequently, all random effects are excluded from the predicted outcomes of each pass. Szczepański and McHale (2016) suggest only setting the player random effects to zero when considering the average player. This is however perceived to be less accurate as specific teams are considered. Players are rated by calculating the ratio between the actual observed success rate for the player over the models' expected success rate.

3 Results and Discussion

The final models were decided after performing Wald tests. The full regression results of the fixed-effect variables and smooth terms can be found in Appendix A. A significance level of 10 % is used for the fixed-effect variables. In work done with preliminary models, this model building approach provided similar final models as when using backward elimination based on AIC scores, but at a much lower computational cost. The final models would be relatively similar if a stricter significance level was used.

Other than the intercepts, no coefficients are appearing to be very high or low in terms of magnitude. Model 1 (difficulty) is built on a total of 565,720 passes. In Model 2 (risk), three passes are omitted due to these having no defined next event in the game, while one observation is missing from the building of Model 3 (potential). This single pass is the last event of the last match to be processed in the data set, which seems to be the reason why it is not recorded properly. However, it is unlikely that the omission has had any influence on the regression result.

Table 2: A sequence of passes made by Rosenborg is used for validation. Rosenborg played against Stabæk in October 2017 on their own home ground covered with natural grass. The score was draw, and Stabæk's Elo rating at the time of play was 1284.13. Refer to Figure 2 for a graphical representation of the starting and ending coordinates. All other variables do not change between events, and are thus not tabulated below. There are no headed passes, and the sequence is initiated right after a tackle in which the first passer is not partaking.

Failed tackle by Stabæk player Ronald Hernández											
No.	*Player*	*Pos*	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	X_1	$\bar{x_2}$	$\bar{y_2}$	x_3	x_4	x_7
1	Samuel Adegbenro	WI	0.871	0.658	0.086	1	63	36	67	67	3
2	Morten Konradsen	CM	0.989	0.895	0.094	2	48	39	67	6	2
3	Tore Reginiussen	CD	0.998	0.894	0.084	3	33	19	67	67	3
4	Vegar Hedenstad	FB	0.854	0.558	0.091	3	43	39	68	68	4
5	Nicklas Bendtner	ST	0.851	0.625	0.104	3	64	26	68	68	4
6	Vegar Hedenstad	FB	0.899	0.549	0.115	4	43	39	68	68	3
7	Pål André Helland	WI	0.984	0.818	0.094	4	77	35	68	68	4
8	Nicklas Bendtner	ST	0.980	0.839	0.132	4	64	26	68	68	2
9	Tore Reginiussen	CD	0.833	0.635	0.192	4	33	19	68	68	8
10	Mike Jensen	CM	0.576	0.315	0.260	4	59	26	68	68	1
Missed shot by Pål André Helland											

3.1 Model Validation

The areas under the ROC and PR curve are used to validate the models. Additionally, the models' fit are tested by evaluating a sequence of passes from a match in the data set.

3.1.1 ROC and PR Curves

In Figure 1, the ROC and PR curves for each model are shown with their resulting AUC values. Considering the guidelines for the AUC-ROC as suggested by Hosmer Jr et al. (2013), Model 1 exhibits an excellent discrimination ability, while Model 2 and Model 3 have an acceptable discrimination ability.

For the PR curves, the first plotted point after the zero recall region has a precision value equal to the highest estimated probability of success given by the model, and the last plotted point has a precision value equal to the percentage of observed successes in the data. Hence, for Model 3, the AUC-PR is low due to a skewed distribution for both the observed outcomes and estimated probabilities for the model's dependent variable. The points in which the curve has to start and stop have quite low precision values, meaning that the curve itself must be concave and extremely curved to produce a high AUC-PR, which is very unlikely given the set of data. The PR curve is in fact convex, but the curvature is not very extreme, and the AUC-PR could thus plausibly be considered as acceptable, actually indicating a good model fit. For Model 1 and Model 2, the good fit as indicated by the AUC-ROCs is supported by the relatively high AUC-PRs.

3.1.2 Sequence of Passes

A sequence of ten passes obtained from a match between Rosenborg and Stabæk is analysed to test the validity of the models. The course of the sequence, which is performed by Rosenborg players, is tabulated in Table 2 and a graphical representation is shown in Figure 2. A failed attempted tackle by an opponent player initiates the sequence, and the sequence is ended with a shot.

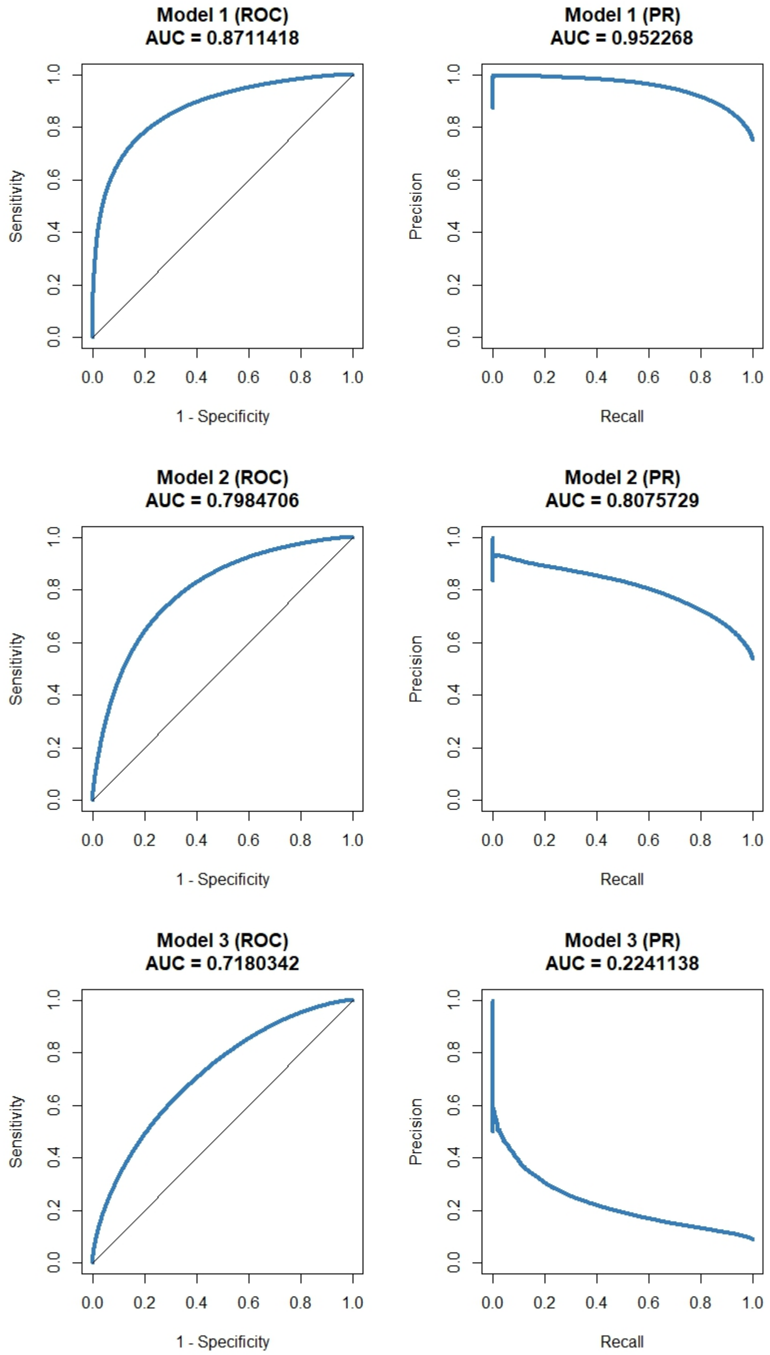


Figure 1: ROC and PR curves for each model

For Model 1, all but the last pass in the sequence are predicted to reach a teammate with a probability of over 80 %, with the probabilities being highest for passes directed backwards. As most of the passes are made close to the sidelines or inside Rosenborg's own half, these values seem to be reasonable. Surprisingly, pass number nine has a quite high predicted probability of reaching a teammate although this pass is made relatively close to Stabæk's goal post, which could be due to the fact that there was a time gap of eight seconds between the pass and the previous event. The last pass in the sequence is directed forwards and the ball moves into the penalty area of Stabæk, which intuitively makes this pass more difficult to attempt. This perceived difficulty is captured by the model as seen from the lower predicted probability.

Intuitive results are also given by Model 2 and Model 3. The last pass in the sequence has the lowest predicted probability of allowing a successful follow-up, corresponding to a high risk of losing possession. The probability of the sequence ending with a shot is increasing as the passes get closer to the opponent's goal post. These results are reasonable, as the pressure from the opponents is usually higher on their own half, and as shots are usually attempted close to the goal.

3.2 Regression Results

3.2.1 Fixed Effects

The coefficients of all significant fixed-effect variables included in the final models are presented in Table 3. Positive signs on the coefficients correspond to an increased probability of success. It is apparent that the first pass in a sequence is the most difficult pass to make. If the first pass is initiated after a tackle and if it is by an involved player, the probability of success increases, while it is opposite in the case of an aerial duel. For tackles, the result is counter-intuitive at first, but different types of tackles are not considered, and some of them, such as a sliding tackle,

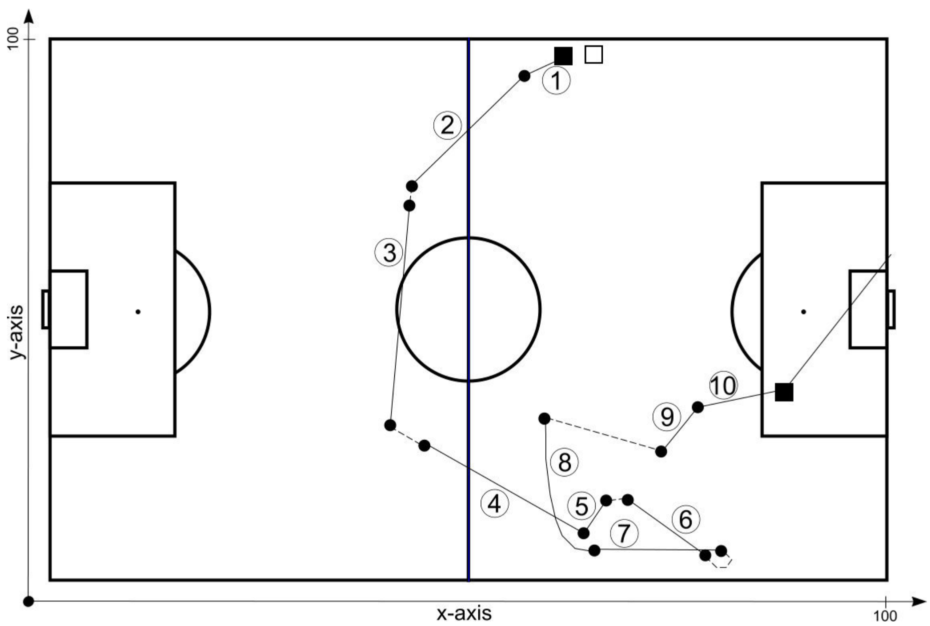


Figure 2: A graphical representation of the sequence of passes used for validation. The start and end of the sequence are represented by filled squares, and a hollow square illustrates where the preceding tackle occurred. Dotted lines are either ball touches or ball carries.

Table 3: Fixed-effect coefficients of the estimated GAMMs. See the note about significance.

Variable	*Short description*	*Model 1*	*Model 2*	*Model 3*
$X_{1.2}$	Pass no. 2	0.351^{***}	0.311^{***}	0.240^{***}
$X_{1.3}$	Pass no. 3, 4 or 5	0.498^{***}	0.471^{***}	0.223^{***}
$X_{1.4}$	Pass no. 6+	0.580^{***}	0.515^{***}	0.223^{***}
X_2	Tackle	0.343^{***}	0.189^{***}	0.325^{***}
X_3	Aerial duel	-0.100^{***}	-0.596^{***}	-0.134^{**}
X_4	Interception	-0.327^{***}	-0.148^{***}	0.091^{*}
$X_{5.1}$	Same player: tackle	0.154^{**}	0.097^{*}	
$X_{5.2}$	Same player: aerial duel		0.532^{***}	
$X_{5.3}$	Same player: interception	0.757^{***}	0.387^{***}	0.338^{***}
X_6	Loose ball	0.515^{***}	0.277^{***}	0.289^{***}
X_7	Previous pass was header	-0.311^{***}	-0.294^{***}	-0.110^{***}
X_8	Headed pass	-1.227^{***}	-0.968^{***}	-0.785^{***}
X_9	Home team advantage	0.138^{***}	0.117^{***}	0.092^{***}
X_{10}	Free kick		-0.181^{**}	
X_{11}	Throw-in	0.184^{***}		-0.106^{***}
X_{12}	Corner		-0.172^{***}	0.363^{***}
$X_{13.1}$	Same team: corner	0.226^{***}	0.120^{*}	$-0.156^{\dagger}$
$X_{13.3}$	Same team: throw-in	$-0.215^{\dagger}$	-0.367^{***}	
X_{14}	Ground surface type	0.123^{***}	0.187^{***}	0.127^{***}
$X_{1.2}{}^{*}X_{10}$	Interaction	0.306^{***}	0.543^{***}	-0.093^{*}
$X_{1.2}{}^{*}X_{11}$	Interaction	$0.225^{\dagger}$	0.447^{***}	
Intercept		1.299^{***}	-0.517^{***}	-3.248^{***}

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

leaves the opponent down, meaning that the pressure is not very high. The effect of interceptions differs between the models. However, it is throughout positive when the same player involved with the interception also passes the ball, which is not surprising as interceptions may happen far from opponents.

As expected, aerial duels in the previous event complicate the task of executing a successful pass. This negative effect is further increased when including the contribution from headed passes. A headed pass will always be the outcome of an aerial duel, and it will always be made by one of the players involved in the duel. Thus, it makes sense that the variable indicating whether the same player is involved becomes insignificant for some of the models. Additionally, the positive effect it has in Model 2 is offset by the more negative effect of the header variable for the same model. In general, headed passes, both in current and previous event, have a negative influence on the success rate. Loose balls in the previous event contribute to a higher probability for the player to successfully execute a pass, which is reasonable as opponents are assumed to not necessarily be nearby during these actions.

For Model 1 and Model 2, the probability of a successful pass increases with the pass number in the sequence, while for Model 3, all passes with pass number higher than one are almost equally likely to succeed compared to the first pass in the sequence. When including contributions from the interaction terms between pass number and free kick, the likelihood of success for the second pass in a sequence is increased for Model 1 and Model 2, while it is decreased for Model 3. Whether a pass is attempted by the same team performing a free kick in the prior event is insignificant for all models. This implies that none of the teams have an advantage over the other team when attempting the pass. For throw-ins, the probabilities of making passes that reaches a teammate and does not lose possession increase when adding the interaction term, as opposed to an unchanged negative effect on the potential of the pass to result in a shot. The positive effect in the first two models is reduced when the same team takes the throw-in and attempts the subsequent pass. Apparently, the team is under high pressure from opponents during such set pieces.

Corners within the five previous events have differing impact on the models. While the difficulty of a pass is not affected by the corner variable itself, it is positively affected when a corner is made by the attacking team, which intuitively is opposite of what is expected. The attacking team is subject to a more confined space with opponents surrounding them, which should have made the attempted pass more difficult to make. Passes made by the defending team are more risky, while the influence on the risk of losing possession is minimal for the attacking team. This is counter-intuitive in the same way as for the pass difficulty. One might expect that the corner effect on pass potential would be more pronounced for the attacking team as they are closer to the opponent's goal post. However, the results suggest otherwise. The defending team has a higher likelihood of success, possibly due to counter-attack possibilities in the open areas ahead after the corner is taken.

All models indicate that teams tend to benefit from playing on their own home ground when passing the ball, as seen from the positive coefficients. Also, the models developed suggest that making passes is easier on artificial grass.

3.2.2 Random Effects

The random-effect coefficients of the teams playing in Eliteserien during the 2014-2016 seasons are shown in Figure 3 for all models. In the figure, the teams' skills of facilitation and hampering, with respect to the aspects of success, are plotted on the vertical and horizontal axis respectively. The higher the coefficient, the better a team is on the specific skill.

For all models, Sogndal, Mjøndalen and Tromsø are placed in the third quadrant, indicating both low facilitation and hampering skills. Interestingly, these three teams have been in the lower

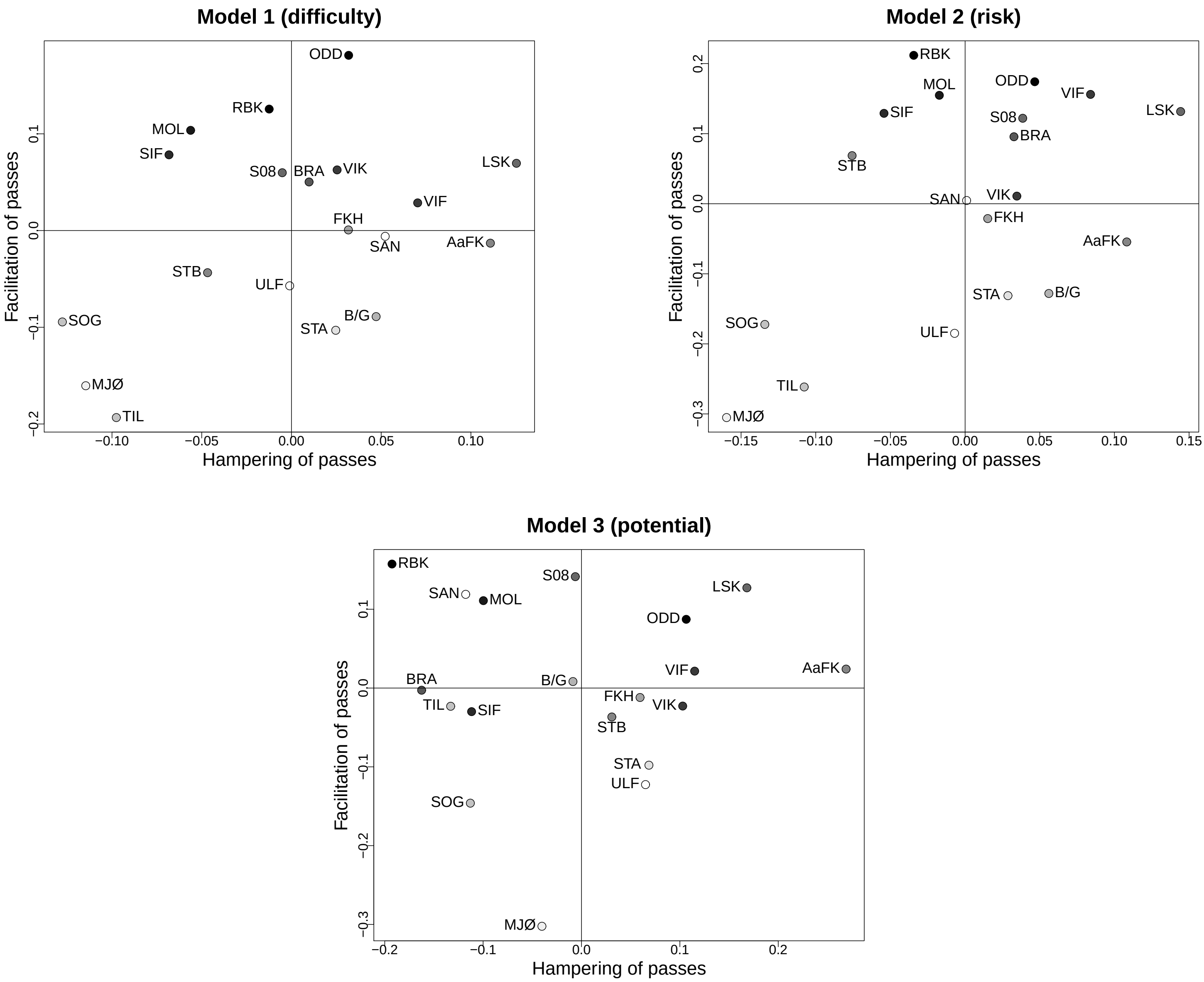


Figure 3: Random-effect coefficients of teams for all models. The coefficients are based on data from the 2014-2016 seasons. The overall team skill of hampering passes is given on the x-axis while the overall team skill of facilitation is given on the y-axis, both skills with respect to the pass aspects considered. Team abbreviations are explained in Table 10. The shade of the circles indicate the average final table ranks, with higher average ranks resulting in darker shades.

half of the table and some have also been relegated from Eliteserien during the seasons considered. Surprisingly, teams such as Rosenborg, Molde and Strømsgodset, all of which always are seen as favourites at the beginning of the seasons, and have performed well during the seasons considered, have negative scores of hampering in all models.

Considering facilitation scores, Odd and Rosenborg are the teams achieving the best scores in terms of pass difficulty and risk, while Rosenborg, followed by Sarpsborg 08, seem to have the best score for pass potential. Statistics from WhoScored.com (2018) state that the top five teams in terms of pass accuracy (passes reaching a teammate) in Eliteserien 2016 were, listed in descending order, Rosenborg, Odd, Strømsgodset, Vålerenga and Molde. Moreover, the teams with highest number of shots per game in the season were Rosenborg, Molde, Sarpsborg 08, Lillestrøm and Strømsgodset. Although being valid only for the 2016 season, the statistics support the findings in terms of the facilitation scores.

The random-effect coefficients of the top ten players during the 2014-2016 seasons in Eliteserien are shown in Table 4 for each model. Central defenders top the list in all models, and only strikers are represented in the bottom rankings for Model 1 and Model 2. While central defenders and central midfielders dominate the top lists for Model 1 and Model 2, more offensive players are included in the top list for Model 3. This is intuitive as shots are usually attempted by offensive players on the opponent's half, and due to many passing sequences being terminated fairly quickly (Hughes and Franks, 2005). On the other hand, some defensive players may seek very safe passing options, either back to the goalkeeper or to another defensive player, and thus more easily get high ratings from Model 1 and Model 2.

To test whether the magnitude of a player's coefficient is influenced by the magnitude of the coefficients of the team the player plays for, the models were run with the team coefficients excluded. The resulting models had higher AICs and were thus not investigated further. However, one should bear in mind a potential correlation between team and player scores.

3.2.3 Smooth Terms

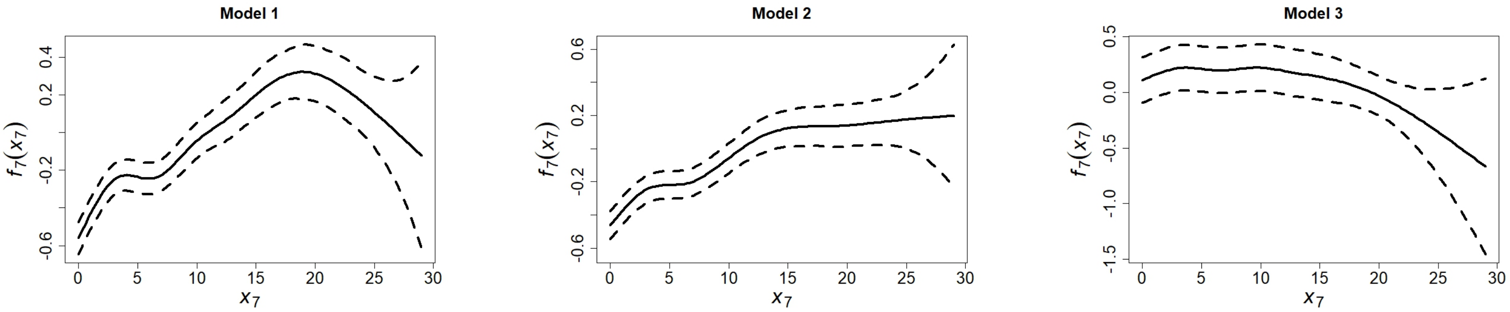
The 1-D smooth functions are illustrated in Figure 4. Panel A illustrates the effect of time passed since the last occurred event measured in seconds. For the models of difficulty and risk, the probability of success is increasing with increased time passed, which is intuitive as more time usually means less pressure. In Model 3, the trend is that passes are associated to a lower potential for leading to shots as the time passed increases. This may perhaps be explained by the fact that the defending team has more time to react when play is slow. For example, with faster ball movement in the forward direction, the likelihood for a counter-attack to be successful could be increased. The negative slope present for the variable in Model 1 and Model 3 after about 18 seconds is probably due to few observations having long time gaps between events.

The effect of game time in minutes is shown in Panel B. For Model 2 and Model 3, the pattern is quite similar. Players have an increasing probability of making low-risk and high-potential passes in the first period of each half of the match, while it is harder, especially when considering the risk of losing possession, to make these passes at the end of the halves, including overtime. Seemingly, the stressing effect of time is captured by these models. For the pass difficulty model, however, a straight line for the function is estimated, giving a linear predictor with a counter-intuitive positive slope. This linear shape is probably due to the introduction of a variable for player-specific game time which is highly correlated to game time ($r = 0.83$) as most players play the entire match.

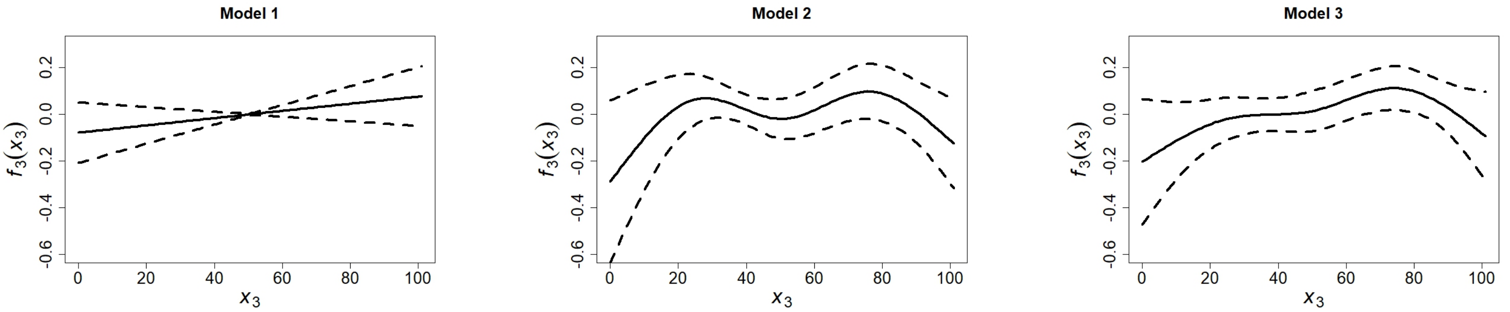
Panel C illustrates the effect of playing time, measured in minutes, for the player passing the ball. For all models, the likelihood of success is highest when a player has been on the pitch between 15 and 85 minutes when attempting the pass, while players having played the entire match seem to have their performance drop in the last minutes of the game. Thus, the anticipated exhaustion

Table 4: The random-effect coefficients of the top ten players during the 2014-2016 seasons according to the models. All player position abbreviations are explained in Table 11, and team abbreviations are explained in Table 10. If a player has played for more than one team in the seasons considered, the team for which the player has played more matches during these years is given.

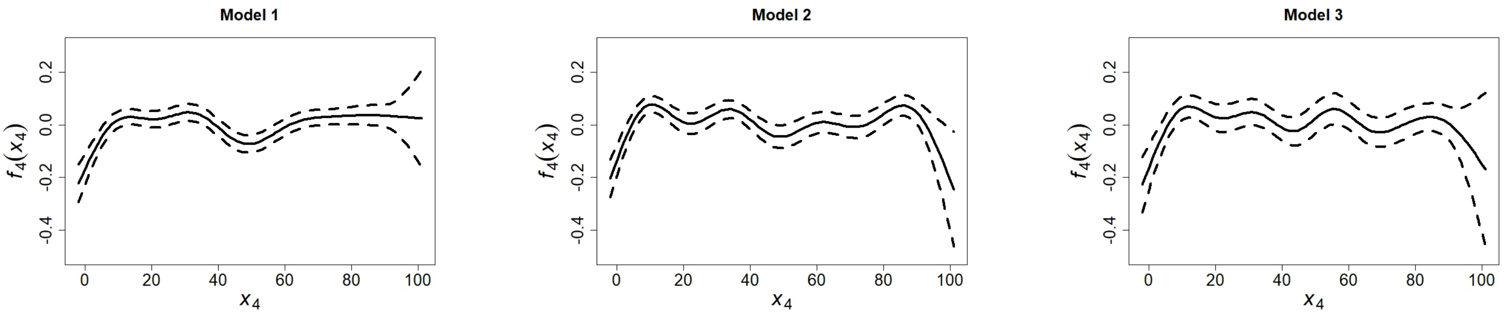
Rank	*Model 1 (difficulty)*				*Model 2 (risk)*				*Model 3 (potential)*			
	Player	*Team*	*Pos*	*Coef*	*Player*	*Team*	*Pos*	*Coef*	*Player*	*Team*	*Pos*	*Coef*
1	Johan Lædre Bjørdal	RBK	CD	0.473	Magnar Ødegaard	TIL	CD	0.278	Joona Toivio	MOL	CD	0.148
2	Martin Ødegaard	SIF	AM	0.422	Anthony Annan	STB	CM	0.275	Magnar Ødegaard	TIL	CD	0.140
3	Christian Grindheim	VIF	CM	0.419	Karol Mets	VIK	CM	0.264	Kristian Brix	B/G	WB	0.137
4	Daniel Berg Hestad	MOL	CM	0.412	Giorgi Gorozia	STB	CM	0.263	Mohamed Ofkir	LSK	WI	0.124
5	Markus Berger	STA	CD	0.401	Johan Lædre Bjørdal	RBK	CD	0.260	Giorgi Gorozia	STB	CM	0.110
6	Joona Toivio	MOL	CD	0.397	Indridi Sigurdsson	VIK	CD	0.252	Jone Samuelsen	ODD	CM	0.104
7	Tomasz Sokolowski	STB	CM	0.394	Anthony Soares	VIK	CD	0.251	Maic Sema	FKH	AM	0.099
8	Johan Andersson	LSK	CM	0.391	Jukka Raitala	SOG	FB	0.247	Ernest Asante	STB	WI	0.099
9	Even Hovland	MOL	CD	0.341	Fredrik Michalsen	TIL	CM	0.238	Eirik Mæland	FKH	CM	0.098
10	Enoch Kofi Adu	STB	CM	0.337	Hólmar Örn Eyjólfsson	RBK	CD	0.237	Erlend Hanstveit	BRA	FB	0.098



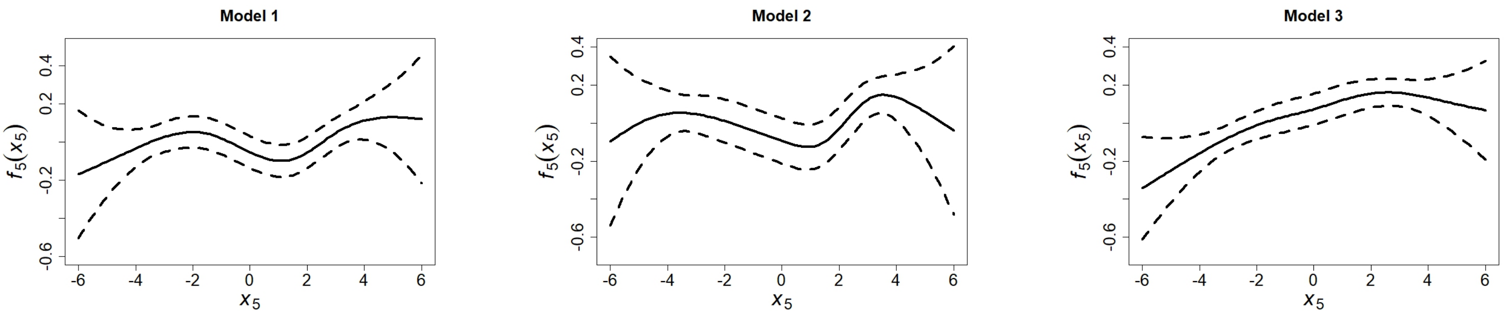
Panel A: Time passed since last event, $f_7(x_7)$



Panel B: Game time in minutes, $f_3(x_3)$

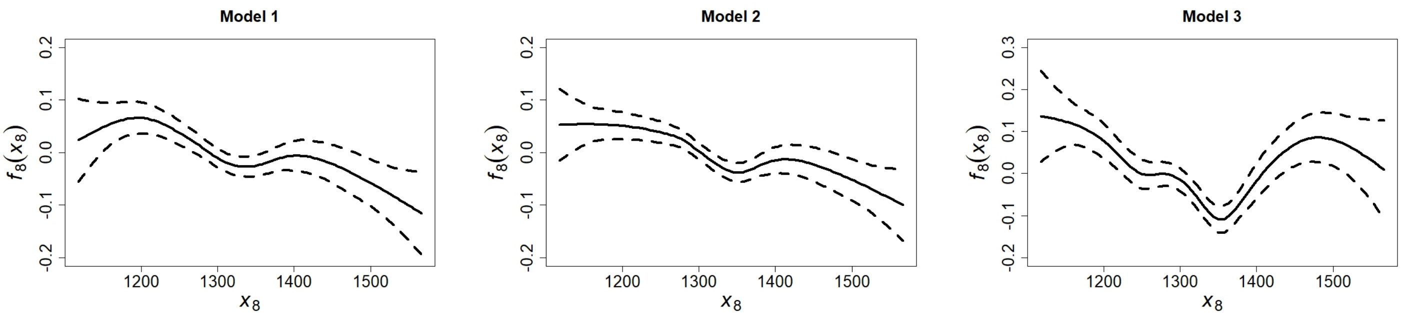


Panel C: Time played by player, $f_4(x_4)$

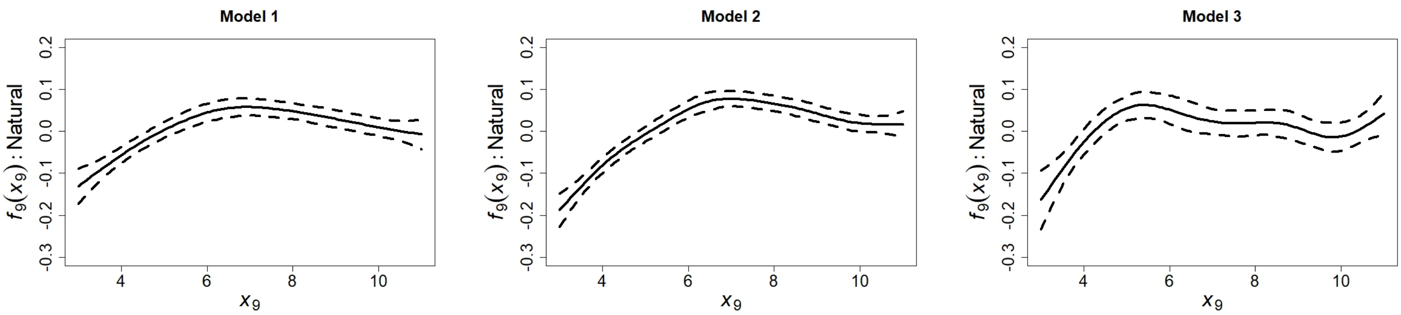


Panel D: Goal difference, $f_5(x_5)$

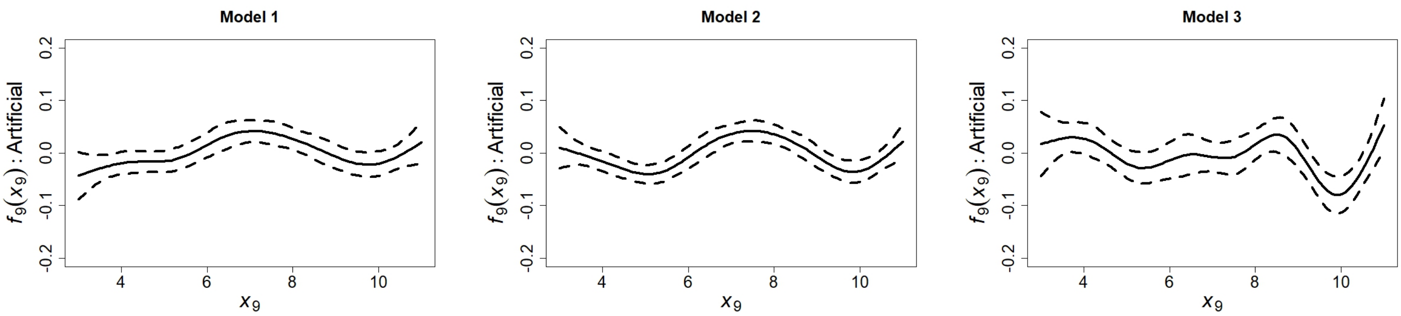
Figure 4: The resulting 1-D smooth functions. The dotted lines indicate the 95 % confidence intervals of the functions. Depicted are panels A-D.



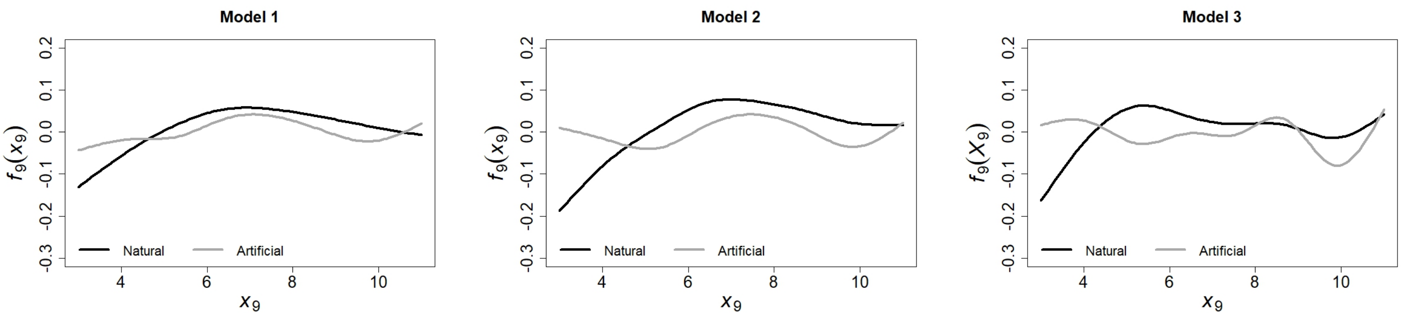
Panel E: Elo rating of opponent team, $f_8(x_8)$



Panel F: Month of play and natural grass, $f_9(x_9)$*X_{14}



Panel G: Month of play and artificial grass, $f_9(x_9)$*X_{14}



Panel H: Artificial versus natural grass

Figure 4: The resulting 1-D smooth functions. The dotted lines indicate the 95 % confidence intervals of the functions. Depicted are panels E-H.

effect seem to be captured by the models, with the effect being more prominent for the risk and the potential of passes.

In Panel D, the influence of the goal difference variable is illustrated. Effects from goal differences higher than four in absolute value should be interpreted with care due to few occurrences in the data set. Considering Model 1 and Model 2, it seems like players have a higher likelihood of success when the goal difference is big, that is, when the match outcome seems to be definite, which is intuitive. In terms of the potential of shots, the teams chasing a goal to even the score are less likely to succeed with their passing sequences, possibly due to a stressing factor of being in need of a goal, or the pressure from the defending team trying to hold on to a narrow lead.

For Model 1 and Model 2, a higher Elo rating of the opponent team makes it harder for a player to achieve success, as seen in Panel E. The same pattern is partly true for Model 3, but when playing against teams with a very high Elo rating, the chance of success is increased.

Panel F and Panel G display the interaction between month of play with natural grass and month of play with artificial grass respectively. The expected behaviour of the month of play variable in regards to fatigue effects is supported by all models when playing on natural grass. Although no clear pattern is seen when the ground surface type is artificial grass, players seem to have a higher probability of success in the middle of the season. Comparing the two smooth functions for month of play in Panel H, it is not apparent that artificial grass is more beneficial throughout the year. However, when also considering the positive fixed-effect variable for artificial grass, the grey line would be shifted upwards, revealing a clear advantage of playing on artificial grass in all models, especially in the months of March and April when the maintenance of natural grass is more challenging than when the season ends in October and November.

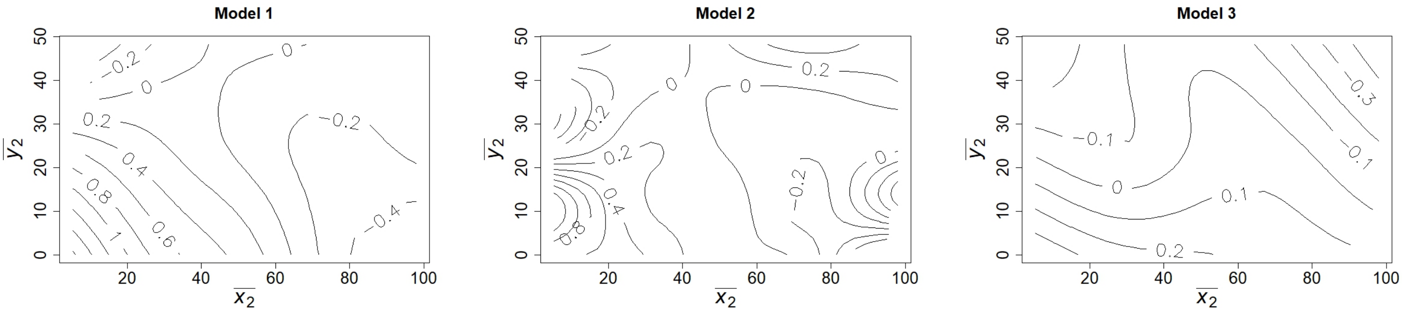
For the multidimensional smooth functions in Figure 5, the resulting contribution is varying between the models. The differing contribution given by the average position of a player is shown in Panel A. When using absolute values for the y-coordinates, the y-axis only ranges from zero to 50, meaning that low y-coordinates correspond to players both on the left and the right side of the pitch. All models indicate that the average position of being in the bottom left corner, where the full backs usually are situated, is the easiest. The bottom right corner, the flanks, and the area in front of the opponent's goal post are associated to a lower probability of success for all models.

For all models, the probability of success in the end minutes is higher for the teams falling behind than the teams being in the lead as seen in Panel B. Thus, the team leading actually has a harder task even if they have a comfortable lead. However, with fewer observations of leads, or defeats, of four or more goals, some of the most extreme contributions must be interpreted more carefully.

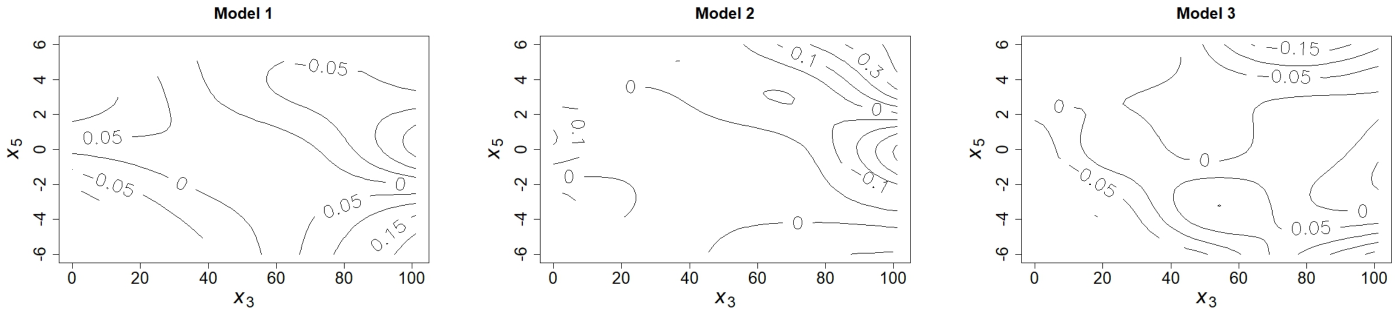
Panel C and Panel D both show results from the 4-D smooth function, covering the starting and ending coordinates of a pass. Given a pair of starting coordinates, the contour lines give the contribution for all end-coordinate possibilities. When passing from the defensive half, Panel C, the likelihood of success is higher when passing backwards compared to passing forward according to the models for difficulty and risk. This is intuitive as the opposition is situated in front of the passer at most times. For the potential of leading to a shot, there is little variation in the magnitude of the contribution, but passing the ball forward gives a higher probability of succeeding. This is also intuitive as the ball is moving closer to the opponent's goal post where shots are more likely to happen. The same intuitive results are present in the case of a pass made from the offensive half in Panel D as well.

3.3 Predictions and Player Ratings

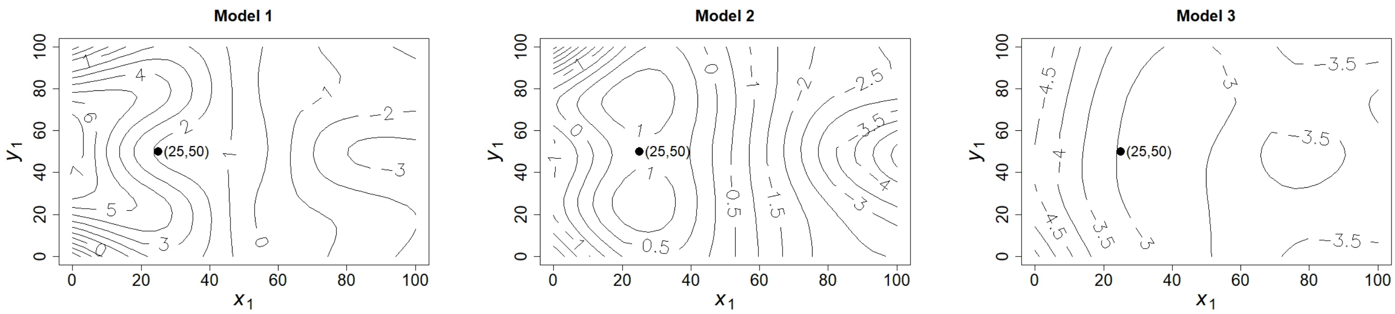
In Figure 6, the predicted success rate, based on full predictions, is plotted against the observed success rate for each model and each player in the 2017 season. Only players attempting a minimum



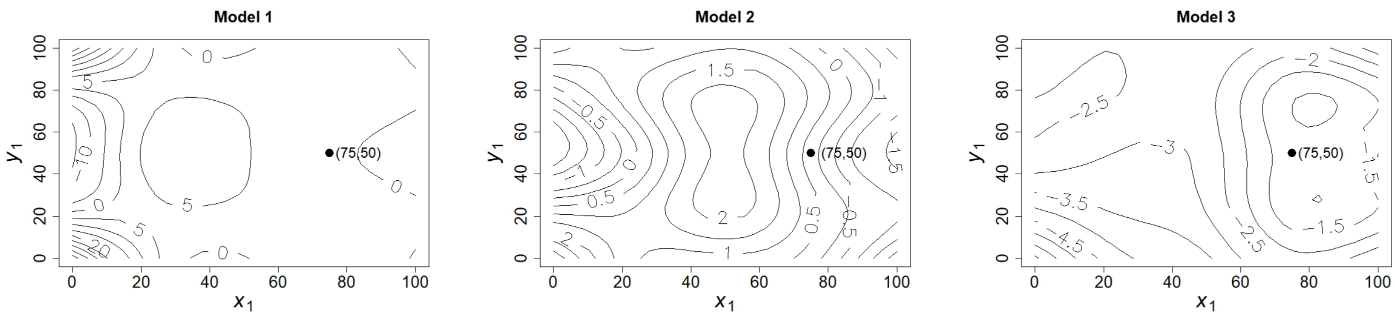
Panel A: Average position of player, $f_2(\bar{x_2}, \bar{y_2})$



Panel B: Interaction between game time and goal difference, $f_6(x_3, x_5)$



Panel C: The 4-D smooth with given start coordinates, $f_1(25, 50, x_1, y_1)$



Panel D: The 4-D smooth with given start coordinates, $f_1(75, 50, x_1, y_1)$

Figure 5: The resulting multidimensional smooth functions are depicted in panel A-D.

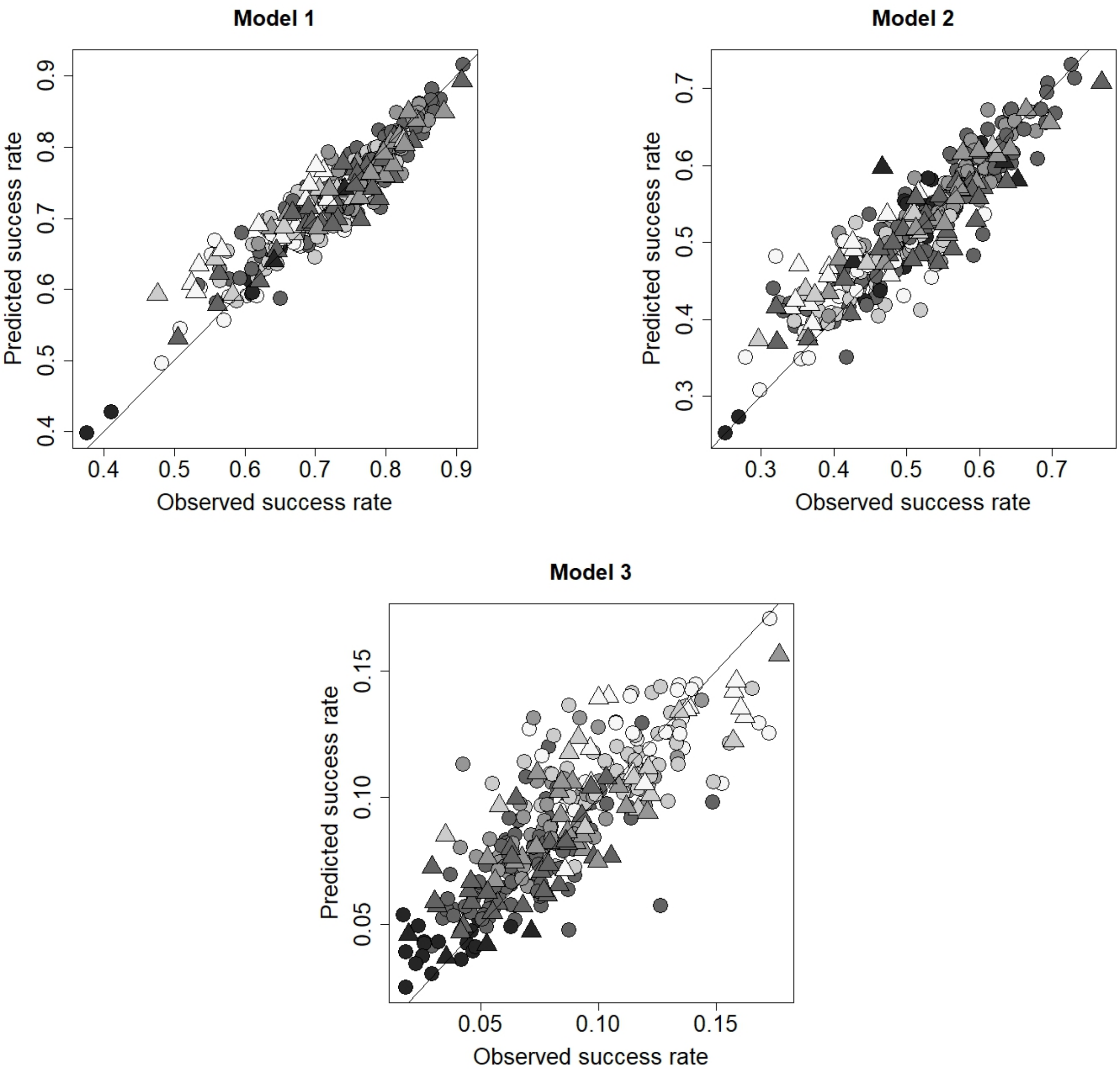


Figure 6: The predicted versus observed success rate for each model and each player in the 2017 season. The different shades of grey represent groups of player positions, and different symbols are used to tell whether the player is new to the 2017 season compared to the 2014-2016 seasons (▲) or not (●). A new player is treated as an average player with no random-effect player coefficient. Shades: ● - goalkeepers, ● - defenders, ● - central and defensive midfielders, ● - wingers and attacking midfielders, ○ - strikers. Only players with more than 100 passes are considered.

of 100 passes are plotted. The different shades of grey represent groups of player positions, and different symbols are used to tell whether a player is new to the 2017 season compared to the 2014-2016 seasons.

For Model 1 and Model 2, there are more strikers on the left side of the diagonal compared to the right side, indicating that the models predict a higher likelihood of success than what is actually observed. The same pattern can be seen for wingers and attacking midfielders, although not as clear. Considering the magnitude of the success rates, defenders, defensive midfielders and central midfielders tend to achieve higher success rates than goalkeepers and more offensive players. It is also interesting that the discrepancy between the predicted and observed success rates tend to be smallest for the case of high success rates. Overall, the largest discrepancy between the rates seems to be present for strikers, as seen from their larger deviance from the diagonal.

Considering Model 3, it is clear that the different groups of player positions cluster together. Moreover, the magnitude of the success rates are increasing in the forward playing direction, which is intuitive due to the nature of the model. In all models, the tendency is that the majority of the players lies on the left side of the diagonal, indicating a higher predicted success rate than what is actually observed. However, the distribution of offensive players in the effectiveness model seems to be more even.

The top ten outfield players in the 2017 season for each model are presented in Table 5. Note that only players attempting more than 209 passes are considered for the analysis. This is the equivalent of having attempted an average number of passes in 20 % of the matches, i.e. six matches, that season.

Defensive players dominate the top list for the difficulty model, while the proportion of offensive players in the lists for the two other models is higher. However, there are still five defensive players in the list for Model 3. As there are many passes made by defenders on their team's own half that have low potential for leading to shots, the model might not be able to give realistic expected values for those defenders that tend to be more active in the offensive play. Hence, offensive defenders are singled out compared to other defenders, although this does not mean that their ratings are fair compared to other player positions.

For comparisons of player ratings in the 2017 season, the developed models were rerun with data from the 2017 season and the resulting random-effect coefficients of the top ten outfield players are shown in Table 6. The idea is to see whether the model predictions can be used to rate players out of sample instead of having to run the models over again for each new season to get new coefficient estimates. When comparing the player ratings across the two approaches used, five, three and four of the same players on the top ten lists are present on both lists according to Model 1, Model 2 and Model 3 respectively. Two of these players are new to the 2017 data set, but still manages to be spotted by the models through predictions. Five more new players are present in Table 6, three of which played for Kristiansund, the newly promoted team. In general, the player position patterns are the same across the models for the two approaches.

4 Concluding Remarks

In this paper, the passing ability of football players in the Norwegian top division Eliteserien has been evaluated through the development of three GAMMs. Three aspects of a pass have been considered, all of which should be taken into account when evaluating the passing ability of a football player. The pass difficulty model, Model 1, is inspired by the model considered in Szczepański and McHale (2016), which is further examined by Tovar et al. (2017) and McHale and Relton (2018). Model 3, which handles pass potential, has a similar nature to what is examined in Power et al. (2017), Brooks et al. (2016) and Mackay (2017), whereas Model 2 is seemingly

Table 5: The top ten outfield players in the 2017 season. Explanation of the abbreviations used for the teams and the player positions can be found in Appendix B. Only players with more than 209 passes, the equivalent of playing six matches, are considered, and if a player has played for more than one team in the season considered, the team for which the player has played more matches is given.

Panel A: Model 1

#	*Player*	*Team*	*Pos*	*Expected*	*Actual*	*Obs*	*Ratio*
1	Daniel Braaten	BRA	ST	0.673	0.741	282	1.102
2	Vegar Hedenstad	RBK	FB	0.744	0.812	1348	1.090
3	Bonke Innocent	LSK	DM	0.757	0.823	294	1.087
4	Espen Ruud	ODD	FB	0.706	0.763	1504	1.081
5	Reiss Greenidge	SOG	CD	0.707	0.764	229	1.081
6	Thomas Grøgaard	ODD	FB	0.762	0.812	1219	1.067
7	Taijo Teniste	SOG	FB	0.681	0.727	673	1.067
8	Michael Haukås	VIK	WI	0.659	0.702	242	1.065
9	Martin Ellingsen	MOL	CM	0.761	0.810	406	1.065
10	Birger Meling	RBK	FB	0.778	0.829	981	1.065

Panel B: Model 2

#	*Player*	*Team*	*Pos*	*Expected*	*Actual*	*Obs*	*Ratio*
1	Daniel Braaten	BRA	ST	0.425	0.496	282	1.167
2	Anders Trondsen	S08	CM	0.532	0.608	1152	1.142
3	Herman Stengel	VIF	CM	0.593	0.677	1007	1.142
4	Jacob Rasmussen	RBK	CD	0.673	0.767	675	1.140
5	Vegar Hedenstad	RBK	FB	0.532	0.605	1348	1.137
6	Kasper Skaanes	BRA	WI	0.482	0.545	297	1.133
7	Enar Jääger	VIF	FB	0.646	0.731	1281	1.132
8	Jonatan Nation	VIF	CD	0.605	0.685	1771	1.132
9	Birger Meling	RBK	FB	0.572	0.646	981	1.131
10	Mathias Normann	MOL	CM	0.531	0.600	255	1.130

Panel C: Model 3

#	*Player*	*Team*	*Pos*	*Expected*	*Actual*	*Obs*	*Ratio*
1	Jostein Gundersen	TIL	CD	0.052	0.087	492	1.699
2	Lasse Nilsen	TIL	FB	0.097	0.149	444	1.535
3	Daniel Braaten	BRA	ST	0.120	0.152	282	1.388
4	Kim André Madsen	SIF	CD	0.055	0.076	476	1.373
5	Deyver Vega	BRA	WI	0.111	0.149	302	1.340
6	Gjermund Åsen	TIL	AM	0.117	0.156	867	1.333
7	Martin Broberg	ODD	WI	0.092	0.122	483	1.316
8	Vegard Bergan	ODD	CD	0.058	0.075	771	1.307
9	Vegard Forren	MOL	CD	0.064	0.084	392	1.307
10	Anthony Ikedi	FKH	CM	0.076	0.100	261	1.303

Table 6: The random-effect coefficients of the top ten players during the 2017 season according to the models. Abbreviations for teams and positions are explained in Appendix B. The team for which the player has played most matches in the season considered is given.

Rank	*Model 1*				*Model 2*				*Model 3*			
	Player	*Team*	*Pos*	*Coef*	*Player*	*Team*	*Pos*	*Coef*	*Player*	*Team*	*Pos*	*Coef*
1	Espen Ruud	ODD	FB	0.383	Henning Hauger	SIF	DM	0.230	Gjermund Åsen	TIL	AM	0.090
2	Vegar Hedenstad	RBK	FB	0.367	Jacob Rasmussen	RBK	CD	0.218	Jostein Gundersen	TIL	CD	0.085
3	Thomas Grøgaard	ODD	FB	0.362	André Danielsen	VIK	CM	0.203	Kaj Ramsteijn	AaFK	CD	0.083
4	Henning Hauger	SIF	DM	0.347	Mikkel Kirkeskov	AaFK	FB	0.195	Liridon Kalludra	KBK	WI	0.083
5	André Danielsen	VIK	CM	0.328	Ulrik Yttergård Jensen	TIL	CD	0.186	Lasse Nilsen	TIL	FB	0.077
6	Taijo Teniste	SOG	FB	0.325	Thomas Grøgaard	ODD	FB	0.185	Christian Grindheim	VIF	CM	0.076
7	Vito Wormgoor	BRA	CD	0.288	Christoffer Aasbak	KBK	WB	0.184	Fredrik Midtsjø	RBK	CM	0.067
8	Bonke Innocent	LSK	DM	0.288	Kristoffer Haraldseid	FKH	WB	0.184	Eirik Hestad	MOL	CM	0.067
9	Jonatan Nation	VIF	CD	0.288	Jonatan Tollås Nation	VIF	CD	0.184	Vegar Hedenstad	RBK	FB	0.058
10	Nikita Baranov	KBK	CD	0.283	Enar Jääger	VIF	FB	0.180	Vegard Bergan	ODD	CD	0.058

unique by incorporating a player's ability to make passes that can be successfully followed up by the recipient.

The coefficients of Model 1 correspond well with the findings in Szczepański and McHale (2016) in terms of both sign and magnitude. Considering the smooth functions, the time between events provides similar results as the time between passes in Szczepański and McHale (2016), except that the former function continues beyond the eight seconds cut-off used in previous research. Furthermore, the effect of game time, the average player positions and the 4-D smooth function, when passing from the defensive region, have similar patterns as found by Szczepański and McHale (2016). The predictions made for pass difficulty in the consecutive season seem to be a bit more accurate in Szczepański and McHale (2016). However, this may be explained by the fact that players that do not appear in the training data are considered to be average players in the predictions. The home team advantage variable, indicating that the home team has a higher probability of making a successful pass compared to the away team, supports the previous findings by McHale and Relton (2018) and Tovar et al. (2017).

Although the research previously done on pass potential is not directly comparable to the approach used for Model 3, both Power et al. (2017); Brooks et al. (2016) support the finding that passes made closer to the opponent's goal post are more likely to lead to shots. In Mackay (2017), where not only passes are considered as events in a possession, all variables concerning the previous event of passes in Model 3, except for free kicks, are the same. Of these, all but one have the same sign on the coefficients as Model 3. Also, similar negative effects are found for headed passes in the current event. Offensive players tend to dominate the top ratings in Power et al. (2017); Brooks et al. (2016); Mackay (2017), which is different from the top list for Model 3 where more variation in player positions is present. This might indicate that Model 3 more properly deals with differences in player positions. Nevertheless, when using the models to rank players, the fact that different player roles and positions may be associated to different passing distributions should be taken into account.

Model 3 (potential) produced a PR curve with an unusual shape. One interpretation is that regression on generated shots is not suitable. Alternative types of models could for example be based on modelling hazard rates (Volf, 2009). This was applied to basketball by Cervone et al. (2014, 2016), to calculate the expected value of possessions, which corresponds fairly well to the idea of Model 3, although considering points scored instead of shots taken.

The three developed models' capabilities of assessing players upon their passing abilities have been demonstrated in this paper. Furthermore, the top ten rated pass makers in Eliteserien, and the teams' skills of facilitating and hampering passes were identified.

The developed models can provide coaches and players with valuable information that can be transferred to training sessions, with the aim of increasing performance. As just one example, being able to identify the best passers on the opposing team, may allow a coach to modify the defensive tactics to better deal with the opponent.

A recurrent theme in many of the findings of the pass potential model is that they might be explained as effects from counter-attacks. This indicates that teams in the Norwegian top division may benefit from putting a low pressure on the opponents and awaiting counter-attack opportunities.

The seasonal effects on ground conditions seem to be captured by the models as the perceived difficulty of maintaining natural grass during the near-winter months was supported. Additionally, passes appear to be more easily made on artificial turf. From this, it may follow that teams should choose a game plan depending on the state of the surface on which the match is played. Players may also choose to carefully decide which types of passes to make on different types of turf. Passes made along the ground may be easier to steer in the right direction when playing on artificial turf, whereas long passes in the air may be relatively more beneficial on natural turf, depending on the ground conditions.

The same type of analysis as shown in this paper can also be performed for any other league for which similar event-data is available, including all the top leagues in Europe. It may be that some of the results and insights will not carry over to other leagues. For example, the use of artificial turf and the challenges of ground conditions are more prominent in the Norwegian league system than in most other countries. In addition, playing styles of teams and the general abilities of players may differ for other leagues, thereby leading to different conclusions regarding which types of passes are more effective. Finally, the models and the corresponding analysis can be extended to other sports where passing between players is essential.

Acknowledgements

The authors are grateful for insightful comments provided during the revision of the paper, and wishes to thank two anonymous reviewers, an associate editor, and an editor. Furthermore, Opta Sports provided access to relevant data from Eliteserien, without which this work could not have been performed. Finally, football coach Hugo Pereira is thanked for his input and feedback during the development of the models.

References

S. Boor, C. Hanson, and C. Ross. Rising stars. `https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html`, 2018.

J. Brooks, M. Kerr, and J. Guttag. Developing a data-driven player ranking in soccer using predictive model weights. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 49–55. ACM, 2016.

D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry. POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data. *Proceedings of the MIT Sloan Sports Analytics Conference*, 2014.

D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry. A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514):585–599, 2016.

ClubElo. API. http://clubelo.com/API, 2018. Accessed on 23/1/2018.

Eliteserien. Stadioner. `https://www.eliteserien.no/stadioner`, 2018. Accessed on 22/1/2018.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

A.E. Goodenough, A.G. Hart, and R. Stafford. Regression with empirical variable selection: description of a new method and application to ecological datasets. *PLoS One*, 7(3):e34338, 2012.

F. Greig. Dexterity, accuracy and FIFA skills – the people behind opta's football data. `https://inews.co.uk/sport/football/opta-data-analyst-stats-football/`, 2017. Accessed on 9/3/2018.

L. Gyarmati and R. Stanojevic. Qpass: a merit-based evaluation of soccer passes. *arXiv preprint arXiv:1608.03532*, 2016.

D.W. Hosmer Jr, S. Lemeshow, and R.X. Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, Hoboken, New Jersey, 2013.

M. Hughes and I. Franks. Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, 23(5):509–514, 2005.

L.M. Hvattum. Playing on artificial turf may be an advantage for Norwegian soccer teams. *Journal of Quantitative Analysis in Sports*, 11:183–192, 2015.

C.M. Judd, G.H. McClelland, and C.S. Ryan. *Data analysis: A model comparison approach.* Routledge, Abingdon, UK, 2011.

X. Lin and D. Zhang. Inference in generalized additive mixed modelsby using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):381–400, 1999.

N. Mackay. Predicting goal probabilities for possessions in football. Master of science, Vrije Universiteit Amsterdam, 2017. URL `https://beta.vu.nl/nl/Images/werkstuk-mackay_tcm235-849981.pdf`.

I.G. McHale and S.D. Relton. Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research*, 268(1):339–347, 2018.

M. Mohr, P. Krustrup, and J. Bangsbo. Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, 21:519–528, 2003.

Opta Sports. World leaders in sports data. `https://www.optasports.com/`, 2018. Accessed on 13/4/2018.

R. Pollard and G. Pollard. Home advantage in soccer: a review of its existence and causes. *International Journal of Soccer and Science*, 3:28–38, 2005.

P. Power, H. Ruiz, X. Wei, and P. Lucey. Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1605–1613. ACM, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

R. Rein, D. Raabe, and D. Memmert. "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science*, 55:172–181, 2017.

L. Steinberg. Changing the game: The rise of sports analytics. `https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics`, 2015. Accessed on 13/2/2018.

Ł. Szczepański and I. McHale. Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2):513–533, 2016.

Total Sportek. 25 world's most popular sports (ranked by 13 factors). http://www.totalsportek.com/most-popular-sports/, 2007. Accessed on 12/2/2018.

J. Tovar, A. Clavijo, and J. Cárdenas. A strategy to predict association football players' passing skills. Technical report, Universidad de Los Andes, 2017.

P. Volf. A random point process model for the score in sport matches. *IMA Journal of Management Mathematics*, 20(2):121–131, 2009.

WhoScored.com. Tippeligaen, 2018. URL https://www.whoscored.com/Regions/165/Tournaments/41/Seasons/6115/Norway-Eliteserien.

S.N. Wood. *Generalized additive models: an introduction with R.* Chapman and Hall/CRC, Boca Raton, Florida, 2006.

S.N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36, 2011.

A Regression Results

Table 7: Model 1 regression result. The random effect coefficients are not given here, they are presented in Table 4. See the note about significance.

Fixed effects		
Variable	*Coefficient(SE)*	
$X_{1.2}$	0.351^{***}	(0.068)
$X_{1.3}$	0.498^{***}	(0.014)
$X_{1.4}$	0.580^{***}	(0.013)
X_2	0.343^{***}	(0.015)
X_3	-0.100^{***}	(0.024)
X_4	-0.327^{***}	(0.024)
$X_{5.1}$	0.154^{**}	(0.052)
$X_{5.3}$	0.757^{***}	(0.043)
X_6	0.515^{***}	(0.021)
X_7	-0.311^{***}	(0.016)
X_8	-1.227^{***}	(0.014)
X_9	0.138^{***}	(0.008)
X_{11}	0.184^{***}	(0.045)
$X_{13.1}$	0.226^{***}	(0.036)
$X_{13.3}$	$-0.215^{\dagger}$	(0.122)
X_{14}	0.123^{***}	(0.011)
$X_{1.2}{*}X_{10}$	0.306^{***}	(0.036)
$X_{1.2}{*}X_{11}$	$0.225^{\dagger}$	(0.122)
Intercept	1.299^{***}	(0.068)

Smooth terms	
Variable	*Sign.*
$f_1(x_0, y_0, x_1, y_1)$	***
$f_2(\bar{x_2}, \bar{y_2})$	***
$f_3(x_3)$	
$f_4(x_4)$	***
$f_5(x_5)$	***
$f_6(x_3, x_5)$	***
$f_7(x_7)$	***
$f_8(x_8)$	***
$f_9(x_9) : Artificial$	***
$f_9(x_9) : Natural$	***

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table 8: Model 2 regression result. The random effect coefficients are not given here, they are presented in Table 4. See the note about significance.

Fixed effects		
Variable	*Coefficient*	*(SE)*
$X_{1.2}$	0.311^{***}	(0.012)
$X_{1.3}$	0.471^{***}	(0.011)
$X_{1.4}$	0.515^{***}	(0.012)
X_2	0.189^{***}	(0.029)
X_3	-0.596^{***}	(0.062)
X_4	-0.148^{***}	(0.021)
$X_{5.1}$	0.097^{*}	(0.042)
$X_{5.2}$	0.532^{***}	(0.066)
$X_{5.3}$	0.387^{***}	(0.036)
X_6	0.277^{***}	(0.017)
X_7	-0.294^{***}	(0.014)
X_8	-0.968^{***}	(0.014)
X_9	0.117^{***}	(0.006)
X_{10}	-0.181^{**}	(0.057)
X_{12}	-0.172^{***}	(0.050)
$X_{13.1}$	0.120^{*}	(0.060)
$X_{13.3}$	-0.367^{***}	(0.109)
X_{14}	0.187^{***}	(0.009)
$X_{1.2}$*X_{10}	0.543^{***}	(0.063)
$X_{1.2}$*X_{11}	0.447^{***}	(0.109)
Intercept	-0.517^{***}	(0.085)

Smooth terms	
Variable	*Sign.*
$f_1(x_0, y_0, x_1, y_1)$	***
$f_2(\bar{x_2}, \bar{y_2})$	***
$f_3(x_3)$	**
$f_4(x_4)$	***
$f_5(x_5)$	***
$f_6(x_3, x_5)$	***
$f_7(x_7)$	***
$f_8(x_8)$	***
$f_9(x_9) : Artificial$	***
$f_9(x_9) : Natural$	***

Note: *p<0.05; **p<0.01; ***p<0.001

Table 9: Model 3 regression result. The random effect coefficients are not given here, they are presented in Table 4. See the note about significance.

Fixed effects		
Variable	*Coefficient*	
$X_{1.2}$	0.240^{***}	(0.019)
$X_{1.3}$	0.223^{***}	(0.018)
$X_{1.4}$	0.223^{***}	(0.020)
X_2	0.325^{***}	(0.036)
X_3	-0.134^{**}	(0.050)
X_4	0.091^{*}	(0.036)
$X_{5.3}$	0.338^{***}	(0.059)
X_6	0.289^{***}	(0.029)
X_7	-0.110^{***}	(0.025)
X_8	-0.785^{***}	(0.026)
X_9	0.092^{***}	(0.010)
X_{11}	-0.106^{***}	(0.028)
X_{12}	0.363^{***}	(0.080)
$X_{13.1}$	$-0.156^{\dagger}$	(0.088)
X_{14}	0.127^{***}	(0.014)
$X_{1.2}{}^{*}X_{10}$	-0.093^{*}	(0.040)
Intercept	-3.248^{***}	(0.118)

Smooth terms	
Variable	*Sign.*
$f_1(x_0, y_0, x_1, y_1)$	***
$f_2(\bar{x_2}, \bar{y_2})$	***
$f_3(x_3)$	
$f_4(x_4)$	***
$f_5(x_5)$	***
$f_6(x_3, x_5)$	
$f_7(x_7)$	***
$f_8(x_8)$	***
$f_9(x_9) : Artificial$	***
$f_9(x_9) : Natural$	***

Note: †p<0.1; *p<0.05; **p<0.01; ***p<0.001

B Abbreviations

Table 10: Team name abbreviations and the seasons of 2014-2017 in which the teams have played in Eliteserien.

Abbreviation	*Explanation*	*Seasons*
AaFK	Aalesunds FK	'14,'15,'16,'17
B/G	FK Bodø/Glimt	'14,'15,'16
BRA	SK Brann	'14,'16,'17
FKH	FK Haugesund	'14,'15,'16,'17
KBK	Kristiansund BK	'17
LSK	Lillestrøm SK	'14,'15,'16,'17
MJØ	Mjøndalen IF	'15
MOL	Molde FK	'14,'15,'16,'17
ODD	ODDs BK	'14,'15,'16,'17
RBK	Rosenborg BK	'14,'15,'16,'17
S08	Sarpsborg 08 FF	'14,'15,'16,'17
SAN	Sandefjord Fotball	'15,'17
SIF	Strømsgodset TF	'14,'15,'16,'17
SOG	Sogndal Fotball	'14,'16,'17
STA	IK Start	'14,'15,'16
STB	Stabæk Fotball	'14,'15,'16,'17
TIL	Tromsø IL	'15,'16,'17
ULF	Sandnes Ulf	'14
VIF	Vålerenga Fotball	'14,'15,'16,'17
VIK	Viking FK	'14,'15,'16,'17

Table 11: Player position abbreviations

Abbreviation	*Explanation*
AM	Attacking midfielder
CD	Central defender
CM	Central midfielder
DM	Defensive midfielder
FB	Full back
ST	Striker
WB	Wing back
WI	Winger