We use the same example for which the *PPCF* was presented in Figure 1. In Figure 5c, we see a probability density map that can be integrated to determine which receivers are most likely. In this case, there is an obvious easy back pass to Red #70 and this player represents the most likely target with a less likely outlet pass to Red #7 or Red #9. It is also possible that a forward through ball will be attempted to Red #12 or a long cross to Red #27.

According to the *opportunity* model presented in Figure 5d, the most dangerous region is to the left of the goal where Red #12 is available to receive a pass in the space that Blue #27 is currently occupying but cannot control because of their speed towards the end line. A shot from distance is available if the ball is passed to Red #9 or Red #7 and the cross to Red #27 on the right side could also result in a score if properly executed. Overall, this is not a scenario that represents a high probability of scoring. The total *opportunity* of scoring with the next ball touch integrates to 1.1%.

3.5. Parameter Estimation

The probability given in Equation 2 allows us to write the total likelihood for each on-ball event in the training set as follows:

$$\mathcal{L}(T|\boldsymbol{\theta}) = \prod_{D\in T} \begin{cases} P(G|D,\boldsymbol{\theta}) & k=1 \\ 1-P(G|D,\boldsymbol{\theta}) & k=0 \end{cases} \tag{8}$$

Where T represents the training set of events, k indicates whether the subsequent on-ball event is a goal $k=1$, or not $k=0$, and $\boldsymbol{\theta}$ represents the vector of model parameters. For fitting, a set of five games training games is reserved from the total of 58 and excluded from the rest of the analysis. Due to the small size of the training data and issues with event data synchronization discussed in Section 2.1, we use a Bayesian approach with normally distributed priors to estimate the maximum a posteriori probability (MAP) for each parameter. Detailed information about the model parameters comprising $\boldsymbol{\theta}$ and their MAPs is found in Table 1.

Parameter	MAP	Units	Description	Prior Parameters	Prior Selection
$\boldsymbol{s}$	0.54	Seconds	Temporal uncertainty on player-ball intercept time.	$\mu=0.5, \sigma=0.1$	Use value from fit in [7]
$\boldsymbol{\lambda}$	3.99	Hz	$1/\lambda$ is proportional to the average time it takes a player to control the ball.	$\mu=4.2, \sigma=2.0$	Use value from fit in [7]
$\boldsymbol{\kappa}$	1.72	None	Defensive advantage, scales control rate for defending players.	$\mu=1.5, \sigma=0.5$	We expect moderate defensive advantage
$\boldsymbol{\sigma}$	23.9	Meters	Related to the mean distance between on-ball events.	$\mu=14, \sigma=10$	Use mean distance computed in Figure 3
$\boldsymbol{\alpha}$	1.04	None	Preference for maintaining possession increases with α.	$\mu=1, \sigma=0.2$	Expect approximate proportionality with PPCF
$\boldsymbol{\beta}$	0.48	None	Small values improve chance of scoring further from the goal.	$\mu=0.5, \sigma=0.5$	Open space improves scoring probability[6].

Table 1. A table describing the model parameters, the prior used and the maximum a posteriori probability (MAP).

4. Validation

In constructing the *off-ball scoring opportunity* model, our objective has been to produce a leading indicator of scoring that is less stochastic than scoring itself and allows us to assign opportunity

[6] In this model, β can be thought of as fudge factor to ensure that the resultant model can be integrated to give expected scoring. In future iterations of the analysis, this parameter can be replaced by using an improved scoring probability model.

creation to players even if no shot or score results. In other words, we want to build a metric that is more predictive of future scoring than scoring itself is. To validate that we have achieved this, we compute the correlation between three leading indicators of future scoring: 1) *OBSO* 2) shots and 3) goals themselves and we compare them to the scoring in a subsequent match on a per-player basis. Looking at Table 2, we see that *OBSO* is more correlated with future goals than either shots or goals themselves are.

i\i+1	OBSO	Shots	Goals
OBSO	0.60	0.37	0.26
Shots	~	0.35	0.17
Goals	~	~	0.12

Table 2. The Pearson correlation coefficient (PCC) between the current (ith) and the subsequent (i+1th) game for players in the 53 game test set for three performance indicators: OBSO, shots, and goals.

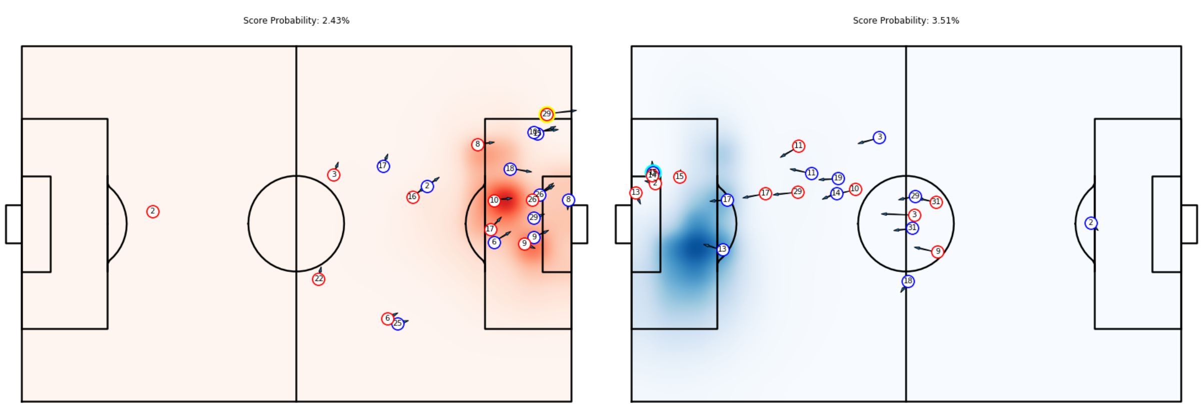
In addition, *OBSO* is correlated with both itself and with future shots. This indicates that *OBSO* can be used as a leading indicator of player performance that is less stochastic than goals or shots themselves are.

5. Applications

In this section, we use the *off-ball scoring opportunity (OBSO*) model to propose applications in four categories: 1) tactical moment analysis 2) match analysis 3) team performance and 4) player performance. Although we believe that the control model (*PPCF),* seen in Figure 5b, and the transition model, seen in Figure 5c, have many interesting applications apart from scoring, these remain outside the scope of this paper. The applications discussed in the subsequent sections focus on *off-ball scoring opportunities*.

5.1. Tactical Moment Analysis

An important aspect of opposition analysis and post-match analysis for the analyst at a soccer club lies in identifying critical moments during the course of the match. This is a time consuming job and requires many hours of video review. The *opportunity* model presented in this paper both *quantifies* and *visualizes* scoring *opportunities*. The integrated magnitude of the *opportunities* can help in choosing which clips to watch while the *opportunity* maps themselves provide additional insight into the *opportunity*. Four examples are shown below.



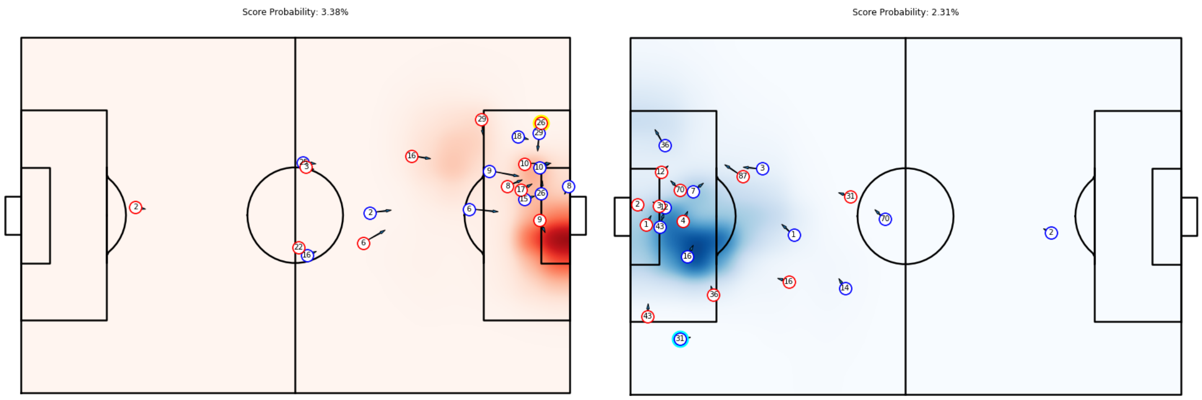


Figure 6. This figure shows selected opportunity maps. a) Upper left: Red #29 has the ball. The integrated scoring chance is 2.43%. b) Upper right: Blue #15, obscured by Red #14, has the ball, integrated scoring chance is 3.51%. c) Lower left: Red #26 has the ball. The integrated scoring chance is 3.38%. d) Lower right: Blue #31 has the ball. The integrated scoring chance is 2.31%.

In Figure 6a, Red #29 controls the ball and is running quickly towards the end line.[7] Meanwhile, Red #10 and Red #17 have created space at the top of the box and are open to receive a lobbed cross. There are other target options including Red #9 who is closer to the goal but because of the defensive coverage from Blue #9 and the distance between Red #29 and Red #9, there is a lower probability of scoring from this pass target.

In Figure 6b, Blue #15 (obscured by Red #14) has the ball and appears to have attracted the attention of the defense. This has left Blue #13 free to make a run into the box in excellent scoring position. The pass will be difficult in a way that is not fully captured by the model due to the proximity of the defenders on Red #14, but if the pass reaches Blue #13, it is an easy score.

In Figure 6c, Red #26 is on the left side of the box. There are a number of teammates near him inside the box, but the best scoring chance comes from the long cross to Red #9 who is wide open on the six-yard line near the far post.

In Figure 6d, Blue #31 has the ball and is relatively free to run into the box himself,[8] but Blue #16 is open and in excellent position to receive a pass and take a shot before being closed down by the defense. The distance from goal and the presence of the keeper limits the overall magnitude of this chance.

5.2. Match Analysis

By integrating the *opportunity* maps throughout the course of the match, we can get a sense of how the match played out spatially and temporally.

[7] One limitation of our model in its current form is that it does directly not account for the difficulty of crossing while running at speed. This has the effect of over-estimating the scoring chance from a scenario such as that in Figure 6a while underestimating the scoring chance for a passer who is stationary and under no defensive pressure.
[8] Note: the opportunity for the player in possession of the ball to score is not considered in the *OBSO* model since this is accounted for during the previous moment before the ball was passed to this player.

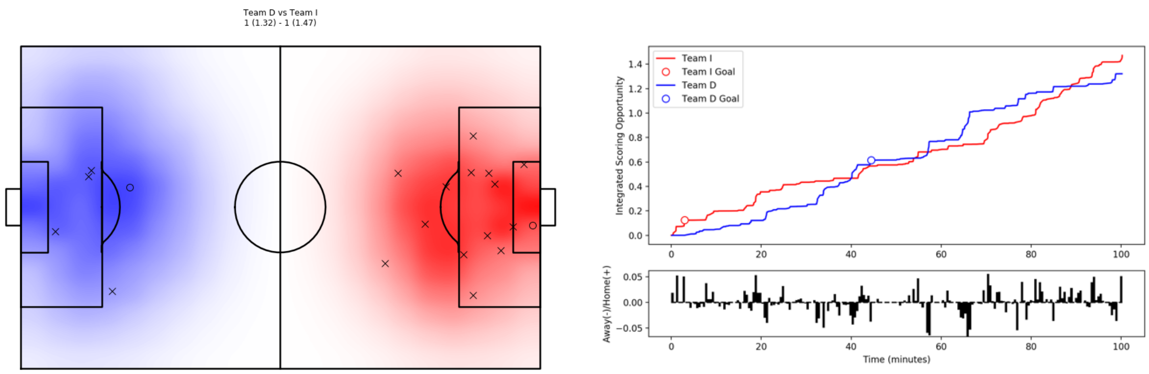


Figure 7. Match analysis figures for a game between Team I (home team in red) and Team D (away team in blue) with a final score line of 1-1. a) The left figure represents the time-integrated spatial opportunity map. A shot is denoted by an "x" and goal is denoted by an "o". b) The right figure demonstrates a space-integrated time varying opportunity map. The upper plot shows the integrated scoring opportunity versus time while the bottom plot shows the instantaneous scoring opportunity per team with positive values for the home team and negative values for the away team. Scores are denoted with an "o".

Figure 7 shows possible visualizations that can be used to represent the *opportunity* model during the course of a match. In this match, Team I dramatically outshot Team D which would generally lead to a mismatch in the number of expected goals in favor of Team I, however, their integrated scoring opportunities are remarkably similar, 1.32 versus 1.47, a similarity reflected in the final score line of 1-1. Notice how in the time-varying plot on the right, Team I creates many opportunities in the first 20 minutes resulting in one score. At the 30-minute mark, the momentum begins to shift to Team D culminating in a score right before the half.

5.3. Team Performance

As with expected goals, *off-ball scoring opportunity* can be used to identify trends at the team and at the player level. Over a large number of games, we would expect the average *OBSO* to regress to the average number of goals scored, but because *OBSO* does not include information about whether the *opportunities* were capitalized on, deviation from the average correlation between goals scored and *OBSO* can be used as a proxy to measure team decision making within and around the penalty area and of finishing skill.

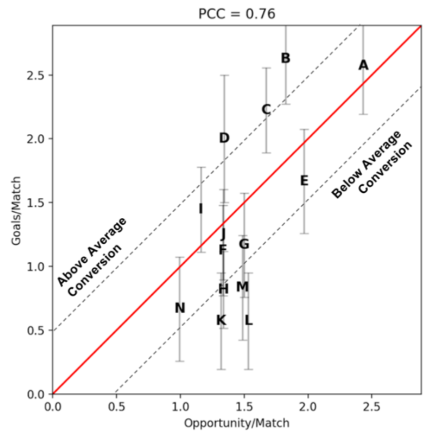


Figure 8. The average goals/match plotted versus the average opportunity/match for the 14 teams in the test set of 53 games. The Pearson correlation coefficient is found to be 0.76.

In Figure 8, we can see how each team's goals per match compares to their *opportunity* per match. As expected, the higher ranked teams tend to have higher goals scored and *opportunity* created per match. Interestingly, there is a cluster of lower ranked teams (Team H, Team K, Team L, and Team M) that all exhibit below average *opportunity* conversion. Apart from the element of randomness in scoring, there are three main reasons that could explain this worse than expected conversion rate: 1) inferior skill in passing or receiving 2) inferior awareness of opportunities and/or 3) inferior finishing.

5.4. Player Performance

As with our analysis of teams, we can highlight player performance and look for trends in player behavior. In Table 3, we can see information about the top 20 players ranked by their per-match *off-ball scoring opportunity*. Unsurprisingly, the majority are center forwards but a few who do not play that position are also in the top 20. Most interesting is the right back from Team A who is clearly an attacking threat as borne out by their scoring production and their mean *OBSO*/match.

Team	Rank	Position	Mean OBSO/Match	Mean Goals/Match	# Matches
Team E	**1**	Center Forward	0.44	0.00	1
Team A	**2**	Center Forward	0.34	0.20	5
Team A	**3**	Center Forward	0.34	0.29	7
Team E	**4**	Center Forward	0.32	0.17	6
Team J	**5**	Center Forward	0.31	0.00	2
Team C	**6**	Center Forward	0.29	1.00	5
Team C	**7**	Right Winger	0.28	0.67	9
Team B	**8**	Center Forward	0.28	0.50	8
Team B	**9**	Center Forward	0.27	0.86	7
Team G	**10**	Center Forward	0.26	0.17	6
Team A	**11**	Center Forward	0.26	0.14	7
Team A	**12**	Attacking Midfielder	0.25	0.43	7
Team C	**13**	Center Forward	0.25	0.56	9
Team L	**14**	Right Midfielder	0.24	0.20	5