Off-Puck Scoring Opportunities: Valuing Offensive Player Movement in Ice Hockey

Jules Lanari-Collard - Imperial College London

Submission for HALO 2026 Hackathon

Statement of Impact

In this project, we develop a model for **off-puck scoring opportunity**, which quantifies the value of offensive player off-puck movement. Movement off the puck is an integral part of an ice hockey player's skillset, but is difficult to analyse without detailed tracking data. The model is shown to have applications in player evaluation, tactical analysis, and breaking down individual in-game sequences.

1 Introduction

Over the course of an ice hockey game, any given player is actually in control of the puck for very little time compared to the overall time on ice. Therefore, a major component of player performance is off-puck behaviour, both on offence and defence. Focusing on the offensive case, we explore the occupation of space, which can be broken down into three questions:

1. **Control**: to what extent does a player control the space they occupy?
2. **Availability**: how likely is it that the player can receive the puck in the space they occupy?
3. **Danger**: how dangerous is the space occupied by the player?

In this project, building on ideas from Spearman [3] in soccer analytics, we exploit the depth of the data provided to develop tracking-based models for each of these questions, which together provide a useful tool for not only analysis and visualisation of in-game sequences, but also quantifying the ability of players to both generate and occupy valuable spaces on the ice. We demonstrate applications to pre- and post-game analysis, as well as player and team evaluation.

For simplicity, we restrict our attention to 5-on-5 play and exclude empty-net situations. It should be noted that similar work has been done by Inayatali and Chan on a sample of NHL data [2], however there are differences with this project in both the type of data and methodology used. In particular, they did not have access to event-level pass data, and their expected goals model did not consider the context of defender positioning.

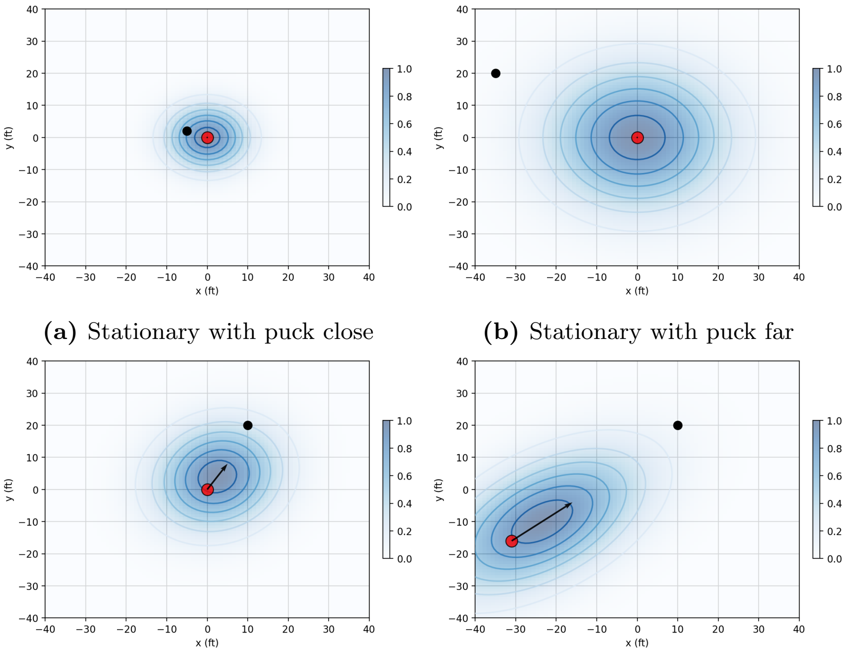
2 Methods

Our goal is to model the scoring threat posed by the off-puck offensive players, which we will call *off-puck scoring opportunity* (OPSO). Specifically, for a given location on the ice, we model this as the probability of the attacking team successfully completing a pass to that location, and subsequently scoring. To give some intuition, here are some examples:

- A team attempting to break out of their defensive zone: many teammates will likely be open and ready to receive a pass (either in the defensive or neutral zones), however they are all unlikely to score immediately from that position, so the overall OPSO should be low.
- A 2-on-1 in the offensive zone, with the defender committed to the puck-carrier: the supporting offensive player is open and likely to score if they receive the puck, generating a high OPSO.

These examples are not exhaustive, but hopefully provide some intuition for the concept. On its own, this is a complex problem to model, so we implement the framework suggested by Spearman [3], which breaks the problem down into

$$\text{OPSO} = \text{Danger} \times \text{Control} \times \text{Availability}.$$



(a) Stationary with puck close **(b)** Stationary with puck far

(c) Moving with puck close **(d)** Moving with puck far

Figure 1: Player Influence

This is a simplification of the full model; see appendix A for the full specification. The *danger* term represents the likelihood of scoring from the given location, *control* represents the likelihood of controlling the puck if it reaches that location, and *availability* represents the likelihood of the puck being passed to that location.

In the following sections, we formulate models for each of these terms, which are then combined to form a full OPSO model. The models can be fit to any event which has player tracking and velocity data.

2.1 Control

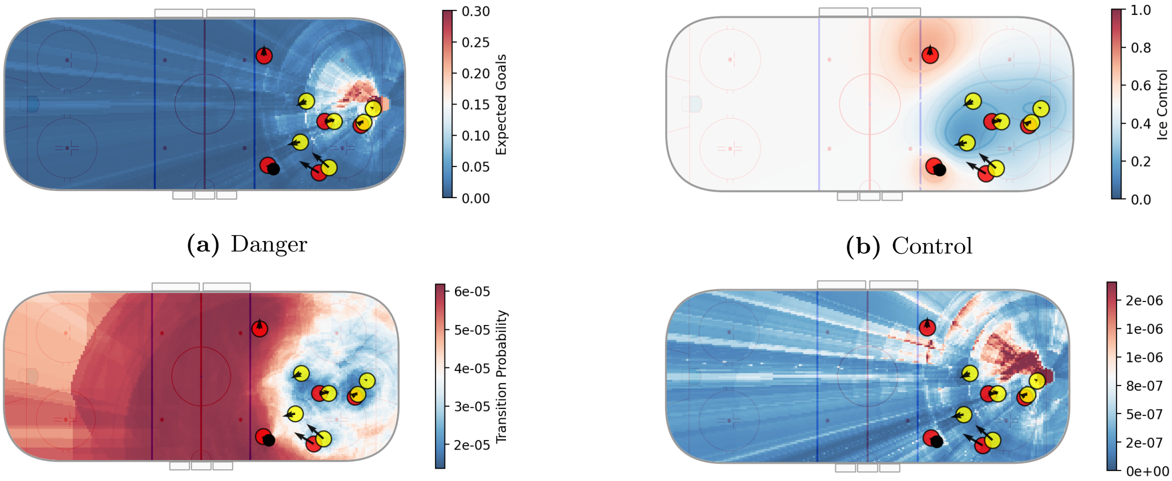
The control model must quantify, for each location on the ice, how likely the offensive team is to control the puck, if a pass is made to that location. We will call the individual control exerted by a player *player influence*, and call the overall team-wide control *ice control*, inspired by the concept of *pitch control* introduced for soccer by Spearman [3].

In his implementation, Spearman [3] leverages highly-detailed tracking data of player and ball motion to develop a complex physics-based model. As the data available to us is event-based, we turn to a simpler model, also developed for soccer by Fernandez and Bornn [1]. We model player influence as an ellipse, whose shape depends on the speed, direction and distance from the puck of the player. The latter aims to capture the notion that the further away from the puck a player is, the more time they have to react and move towards it. Illustrations of the model are given in figure 1, where we see how the shape and size of a player's influence area varies in different situations. When combined across all players on the ice, we obtain a good representation of which teams control which areas of the ice, as in figure 2b.

2.2 Availability

Given any on-ice state, the transition model aims to predict which locations (and hence players) are most likely receive a pass from the puck carrier. In effect, which players are most 'open' to a potential pass? We use the successful and unsuccessful passes in the event data to train the model, considering the position of the passer, potential target, distance to the target, and relative locations of defensive and offensive players.

A key assumption is that the probability of a pass being to a particular location is proportional to its success probability. This seems intuitive, although may not hold in practice. However, removing this assumption increases the complexity of the problem significantly, and is beyond both the time and computational scope of this project. Even if the assumption does not hold, it likely provides a reasonable approximation of passing tendencies.



(a) Danger

(b) Control

(c) Availability

(d) Combined Model (OPSO)

Figure 2: Visualisation of OPSO Model

2.3 Danger

The danger model estimates the likelihood of scoring from different areas of the ice, given that the attacking team successfully controls the puck at that area. This is tantamount to an *expected goals* (xG) model. We use the even-strength shot data to fit an xG model which considers not only the distance and angle from the goal, but also the presence of defenders in the shooting lane and their proximity to the shooter.

2.4 Combined Model

Finally, we can combine the models to determine the off-puck scoring opportunity (OPSO) across the ice surface. For illustration, we choose a particular event which reflects a controlled possession in the offensive zone, with the puck marked as a black circle (located on the right of the blue line). The surfaces for each separate model and the combined OPSO model are shown in figure 2.

The danger model recognises that, due to the positioning of the defenders (yellow), most of the shooting lanes on the right side (offensive point of view) are blocked, and the only dangerous shots available would come from the left side. The control model provides a useful visualisation of the state of the ice. The two offensive (red) players in front of the net and in the slot are completely nullified by the defenders surrounding them, and the only areas of control for the offensive team are at the blue line. The availability model provides a similar picture, inferring that the only offensive player open for a safe pass is the player on the left side of the blue line.

In combination, we see that most of the offensive OPSO is generated by the player on the left side of the blue line. Although it is not controlled by any offensive players, the area in front of the goal is deemed high value as it is highly dangerous and not well-covered by the defensive players. The team-wide OPSO is a relatively low 0.77%, reflecting the fact that the defensive team is well-positioned and containing the offense effectively.

3 Applications

One of the main attractions of the OPSO model is that we can find dangerous scoring opportunities from the data, regardless of whether those opportunities actually resulted in a shot on net. One of the limitations of xG as a metric for chance quality is that it only values shots, and cannot value opportunities which *almost* happened, such as a player whiffing on a slap shot, the puck being intercepted just before it reaches a wide-open potential shooter, or the player in possession not identifying open teammates. Such opportunities do not show up in the match data, but are important in telling the story of a game.

Another benefit is that OPSO can be interpreted both on a granular play-by-play level, and also across games and seasons. This provides applications in coaching, pre- and post-game analysis, and front office decision-making.

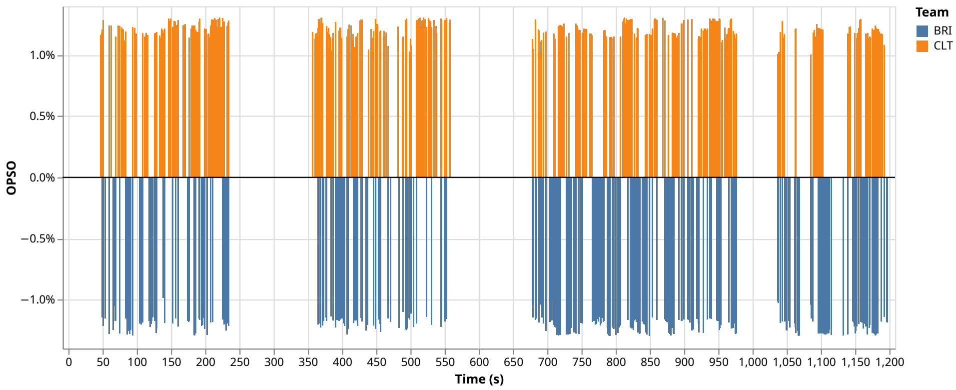


Figure 3: OPSO Timeline Example (BRI @ CLT 2023-10-27, 1st Period)

3.1 Game Analysis

Each tracked event has an associated OPSO value, so we can monitor the OPSO generated by each team over the course of a game, which can be used to visualise match momentum. An example of this for a single period is shown in figure 3, with the orange and blue bars representing dangerous opportunities for the home and away teams respectively. We can identify strong spells for each team by the clusters of bars on the graph; the home team created a lot of danger in the 200-230, 500-550 and 925-975-second ranges in particular.

The gaps in the graph represent times in the game where the tracking data is not complete. The chart also illustrates a limitation of the model in its current form, as the OPSO for each event does not vary significantly - this is discussed further in section 4.

This type of analysis can also be aggregated across games to infer team strength. In their limited-sample study, Inayatali and Chan find that their implementation of OPSO correlates with both shot attempt and shot on goal rates [2], suggesting that it is an informative metric in determining match momentum and team strengths. Computational cost and time restrictions prevent us from verifying if the results hold for this particular model, so we leave this for further study.

3.1.1 Identifying Defensive Breakdowns

We can also use the highest OPSO events to identify defensive breakdowns, for example. This use case is particularly valuable to coaching staffs and players for post-match analysis. One such play is shown in figure 4. The offensive team has the puck behind the goal-line, and the OPSO plot in figure 4d shows a high danger area right in front of the net, which is occupied by an offensive player. The overall OPSO on this play is 1.08%, which is moderately high. The separate models allow us to identify exactly how this danger is created; the passing lane is somewhat blocked as the pass probability in the high danger area is only moderate, however the blown coverage is shown clearly in the control plot. The true result of this event, in fact, was an attempted pass to the open player in the slot, however it was blocked, and no shot resulted from the play.

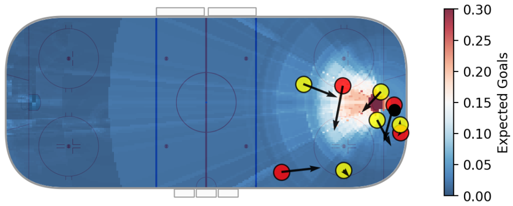
3.2 Player and Team Tendencies

Another use case for coaching staffs is pre-game preparation. OPSO maps can be generated for individual opposition players across multiple games, which can be used to identify their movement tendencies, for which defensive strategies can be developed. Similar maps can be used on a team level to identify where opposition teams generate danger, and plan defensive schemes accordingly.

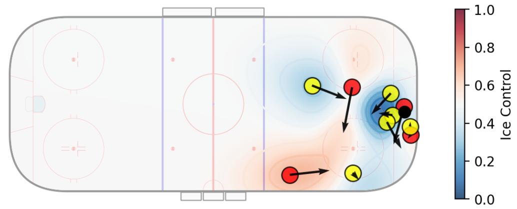
3.3 Player Evaluation

For front offices, there are a number of ways that OPSO can be used for player evaluation. Firstly, we can aggregate OPSO across a player's puck possessions to assess how well players manufacture dangerous situations with their on-puck actions. We hypothesise that players such as Quinn Hughes and Lane Hutson would shine in this metric, as they are known for their elite movement with the puck, skating around the offensive zone and pulling in defenders until space opens up for a teammate.

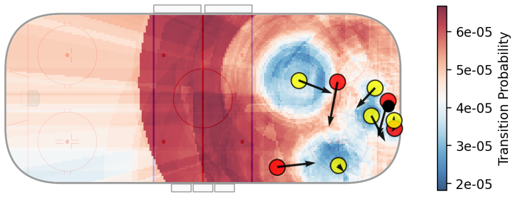
Inayatali and Chan termed this metric *on-puck space generation* (OPSG), and showed it to have positive correlation with goal and assist rates [2]. Our model is similar in structure but varies in methodology, so more work is required to determine whether these results would hold. Such work involves fitting the



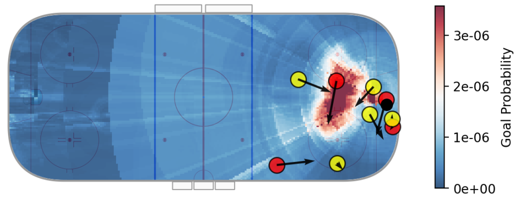
(a) Danger



(b) Control



(c) Availability



(d) Combined Model (OPSO)

Figure 4: Defensive Breakdown Example (BRI @ CLT 2023-10-27, 1st Period, 156.1s)

OPSO model on tens of thousands of events, so this is left for further study, due to both computational and time constraints.

Off the puck, we can evaluate the quality of players' off-puck movement by aggregating the OPSO of the positions they take up. Fernandez and Bornn developed a similar metric in soccer, termed *space occupation gain/loss* [1]. This may require more granular tracking data, however, and is also computationally costly and thus left for further investigation.

There is also scope for analysing defensive play. We have already seen the applications for analysing defensive breakdowns in section 3.1.1, but we can also attempt to infer individual players' defensive ability. Quantities of interest could be OPSO allowed whilst on the ice, which could be weighted to assign responsibility for different areas between players and distribute credit or blame for the defensive coverage.

4 Conclusion

There are limitations to the model in its current form. As noted before, there is not much variation in total OPSO between events. This is likely due to the danger model not always assigning value in an intuitive way. For example, if the play is in the neutral zone, the danger model notices that there are no defenders in front of the goal and assigns very high xG values to that unoccupied area, despite it not being a relevant area with respect to the position of the puck. Potential solutions are available, involving a more context-aware danger model used by Fernandez and Bornn [1], which is discussed further in appendix B. Improvements are also possible to both the control and availability models.

The key benefit of the model lies in its interpretability. Not only does the final OPSO field provide an easy-to-understand visualisation of the danger posed by the players' positions, but each constituent model is also interpretable on its own. This quality gives the model exciting applications for players and coaches, as demonstrated in section 3. With sufficient computational resources [1], the model can be scaled for player and team evaluation in front offices.

This report also does not explore the full possibilities of this model. As discussed in section 3, now that the models have been developed, computing OPSO over a large sample of games would allow analysis of team and player trends, as well as evaluation of the metric's predictive power and correlation with goal-scoring. Quantifying individual defensive ability remains a key problem in hockey analytics, as it is difficult to separate credit between team and players. The OPSO framework offers a novel approach to this problem, and is an exciting avenue for future research.

This author is excited by the potential for further work in this area, both through refining the constituent models and exploring the proposed use cases. Due to scarcity of public tracking data, spatial analysis of ice hockey is not often seen publicly, and inspiration for further research can be taken from the more extensive literature on spatial analysis in soccer. The author thanks Sportlogiq, Teamworks, and the HALO 2026 Hackathon team for the opportunity to work with this dataset, as well as the authors of the three key articles which inspired this project ([1], [2], [3]).

[1] and a more efficient implementation - the code written for this project was optimised for event-by-event analysis and ease of use, so does not yet scale well.

References

[1] Javier Fernandez and Luke Bornn. “Wide Open Spaces: A statistical technique for measuring space creation in professional soccer”. In: MIT Sloan Sports Analytics Conference. 2018.

[2] Hassaan Inayatali and Timothy Chan. “Evaluating Space Creation in the National Hockey League using Puck and Player Tracking Data”. In: *Linköping Hockey Analytics Conference* (July 12, 2024), pp. 13–25. ISSN: 1650-3740. DOI: `10.3384/ecp209002`. URL: `https://ecp.ep.liu.se/index.php/linhac/article/view/1036` (visited on 02/07/2026).

[3] William Spearman. “Beyond Expected Goals”. In: *Beyond Expected Goals*. MIT Sloan Sports Analytics Conference. 2018.

A Model Specification

The basis for splitting up the model into the three separate components is given by Spearman [3]. Recall that we define OPSO at a given location $r \in \mathbb{R}^2$ as the probability that the attacking team successfully completes a pass to r and subsequently scores from that location. Encoding the current state of the ice (player & puck positioning) as D, we write

$$\text{OPSO}_r = P(G_r|D),$$

where G_r represents the next on-ball event being a goal scored from position r. We introduce the following quantities:

- T_r (transition): the next on-ball event occurs at location r.
- C_r (control): the puck at r is controlled by the offensive team.
- S_r (score): a goal is scored from position r.

All three of the above occuring is equivalent to G_r, so we write

$$\begin{aligned} P(G_r|D) &= P(S_r \cap C_r \cap T_r|D) \\ &= \underbrace{P(S_r|C_r, T_r, D)}_{\text{score model}} \cdot \underbrace{P(C_r|T_r, D)}_{\text{control model}} \cdot \underbrace{P(T_r|D)}_{\text{availability model}} \end{aligned}$$

The overall OPSO is then calculated by summing over the ice surface, denoted $\mathcal{R} \subset \mathbb{R}^2$, with

$$\text{OPSO} = P(G|D) = \sum_{r \in \mathcal{R}} P(S_r|C_r, T_r, D) \cdot P(C_r|T_r, D) \cdot P(T_r|D).$$

Our implementation uses a grid of size 200×100.

A.1 Control Model

The player influence area demonstrated in figure 1 is derived from the method implemented by Fernandez and Bornn [1], who model the influence area as a 2-dimensional Gaussian p.d.f. For a player i, we define their influence I_i on position $p \in \mathcal{R}$ as

$$I_i(p) = \frac{\mathcal{N}(p; \mu_i, \Sigma_i)}{\mathcal{N}(\mu_i; \mu_i, \Sigma_i)} \in (0, 1],$$

where $\mathcal{N}(\cdot; \mu, \Sigma)$ is the two-dimensional Gaussian p.d.f. centered at μ with covariance matrix Σ. The denominator is included to standardise the value between 0 and 1.

For a player i located at r_i with velocity vector $\vec{v}_i$, we choose

$$\mu_i = r_i + 0.5 \times \vec{v}_i,$$

being the location of the player in 0.5 seconds, if they continue at their current trajectory.

The shape of the ellipse is determined by the covariance matrix Σ_i. It can be decomposed into products of a rotation matrix R_i (to incorporate the player's direction) and a scaling matrix S_i (to incorporate their speed), with $\Sigma_i = R_i S_i S_i R_i^{-1}$. For a player travelling at angle θ, we choose

$$\begin{aligned} R &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ S &= \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}. \end{aligned}$$

s_x and s_y are modelling choices. Fernandez and Bornn suggest

$$s_x = s_y = \frac{\text{radius} + \text{radius} \times \text{speed\_ratio}}{2}.$$

We choose speed_ratio $= \frac{||v_i||^2}{36^2}$, with 36 ft/s chosen to be a maximum attainable speed. We also choose

$$\text{radius} = \min\left\{\exp\left(\frac{d_i}{10} + 12\right), 30\right\},$$

where d_i is the distance of the player from the puck. That is, the radius of influence increases exponentially with distance from the puck, up to a maximum value of 30 feet.

These parameters were selected through experimentation, so that the control field provided the most intuitive visualisation of the ice. A data-driven approach to selecting these parameters would be a reasonable next step.

Finally, the overall team-wide pitch control at a location p is calculated by summing the influence of offensive players, subtracting the influence of defensive players, and applying the expit function σ to normalise between 0 and 1:

$$\text{PC}(p) = \sigma\left(\sum_i I_i(p) - \sum_j I_j(p)\right).$$

A.2 Availability Model

For model feasibility, we implement the simplifying assumption proposed by Inayatali and Chan [2], that transition probability to a given location is proportional to the probability of a pass to that location being successful. With this assumption, we can train a model which learns the probability of any given pass being successful, and then normalise to obtain a transition probability.

We trained an XGBoost binary prediction model on the following features:

- Distance to target.
- x and y distance to target.
- For each of the (up to) 3 closest defenders to the passer:
  - Defender distance to the passer.
  - Defender distance to the target.
  - Angle of the defender to the intended pass vector.

The model was trained on over 200,000 even-strength passes, and obtained a cross-validated ROC AUC of 0.75.

A.3 Danger Model

The danger model is equivalent to an xG model, which learns the probability of any given shot resulting in a goal. We fit an XGBoost binary prediction model on the following features:

- Angle of the shot to the goal.
- Distance of the shot to the goal.
- For each of the (up to) 2 closest defenders to the intended shot path:
  - Defender distance to the goal.
  - Perpendicular distance of the defender to the intended shot path (i.e. 0 implies the defender is blocking the line of the shot to goal).

The model was trained on over 32,000 even-strength shots, and obtained a cross-validated ROC AUC of 0.78.

B Potential Improvements

As discussed in section 4, the main limitation of the OPSO model currently is the lack of variability in overall OPSO between events. This is likely due to the danger model dominating the other models by consistently assigning high danger to the slot, even when the play is not in the offensive zone.

An attractive solution is that proposed by Fernandez and Bornn [1], where they suggest defining high danger areas as the areas that defenders position themselves. In their words, “considering a sufficiently high number of situations, the defending team distributes itself throughout the field in a manner which covers high value spaces.” We could train a danger model which learns the usual positions of defenders (potentially through defensive influence from the control model) from a given situation. In this way, for a defensive zone breakout, the high danger areas would likely be center-ice and the neutral zone, and as the puck moves closer to the goal, the model would become closer and closer to an xG model. Such a context-aware danger model would greatly improve the flexibility of the OPSO model, making it more interpretable in a wider range of situations. This author did attempt to implement such a model, but it proved too complex given the time constraints, so a more conventional xG model was chosen.