xSFR - Expected Shot From Recovery

Non-technical report

Jedrzej Sarna

October 2024

1 Introduction

Manchester City is often synonymous with tiki-taka and Pep Guardiola's distinctive possession-based style, which emphasizes fluid ball movement and dominance of play. However, the club's success is not solely attributable to maintaining possession but also to their remarkable ability to swiftly recover the ball and execute precise actions in the crucial moments that follow.

In this context, the introduction of **Expected Shot From Recovery (xSFR)** offers a novel and insightful metric that captures a team's efficiency in transitioning from defense to attack. Specifically, it measures the likelihood of a shot occurring within 10 seconds after regaining possession. By doing so, xSFR provides a deeper understanding of how teams capitalize on defensive recoveries, as well as offering insights into individual players ability to recover the ball and assess whether their subsequent actions lead to dangerous opportunities within a short period.
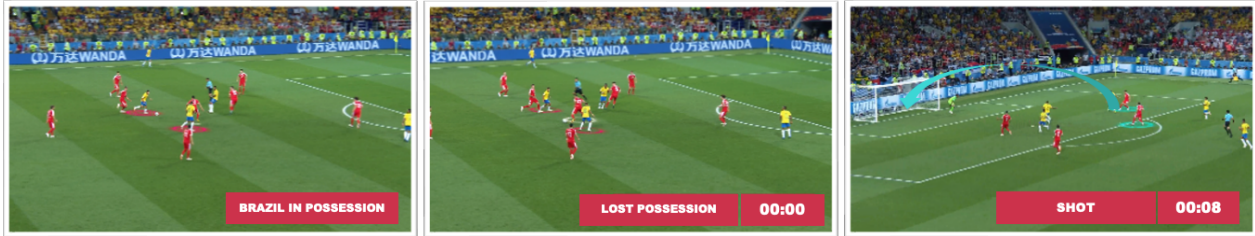

Figure 1: Example of sequence of play captured by the xSFR metric (Brazil vs. Serbia match). Initially, Brazil holds possession, but Serbia quickly regains control, triggering the start of the xSFR time frame. Within 8 seconds, Serbia attempts a shot, demonstrating how the metric tracks the transition from recovery to shot attempt.

2 Metric explanation

2.1 10 seconds interval

What is crucial, yet not mentioned in the name of the metric, is the defined 10-second window for a shot to occur. This duration was selected based on thorough analysis, which demonstrated that shorter time frames resulted in a very small percentage of recoveries leading to shots, as it is often physically impossible to achieve in such a brief span. Additionally, based on available data, in football match the average duration of team retaining possession is around 9 seconds, reinforcing the idea that a 10-second interval is appropriate. On the other side, longer periods may not be as closely related to the initial moment of ball recovery, which is the key event analyzed within the metric.

2.2 Features included

The metric was formulated as a machine learning problem. More specifically, based on data from the 2017/18 Premier League season, a model was developed to identify the factors influencing the likelihood of a shot occurring after regaining possession. A logistic regression model was chosen, where the output can be interpreted as the probability of desired event. Let's now examine the variables that were considered in the final model.

- **Distance to goal**: The distance between the location of ball recovery and the opponent's goal. A shorter distance increases the likelihood of a shot, as teams are in a more advantageous position to attempt a quick attack.

- **Minute of match**: The current minute of the match. This reflects match context, as teams may adjust their tactics based on the stage of the game. Analysis revealed that later stages of the game see a modest increase in the frequency of shots following recoveries, which might be due to teams taking more risks.

- **Pass distance**: The distance of passes made after ball recovery. For this metric, shorter passes turned out to increase the probability of a shot within the 10-second window. Longer passes may result from a sense of urgency and can often be inaccurate, whereas a short pass to a nearby teammate can prove to be a more effective option.

- **Anticipation or interception**: Indicates whether the player regained possession through anticipation or interception, as opposed to other methods such as a sliding tackle or less easily defined recoveries. As expected, anticipation or interception has a significant impact on the likelihood of a shot occurring. This is because the player typically has better control of the ball after these types of recoveries, allowing them to make more deliberate decisions about the next actions.

- **Result**: The current result of the match. While this may not directly affect the immediate chance of a shot, teams that are winning or losing may adjust their play style, influencing the likelihood of attempting a shot after recovering possession.

- **Central area**: Indicates whether the ball recovery occurred in a central area of the pitch. Recoveries in central areas are more likely to lead to a shot due to the strategic advantage of these positions in launching attacks.

By combining all these variables, we obtain a model that can, with a high degree of accuracy, estimate the likelihood of a shot being taken within the next 10 seconds after a ball recovery.

3 Results explanation

To provide context, during the 2017/18 Premier League season, approximately 5 percent of ball recoveries resulted in a shot within the following 10 seconds. The model is designed to identify players who significantly exceed this baseline. To fairly compare players, it is important to account for the number of minutes played, so normalization per 90 minutes is applied. Additionaly only players that played more than 60% of all possible minutes were considered. The top-performing players according to this metric achieve values in the range of around 25%-35%, representing a substantial improvement over the league average. It's no surprise as well, that defensive midfielders, who are primarily responsible for ball recoveries, dominate the rankings.

4 Recommendations for Manchester City

This metric was largely developed to help identify a replacement for the aging Fernandinho in the team, who was mainly responsible for this role before. When looking in the remaining top five European leagues, one standout player is Frank Anguissa of Olympique Marseille with an impressive 0.40 xSFR per 90. At just 22 years old, Anguissa had an outstanding season, including reaching the Europa League final, which was one of the highlights of the previous season. Currently, he is valued at only €18 million.