



optapro

**Beyond shots:
A new approach to quantifying
scoring opportunities**

Dan Altman

@NYAsoccer
#OptaProforum

optapro

A PERFORM GROUP COMPANY

Beyond shots – Dan Altman

Why quantify scoring opportunities?

- **Basing player ratings on points is difficult**

It's hard to tell which goals and plays were pivotal

Shapley values offer one possibility, but they depend on hypotheticals

- **Basing player ratings on goals is difficult, too**

Goals can be idiosyncratic, and many players aren't directly involved in them

- **We need less noisy metrics for the performance of players all over the pitch**

Metrics for individual players should sum to a team metric correlated with results

Metrics for players should be consistent from year to year and give good incentives

Metrics for teams should pick up predictive power early in the season

Beyond shots – Dan Altman

The status quo: expected goals based on shots

- **Why use shots?**

Shots are more common than goals, and creating shots seems to be a more consistent part of performance than scoring goals

- **How are expected goals calculated?**

Each shot has an estimated chance of becoming a goal based on historical averages in the same location and situation

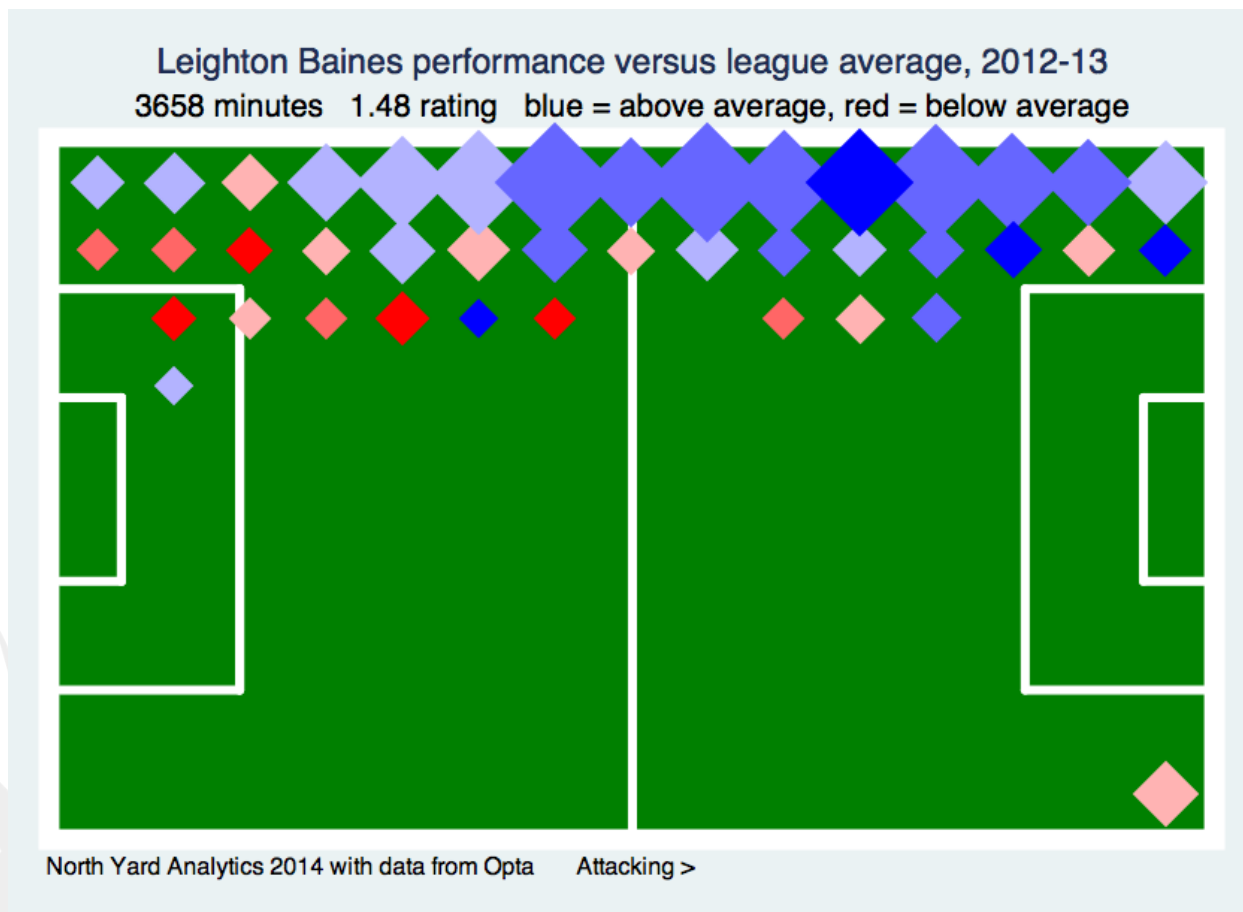
- **What can we use expected goals for?**

Breaking down expected goals into contributions by individual players results in player ratings that sum to a team's expected goals and are persistent over time

Early in the season, a team's expected goals are a better predictor of its final position in the table than many other metrics, including shot ratios

Beyond shots – Dan Altman

The status quo: expected goals based on shots



Ratings give credits for contributions to own team's expected goals and demerits for responsibility for opposing team's expected goals

Expected goal difference was correlated at 0.83 with final positions in the table in 2012-13

Beyond shots – Dan Altman

Is there an alternative?

- **A shot is just one kind of situation on the pitch**

The estimated chance of scoring a goal within a given time period can be attached to any class of situations that can be measured consistently over the season

- **Player ratings can be based on situations as well**

Just as for shots, contributions to situations can be divided up among players, and arbitrary rules of thumb for dividing credit can be avoided

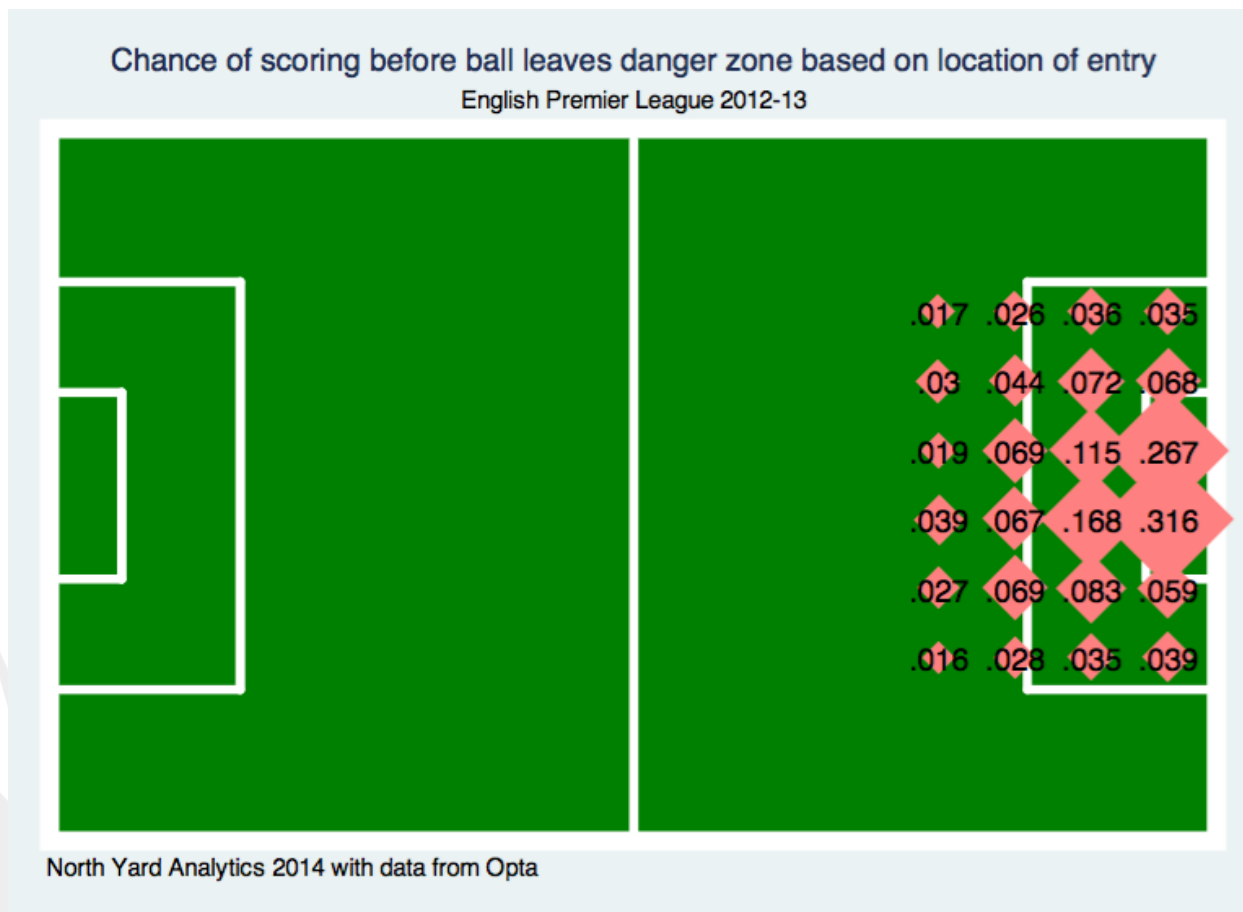
- **Looking behind the shots may yield more useful information**

Simply measuring the chances of scoring from shots says nothing about how the shots were created

If certain situations usually lead to high-quality shots and goals, they may be a better basis for ratings

Beyond shots – Dan Altman

An early example: danger zone entries



When the ball enters this “danger zone”, what is the chance of scoring before it leaves or is lost?

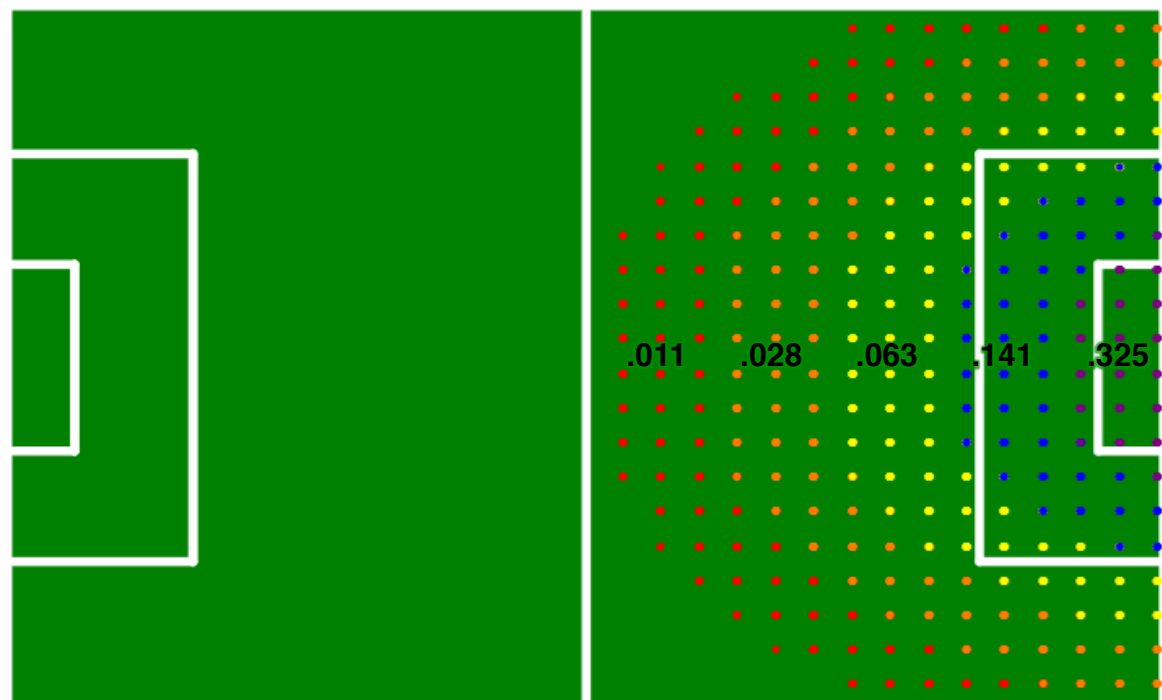
Expected goal difference was correlated at 0.83 with final positions

Top 10 players with 2,000+ minutes:
Suárez, Carroll, Van Persie, Agüero, Lukaku, Silva, Bale, Snodgrass, Benteke, Lambert

Beyond shots – Dan Altman

Further development: a five-zone system

Five-zone system for assessing expected goals



North Yard Analytics 2015 with data from Opta

Each zone contains all zones closer to goal

Players receive credit for advancing the ball between zones based on differences in expected goals

Expected goal difference from all zone entries was correlated at 0.82 with final positions

Beyond shots – Dan Altman

Further development: a five-zone system

- **These ratings use simple zones for offensive production only**

Could assign demerits to defenders using same system as for shot-based ratings

A better solution would be to use tracking data to assign demerits

Zones could also be designed to conform to common attacking channels

- **But zones that cover more of the field allow the rating of more players**

Top 10 attackers for offensive production per minute in 2012-13: Suárez, Silva, Cazorla, Hazard, Mata, Tévez, Agüero, Carroll, Bale, Lukaku

Top 10 central midfielders for offensive production per minute in 2012-13: de Guzmán, Touré, Gerrard, Davis, Osman, Dembélé, Ki, Cabaye, Lampard, Noble

Top 10 fullbacks for offensive production per minute in 2012-13: Baines, Johnson, Coleman, Rangel, Gibbs, Enrique, Azpilicueta, Walker, Evra, Zabaleta

Beyond shots – Dan Altman

Can we do better?

- **Tracking data allow more complex definitions of situations leading to goals**

TRACAB data from the Premier League track each player, the ball, and referees 25 times per second on a pitch with roughly 1,000,000 possible locations

These data can be synched with Opta f24 files to merge positional information and description of events in the match

- **Preliminary work suggests algorithms can identify important situations**

In a test, a prototype algorithm correctly spotted every 2-on-1 and 3-on-2 advantage in front of goal during several matches

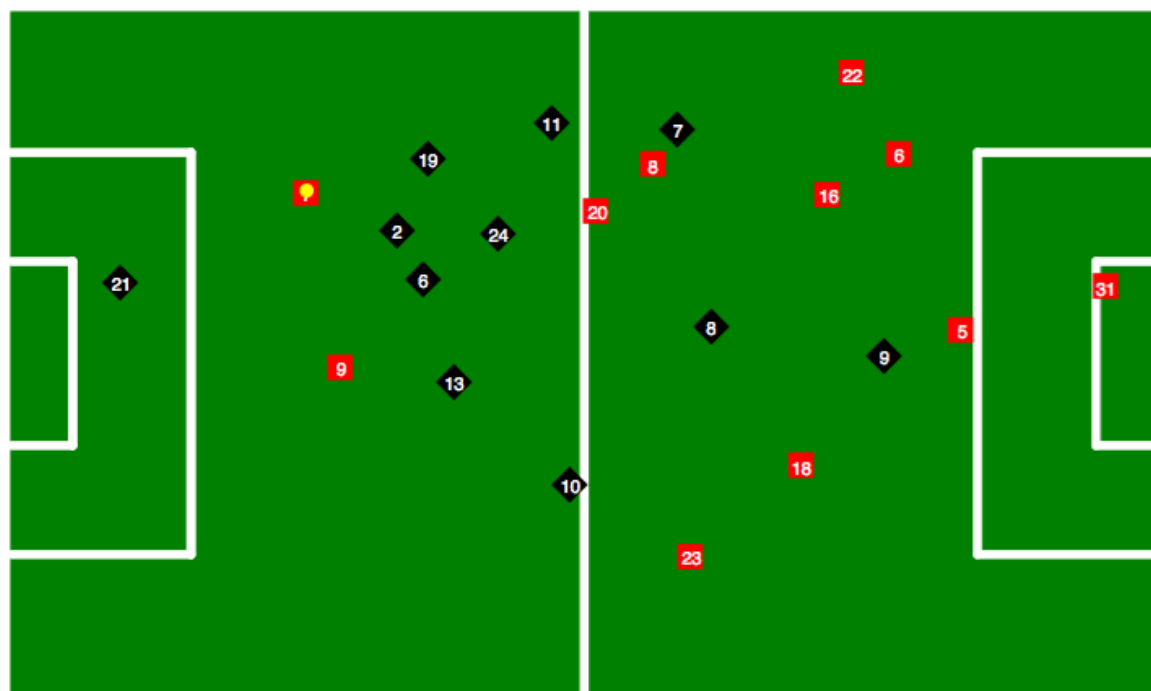
Over a season, each situation can be assigned a probability of leading to a goal based on the actual outcomes listed in the f24 files

Credits for creating situations can be split between the runners, passers, decoys, and so on in any combination; demerits for allowing them go to the defenders

Beyond shots – Dan Altman

Can we do better?

A 2-on-1 break from the 2013-14 Premier League season
(Player #9 on the red team scores)



North Yard Analytics 2015 with data from TRACAB via the Premier League and Opta

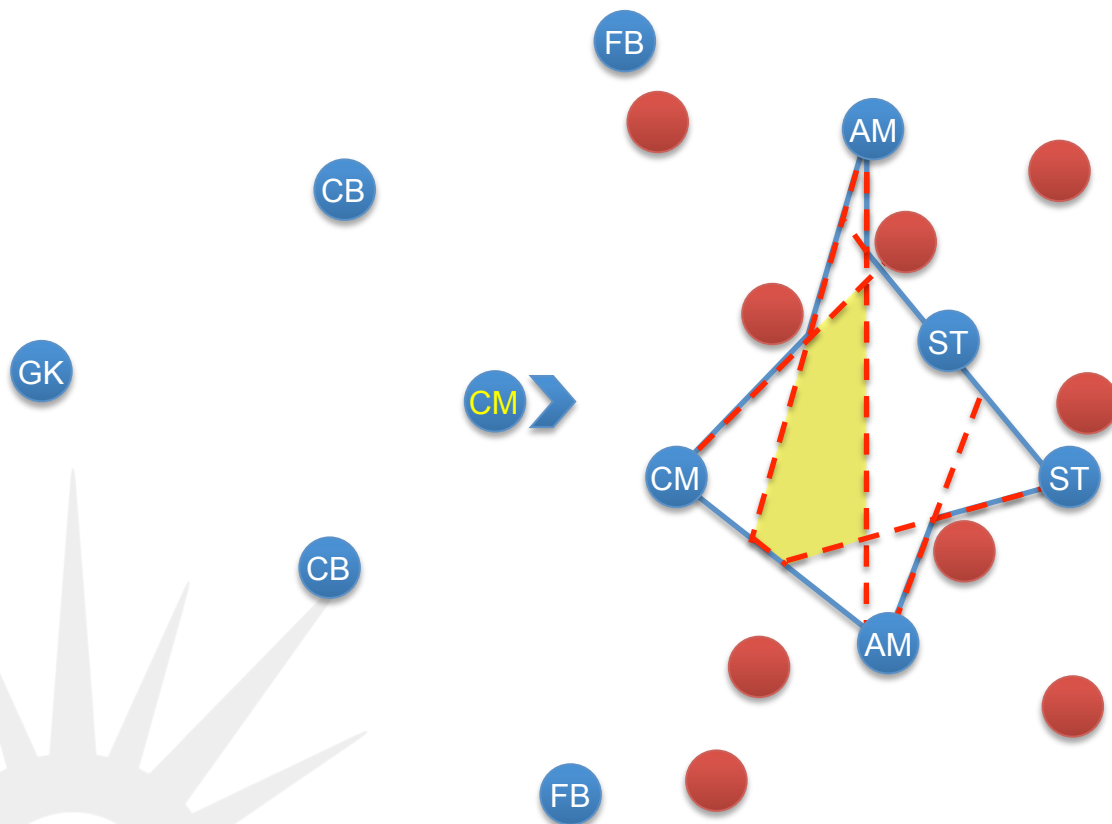
Criteria were two players behind the penultimate defender and in the middle of the pitch, of whom one had possession

Manual checks for corners, celebrations

Of two 2-on-1 situations in this game, one led to a goal; one player (#7) received the pass in one and was the runner in the other

Beyond shots – Dan Altman

Going beyond: the pass before the through ball



From the yellow area, the passing lanes to and between all the other players are visible

In mathematics, this area is called the kernel of a star-shaped space

Do moves preceded by passes into the kernel of attacking players lead to better opportunities to score?

Beyond shots – Dan Altman

How might all of this be useful?

- **Algorithms can quickly identify players who are influential on the pitch**

Even the tracking algorithm runs in a few minutes, versus watching hours of video

- **The same algorithm can be used for leagues around the world with f24 data**

Players in known leagues can be used as reference and compared to players in other leagues, e.g. “Find me the Eden Hazard of Brazil’s Serie A”

- **Individual player ratings sum to team ratings that are correlated with results**

This means players who improve their ratings will also improve the team’s results; it sets good incentives and enhances the predictability of results and finances

- **Team ratings pick up information quickly on likely final positions in the table**

An early read can be helpful for planning transfer spending and future finances

Thank you

Dan Altman
@NYAsoccer
#OptaProForum



optapro

A PERFORM GROUP COMPANY