

Smart Data Scouting in
Professional Soccer:
Evaluating
Passing Performance based on
Position Tracking Data

Makoto Saotome

1.Introduction

In the U.S., quantitative analysis developed rapidly in the early 2000s, especially for the four major sports like baseball and basketball.

On the other hand, in soccer, where quantitative analysis is difficult due to the complex nature of the game, until the early 2010s, "classic indicators" such as the number of shots, success rate of passes, and ball retention rate, which do not lead to the evaluation of players or teams, were used not only in the press and on the field but also in academia. These "classic metrics" were used in the press and on the field, as well as in academia.

Nevertheless, in recent years, researchers, data companies, and even tactical enthusiasts have been proposing new metrics to quantitatively analyze soccer. Expected Goals (xG) is a well-known example of a modern indicator. This indicator, which takes into account not only the position of the shot, but also various contexts such as body parts, the process of shooting, etc., has given rise to other indicators related to shooting, such as Expected Assists and Expected Saves.

LIVERPOOL



1 : 2



REAL MADRID

46:26

SUMMARY

POSSESSION

SHOOTING

PASSING

DEFENDING

EVENTS



DRIBBLE SUCCESS RATE



SHOT ACCURACY



PASS ACCURACY

48

Possession %

52

9

Shots

7

1.9

Expected Goals

2.0

82

Passes

87

10

Tackles

9

1

Tackles Won

1

6

Interceptions

9

3

Saves

3

1

Fouls Committed

0

1

Offsides

1

2

Corners

0

1

Free Kicks

1

0

Penalty Kicks

0

1

Yellow Cards

0

0

Red Cards

0



DRIBBLE SUCCESS RATE



SHOT ACCURACY



PASS ACCURACY

⊗ Advance



However, it is said that shooting accounts for less than 2% of all actions with the ball that occur during a soccer match. In other words, defensive midfielders and defenders, who are not often involved in shooting, are still not easily evaluated by indices such as xG. Therefore, researchers and data companies have proposed models over the last five years to quantify actions other than shooting.

Here are only the models I was interested in.

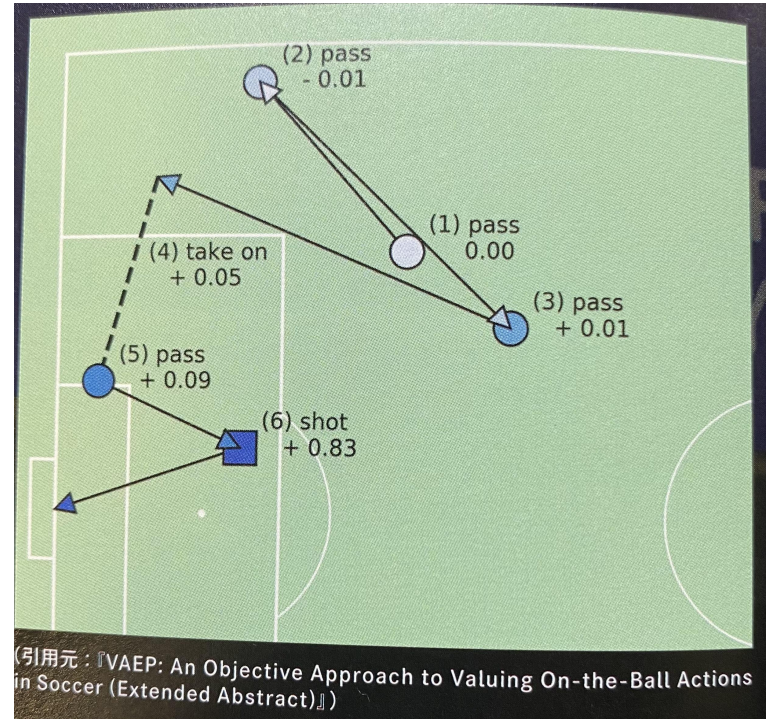


2.Example

The models proposed so far are varied.

No.1

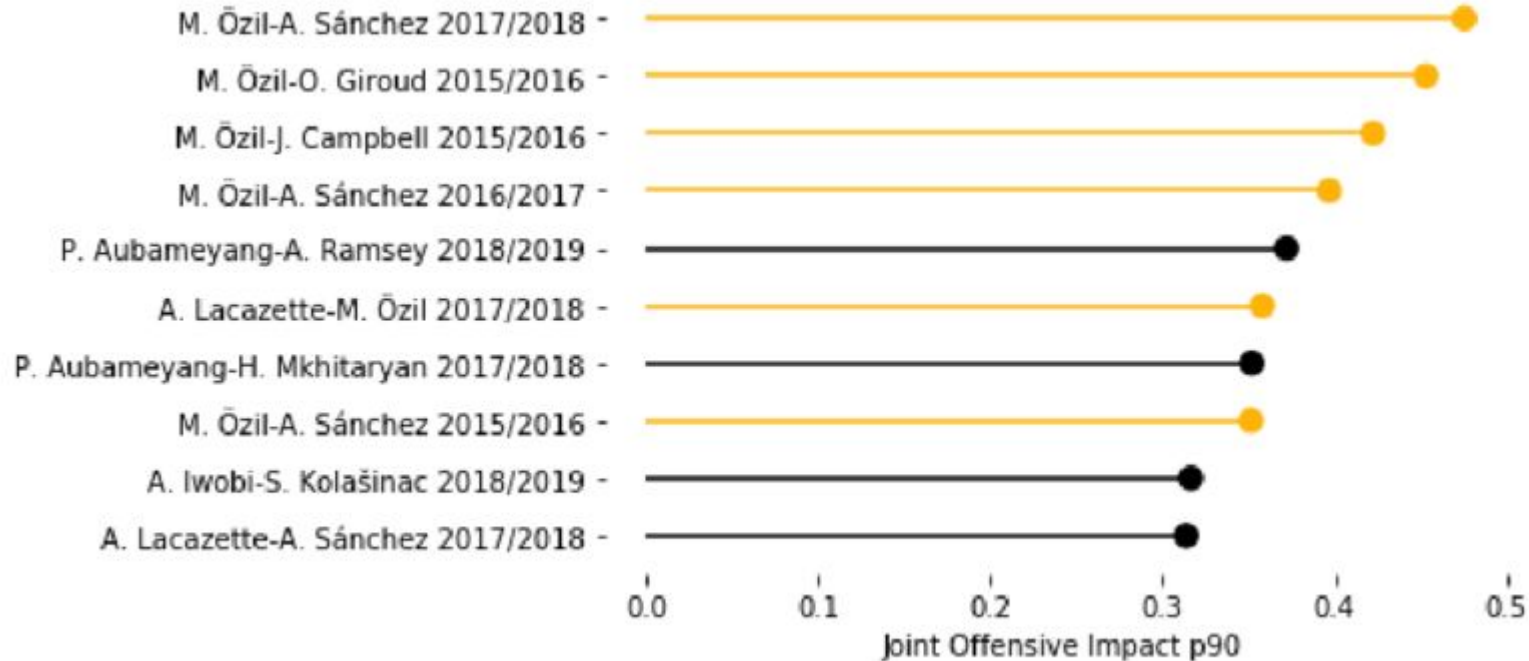
The "Valuing Actions by Estimating Probabilities" (VAEP) metric, which evaluates ball actions using event data that includes information on the time and location of actions occurring in a match.



No.2

VAEP has evolved into the Joint Offensive Impact (JOI) and Joint Defensive Impact (JDI) by the joint research team to quantify the interaction between two players. The JOI is derived from the aggregate VAEP ratings of the interactions, such as passing and dribbling, between the players. The JDI calculates whether the play of the two players and their opponents exceeded or fell short of expectations.

The top-ten-ranked Arsenal player pairs in terms of Joint Offensive Impact per 90 minutes since the start of the 2015/2016 season.



3.Main

Floris Hose and his colleagues at the University of Groningen attempted to quantify passes from a different perspective than the previously mentioned indicators, such as passes, which are calculated in relation to the probability of scoring a goal or taking a shot. Hose focused on the defensive movement against the pass. They defined the objective of the pass as the creation of space by disrupting the opponent's defensive organization. They proposed "I-Mov" and "D-Def" as indicators of the degree to which this is achieved.

Validating D-Def

$$\text{D-Def} = |\text{PC1}| + |\text{PC2}| + |\text{PC3}|$$



Longitudinal

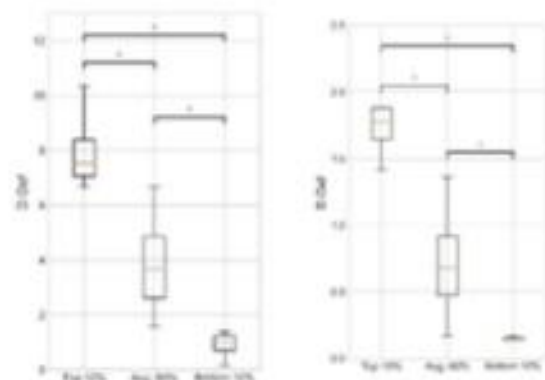
Lateral

ORIGINAL ARTICLE

Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches

Floris R. Goerl^{1*}, Matthias Kempe¹, Laurentius A. Meeuwis¹ and Koen A.P.M. Lemmens¹

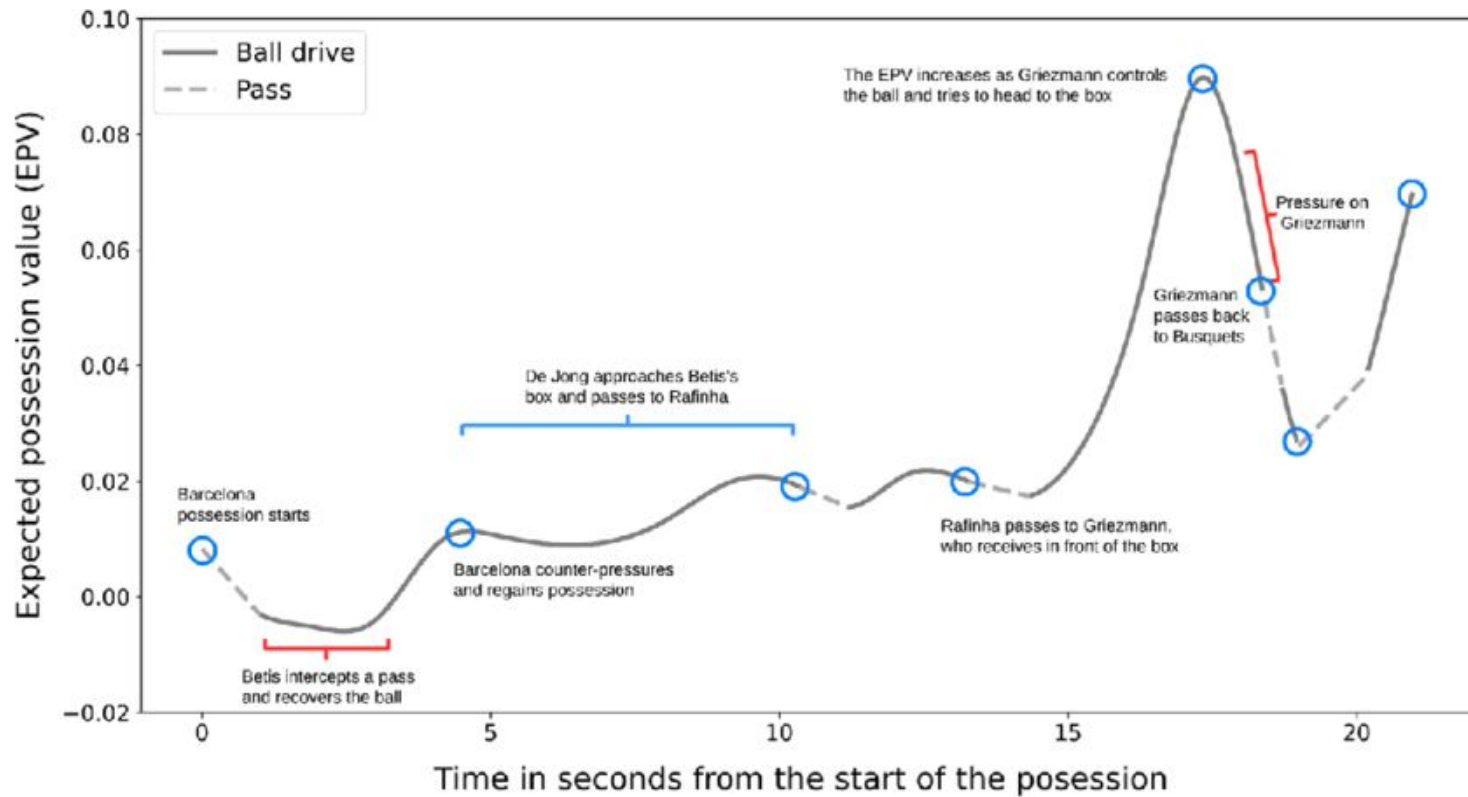
Differentiate Passes & Players



I-Mov is the total displacement of the individual players on the defending team. The total movement of an individual is obtained by summing the displacements on the X and Y axes on the coordinates of all players from the moment the pass is issued until 3 seconds later. D-Def refers to the degree of division of the defensive organization. For quantification, the centers of gravity of the three lines (DF, MF, and FW) were calculated and the displacement of each center of gravity on the coordinates at 3 seconds after the bus. In addition, the surface area of the defensive formation and the spread between the players are calculated, and a principal component analysis is performed to create a composite measure of the total D-Def.

The Expected Possession Value (EPV) presented in 2019 by Javier Fernandez, Barcelona's head of analytics, uses tracking data to incorporate a great deal of contextual information into a model, including the position and movement of all players, the risk and reward of actions, defensive organization, and pitch control. The EPV uses tracking data to incorporate a great deal of contextual information into a complex computational model. The probability of a team scoring or conceding a goal at any given moment is estimated and expressed as a number between 1 and -1, as an indicator that includes various components such as play selection, probability of success or expected value of a pass, dribble, or shot, etc.

Evolution of the expected possession value (EPV) from the perspective of FC Barcelona during a match against Real Betis in La Liga season 2019/2020



EPV can be applied not only to analyze whether a single pass can increase a team's chances of scoring, but also to evaluate decision-making in play selection and positioning before receiving the ball, since it represents potential play options and the expected value of each. Last year, Fernández, together with researcher Adrià Alves of the Universidad Pompeu Fabra in Barcelona, introduced a more sophisticated model that combines EPV with data on players' body orientation estimated using a posture model from match footage.



Background work on EPV

A theoretical framework for the expected possession value

EPV Basketball approach

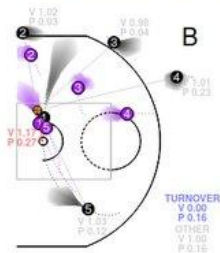


Image from: <https://arxiv.org/pdf/1408.0777.pdf>

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

Markov-chain based EPV



Image from: <https://karun.in/blog/expected-threat.html>

Rudd S (2011) A framework for tactical analysis and individual offensive production assessment in soccer using markov chains. In: New England Symposium on Statistics in Sports.

Singh K (2019) Introducing expected threat (xt). <https://karun.in/blog/expected-threat.html>

Action-value models

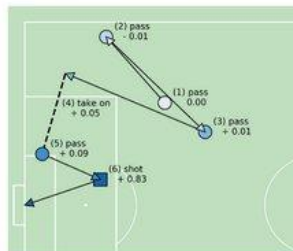


Image from: <https://arxiv.org/pdf/1802.07127.pdf>

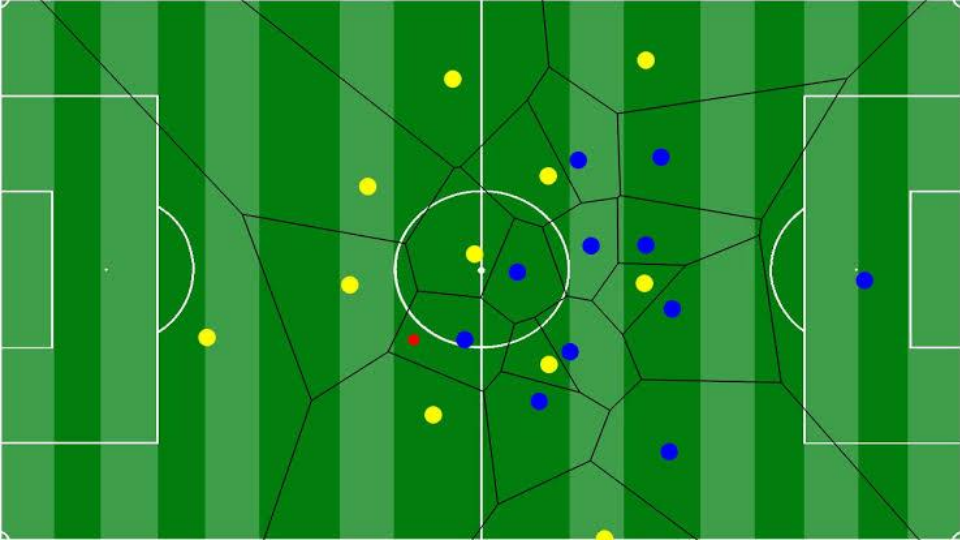
Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2020). VAEp: An Objective Approach to Valuing On-the-Ball Actions in Soccer. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20* (pp. 4696-4700). International Joint Conferences on Artificial Intelligence Organization.

American Soccer Analysis. (2020) Goals added and the great possession shift. https://www.americansocceranalysis.com/home/2020/4/22/37Ucr0d5unxtrym2cfhzormdzip_hq

StatsPerform (2020). Introducing a possession value framework <https://www.statsperform.com/resource/introducing-a-possession-value-framework/>

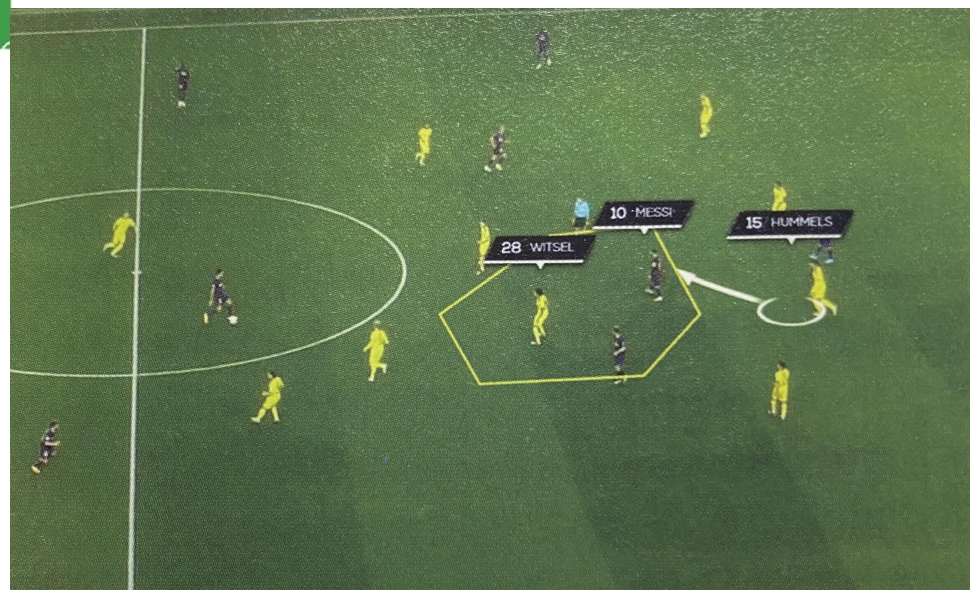
- Rigid vs fluid notion of possessions
- Discrete vs continuous state representation
- Dynamic programming and explicit transition matrix vs nonlinear function approximation
- Event data vs tracking data to exploit relationships between the 22 players and ball
- Missing off-ball performance evaluation
- Single-location vs field-wide action destination (e.g. passes)
- EPV-only interpretation vs component-level interpretation

EPV can also be applied to defensive evaluation. The Barcelona team used a Voronoi diagram based on average positions during the two minutes of off-ball possession to assign a defensive area of responsibility to each player on the defensive team. When an opponent is able to receive a pass in his area of responsibility and closer to the goal than he is, the increase in EPV of the opponent in response to that pass is counted as the responsibility of the defender in charge of that area. By quantifying the responsibility for passes that may pose a threat to the defender, it becomes possible to analyze the positions that need to be improved in the team's defense and the places that the opposing team should target.



Mats Hummels pressured Messi as a bus passed Messi standing in Axel Witsel's area of responsibility during the 2019-20 CL season, Barcelona vs. Dortmund ↓↓↓

↑ This is Voronoi diagram



4. Conclusion

Four weeks ago, there was an introduction to tracking data in football in this laboratory, and the current situation is that the operation of tracking data is very expensive. However, I hope that one day something will happen and players at all positions will be legitimately evaluated using metrics, not just subjective evaluations by those at the top.

Thank you very much for your attention.

