Anatomy of Receiving and Turning with the Ball

Soumyajit Bose: Cornell University, sb2293@cornell.edu
Manas Saraswat: Gartner, manas.saraswat@gmail.com

Abstract

StatsBomb's latest product launch provides new exciting and unique opportunities to work with high quality 360 data, bringing more context to game events. The location of players in the field of view of broadcast camera allows one to innovate and build new performance evaluation metrics. In this work, we focus on a subset of such possible contextualized metrics. We study the ability/proclivity of players to receive and turn with the ball, unpressed or under pressure, and suggest a detailed step-by-step way for clubs and analysts to study the same.

1. Introduction

The evolution of data quality and context in football and the study/analysis henceforth has come a long way since the days of inception of the sport. Manual collection of data is being replaced by highly sophisticated and ever-improving Computer Vision (CV) techniques. The information provided by event-only data - the x and y coordinates of events taking place on the field (and sometimes the end coordinates of the event as well, if different) - while valuable and holds a lot of promise, is limited and incomplete. Companies such as Metrica and Tracab among several others have been in the business of supplementing the event-only data set with more context by recording "tracking data" - snapshots of the entire field at given intervals of time. This provides us with a plethora of new information - the location and the velocities of the other players and the ball itself. In 2020, StatsBomb launched their own product - 360 data which provides tracking data + event data from broadcast camera footage. This, in conjunction with their high quality and rich event data set, allows clubs to study and create analytic models at an industrial scale. So, what are certain ideas one could possibly explore with this new stream of data ?

As the StatsBomb Conference 2021 came to its end, it became very clear that there is plenty that can be explored, like Line-breaking passes, Pitch Control models to study potential space-exploiting opportunities, and how to pressurize or disrupt opponents, among others. This study employs the 360 data as well to study individual skills/tendencies of players - in particular - who receive the ball and turn to progress.
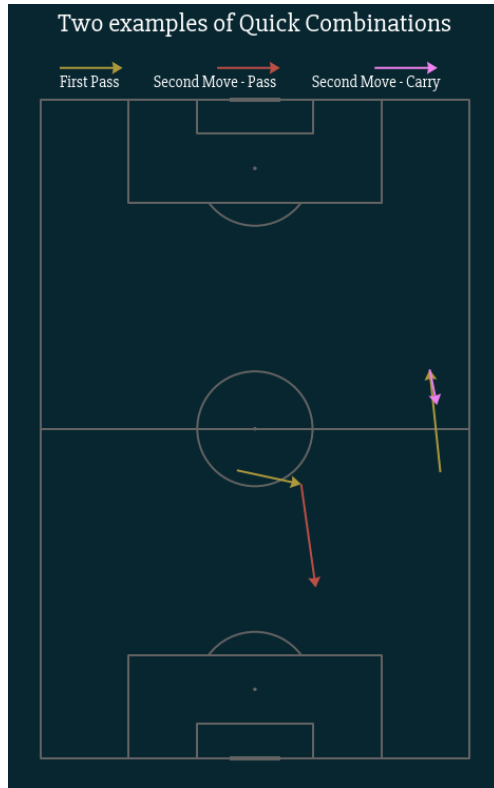
2. The problem statement

Figure 1. Examples of quick combinations

To study the abilities of a player to turn with the ball, we need to look at the combination plays the player is involved in - receive a pass from a team-mate, then turn and pass the ball to a team-mate or carry the ball oneself. We identify and isolate such sub-sequences of actions from the possession sequences provided in the StatsBomb data. In fact, we keep three successive actions instead of two if performed by the same team - the first pass by a team-mate that the player in focus receives, the subsequent action performed by the player (pass or ball carry), and the next action if the second action was a pass to a team-mate. For the purposes of this study we will refer to the player in question as the "receiver".

The pass these receivers receive will be referred to as the "first pass". The subsequent action the receivers make upon receiving the pass will be referred to as the second move. If the second move is a successful pass, the next action performed by a teammate will be referred to as the "third move" - these will be our reference terms in the rest of the paper.

A couple of points need to be stressed right away. Firstly, we do not have body pose data available in football yet at a large scale. In the absence of such data, what we have really done is use a proxy for "turning with the ball". We study the angle between the first pass and the second move as our key variable to quantify turning. Secondly, even with our highly focused choice of research topic, we have further chosen to focus on the first pass being played from the back of the receiver.
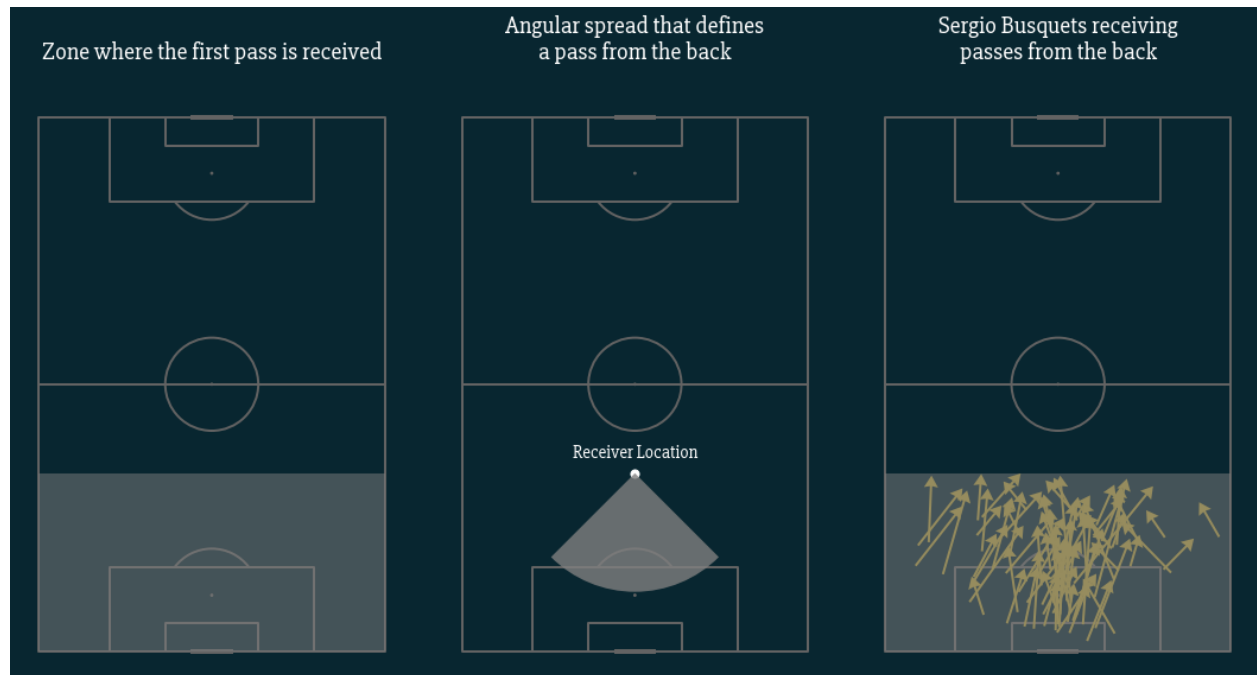

Figure 2. Zone of the pitch and angular illustration of "back", with some examples of the first pass

The second point immediately raises a follow-up question - what exactly defines the "back" of the receiver? For that, we have chosen to work with a certain angular spread as shown below - if the first pass is received from within the angular spread of the shaded cone, it's a pass from the back. Also, as shown, with sufficiently large data sets, the study can be limited to very specific regions of the field. For eg. to truly evaluate central defensive midfielders, one would preferably focus on the central regions of the defensive+middle third or own half of the pitch. The data set provided is limited in size - with 540 select games from La Liga 2020-21 and Bundesliga 2020-21, hence our study looks at the entire field and a fairly broad angular spread to define "back", ranging from -45° to +45°.

3. Methodology for Evaluation

To study the turning skill of players, one could ask a wide variety of questions. We have posed several such questions in this study and tried to answer them in order to quantify the skill. The questions are organised in a sequential manner as follows :

a) how much can the receiver turn to perform the next task upon receiving? What is the angular distribution of the second move?

b) how risky on average is the zone where the receiver operates? how tight are the spaces where the receiver typically receives to turn?

c) how successful is the receiver in completing the next move?

d) how much threat can the receiver directly create with the ball? how much threat does the receiver allow the next player to create? how close to the opposition goal does the receiver progress the ball?

e) coming back to the question of completion, how successful is the receiver at completion above expectation? (expectation = completion rate of an average player)

f) how (un)predictable is the receiver with the turns?

g) how (un)perturbed is the receiver when pressed while turning?

3.1 Relative Angle Sonars

To answer all these questions, we need a few building blocks. We calculate the angle between the first pass and the second move. This is not to be treated as the tail-to-tail angle between the two vectors, rather simply as the head-to-tail angle between the two vectors (a fixed convention of clockwise angle measure is also chosen). Shown below are a few examples. Henceforth, this angle will be referred to as the "relative angle" between the first pass and the second move. Using this relative angle, we can create our first quantification - the relative angle sonar.

The idea for relative sonar comes from John Muller in his article [1]. The usual passing sonar uses a polar/angular histogram to depict passing tendencies of players. For each pass, its angle with respect to some fixed line on the pitch is calculated, and the distribution of pass frequency with respect to this angle is shown by the sonars. Often, added information like median pass length is added as a visual cue/element in the sonars. The relative sonars are built on the same principle as well, but replaces the absolute angle with the relative angle described above. On these sonars, an angle of 0° refers to the case where the receiver sends the ball back to the first passer. An angle of 180° refers to the case where the receiver is able to turn completely with the ball and progress it vertically. Vertical here is a mnemonic for "progression parallel to the direction of the first pass".

The relative sonars thus differ from the usual sonars in the interpretation of the angle. Usual sonars give the absolute angle distribution of the passes, i.e, where the passes ended up. The

relative sonars give us a depiction of the turn angle between two successive moves. For example, Ivan Rakitic receiving a pass from Marcos Acuna/Diego Carlos on the left and then turning to hit a long diagonal to the other flank - will show up as a diagonal pass in the usual sonar but a vertical entry in the relative sonar.
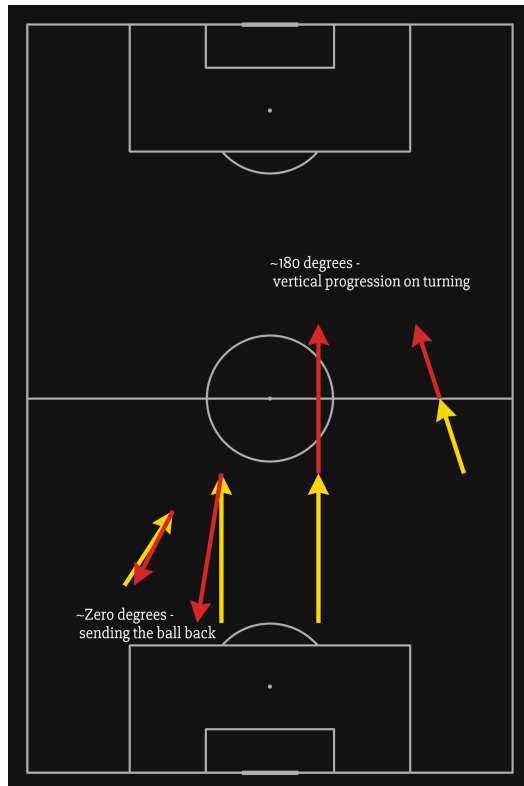

Figure 3. Illustration of the relative angle between the first pass and the second move
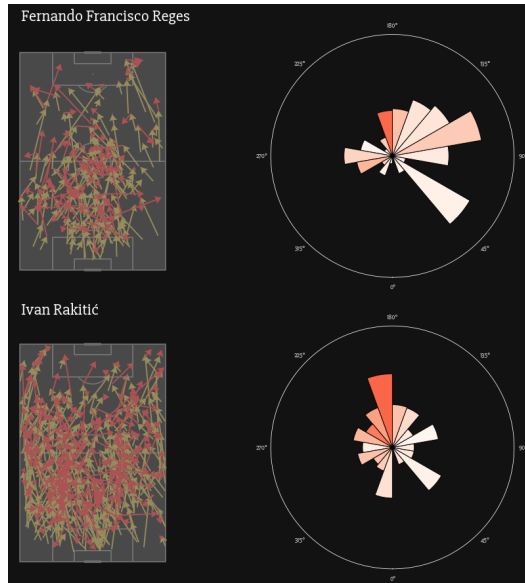
Figure 4. Illustration of the relative angle sonars for two Sevilla midfielders - Fernando and Rakitic

3.2 Expected Move Completion Model

Here we propose another approach to quantifying a player's ability to receive and turn under pressure. While most of the metrics proposed thus far make use of the pressure event, a data point available in the StatsBomb Event data, we use StatsBomb's 360 data to build our own pressures for the purposes of this model.

The idea is to look at the completion probabilities of passes and carries. By adding carries to the mix, we can build a more holistic view of a player's ability to negotiate the press. To add pressure data into the model, we decided to add the number of pressors within a 5-yard radius of the player on the ball. This is done by using 360 data, and calculating the Euclidean Distance from the 'Actor' (StatsBomb 360 terminology of the player performing the event). We excluded events where there was no 360 Frame available, and also excluded GoalKeeper information, to focus solely on outfield players for the purposes of this study. Additionally, passes from Regular Play are used.

The objective is to classify whether the carry or pass was successful or not. For this we chose the HistGradientBoostingClassifier, as we had enough quantum of data to use a boosting methodology. For Carries, if the subsequent action is a dispossession, it is classified as unsuccessful since the data doesn't have a clear classification for it.

The predicted classes for completions of passes and carries from the model were joined back with the original data set to study how players perform under pressure, compared to their expectation level.

3.3 Threat Creation and Facilitation

To assign threat values to actions, we use Karun Singh's Expected Threat model [2], or xT. The xT model is a Markov chain model that, given a pass or a ball carry from a certain zone of the pitch to another, looks at the probability of scoring in the next few moves. Simply speaking, it manifests in two tenets : (i)the closer the action happens to the opposition goal, the higher the threat potential, and (ii)the more vertical/progressive the pass or carry is, higher is its threat potential. xT is used to quantify the threat potential of the second move and the third move in the sub-sequences being studied.

Mathematically, this idea is captured by dividing the pitch into specific zones and assigning probabilities to different events happening in those zones. Following [2], the pitch is divided into $12*8$ grid. In each zone, there is a finite probability that a player in possession decides to move the ball (to another zone or stay in the same zone) or shoot the ball, denoted by $m_{x,y}$ and $s_{x,y}$ respectively, where x and y refer to the index of the zone. If the player decides to shoot, there is a finite probability of scoring, indicated by $g_{x,y}$. If the player decides to move, there is a finite transition probability of moving to any of the available 96 zones, given by $T_{(x,y)->(z,w)}$. Then, a threat value can be assigned to each zone of the pitch by iteratively solving the following equation

$$xT_{x,y} = s_{x,y} * g_{x,y} + m_{x,y} * \sum_{z,w} T_{(x,y)->(z,w)} xT_{z,w}$$

with the initial value of 0 assigned to each zone to start the iteration process. This iterative structure has a very nice interpretation of xT being equal to the probability of scoring within the next n moves (n being the number of iterations it takes to converge). Once completed, the iterative process assigns threat values to all locations on the pitch which can be visualized by the heatmap shown below

Figure 5. Heat map of expected threat - brighter region (bright yellow) on the right implies higher threat.

xT can be used to look at the threat value of the second move. This gives us Direct Threat - a measure of how much the receiver is being able to/allowed to create with the ball directly. xT of the third move gives us xT Facilitated - a measure of how much the receiver facilitates the next player to create. This typically shows up in the "up-back-and-through" type of buildup/attack sequences, where the receiver receives the first pass with a lot of pressure from the back. Instead of turning the receiver lays off the ball to the next player to play a defence-splitting pass or carry the ball up field.

3.4 Risk and Tightness

Risk is evaluated using xT as well, but from the perspective of the opposition. Simply said, risk is equal to the threat the opposition could create by forcing a turnover on a certain point of the pitch. To evaluate this, the xT zone map is simply reversed - regions closer to own goal now have higher values, corresponding to higher risk. The value of risk is simply set equal to the xT of the zone of action.

Figure 6. Heatmap of risk - flipping the xT heat map.
Brighter now refers to more risky zones - if the ball is lost there, opposition has higher chance of scoring.

Tightness is measured using 360 Data input. The visual area is Voronoi-tessellated using the location of all the players. Tightness is directly proportional to the area of ball control of the receiver,i.e, the area of the Voronoi cell corresponding to the receiver. We realize that this creates a bit of counter-intuitive metric to define tightness - since it's directly proportional to the area, tightness value is small if the receiver has to control in tight spaces. An alternative could perhaps be to divide the total pitch area by the Voronoi cell area of the receiver; in that case, tightness increases as the space reduces, and this perhaps is the more intuitive definition for tightness.

3.5 (Un)Predictability

Upon receiving the first pass, the receiver can choose to turn into any direction. What makes the receiver unpredictable is the variance in the turning angle. More concretely, player A receiving and then carrying the ball back or passing it back is a predictable pattern easy to identify and exploit, while player B receiving and then twisting and turning in all sorts of directions makes him unpredictable. This very nature can be captured into a metric - named Entropy aptly - by measuring the "Shannon entropy" of the angular distribution of the turns. The turns are divided into several relative angle bins - 18 specifically, with spread of 20° each, and the total number of turns into each bin is evaluated as a fraction of the total number of turns. Call this quantity f_i for the $i-th$ bin, where i ranges from 1 to 18. Then the entropy En of turns for a particular player is given as

$$En = -\sum_{i=1}^{18} f_i \log(f_i)$$

Note that there is nothing sacred about the number of bins being 18; any number of bins can be chosen to evaluate entropy.

3.6 Pressure Resistance

Press resistance is an extremely important quantity especially for players playing in the deeper zones, for eg, the CDMs. Being able to turn under pressure and play the ball out, or attract pressure and release to an unpressured teammate can be an extremely valuable skill set to have for these players. As such, pressure resistance is something that needs to be measured for the receivers - albeit via a whole set of proxies. The proxies involve studying the above metrics but specifically focusing on under pressure turns - comparing the success rate, xT created, xT facilitated, progression percentage etc without and under pressure.

Along with these, a single metric can also be created to act as a proxy for press resistance. This also involves the study of distributions. In particular, we look at the angular distribution of turns without and with pressure, and measure the "distance" between the two distributions using Jensen-Shannon Divergence - hence the metric is named JSD. If the JSD between the two distributions is large, then the said receiver is perturbed by pressure, for good or worse. However, it must be said that the typical angular pattern for players follows pretty much a clear pattern - without pressure, there is a lot of vertical turns. Under pressure, it gets distributed into several bins. To compute JSD, we first take a look at a similar measure called the Kullback-Leibler Divergence between two probability distributions P and Q (discrete version):

$$D_{KL}(P||Q) = \sum_{i=1}^{18} P_i \log\left(\frac{P_i}{Q_i}\right)$$

Jensen-Shannon divergence is a symmetrized version of the above. Define $M = 0.5 * (P + Q)$. Then, JSD is given as (discrete version):

$$JSD(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)$$

In the two formulae above, P_i and Q_i refer to the relative angular distribution or fraction of moves without and under pressure respectively.

4 Results - Quantification via Metrics

In this section, we look at some results produced from applying the afore-mentioned metrics to the StatsBomb 360 data set. For now, we have mainly focused on center and defensive midfielders, although it must be noted that many of the players have played in several different roles for their clubs. There are also a few players included who have decidedly played a more offensive role than the rest, and are usually viewed as attacking midfielders. We have also chosen to highlight 10 well-known names : Frenkie de Jong, Sergio Busquets, Pedri, Toni Kroos, Luka Modric, Koke(Jorge Resurreccion), Dani Parejo, Joan Jordan, Joshua Kimmich and Florian Wirtz.

We first give the risk-tightness profile of the selected players. As shown, players operating in more risky zones are the defensive or center midfielders rather than the attacking midfielders. But they also usually have more space to operate in (remember our slightly counter-intuitive use of "tightness" - it is directly proportional, not inversely, to the space of operation).
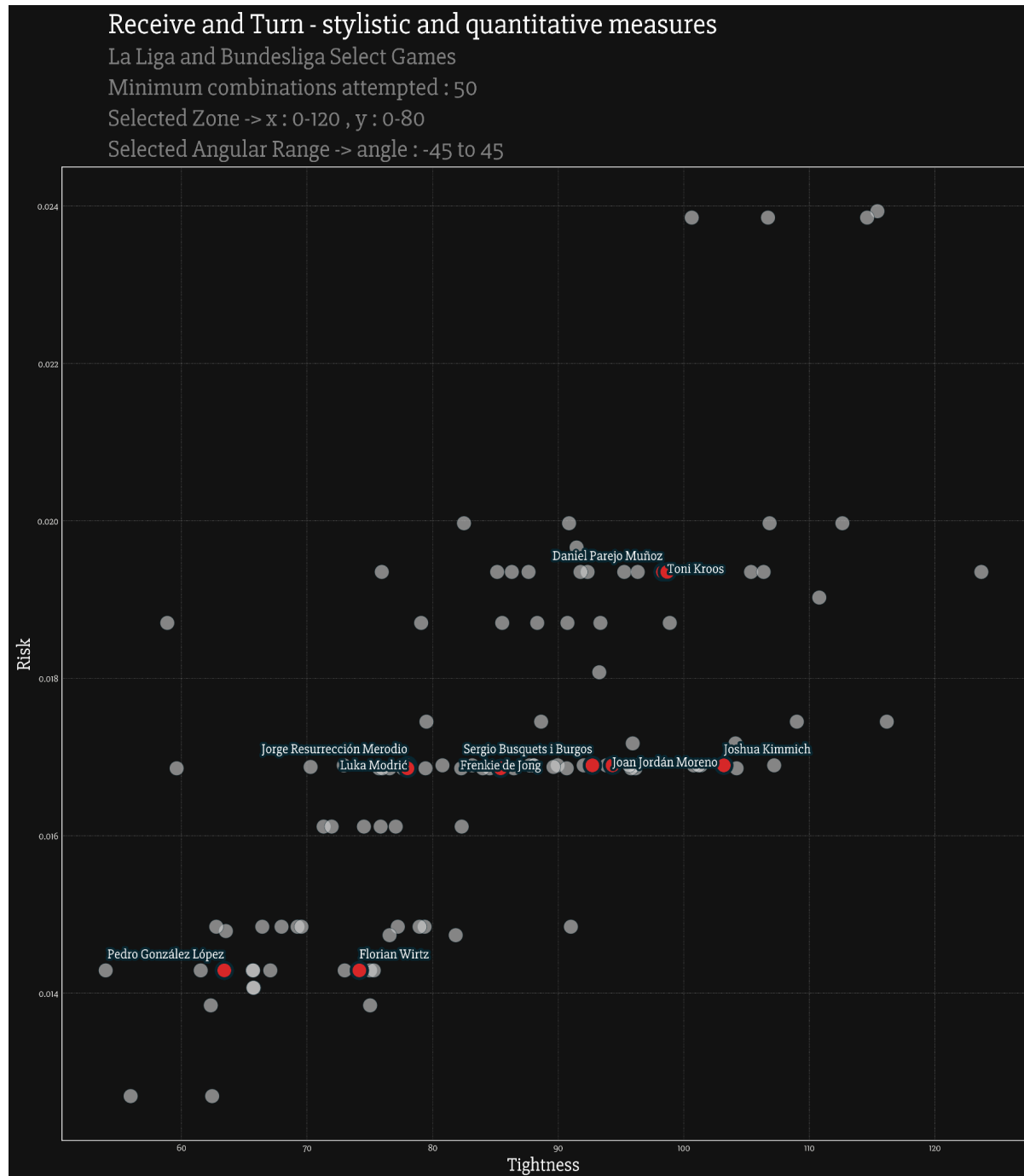
Figure 7. .Attacking midfielders play in less risky zones but have to operate in small areas.

Next, we look at the combination volume and success percentages for these midfielders. First up, we consider all actions - pressured or unpressured. Our highlighted midfielders are decidedly superb on the ball and do well in keeping possession even at high volume. At least 100 combinations are required to be considered for this graph.
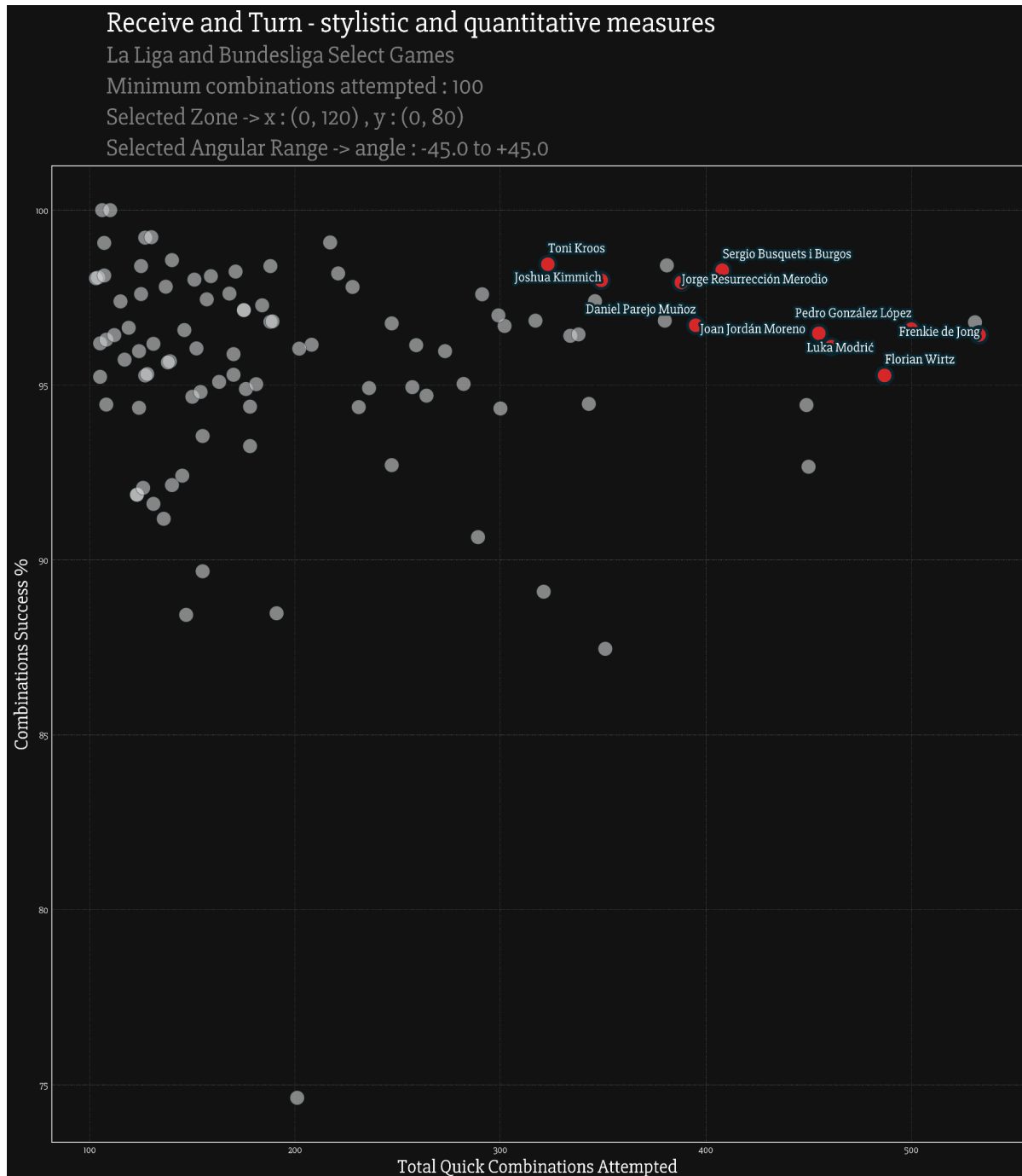
Receive and Turn - stylistic and quantitative measures
La Liga and Bundesliga Select Games
Minimum combinations attempted : 100
Selected Zone -> x : (0, 120) , y : (0, 80)
Selected Angular Range -> angle : -45.0 to +45.0

Figure 8. Combinations success

When focusing only on the pressured combinations, the highlighted midfielders still perform really well. At least 50 combinations are required to be considered for this graph.
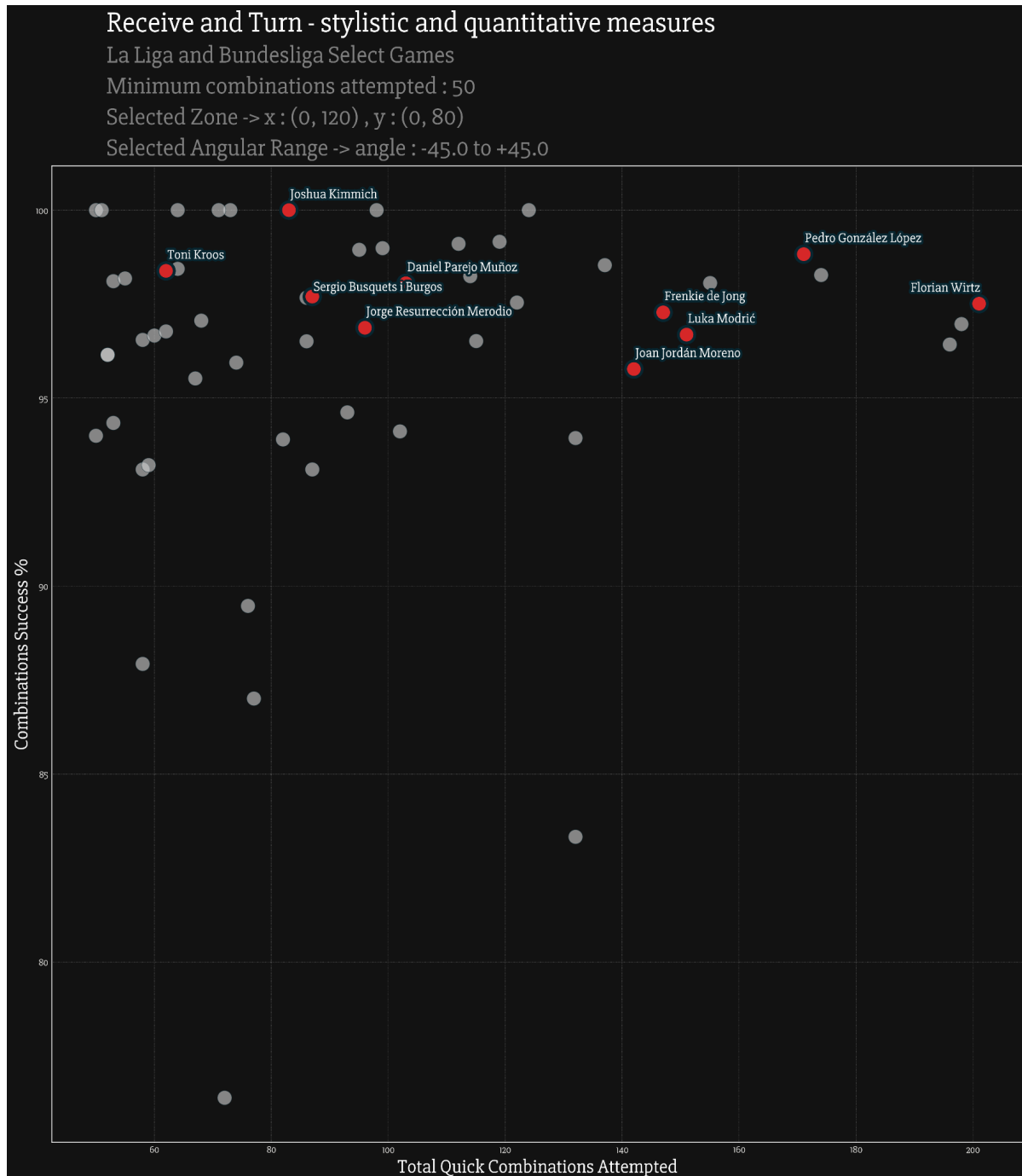
Figure 9. Combinations success under pressure

Once we have a glimpse of success percentages, we are ready to assess the direct threat provided by these receivers via the second move. This is shown below using xT per 100 combinations. The normalization is necessary as players playing for possession heavy sides would typically have more combinations and hence more chance to rack up xT simply due to a volume effect.
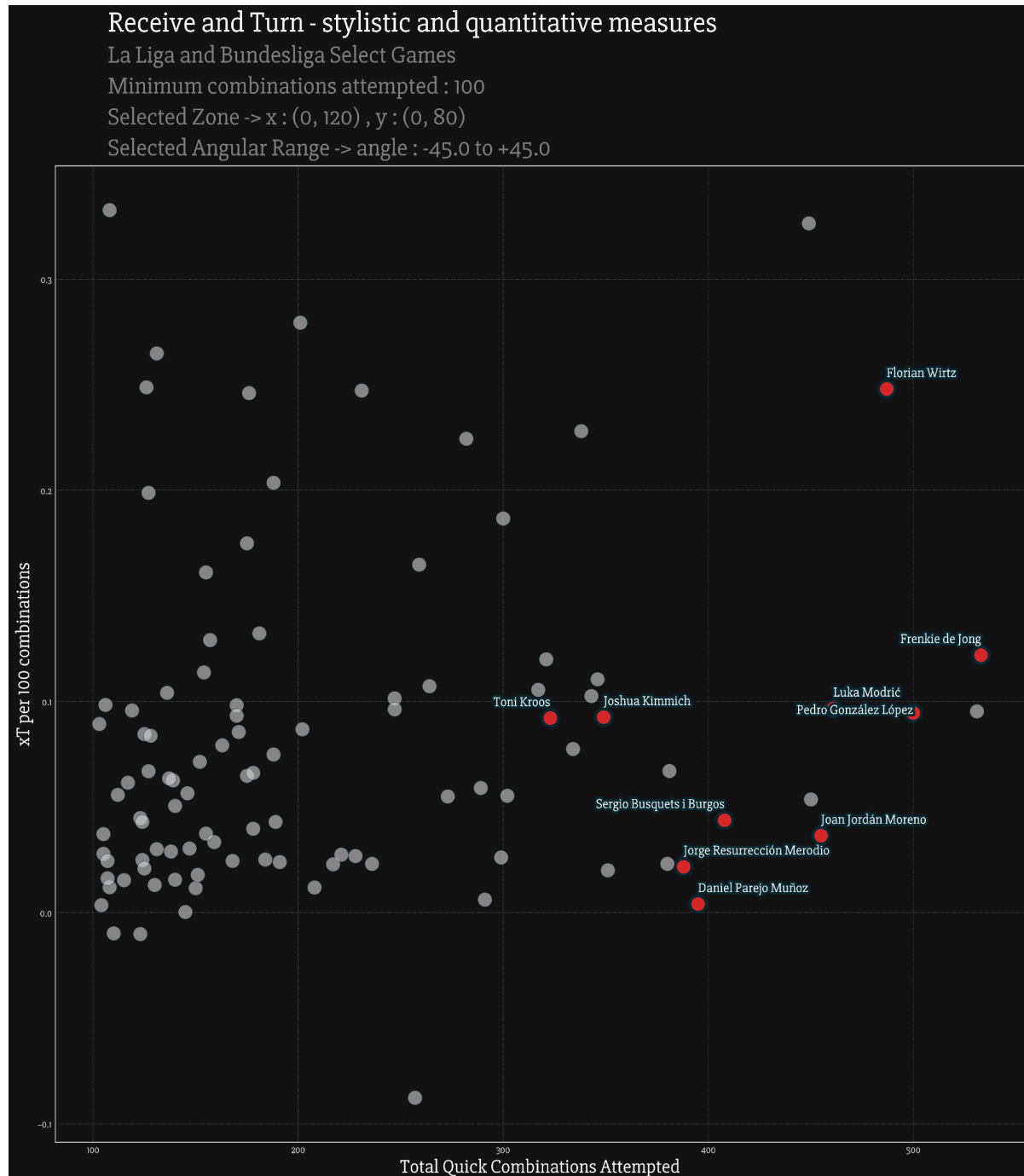
Figure 10. Direct threat from all actions.

We repeat the same focusing on the under-pressure actions only. There is a curious feature here - the xT under pressure seems to be larger than xT of all actions, for most of the players. This could be a simple noise in the data, or could be that under-pressure actions are typically against unsettled defences and hence better at picking out space, or could be that under-pressure actions are more of the "hit-and-hope" kind and the players got lucky in this particular sample set. More investigation needs to be done to examine this.
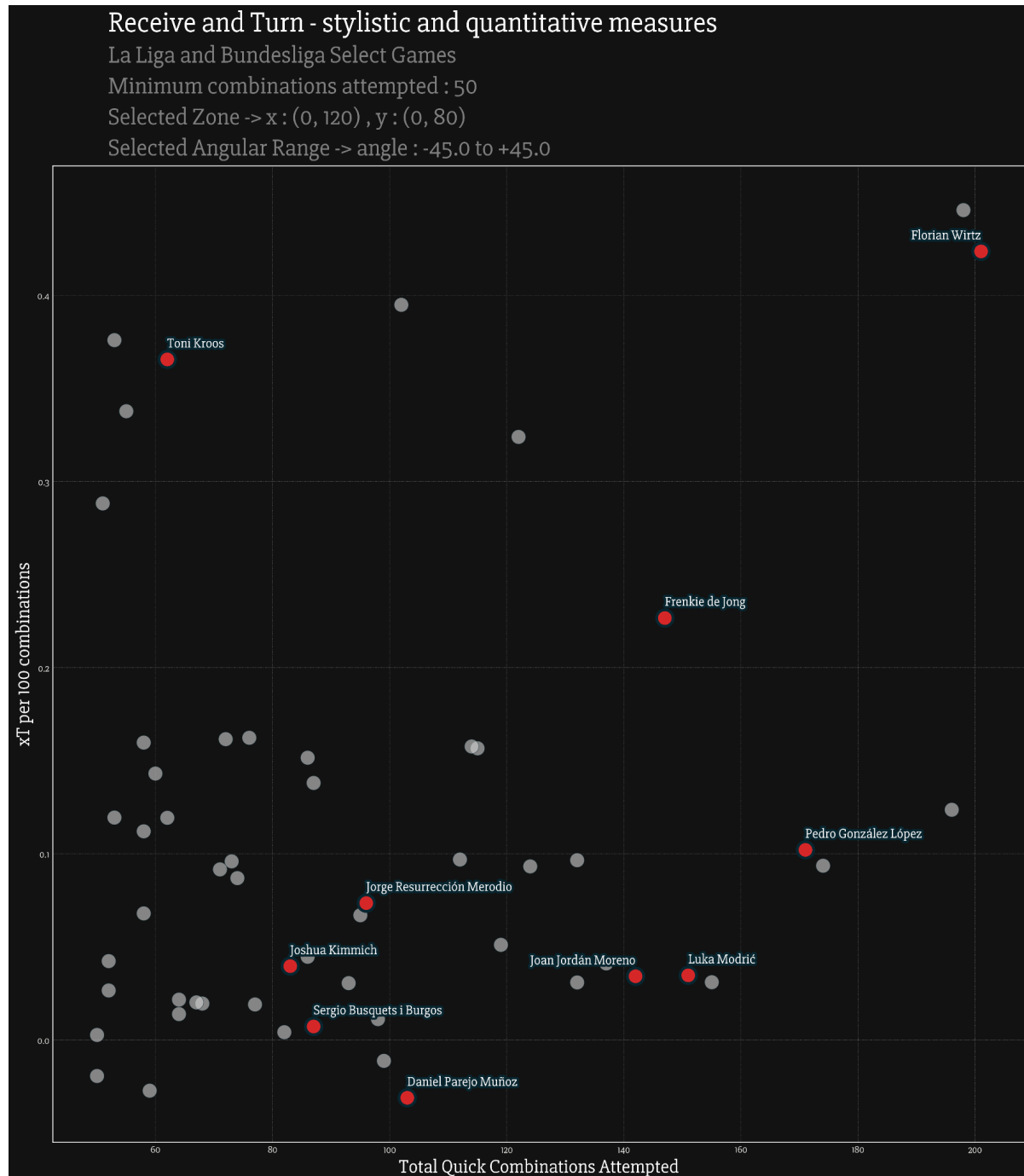
Figure 11. Direct threat from under-pressure actions.

We repeat the above, for all actions and then focusing on under-pressure actions, for xT facilitated per 100 combinations, and average ball progression % as well. One theme that remains common throughout - some of the highlighted players are absolute elite. Frenkie de Jong, Kroos, Kimmich, Modric, Busquets - they are elite at ball progression and threat creation or facilitation.
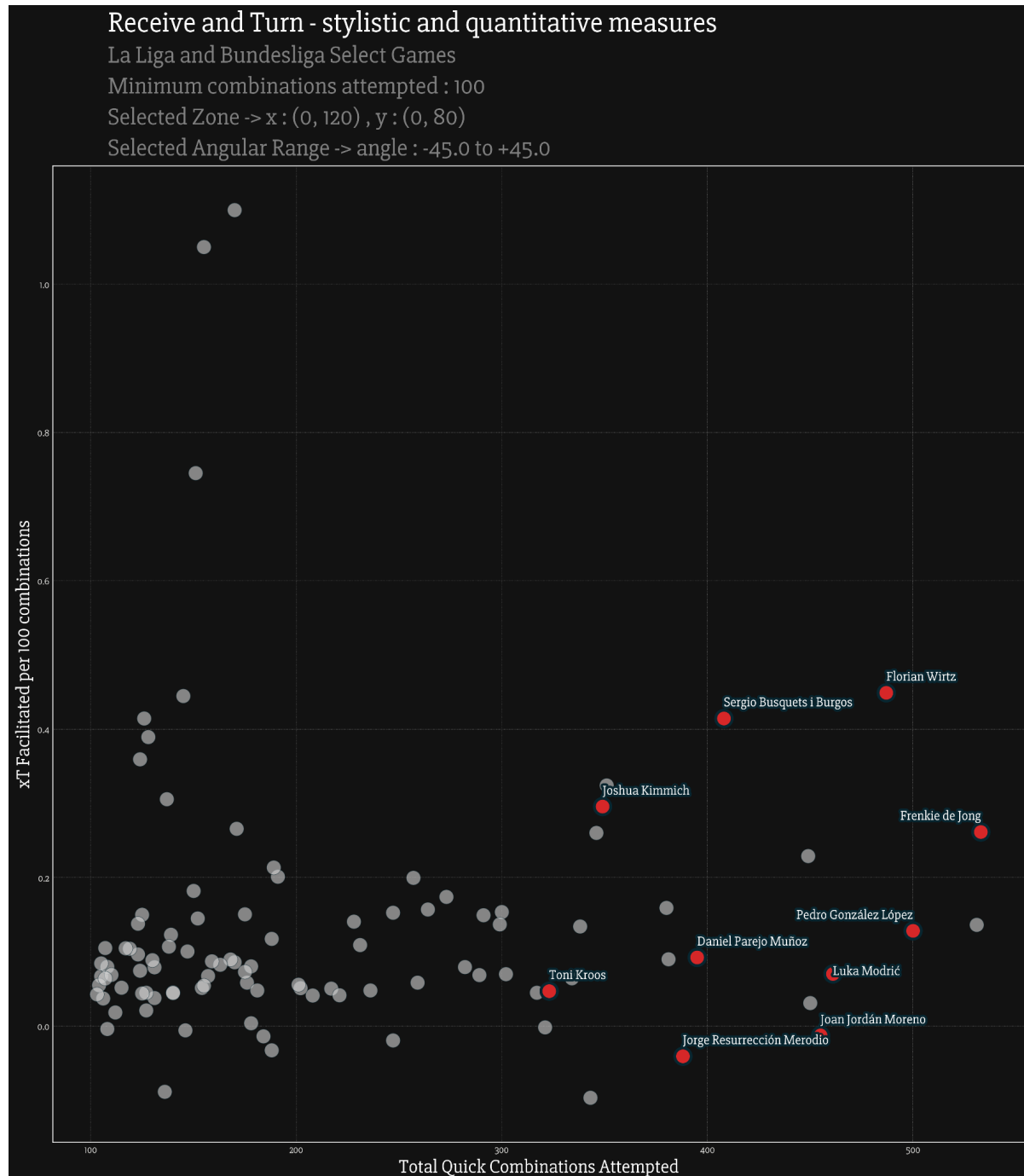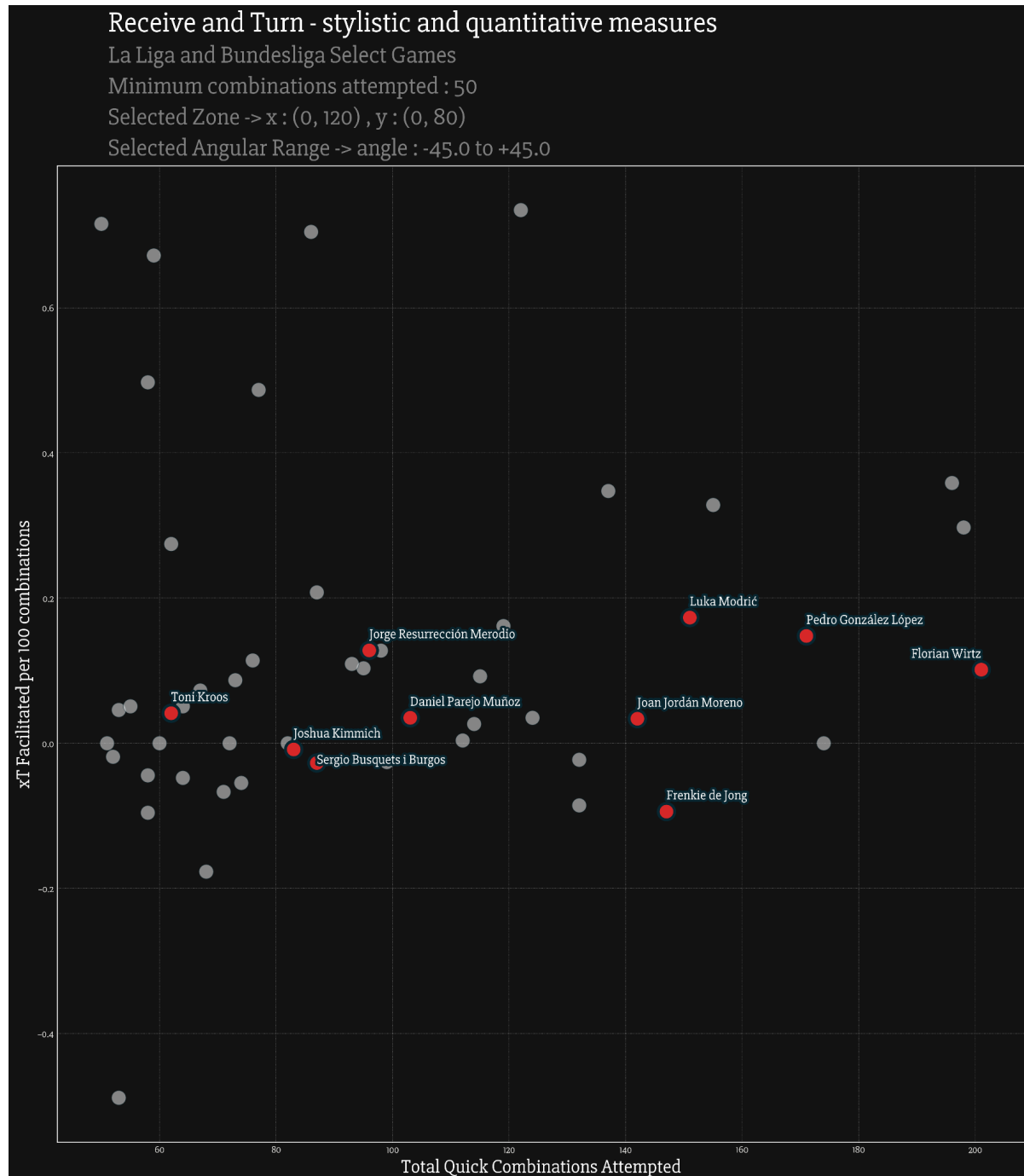
Figure 12. Threat facilitated from all combinations.

Receive and Turn - stylistic and quantitative measures
La Liga and Bundesliga Select Games
Minimum combinations attempted : 50
Selected Zone -> x : (0, 120) , y : (0, 80)
Selected Angular Range -> angle : -45.0 to +45.0

Figure 13. Threat facilitated from under-pressure combinations only.
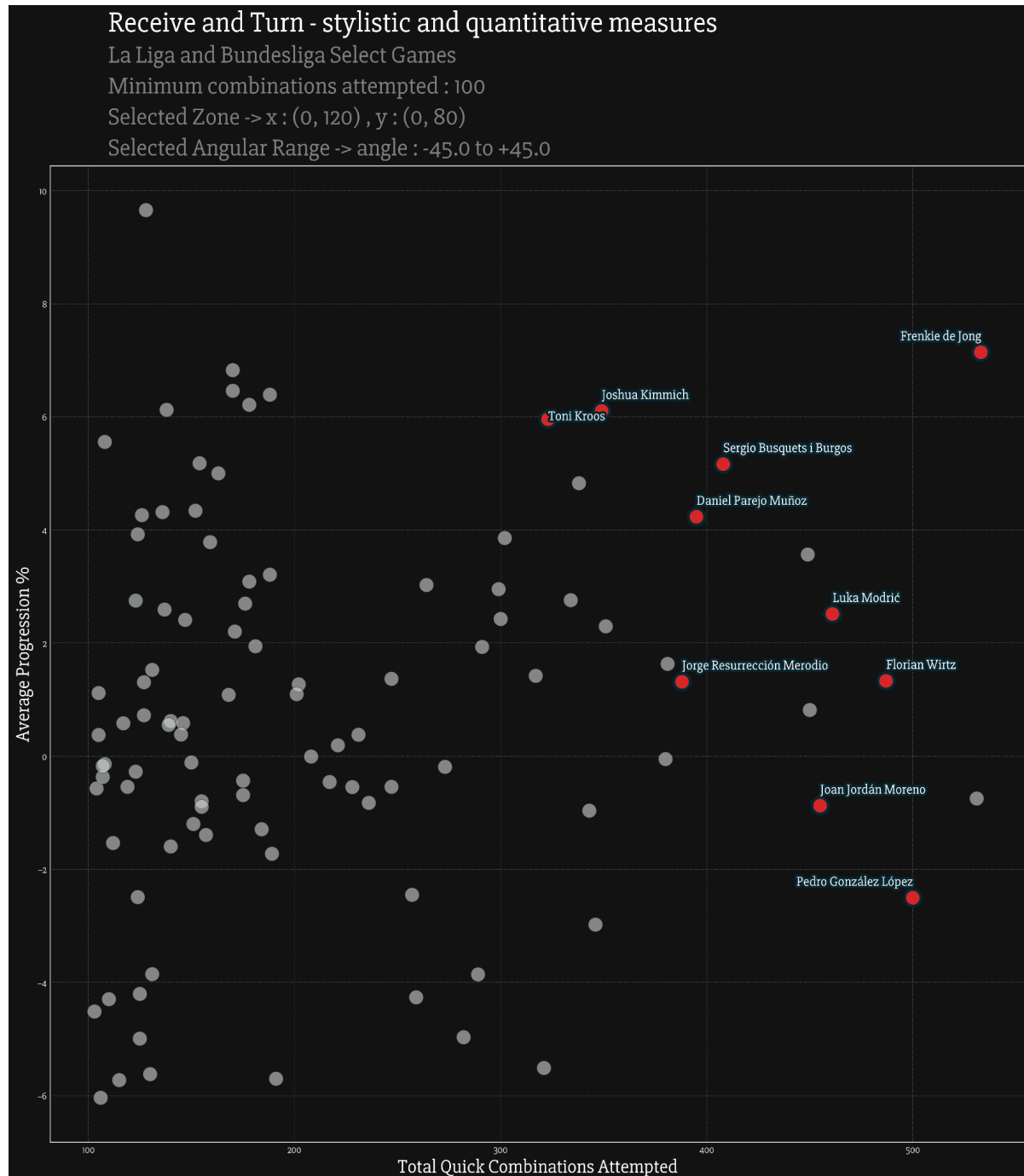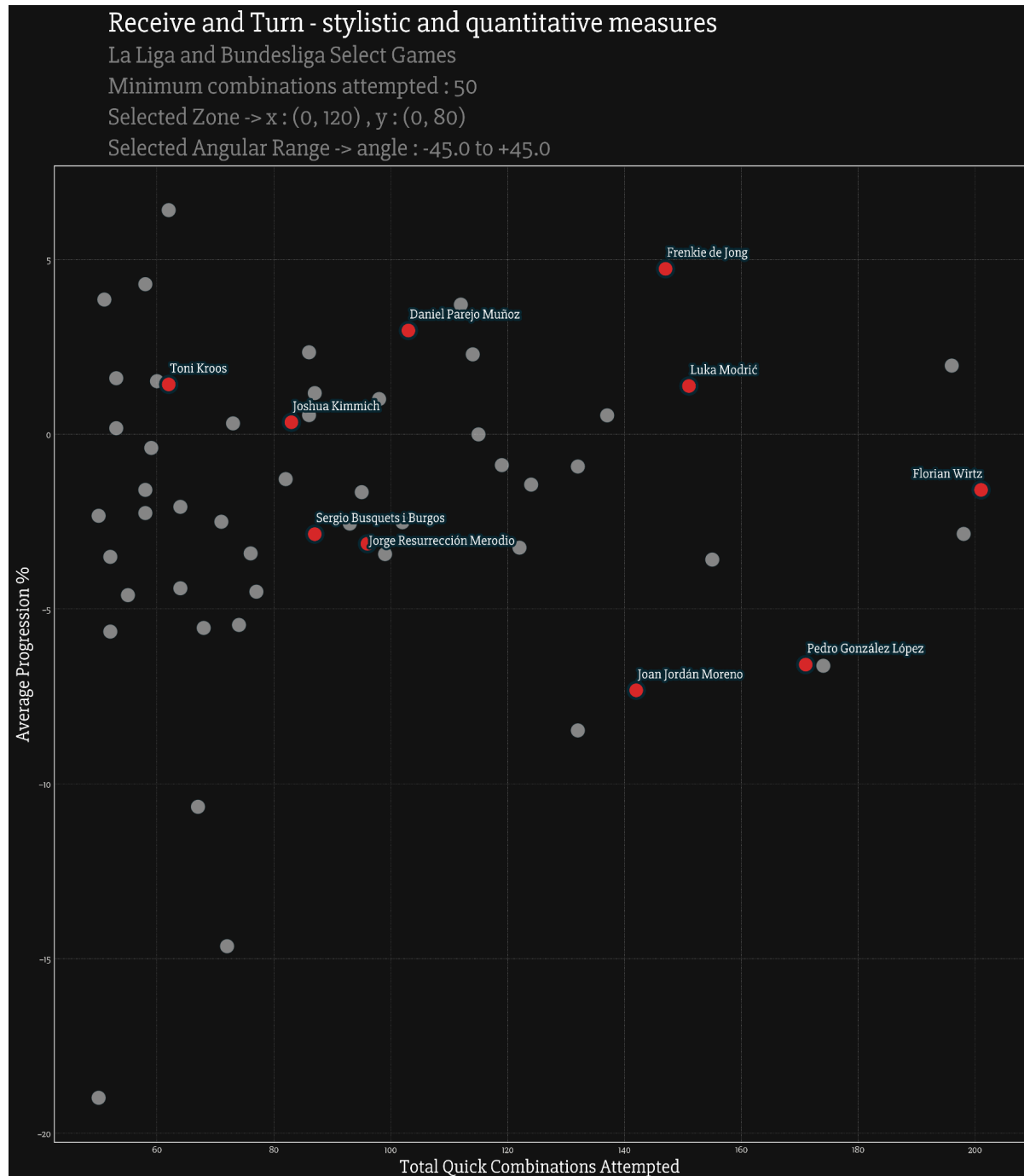
Figure 14. Ball progression from all combinations.

Figure 15. Ball progression from under-pressure combinations only.

Next, we have a quick look at (un)predictability - as measured by entropy - for all combinations for the selected players. This gives an idea of how varied a set of turns a particular receiver can produce on receiving the ball. As always, we compare it with the entropy for under-pressure actions only. Elite players remain pretty much unaffected by pressure.
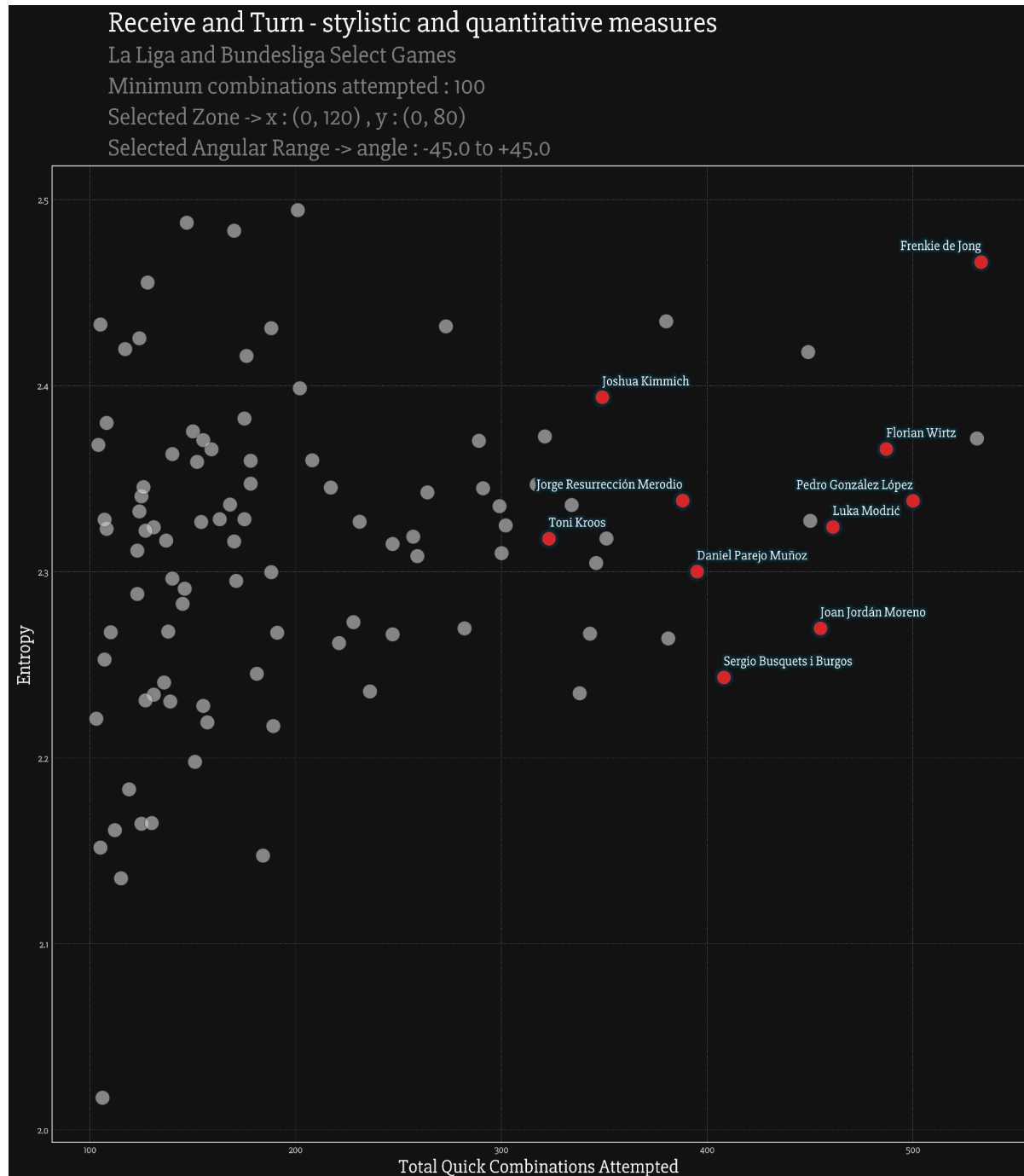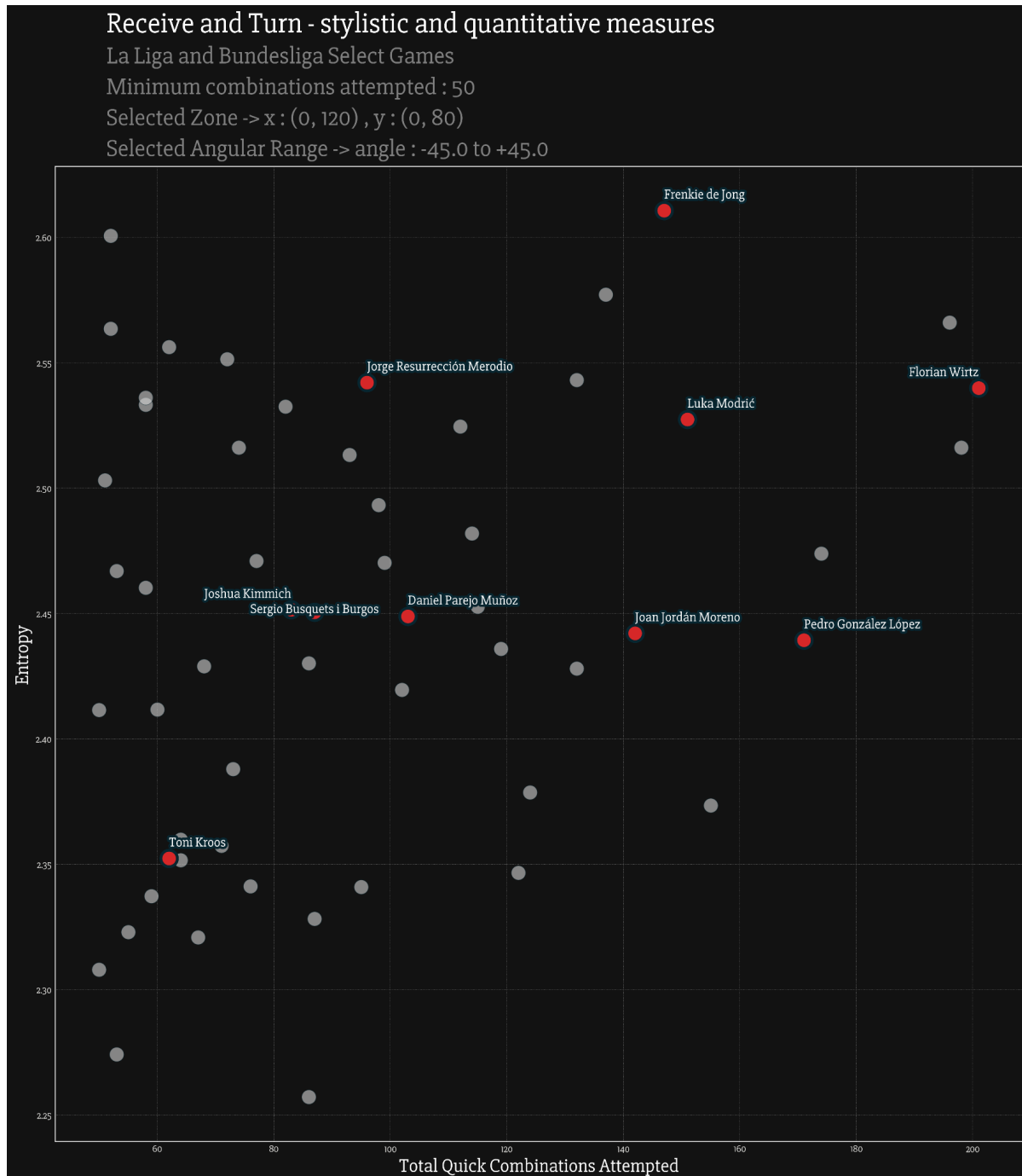
Figure 16. Unpredictability for all combinations.

Figure 17. Unpredictability for under-pressure combinations.

To further solidify the idea of players remaining unaffected by pressure, we look at the Jensen-Shannon divergence between the unpressured actions angular distribution and the pressured actions angular distribution. Elite players register comparatively low divergence, indicating that opposition press don't affect their game as much as others.
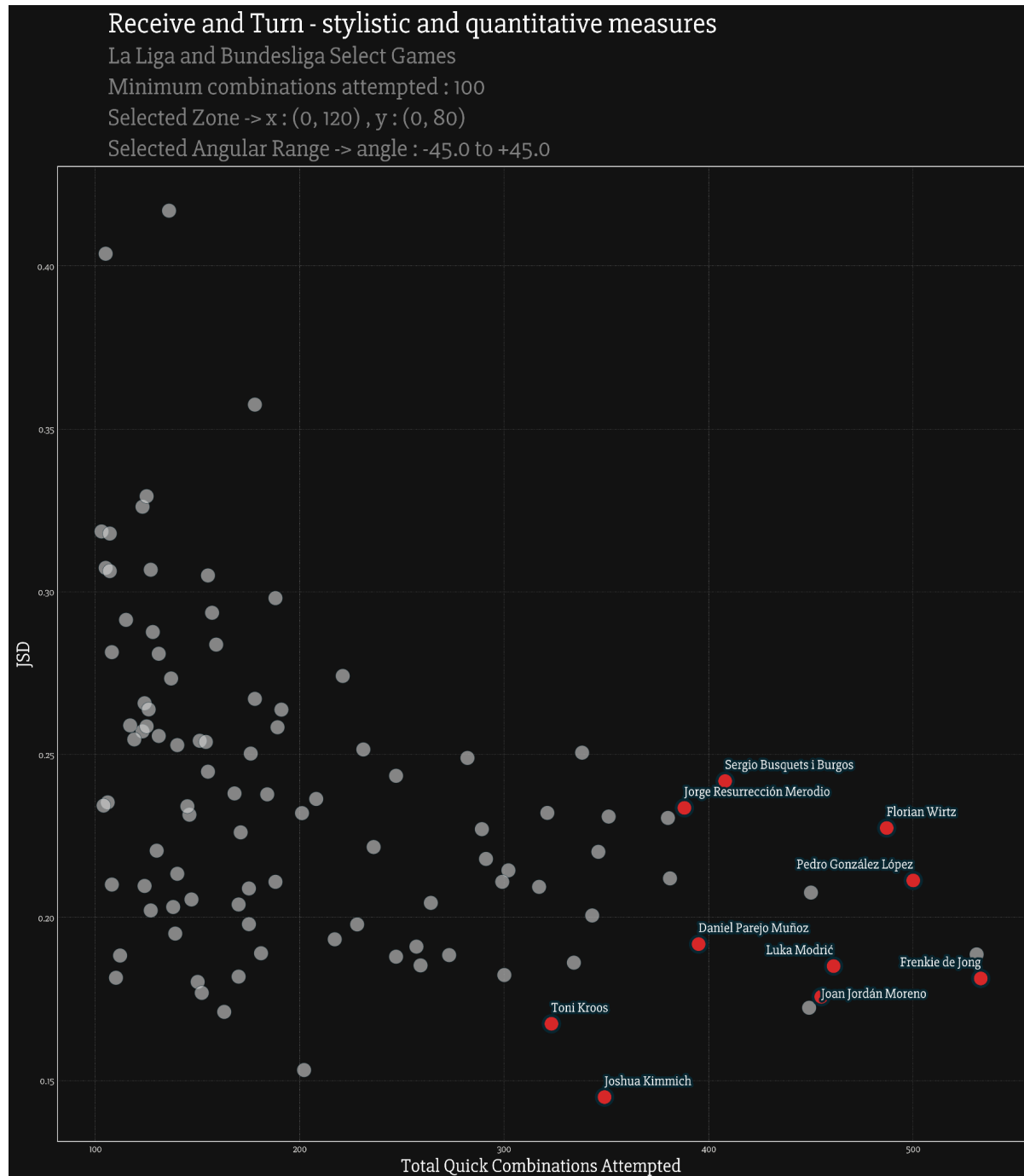
Figure 18.

Finally, we present some results from the expected move completion model, and compare actual completion percentages with expected completion percentages for several players (forwards are included in this graph).
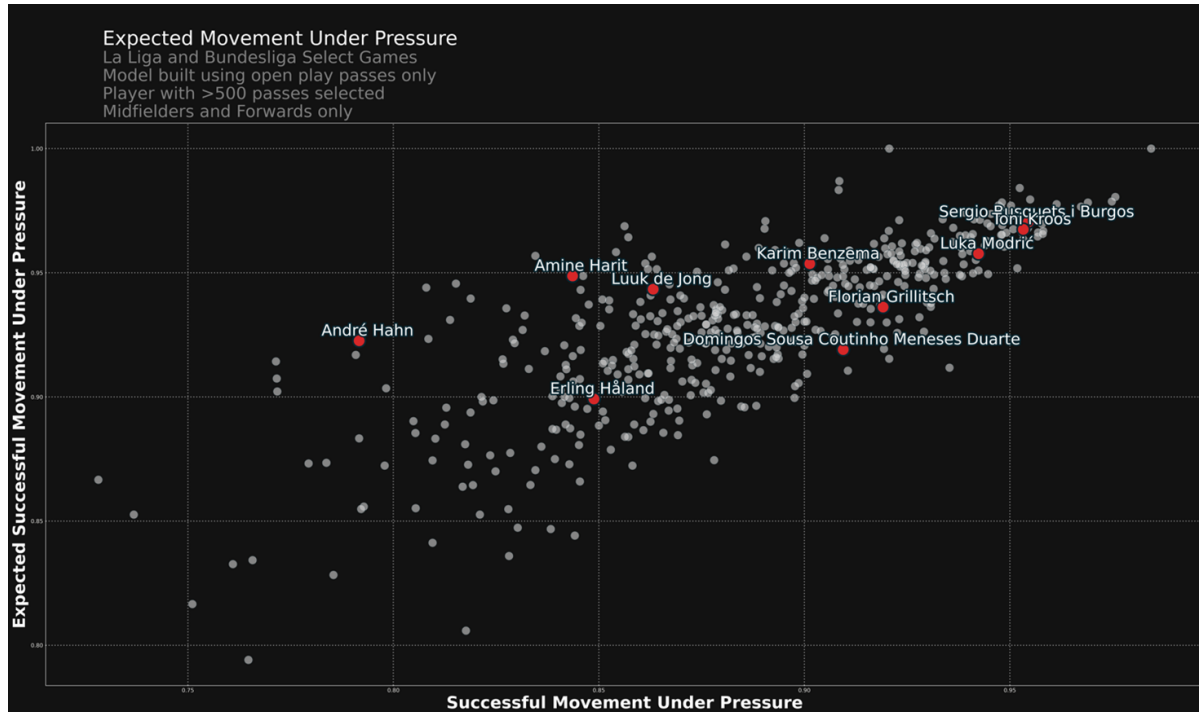
Figure 19. Move completion above expectation.

The cluster on the top right is the 'Elite Player' zone, players who are really unperturbed by pressures. The middle zone is mostly strikers, as they come under severe pressure when receiving the ball in the final third, a standout is Karim Benzema highlight below who performs exceptionally under pressure.

5 Uses of the formalism - actionable insights

This study can potentially have quite a few actionable insights. From a coaching point of view, especially for young players, this could be used as a tool to identify trends and act on. For eg. if a player is consistently sending a pass back upon receiving, is it because they are reacting late/arriving late to the first pass ? Is it because they are risk-averse ? Could there be a confidence issue ? An in-depth study with body pose data (mentioned in the next section) and subsequent coaching could really help the players grow.

It can definitely work as part of recruitment. Data scientists/analysts can in principle include some of the insights generated from the metrics explored above in their data-based reports for potential recruitment.

Finally, from an opposition analysis point of view, this could be used a data-based perspective to identify potential pressing traps. Examples would include forcing the ball to someone who is known to be a risk-averse passer, and force the ball deeper from there. Or, if the entropy metric indicates a proclivity towards turning in specific ways/directions, it can become a valuable information for the defending team when marking the said player.

6 Future Directions

In this entire study, we have used the angle between two successive events as a proxy for judging the turning abilities of the ball. The availability of body pose data in future will make this study more rigorous and complete. We will be able to study how exactly the players receive the ball (half-turn, back to goal etc) and based on posture and reaction time data, make much more accurate claims about turning.

Secondly, as pointed out in the Results section, the expected Threat value for the turning actions performed under pressure seem to be larger on average for the selected players, and several possibilities (not explored in this paper) were put forth. An immediate extension of this work could be to study those possibilities. It could very well turn out to be simply a "small sample size" effect. However, if it's a real effect observed in larger data sets, then it might be worth switching xT with another possession value metric that takes more context into account. For eg, possession value models built by Javi Fernandez et al taking tracking data into account, or StatsBomb's own OBV models could provide better insights than the simple xT model used here. Or, the xT model itself could be augmented by adding pressure and/or 360 insights.

7 Caveats

There are a couple of caveats associated with this analysis that needs to be kept in mind. Firstly, its not a huge data set. Some of the sample sizes are clearly small as shown in the Results section. As such, the exact numbers and rankings are not the most important aspect of the study - this paper is to be treated as a methodology paper, where the ideas can be extracted from and applied to a bigger data set if used to evaluate players for purposes of scouting. Further complications arise in the 360 frames where the passer is at the edge of the visual frame, and hence, his Voronoi cell of control or amount of pressure on the passer can't be accurately determined.

Acknowledgments

References

[1] John Muller, https://spacespacespaceletter.com/how-passing-vertex-angles-explain-chelsea/,

[2] Karun Singh, https://karun.in/blog/expected-threat.html

[3] American Soccer Analysis
https://www.americansocceranalysis.com/home/2020/7/27/gboostmeasuring-what-happens-after-the-pass