Evaluating players: Chelsea midfielders

Summer 2018 scouting report to analyse defensive metrics for central midfielders

Objective

We aim to analyse the current (2017-2018 season) Chelsea midfielders for their contribution or the lack of it to the defensive function of the team. To give context, N'golo Kanté had another stellar season used as a double pivot in the middle of the pitch. On the contrary, Cesc Fàbregas and Timoeue Bakayoko couldn't form a good partnership with Kanté leading to a string of poor defensive displays where the Blues were far too easy to play through their midfield block. In the summer of 2018, Chelsea made the call to replace Antonio Conte with Maurizio Sarri who, with Napoli, had managed to make a name for possession and high pressing football. With this study, we want to analyse the underlying stats for Serie A midfielders using some of the defensive metrics and gauge a suitable player to join Sarri at Chelsea this summer and partner Kanté for the upcoming season.

Defensive metrics

We have implemented the following 6 defensive metrics which are then visualized for each of the current midfielders at Chelsea to give us a comparative overview of player's respective strengths and weaknesses solely from a defensive (duelling and intercepting) perspective.

1. **Middle 3rd Def Duels Won**: This is a standard metric and we simple aggregate the volume of successfully won defensive duels in the middle 3rd by a player across all the matches played in the season. [1] A defensive duel is defined as "when a player attempts to dispossess an opposition player to stop an attack progressing".
2. **Middle 3rd successful Interceptions**: Another standard metric where we count the aggregated number of successful interceptions ([1] defined as "an act of player actively intercepting the ball by anticipating its movement when the opponent is shooting, passing or crossing) in the middle 3rd.
3. **Middle 3rd Fouls committed**: This too is a standard metric where we specifically look at the negative aspect of defending in fouls and aggregate the volume of fouls committed in the middle 3rd to allow us to gauge the defensive discipline of a player. The aim of adding in this metric to provide the whole picture where a player might have more defensive duels won and more interceptions but also giving away too many fouls in dangerous areas.
4. **Middle 3rd Interceptions xT avoided**: This has been developed using position-based [2] xT (Expected Threat) model. We associate the xT value avoided by interceptions in the middle 3rd based on the location and attribute it to the player who made the interception. The
5. **Middle 3rd Def Duels Counterattack xT**: This represents the xT value associated with potential counterattacks initiated from defensive duels won in the middle 3rd of the pitch. This metric also relies on the position-based xT model where we perform a probabilistic search moving from when the defensive duel action was performed that led to a series of passes before resulting into a counterattack.
6. **Def Duels Possession Chain xT**: This calculates the xT associated with defensive duel actions in isolated possession chains thereby rewarding players for these actions that allowed their teams to maintain possession. Xgboost model was trained on Bundesliga possession chains and scored for Premier League / Serie A midfielders. We filtered down the training and scoring set to use only 'Ground defending duel' actions only as the model was skewed towards tall forwards like Christian Benteke and Romelu Lukaku owing to them winning the attacking duels in dangerous areas in and around the box.

Chelsea midfielders

Radar charts provided below for the four players Chelsea deployed in midfield in the 2017-2018 season. We leave out Ross Barkley given he'd joined in January and also rely less on the chart for Danny Drinkwater as he played only roughly 700 minutes most of which were substitute appearances. That leaves us to analyse the charts below for Kanté, Fàbregas and Bakayoko. It is immediately clear that Kanté leads out for the defensive duel metrics with 97th and 98th percentiles for the standard defensive duels won and defensive duels counterattack xT metrics respectively.

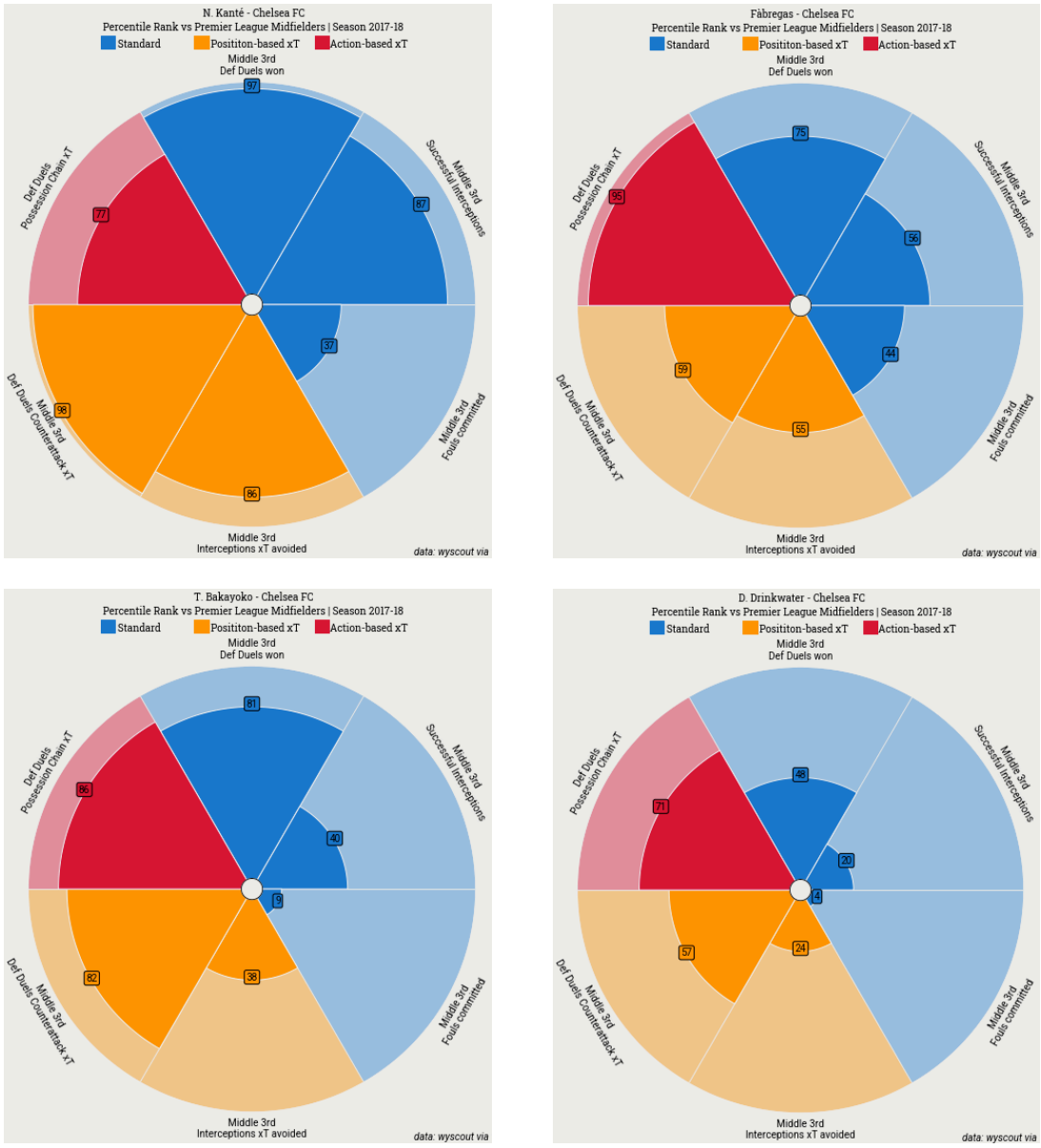

Figure 1: Chelsea midfielders for 2017-2018 season - percentile radar charts comparing the 5 metrics

Data assumptions:

Middle 3rd event's locations filtering has been applied for metrics 1-5. [1] "Midfielder" role name filtering is applied to calculate percentile rankings for all the metrics. Normalization for minutes played (all metrics presented as "{metric} per 90 minutes"), opposition possession % (metrics 1-5) and own team possession % (metric 6). As mentioned above we rely on a percentile-based ranking score and hence we use a negative axis for metric 3 (given that we hope to minimize the fouls committed metric).

Seria A midfielders

We narrow down our search on one metric to analyse players in Serie A which is metric 6: "Action-based Expected Threat for successful defensive duels in isolated possession chains". Please refer the below table for the top 10 players for this metric in this league for their performance in the 2017-2018 season. Again, all the data assumptions mentioned above hold valid for the numbers in this table. Standout players who deserve to be named include Allan of Napoli, Rafinha of Inter Milan, Manuel Locatelli of AC Milan and Zieliński of Napoli. Especially Locatelli aged just 20 was showing some early promising signs of a stellar defensive midfield display. Next, we discuss how with more contextual knowledge we are able to narrow down our search to only a couple of players before making the final recommendation.

shortName	def_mid_duels_won_adjusted_per90	def_mid_successful_interceptions_adjusted_per90	mid_fouls_adjusted_per90	xT_saved_adjusted_per90	xT_added_adjusted_per90	xT_action_based_adjusted_per90
V. Verre	1.858001	0.929000	2.787001	0.012616	0.046729	0.054907
F. Depaoli	2.107023	0.351171	0.936455	0.010791	0.026204	0.030496
Allan	5.889238	1.334894	1.806033	0.029513	0.089558	0.020866
Rafinha	4.001308	0.200065	1.200392	0.008027	0.062444	0.020758
J. Obi	3.256668	1.860953	2.636350	0.044862	0.078348	0.019403
R. Saponara	2.016659	0.201666	2.218325	0.002429	0.040618	0.018913
A. Acquah	1.886639	1.029076	1.200589	0.023913	0.038736	0.018691
M. Locatelli	2.998633	0.545206	1.908221	0.010145	0.056397	0.018474
A. Barberis	2.825458	0.730722	0.682007	0.014560	0.057470	0.017553
P. Zieliński	2.818865	0.939622	0.939622	0.019099	0.042547	0.017543

Figure 2: Top 10 Seria A midfielders for 'Def Duels Possession Chain xT' metric

We look at the radar chart comparison between the two Napoli midfielders who played together under Sarri in the 2017-2018 season. Again, to provide some context here, Chelsea ended up signing Jorginho that summer and we aim to rely on the defensive metrics to validate our hypothesis that Allan would have been a better choice? Focusing on metric 6 chosen, Jorginho ranks in the 53rd percentile compared to Allan's 98th percentile for this metric. In fact, as seen in the table above, Allan is the 3rd best midfielder for this metric across all Seria A midfielders. It is worth noting that even though Jorginho ranks 99th and 100th percentile for counterattack feeding defensive duels and position-based xT avoided with interceptions that is something Chelsea have already got covered in the best-in-class Kanté. Also, the fact that Chelsea aren't necessarily a counterattacking team and giving preference for action-based xT over position-based xT, we believe Allan would be better suited to over Jorginho to provide the ideal midfield partnership with Kanté.

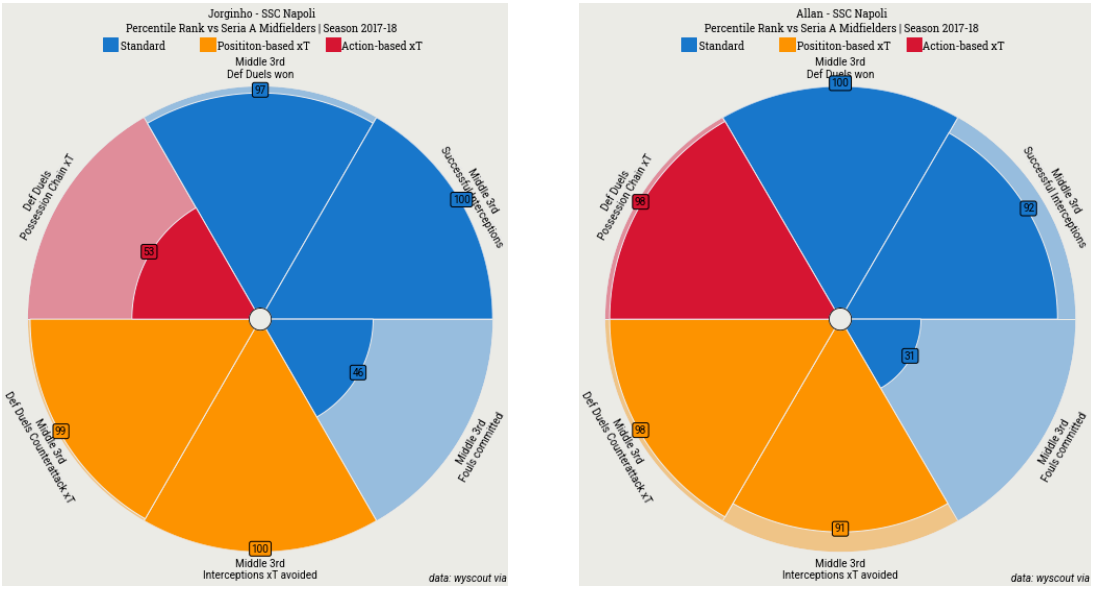

Figure 3: Jorginho vs Allan for SSC Napoli 2017-2018 season - percentile radar charts comparing the 5 metrics

Search criteria: In narrowing down our search to focus on Allan, we present a direct comparison with Jorginho based on the following factors which we believe are fair given that they played for the same team:

a. **Age**: Aged, 27 (Allan) and 26 (Jorginho) both players are heading into the [3] peak ages for midfielders. Minutes played stats stand at 2949 (Allan) vs 2722 (Jorginho) so nothing to separate
b. **Playing style**: Both players are accustomed to playing in a double pivot deep-lying playmaker's role who supply the team with quick exchanges with the wing players in a Maurizio Sarri managed Napoli team that has dominated over 60% of the possession in the recently completed season and hence would be a natural fit in transitioning over to Chelsea and their playing style.
c. **Market value context as seen in the graph below**: Allan was signed by Napoli and Sarri for €10m in 2015. Currently he's valued at €25m compared to Jorginho's €40m. Given a better value for money and being a year older in his peak age we believe Allan just pip's Jorginho solely based on these three factors. To provide another interesting aspect and given we have data looking back from the present day, see below table for how both player's valuation evolved over time since.

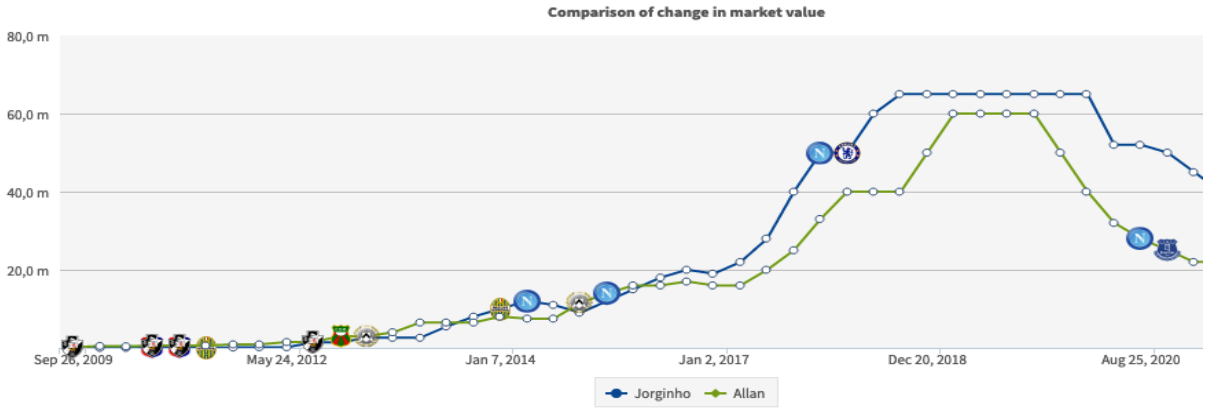

Figure 4: Jorginho vs Allan Transfermarkt comparison of market value over time

Final recommendation:

There was a common consensus with the group given the detailed investigation of the current Chelsea situation as was observed in the summer of 2018. Emphasis of this analysis was placed on the search for a midfielder who could ably partner Kanté at the heart of Chelsea's midfield and this idea was concurred by the group. We discussed that, Allan would not only proivde defensive solidarity but also creativity in spawning attacks and counterattacks from deep making him an all-round number 8 / number 6 prospect. Given there was a split between defenders and midfielders analysed within the group, we realised that the underlying data would need to be calibrated across different metrics for performing a fair comparison. Additionally, all of us had analysed players from different leagues which also suggestss the need for normalization for relative league strength. Owing to all of these complexities and citing the most relevance in terms of forming that strong parternship with Kanté in the midfield, the group agreed that Allan was the most suitable midfielder amongst the top 10 players presented (sorted by the narrowed down action-based xT metric).

Code:

1. Create possession chains and train action-based xT model using this notebook: https://github.com/yashkarle/soccer-ds-stats/blob/main/Projects/4UppsalaMMS/4_ValuingActions/possession_chains.ipynb
2. Create final metrics and plot radar charts using this notebook: https://github.com/yashkarle/soccer-ds-stats/blob/main/Projects/4UppsalaMMS/3_StatsScouting/assignment_2_evaluating_players.ipynb

Data:

Wyscout data via figshare, supporting data json dumps and pre-trained models can be found here: https://github.com/yashkarle/soccer-ds-stats/tree/main/Projects/4UppsalaMMS/data/Wyscout

Future work:

We can look into more details of the under the hood implementations of the xG model used for the action-based xT metric calculations. This will allow us to customize the feature set for midfielders looking specifically from defensive duels point of view. We can also utilize feature attribution techniques like Shapley values to understand how exactly the xG model relies on individual features to better tweak the same for our specific use case. Lastly, we can use of clustering techniques to add player roles and position-based groupings in the data containing the metrics when doing the final ranking better in an objective manner.

References:

1. Wyscout data glossary: https://dataglossary.wyscout.com/
2. Expected Threat model: https://karun.in/blog/expected-threat.html
3. Peak age of players based on their position: https://theathletic.com/2935360/2021/11/15/what-age-do-players-in-different-positions-peak/