

## Profilage technique des joueurs de football

Drs Raffaele Poli, Loïc Ravenel et Roger Besson

### 1. Introduction

Ce rapport mensuel présente la méthode développée par l'Observatoire du football CIES pour déterminer le profil technique des footballeurs, les classer dans différents groupes et établir des hiérarchies au sein de ces derniers. Pour ce faire, nous nous sommes basés sur les données relatives aux actions effectuées par les joueurs (tirs, centres, interceptions, etc.) produites par notre partenaire [InStat](#).

L'analyse a été effectuée sur un échantillon de 7'215 joueurs ayant disputé au moins 750 minutes de championnat lors de la saison 2021 ou 2021/22 (jusqu'à fin janvier 2022) au sein de 36 ligues de première ou deuxième division d'associations membres de l'UEFA. Du fait de la spécificité de leur poste, les gardiens de but n'ont pas été inclus dans l'analyse.

Figure 1 : ligues prises en compte et nombre de joueurs

 Bundesliga (AUT)	138 joueurs	
 Pro League (BEL)	236 joueurs	
 Premier League (BLR)	235 joueurs	
 First League (BUL)	166 joueurs	
 1. HNL (CRO)	122 joueurs	
 1. Division (CYP)	127 joueurs	
 First League (CZE)	191 joueurs	
 Superliga (DEN)	129 joueurs	
 Premier League (ENG)	257 joueurs	
Championship (ENG/2)	348 joueurs	
 Liga (ESP)	257 joueurs	
Segunda División (ESP/2)	309 joueurs	
 Veikkausliiga (FIN)	181 joueurs	
 Ligue 1 (FRA)	260 joueurs	
Ligue 2 (FRA/2)	253 joueurs	
 Bundesliga (GER)	217 joueurs	
2. Bundesliga (GER/2)	211 joueurs	
 Super League (GRE)	163 joueurs	
 NB I (HUN)	123 joueurs	
 Ligat haAl (ISR)	165 joueurs	
 Serie A (ITA)	269 joueurs	
Serie B (ITA/2)	245 joueurs	
 Eredivisie (NED)	211 joueurs	
 Eliteserien (NOR)	230 joueurs	
 Ekstraklasa (POL)	202 joueurs	
 Primeira Liga (POR)	213 joueurs	
 Liga I (ROM)	199 joueurs	
 Premier League (RUS)	187 joueurs	
 Premiership (SCO)	145 joueurs	
 Super Liga (SRB)	202 joueurs	
 Super League (SUI)	113 joueurs	
 Super Liga (SVK)	126 joueurs	
 1. SNL (SVN)	107 joueurs	
 Allsvenskan (SWE)	237 joueurs	
 Süper Lig (TUR)	268 joueurs	
 Premier League (UKR)	184 joueurs	

## 2. Variables et axes de classification

Parmi les très nombreux indicateurs de jeu collectés par InStat, onze ont été retenus pour déterminer le profil technique des joueurs. Ils renvoient à des actions tant défensives qu'offensives. La sélection a été effectuée afin de limiter les redondances et éliminer les variables trop concentrées sur quelques individus en raison du faible nombre d'événements. Par exemple, les tirs ont été préférés aux buts, les premiers étant fortement corrélatés aux seconds, tout en étant mieux répartis entre les joueurs.

Afin de déterminer des profils techniques indépendamment du niveau ou style de jeu du club d'emploi, les valeurs attribuées aux joueurs pour chacune des onze variables sélectionnées ont été définies en référence à la valeur moyenne des autres membres de l'équipe, à savoir comme un ratio entre la valeur du joueur et celle de l'ensemble de ses coéquipiers (gardiens exclus).

Par exemple, une valeur égale à deux sur le plan des tirs indique que le joueur réalise deux fois plus de tirs que ses partenaires. De cette manière, un joueur évoluant dans une équipe de faible niveau ne voit pas ses valeurs structurellement réduites par rapport aux joueurs d'équipes mieux classées. Ainsi, nous pouvons véritablement analyser un profil de jeu plutôt qu'une performance fortement liée à la force des équipes.

À partir des valeurs relativisées, nous avons réalisé une analyse en composantes principales (ACP). Les résultats sont exprimés visuellement dans la Figure 3 sous la forme d'un plan factoriel avec les onze variables retenues représentées par des flèches. Plus une flèche est longue et proche d'un axe, plus la variable en question est importante dans la définition de ce dernier.

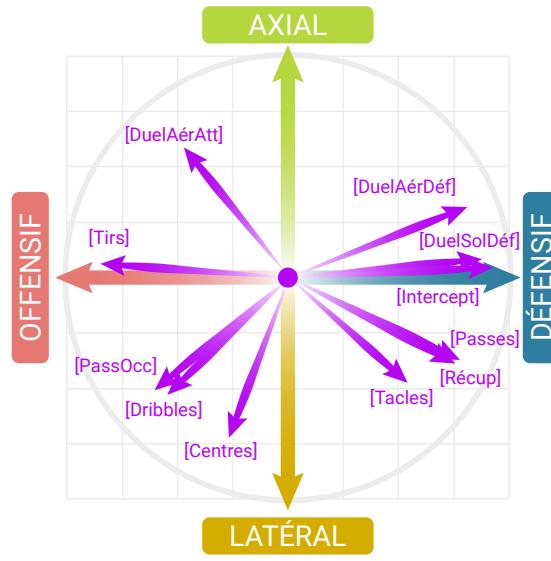
La variable interception est fortement impliquée dans la formation de l'axe horizontal, tout comme celle des tirs à son opposé. Cet axe définit donc le penchant défensif ou offensif des joueurs. Les deux variables les plus discriminantes au niveau de l'axe vertical sont les centres, qui sont surtout le fait de joueurs de couloir, et les duels aériens, qui sont surtout le fait des avant-centres et des défenseurs centraux. Cet axe renvoie ainsi aux différents positionnements des joueurs du même compartiment de jeu (défense, milieu ou attaque).

Les deux axes principaux de l'ACP expliquent presque 70% de la variance totale, l'axe défensif-offensif expliquant à lui seul presque la moitié de cette dernière. Cela signifie que les onze variables retenues, ainsi la relativisation par rapport aux coéquipiers des valeurs attribuées aux joueurs, permettent de rendre largement compte des différences dans les profils techniques des footballeurs.

Figure 2 : indicateurs de jeu retenus pour le profilage

Vocation	Variable	Id
Défensive	Duels aériens gagnés - défense	[DuelAérDéf]
	Interceptions	[Intercept]
	Récupération	[Récup]
	Duels au sol gagnés - défense	[DuelSolDéf]
	Tacles réussis	[Tacles]
Neutre	Passes	[Passes]
Offensive	Duels aériens gagnés - attaque	[DuelAérAtt]
	Centres	[Centres]
	Dribbles réussis	[Dribbles]
	Passes pour occasion	[PassOccas]
	Tirs	[Tirs]

Figure 3 : le plan factoriel pour le profilage des joueurs



### 3. Proximité des joueurs

Cette méthode de profilage permet de calculer des distances entre joueurs. À partir d'un footballeur de référence, il est ainsi possible de déterminer les joueurs qui lui sont statistiquement les plus proches. Par exemple, parmi les footballeurs du big-5, celui ayant le profil de jeu le plus proche de Kylian Mbappé est James Maddison (Leicester City). Si on restreint l'analyse aux joueurs de Ligue 1, c'est Steph Mavididi (Montpellier HSC) qui se rapproche le plus du champion du monde.

Cet exercice peut être effectué en prenant n'importe quel joueur comme référence. A titre d'exemple, toujours au niveau du big-5, Dušan Vlahović apparaît comme le joueur le plus proche d'Erling Haaland, Paulo Dybala de Lionel Messi, Romain Faivre de Neymar Júnior, Dominik Szoboszlai de Kevin de Bruyne, Sadio Mané de Raheem Sterling, Remo Freuler de Jorginho Frello ou encore Jonathan Tah de Virgil van Dijk.

### 4. Classification des joueurs

Au-delà du calcul des distances entre joueurs, l'algorithme des k-medoids permet de les classer en groupes. Cette méthode, dérivée des nuées dynamiques (k-means en anglais) se base sur le choix de joueurs de référence servant de profils-type pour l'élaboration de classes auxquelles s'agrègent, par proximité statistique successive, tous les individus. Six joueurs de référence aux profils et postes divergents ont été sélectionnés dans le cadre de ce rapport (Figure 5).

**Figure 4 : les cinq joueurs de Ligue 1 et des autres ligues du big-5 au profil de jeu le plus proche de Kylian Mbappé**



**Figure 5 : joueurs-type utilisés pour la création de classes**

	Virgil van Dijk Liverpool FC	Défenseur central
	Marcos Alonso Chelsea FC	Défenseur latéral
	Jorginho Frello Chelsea FC	Milieu défensif
	Bruno Fernandes Manchester United	Milieu offensif
	Raheem Sterling Manchester City	Ailier
	Romelu Lukaku Chelsea FC	Avant-centre

La Figure 6 illustre les gestes techniques dominants pour chacune des six classes construites à partir des joueurs de référence. À titre d'exemple, les joueurs de la classe bâtie à partir de Virgil van Dijk gagnent en moyenne 2,62 duels aériens défensifs de plus que leurs co-équipiers, ceux de la classe construite à partir de Marcos Alonso centrent 2,62 fois plus, et ainsi de suite.

### Classe Virgil van Dijk

La « classe Van Dijk » regroupe des joueurs dont les valeurs au niveau de toutes les variables défensives sont supérieures à celles de leurs coéquipiers. Il s'agit surtout de défenseurs centraux se distinguant par leur forte présence dans les duels, tant aériens qu'au sol. Cette classe réunit 21.7% des joueurs de notre échantillon.

### Classe Marcos Alonso

La « classe Marcos Alonso » identifie aussi des joueurs à vocation plutôt défensive, mais très actifs aussi sur le plan offensif au niveau des centres. Il s'agit surtout de défenseurs latéraux ou de milieux évoluant dans le couloir dans le cadre d'un schéma tactique « 3-5-2 ». Cette classe regroupe 15.7% des footballeurs pris en compte.

### Classe Jorginho Frello

La « classe Jorginho Frello » regroupe aussi des joueurs à vocation plutôt défensive, mais relativement actifs aussi offensivement. Il s'agit surtout de milieux axiaux, tant défensifs que relayeurs. Cette classe est non seulement celle qui réunit le plus de joueurs, 26.5% des effectifs analysés, mais aussi la plus hétérogène.

### Classe Bruno Fernandes

La « classe Bruno Fernandes » réunit des joueurs plus actifs offensivement que défensivement. Leur spécificité réside surtout dans la capacité à créer des occasions pour les co-équipiers et à l'importance de leur participation à l'animation offensive générale (dribbles, tirs, centres, etc.). Cette classe est numériquement la plus petite : elle ne regroupe que 8.6% des joueurs.

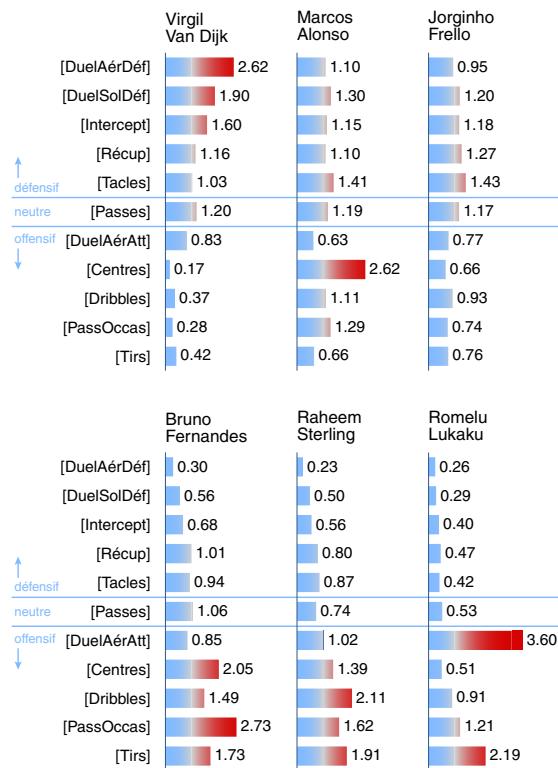
### Classe Raheem Sterling

La « classe Raheem Sterling » identifie des joueurs au profil voisin de celui de la classe Bruno Fernandes, mais se différenciant par une propension plus importante aux tirs et aux dribbles et moins importante aux passes décisives et aux centres. Cette classe réunit 18.0% des joueurs de notre échantillon.

### Classe Romelu Lukaku

La « classe Romelu Lukaku » identifie surtout des footballeurs jouant en tant qu'avant-centres. Les joueurs qui en font partie sont relativement peu présents dans le jeu et concentrent leurs efforts dans deux domaines qui constituent leur spécificité : la finition et les duels aériens offensifs. Cette classe regroupe 9.8% des joueurs.

**Figure 6 : fréquence relative moyenne des gestes techniques par classe (par rapport aux co-équipiers)**



## 5. Une hiérarchisation des joueurs selon les profils définis

Toute hiérarchisation de joueurs se heurte non seulement au problème très souvent sous-estimé de l'impact sur les performances individuelles du différentiel de force collective entre adversaires, mais aussi à la difficulté d'établir des profils de joueurs au style de jeu assez proche pour qu'une comparaison ait du sens.

La relativisation des indicateurs de performance par rapport aux co-équipiers et la création de profils-type à partir de joueurs de référence sont autant de démarches qui limitent ces problèmes, sans cependant les évacuer totalement. En effet, toute classe, aussi homogène soit-elle, comprend toujours des marges où se situent des joueurs au profil atypique ou inter-classe.

Une solution réside à élargir le nombre de classes par rapport au schéma de départ en déterminant par l'outil statistique de l'indicateur de silhouette le degré par lequel la classe à laquelle les joueurs ont été attribués les représente. Ce faisant, les footballeurs bien décrits restent affectés à leur classe, tandis que de nouvelles classes peuvent être créées par l'assemblage de joueurs au profil inter-classe.

Par exemple, si un joueur comme Thomas Partey d'Arsenal est bien défini comme membre de la « classe Jorginho », Nemanja Matić se situe dans une position intermédiaire entre cette classe et la « classe Van Dijk ». Nous pouvons ainsi établir une nouvelle classe regroupant tous les individus statistiquement plus proches de Matić que de Jorginho ou van Dijk. Au final, deux classes intermédiaires (Matić et Trippier) ont été ajoutées aux six classes de départ.

Figure 7 : les neuf classes utilisées pour la hiérarchisation

Classes	Variables les plus représentatives
1 Virgil Van Dijk	1 Duels aériens gagnés - défense 2 Duels au sol gagnés - défense 3 Interceptions
2 Nemanja Matić	1 Duels aériens gagnés - défense 2 Duels aériens gagnés - attaque 3 Interceptions
3 Marcos Alonso	1 Centres 2 Tacles réussis 3 Passes pour occasion
4 Kieran Trippier	1 Centres 2 Tacles réussis 3 Duels au sol gagnés - défense
5 Jorginho Frello	1 Tacles réussis 2 Récupération 3 Passes
6 Bruno Fernandes	1 Passes pour occasion 2 Centres 3 Tirs
7 Raheem Sterling	1 Dribbles réussis 2 Tirs 3 Passes pour occasion
8 Romelu Lukaku	1 Duels aériens gagnés - attaque 2 Tirs 3 Passes pour occasion

Pour chacune des huit classes, nous avons identifié les trois variables où les joueurs s'écartent en moyenne le plus de leurs co-équipiers toutes classes confondues (voir Figure 7), et établi des hiérarchies sur la base de ces trois variables. Par exemple, les joueurs de la « classe Van Dijk » ont été classés en prenant en compte leurs écarts par rapport aux co-équipiers au niveau des trois variables suivantes : duels aériens défensifs, duels au sol défensifs et interceptions.

Les Figures 8a et 8b présentent le top 10 des footballeurs pour chacune des neuf classes toutes ligues confondues. Néanmoins, il est aussi possible et judicieux de prendre en compte le niveau de la ligue dans laquelle les footballeurs évoluent afin de comparer seulement des joueurs de championnats de force relativement similaire, le big-5 par exemple.

**Figure 8a : joueurs les plus actifs par rapport aux co-équipiers**

**Classe 1** | Virgil Van Dijk / [DuelAérDéf], [DuelSolDéf], [Intercept]

1	Haiderson Hurtado	26.3 ans	ŠKF Sered' (SVK)
2	Maximilian Wöber	24.1 ans	RB Salzburg (AUT)
3	Leonardo Balerdi	23.1 ans	Olympique Marseille (FRA)
4	Simone Canestrelli	21.5 ans	FC Crotone (ITA/2)
5	Juuso Hämäläinen	28.2 ans	Inter Turku (FIN)
6	Frank Boya	25.7 ans	Zulte Waregem (BEL)
7	Oliver Abildgaard	25.7 ans	Rubin Kazan (RUS)
8	Frederik Sørensen	29.9 ans	Ternana Calcio (ITA/2)
9	Nicolas Wimmer	27.0 ans	Austria Klagenfurt (AUT)
10	Gabriele Angella	32.8 ans	AC Perugia (ITA/2)

**Classe 2** | Nemanja Matić / [DuelAérDéf], [DuelAérAtt], [Intercept]

1	Andrian Kraev	23.0 ans	Levski Sofia (BUL)
2	Saša Tomanović	32.4 ans	TSC Bačka Topola (SRB)
3	Sergiy Petko	28.1 ans	Veres Rivne (UKR)
4	Benjamin André	31.6 ans	LOSC Lille (FRA)
5	Marco van Ginkel	29.2 ans	PSV Eindhoven (NED)
6	Krzysztof Kubica	21.8 ans	Górnik Zabrze (POL)
7	Tomáš Souček	27.0 ans	West Ham United (ENG)
8	Marshall Munetsi	25.7 ans	Stade de Reims (FRA)
9	Boubacar Fofana	32.3 ans	Sepsi OSK (ROM)
10	Oleksiy Dovhyi	32.3 ans	FK Lviv (UKR)

**Classe 3** | Marcos Alonso / [Centres], [Tacles], [PassOccas]

1	Alfonso Pedraza	25.9 ans	Villarreal CF (ESP)
2	Iasmin Latovlevici	35.8 ans	FC Argeș (ROM)
3	Albert Adomah	34.2 ans	QPR (ENG/2)
4	Illia Putria	23.8 ans	FK Chornomorets (UKR)
5	Leonardo Lukačević	23.1 ans	Admira Wacker (AUT)
6	Krum Stoyanov	30.6 ans	CSKA 1948 (BUL)
7	Rasmus Carstensen	22.2 ans	Silkeborg IF (DEN)
8	Damián Suárez	33.8 ans	Getafe CF (ESP)
9	Yanis Hamache	22.6 ans	Boavista FC (POR)
10	Pedro Porro	22.5 ans	Sporting CP (POR)

**Figure 8b : joueurs les plus actifs par rapport aux co-équipiers****Classe 4** Kieran Trippier / [Centres], [Tacles], [DuelSolDéf]

1	João Aurélio	33.5 ans	Pafos FC (CYP)
2	Ahmet Oğuz	29.1 ans	Sivasspor (TUR)
3	Óscar de Marcos	32.9 ans	Athletic Club (ESP)
4	Ignacio Guerrico	23.6 ans	NK Maribor (SVN)
5	Luis Carlos Murillo	31.4 ans	HJK Helsinki (FIN)
6	Yvan Dibango	20.0 ans	FC Isloch (BLR)
7	Danilo Soares	30.3 ans	VfL Bochum (GER)
8	Ilya Samoshnikov	24.3 ans	Rubin Kazan (RUS)
9	Rubén Duarte	26.4 ans	CD Alavés (ESP)
10	Justas Lasickas	24.4 ans	FK Voždovac (SRB)

**Classe 5** Jorginho Frello / [Tacles], [Récup], [Passes]

1	Antonio Vacca	31.8 ans	Venezia FC (ITA)
2	Theofanis Tzandaris	28.7 ans	PAS Lamia (GRE)
3	Ali Mohamed	26.4 ans	Maccabi Haifa (ISR)
4	Obinna Nwobodo	25.2 ans	Göztepe SK (TUR)
5	Cameron Devlin	23.7 ans	Heart of Midlothian (SCO)
6	Mikhail Bashilov	29.1 ans	Energetik-BGU (BLR)
7	Nikita Korzun	27.0 ans	Shakhtyor Soligorsk (BLR)
8	Miguel Mellado	29.0 ans	OFI Crete (GRE)
9	Juan Munafó	33.9 ans	Asteras Tripolis (GRE)
10	Óscar Valentín	27.5 ans	Rayo Vallecano (ESP)

**Classe 6** Bruno Fernandes / [PassOccas], [Centres], [Tirs]

1	Ronaldo Deaconu	24.8 ans	Gaz Metan Mediaş (ROM)
2	Shahboz Umarov	23.0 ans	Energetik-BGU (BLR)
3	Ricardo Quaresma	38.4 ans	Vitória Guimarães (POR)
4	Quentin Cornette	28.1 ans	Le Havre AC (FRA/2)
5	Alexis Flips	22.1 ans	Stade de Reims (FRA)
6	Filip Kostić	29.3 ans	Eintracht Frankfurt (GER)
7	Yuliyan Nenov	27.3 ans	Botev Vratsa (BUL)
8	Fabian Reese	24.2 ans	Holstein Kiel (GER/2)
9	Kristijan Lovrić	26.2 ans	HNK Gorica (CRO)
10	Yannick Bolasie	32.8 ans	Çaykur Rizespor (TUR)

**Classe 7** Raheem Sterling / [Dribbles], [Tirs], [PassOccas]

1	Adama Traoré	26.1 ans	Wolverhampton FC (ENG)
2	Heorhii Tsitaishvili	21.3 ans	FK Chornomorets (UKR)
3	Marko Livaja	28.5 ans	Hajduk Split (CRO)
4	Gaëtan Laura	26.6 ans	Paris FC (FRA/2)
5	Rabbi Matondo	21.5 ans	Cercle Brugge (BEL)
6	Antoine Semenyo	22.1 ans	Bristol City (ENG/2)
7	Joseph Hungbo	22.1 ans	Ross County (SCO)
8	Chidera Ejike	24.2 ans	CSKA Moskva (RUS)
9	Khvicha Kvaratskhelia	21.0 ans	Rubin Kazan (RUS)
10	Jens Odgaard	22.9 ans	RKC Waalwijk (NED)

**Classe 8** Romelu Lukaku / [DuelAérAtt], [Tirs], [PassOccas]

1	Milan Djurić	31.8 ans	US Salernitana (ITA)
2	Tobias Lauritsen	24.5 ans	Odds BK (NOR)
3	Matt Smith	32.7 ans	Millwall FC (ENG/2)
4	Mostafa Mohamed	24.2 ans	Galatasaray SK (TUR)
5	Roberts Uldrikis	23.9 ans	SC Cambuur (NED)
6	Joselu Mato	31.9 ans	CD Alavés (ESP)
7	Rok Kidrič	26.8 ans	NK Aluminij (SVN)
8	El Bilal Touré	20.4 ans	Stade de Reims (FRA)
9	Andy Delort	30.4 ans	OGC Nice (FRA)
10	Aliou Badji	24.4 ans	Amiens SC (FRA/2)

## 6. Conclusion

Les indicateurs de jeu collectés par InStat constituent un trésor à partir duquel il est possible de développer de multiples recherches innovantes avec des applications très pratiques. Toujours soucieux de jeter des ponts entre science et industrie, ce rapport a été conçu dans cette perspective, qui nous tient particulièrement à cœur.

La méthode de profilage et de classification détaillée dans cette étude est particulièrement utile non seulement dans une perspective descriptive pour comprendre les différents rôles remplis par les joueurs au sein d'une équipe, mais aussi sur le plan de la détection. À ce niveau, le calcul de distances statistiques entre joueurs est un très bon outil pour cibler des recrues potentielles par lesquelles remplacer des joueurs sur le départ.

Le choix de profiler et hiérarchiser les joueurs non pas sur la base de leurs statistiques brutes mais par rapport aux co-équipiers est aussi particulièrement fécond dans une perspective de détection. En effet, cela limite le semipiternel problème de l'impact de la force collective d'une équipe sur la performance individuelle. Ce procédé permet d'identifier des joueurs qui ne se démarquent pas forcément sur un plan absolu, mais dont la productivité est bien supérieure à celle de leurs co-équipiers.

Ce rapport ne constitue qu'un pas supplémentaire dans la direction d'une pleine exploitation des très larges possibilités offertes en termes de recherche et de développement par l'analyse de données techniques de jeu telles que celles récoltées par InStat. Nous nous réjouissons de pouvoir creuser cette question encore davantage et d'être en mesure de diffuser gratuitement les nouvelles avancées auprès de tous les passionnés du jeu.