

# Coaching, Intelligence Artificielle et Science des Données

Cédric Damien, Omar El Euch, Théo Guillaumot,  
Othmane Mounjid, Mathieu Rosenbaum

École Polytechnique  
19 avril 2018



## Disclaimer

Attention nous nous contentons de faire parler les données. Ces résultats doivent être seulement considérés comme un outil d'aide à la décision !

# Plan

- 1 On refait le match (comment se passer des données de tracking)
- 2 On anticipe le match (modélisation)
- 3 On joue le match (coaching optimal)
- 4 La coupe du monde 2018
  - Comment jouer à la coupe du monde ?
  - La liste des 23 optimale

# Plan

- 1 On refait le match (comment se passer des données de tracking)
- 2 On anticipe le match (modélisation)
- 3 On joue le match (coaching optimal)
- 4 La coupe du monde 2018

## Introduction et motivation

### Les données d'Opta F24

Opta fournit tous les événements **avec ballon** survenus pendant un match :

- Événements d'attaque : Passes, tirs, dribbles...
- Événements de défense : Tacles, récupérations de balle, duels aériens...
- Coups de pied arrêtés/remises en jeu : Fautes, touches, corners...
- Autres événements : Cartons, changements de joueur, changements de stratégie...
- Les positions et identités des joueurs impliqués dans chaque événement sont données.

## Introduction et motivation

### Les données d'Opta F24

- Application : Calcul de plusieurs métriques, par exemple la probabilité qu'un joueur réussisse une passe ou un tir dans une certaine zone du terrain.
- Mais : ces données sont seulement centrées autour de la balle.
- Pas d'information sur la position des joueurs ne touchant pas le ballon.
- On ne peut par exemple pas calculer d'indicateur mesurant les compétences défensives ou les qualités de placement d'un joueur. Impossibilité d'utiliser le flux TV pour cela.
- **Comment inférer la position de tous les joueurs sur le terrain en ne connaissant que celle des joueurs touchant la balle ?**

## Un premier replayer naïf

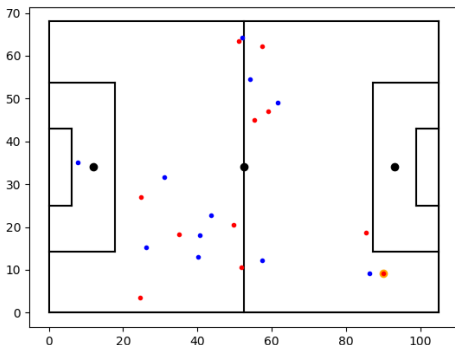
### Algorithme

- Pour un joueur donné, on collecte la suite des instants et des positions où il touche le ballon.
- On suppose qu'il se déplace linéairement entre ces instants avec une vitesse constante.

### Résultats

- Joueurs souvent statiques.
- Répartition non réaliste des équipes.

## Un premier replayer naïf



Screenshot du replayer naïf appliqué à Marseille-Toulouse (2016).  
Bleu : Marseille, Rouge : Toulouse, Orange : ballon.

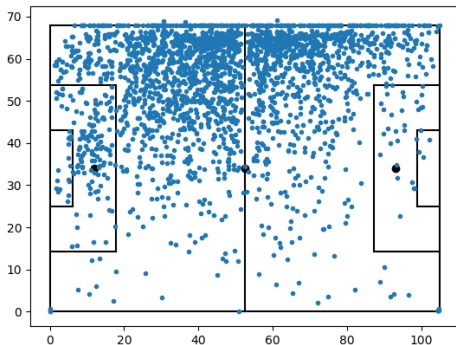


## Replayer quantiles

### Idée

- Grâce aux données Opta, on connaît la stratégie de chaque équipe (ex : 4-4-2, 4-3-3...) et le poste de chaque joueur.
- On calcule la densité de positionnement d'un joueur évoluant dans un poste dans une certaine stratégie.
- Quand un joueur donné possède la balle : On calcule son quantile selon la largeur et la longueur du terrain (déviation par rapport à la position moyenne de son poste).
- Hypothèse : Les joueurs d'une même équipe sont dans le même quantile.
- Ceci permet d'inférer la position de tous les joueurs sur le terrain.

## Replayer quantiles



Nuage de point de toutes les positions d'un arrière gauche dans une stratégie 4-4-2 en Ligue 1 2016/2017 lorsqu'il touche la balle.

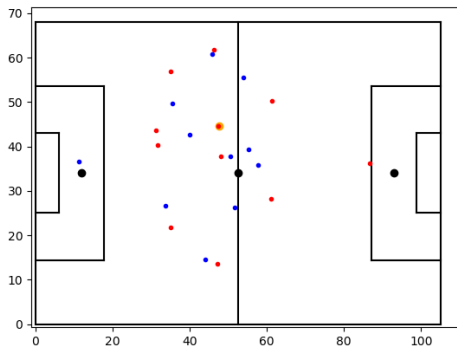


## Replayer quantiles

### Résultats

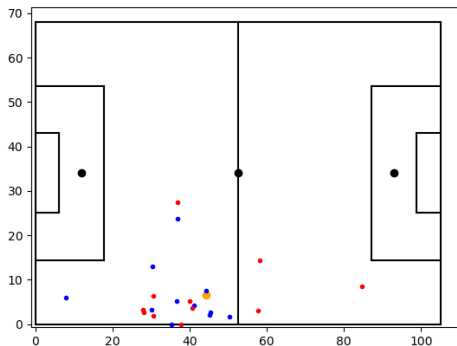
- Mouvements très dynamiques des joueurs.
- Répartition réaliste des équipes lorsque le quantile réalisé n'est pas extrême.

## Replayer quantiles



Screenshot du replayer quantiles appliqué au match  
Marseille-Toulouse (2016) avec un quantile réalisé non extrême.  
Bleu : Marseille, Rouge : Toulouse, Orange : ballon.

## Replayer quantiles



Screenshot du replayer quantiles appliqué au match  
Marseille-Toulouse (2016) avec un quantile réalisé extrême.  
Bleu : Marseille, Rouge : Toulouse, Orange : ballon.

## Replayer quantiles

### Solution

- Prendre un replayer combinaison linéaire du replayer naïf (statique) et du replayer quantiles (dynamique).
- Résultat : Replayer qui permet d'inférer la position des joueurs avec une répartition et une dynamique raisonnables.
- Mais : On ne modélise pas le déplacement des joueurs pour effectuer un pressing/proposer des solutions à un coéquipier → Méthode de Voronoï.

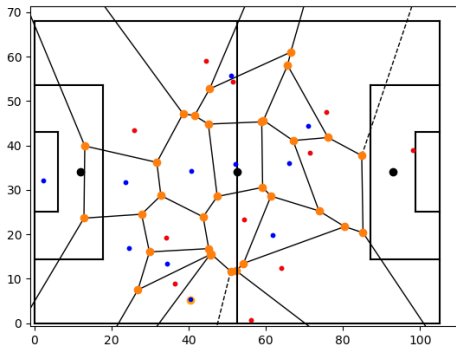
## Méthode de Voronoï

### Définition des cellules de Voronoï

- On suppose que l'équipe  $E_1$  possède la balle et joue contre  $E_2$ .
- On définit la cellule de Voronoï de chaque joueur comme étant la région d'influence du joueur sur le terrain (il y a une définition mathématique rigoureuse).
- L'équipe  $E_2$  qui défend cherche à réduire la surface d'influence de  $E_1$ .
- L'équipe  $E_1$  qui attaque cherche à élargir sa surface d'influence.
- Cette surface d'influence est la surface de l'ensemble pondéré des régions d'influence des joueurs de  $E_1$ , avec un poids important sur les régions d'influence dangereuses (proches des buts) et la région d'influence du joueur qui possède la balle.

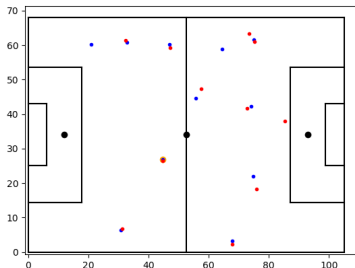
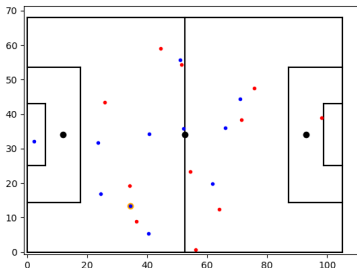


## Méthode de Voronoï



Exemple de cellules de Voronoï appliquées à un screenshot du match Marseille-Toulouse (2016). Bleu : Marseille, Rouge : Toulouse.

## Méthode de Voronoï



Evolution des joueurs après 10 secondes en utilisant la méthode de Voronoï. Bleu : Marseille, Rouge : Toulouse, Orange : ballon.

## Replayer final

### Replayer final

- Une combinaison entre le replayer naïf, le replayer quantiles et la méthode de Voronoï.
- Les positions des joueurs lorsqu'ils touchent la balle correspondent nécessairement aux positions observées en match.
- Exemple : Real Madrid - Barcelone (17<sup>ème</sup> journée de la Liga).

# Plan

- 1 On refait le match (comment se passer des données de tracking)
- 2 On anticipe le match (modélisation)
- 3 On joue le match (coaching optimal)
- 4 La coupe du monde 2018

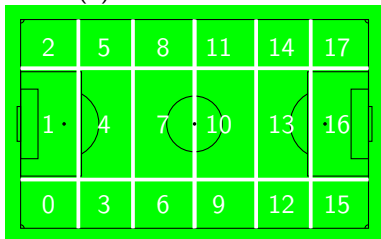
## On anticipe le match

### Objectif

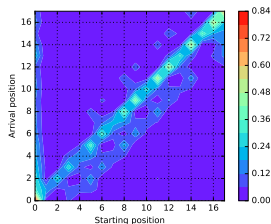
- Modéliser un match entre 2 équipes données.
- But : pouvoir faire des calculs prédictifs : score, possession, noms des buteurs...
- Ceci permettra d'optimiser les compositions de départ selon l'adversaire et la stratégie en fonction de l'avancement du match.

# Positionnement des joueurs et état du match

(a) Zones du terrain



(b) Probabilités de transition



Evt	Possession (%)	Passes réussies (%)	Tirs cadrés (%)	conservation ballon (secs)
remplaçants = 0	49.7	75.4	56.3	3.51
remplaçants = 1	53.4	66.0	35.5	4.42
remplaçants $\geq 2$	48.5	65.5	35.3	3.44
score $\leq -2$	38.5	73.3	32.5	4.3
score = -1	46.6	74.0	45.3	3.8
score = 0	50.0	74.5	48.0	3.33
score = 1	53.4	74.6	74.3	3.78
score $\geq 2$	62.3	75.6	79.0	3.74

## Description du modèle

### Le modèle

Chaque joueur  $i_1$  en zone  $j_1 \in \{0, \dots, 17\}$  peut :

- Transmettre le ballon à un coéquipier  $i_2$  en zone  $j_2$ .
- Transmettre le ballon à un joueur adverse  $i_2$  en zone  $j_2$ .
- Marquer un but.

La probabilité que l'un de ces événements se réalise dépend de :

- L'état actuel du match : positions des joueurs, score, temps...
- La capacité du joueur  $i_1$  à transmettre la balle en zone  $j_2$ .
- La probabilité de présence du joueur  $i_2$  en position  $j_2$ .
- La capacité du coéquipier (resp. joueur adverse)  $i_2$  à recevoir (resp. intercepter) le ballon en zone  $j_2$ .
- La capacité de l'équipe adverse à intercepter le ballon en  $j_2$ .

## Calibration du modèle

### Intelligence artificielle/machine learning

- Quand un joueur donné a la balle à un endroit donné, son action suivante et sa réussite ou son échec est décidée de manière aléatoire en fonction des capacités du joueur et de son environnement, selon un modèle mathématique calibré sur les données Opta des matchs passés.
- Méthode de calibration : Algorithmes d'intelligence artificielle/machine learning.
- Exemple : calibration sur les matchs allers de Ligue 1 de la saison 2016-2017 ; simulation de la 2e partie du championnat (100 simulations).



## Classement en simulation, phase retour 2016/2017

Équipe	Classement moyen	Écart-type du classement
Monaco	2.2	1.2
PSG	4.0	2.9
Nice	4.1	2.7
Bordeaux	4.6	2.2
Marseille	6.3	2.0
Angers	7.5	4.9
Lyon	8.7	3.9
Nantes	9.9	4.3
Lille	9.9	4.8
Dijon	11.0	4.5
Bastia	11.6	4.2
St-Etienne	12.6	4.7
Rennes	12.9	3.7
Montpellier	13.6	3.3
Guingamp	14.1	5.1
Lorient	14.5	3.7
Nancy	14.7	2.0
Metz	14.7	3.0
Toulouse	15.7	5.7
Caen	17.4	1.9

# Plan

- 1 On refait le match (comment se passer des données de tracking)
- 2 On anticipe le match (modélisation)
- 3 On joue le match (coaching optimal)**
- 4 La coupe du monde 2018

# Problématique

## Problème de l'entraîneur

- Pour une composition de l'équipe adverse donnée, l'entraîneur choisit sa tactique et composition de départ et peut modifier sa stratégie tout au long du match.
- Une modification de stratégie peut être :
  - Le remplacement d'un ou plusieurs joueurs.
  - La modification du poste d'un ou plusieurs joueurs sur le terrain.
- **Quelle est la suite de décisions que doit prendre l'entraîneur pour maximiser sa probabilité de gagner le match ?**

## Application : Stratégie optimale dans un match

### Exemple simple

- Match : huitième de finale retour de ligue des champions 2017/2018 : PSG-Real Madrid.
- Sortie de Benzema et rentrée de Bale à la 76<sup>ème</sup>.
- D'après notre modèle : changement pertinent, à l'instant quasi-optimal, permettant d'améliorer la probabilité de gagner d'environ 4%.

# Plan

- 1 On refait le match (comment se passer des données de tracking)
- 2 On anticipe le match (modélisation)
- 3 On joue le match (coaching optimal)
- 4 **La coupe du monde 2018**
  - Comment jouer à la coupe du monde ?
  - La liste des 23 optimale

# Comment jouer à la coupe du monde ?

## Méthodologie

- On considère 8 équipes (autres que la France) participant au mondial dont les 11 types sont constitués de joueurs de Bundesliga, Calcio, Liga , Ligue 1, Premier League : Allemagne, Angleterre, Argentine, Belgique, Brésil, Espagne, Portugal, Uruguay.
- On construit les onze types de ces équipes et estimons les paramètres du modèle pour ces équipes grâce aux données Opta (saison 2017/2018 jusqu'à fin mars).
- Pour près de 200 formations possibles pour l'équipe de France, on calcule les probabilités de victoire contre chacune de ces 8 équipes.
- On retient à chaque fois l'équipe pour laquelle la probabilité de victoire est la plus grande.

# Comment jouer à la coupe du monde ?

## Méthodologie

- Les compositions de l'équipe de France sont construites à partir des 17 joueurs suivants (joueurs quasi-certains de participer à la coupe du monde) :
  - Lloris, Mandanda.
  - Kimpembé, Koscielny, Mendy, Sidibé, Umtiti, Varane.
  - Kanté, Lemar, Matuidi, Pogba, Tolisso.
  - Dembélé, Giroud, Griezmann, Mbappé.

## Comment jouer à la coupe du monde ?

### Face au Brésil (Équipe type pour gagner le mondial)

- Lloris/Sidibé-**Varane-Koscielny**-Mendy/Tolisso-Kanté-Matuidi/Mbappé-Griezmann-Lemar.

### Face à l'Allemagne ou l'Espagne

- Lloris/Sidibé-**Koscielny-Umtiti**-Mendy/Tolisso-Kanté-Matuidi/Mbappé-Griezmann-Lemar.

### Face au Portugal

- Lloris/Sidibé-**Varane-Koscielny**-Mendy/Mbappé-Kanté-Matuidi-Lemar/Griezmann-**Giroud**.

### Face à l'Uruguay

- Lloris/Sidibé-**Varane-Umtiti**-Mendy/**Pogba**-Kanté-Matuidi/**Dembélé-Mbappé-Lemar**.



# Quelle chance de l'emporter pour la France ?

## Méthodologie

- Difficile d'utiliser les valeurs brutes des probabilités obtenues car plusieurs facteurs de biais : seulement 8 équipes, optimisation de l'équipe de France et pas des autres équipes...
- Après débiaisage, ordre de grandeur de **14%**.
- À prendre avec des pincettes...

## La liste des 23

### Méthodologie

- On se pose la question de la présence des joueurs suivants dans la liste :
  - Areola, Costil, Maignan, Ruffier.
  - Debuchy, Pavard.
  - Amavi, Digne, Hernandez, Kurzawa.
  - Aouar, Bakayoko, Doucouré, Kondogbia, Nzonzi, Rabiot, Sissoko.
  - Ben Yedder, Benzema, Lacazette.
  - Coman, Fekir, Martial, Payet, Ribery, Thauvin.
- On remplace (poste par poste) les joueurs de l'équipe de France type par ceux-ci et on conserve ceux dégradant le moins la probabilité de victoire finale.

## La liste des 23

### Liste des 23 optimale

- Lloris, **Maignan**, Mandanda.
- **Amavi**, **Debuchy**, Kimpembé, Koscielny, Mendy, Sidibé, Umtiti, Varane.
- **Aouar ou Doucouré ou Kondogbia**, Kanté, Lemar, Matuidi, Pogba, Tolisso.
- Dembélé, **Fekir**, Giroud, Griezmann, **Lacazette**, Mbappé.