



HAL
open science

Le sport face aux jeux (de données)

Christophe Biernacki

► **To cite this version:**

Christophe Biernacki. Le sport face aux jeux (de données). Session Olympique Universitaire (SOU), Sep 2021, Lille, France. hal-03505657

HAL Id: hal-03505657

<https://hal.science/hal-03505657v1>

Submitted on 31 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le sport face aux jeux (de données)

Christophe Biernacki

Inria



 Université
de Lille

Événement organisé par



En association avec



Dans le cadre de



Take-home message

Les données sont désormais omniprésentes dans tous les domaines d'activités humaines car là où elles sont analysées, la performance est décuplée.

Le sport est une activité humaine, donc les données s'invitent.

Événement organisé par



En association avec



Dans le cadre de



Collecte des données en sport

Événement organisé par



En association avec



Dans le cadre de



Collecte de données : un phénomène ancien

- **Fin du 19ème siècle** : Henry Chadwik, un journaliste sportif et statisticien passionné de baseball, invente la « moyenne au bâton » (un indicateurs sur les frappeurs)

$$\text{Moyenne au bâton} = \frac{\text{Coups sûrs}}{\text{Présences au bâton}}$$

- **Années 60** : la Sabermétrie (toujours en baseball) multiplie les indicateurs statistiques, du recrutement des joueurs au jeu lui-même

La collecte s'emballe...

- **Formule 1** : 300 capteurs/voiture, 500 Go/course
- **Football** : terrains criblés de capteurs et de caméras en tout genre, joueurs portant des boîtiers fixés dans leur dos à l'aide de brassières, ballons équipés de capteurs
- **Tennis** : caméras autour des terrains, capteurs (raquette, bracelet...)

Cette emballement est d'autant plus rapide qu'on se rapproche de « l'industrie du sport » (le football représente 400 milliards €)

Événement organisé par



En association avec



Dans le cadre de



Objectifs de cette collecte

1. **Analyse/amélioration de performance** : identifier les faiblesses d'un athlète pour les corriger, entraînement personnalisé...
2. **Prévention de la blessure** : mettre en lumière les mauvaises habitudes d'un athlète, prévenir l'apparition de pathologies...
3. **Elaboration d'une stratégie** : optimiser la composition d'une équipe, influencer la manière dont un sport se pratique (panier à trois point en basket)...
4. **Repérage de talents** : savoir à l'avance quel jeune est susceptible d'exploser...
5. **Inclusion des spectateurs** : se sentir plus proche des sportifs...

Les questions qui se posent

- Acquisition des données (choix des mesures, technologie, coût)
- Analyse des données (comprendre, prédire)
- Interaction avec le coach/sportif/spectateur
- Confidentialité, vie privée du sportif

Des chercheurs de nombreuses disciplines sont impliqués

Événement organisé par



En association avec



Dans le cadre de



Groupe Statistique et Sport de la SFdS

Objectifs

Le groupe Statistique et Sport, créé lors des 50èmes Journées de Statistique, joue le rôle d'une entité fédératrice rassemblant des professionnels du sport, des statisticiens, des chercheurs et des enseignants. L'objectif de ce groupe est de disposer d'un langage commun afin de construire un dialogue entre statisticiens et praticiens avec un choix de thèmes communs.

Les trois voies potentielles sont l'enseignement, le service et la recherche :

- développer des cours pour augmenter les compétences statistiques dans le sport ;
- fournir des outils pour la prise de décisions dans les clubs, les fédérations, le marketing sportif et événementiel ;
- s'inspirer du sport pour le développement de méthodes.

Bureau

- Christian Derquenne, EDF R&D (Président)
- Arthur Leroy, Paris 5 (Vice-président)
- Marta Avalos-Fernandez, Université de Bordeaux (Secrétaire générale)
- Geoffroy Berthelot, INSEP (Webmestre)
- Brigitte Gelein, ENSAI (Trésorière)
- Guillaume Saulière, INSEP (Secrétaire-adjoint)

Événement organisé par



En association avec



Dans le cadre de



Inria : le numérique en vue des JO 2024

Revea

Soutenu par les fédérations d'athlétisme, boxe et gymnastique, ce projet met en avant l'utilisation de la **réalité virtuelle pour améliorer les performances motrices** des athlètes tout en limitant les risques de blessure.

Partenaires : Inria, Laboratoire M2S - Université Rennes 2, PSMS et CReSTIC - Université Reims Champagne Ardennes, Institut Sciences du Mouvement - Aix-Marseille Université, CIAMS - Université Paris-Saclay, INSEP, Fédération Française de Boxe, Fédération Française d'Athlétisme, Fédération Française de Gymnastique.

PerfAnalytics

Comment **analyser automatiquement** les mouvements d'un sportif à partir de l'outil vidéo pour objectiver ses performances ? C'est l'enjeu de ce projet, soutenu par les fédérations de boxe, cyclisme, gymnastique, lutte et escalade qui se mène au plus près des sportifs pour capter leurs mouvements et cibler leurs entraînements.

Partenaires : Inria, Université Grenoble Alpes, Université de Poitiers / CNRS DR08, Université Gustave Eiffel, Aix-Marseille Université, Fédération Française Montagne et Escalade, Fédération Française de Cyclisme, Fédération Française de Gymnastique, Fédération Française de Boxe, Fédération Française de Lutte, CREPS de Poitiers, TSF Voiron.

Best-Tennis

Son objectif est d'**optimiser la performance du service et du retour de service** des joueuses et joueurs de tennis. L'approche scientifique utilise des expertises en biomécanique pour réduire le risque de blessure, en neurosciences sur la prise d'informations et l'intelligence artificielle à des fins stratégiques.

Partenaires : Inria, Laboratoire M2S - Université Rennes 2, CNRS, CEA, Institut Supérieur de l'Aéronautique et de l'Espace, Institution Nationale des Invalides, Université Sorbonne Paris Nord, Ecole Nationale des Sports de Montagne, Tennis Stat, Trinoma, Volodalen, Fédération Française de Tennis, CREPS de Toulouse

Événement organisé par



En association avec



Dans le cadre de



Focalisation de cet exposé

Analyse des données...

Événement organisé par



En association avec



Dans le cadre de



Analyse de données complexes

Événement organisé par



En association avec



Dans le cadre de



Objectif 1 : prédire (apprentissage supervisé)

Exemple : prédire une performance future

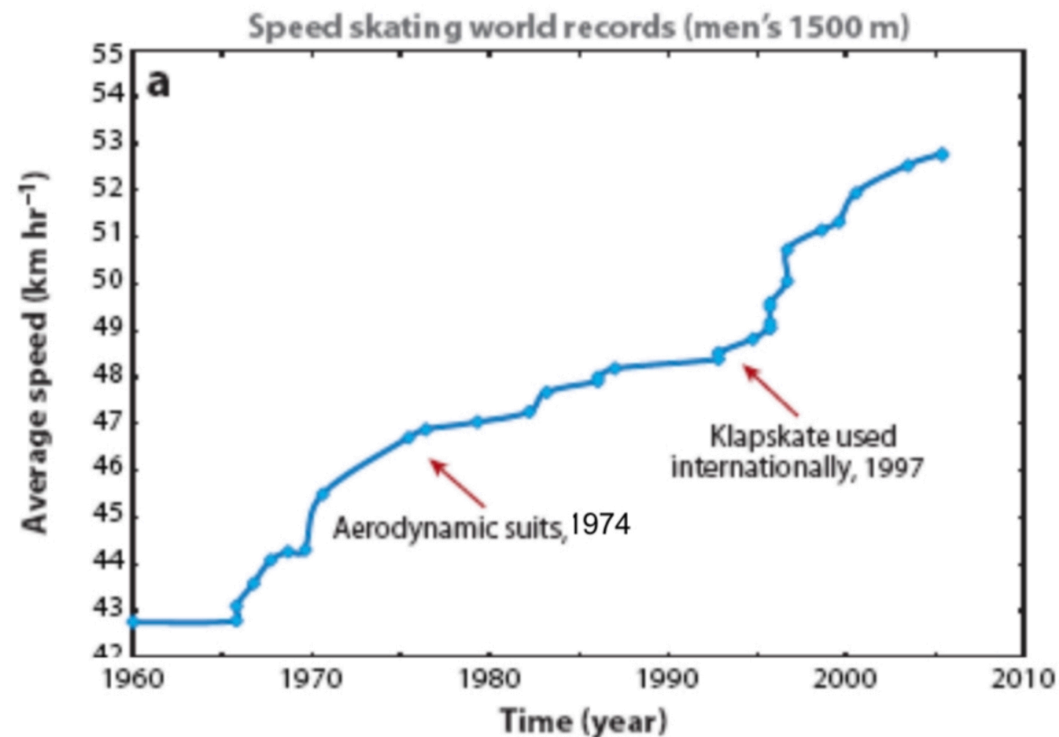


Figure H. Influence des avancées technologiques et des équipements sur les records du monde de vitesse en patinage depuis 1960 dans l'épreuve du 1500m hommes [20].

Événement organisé par



En association avec

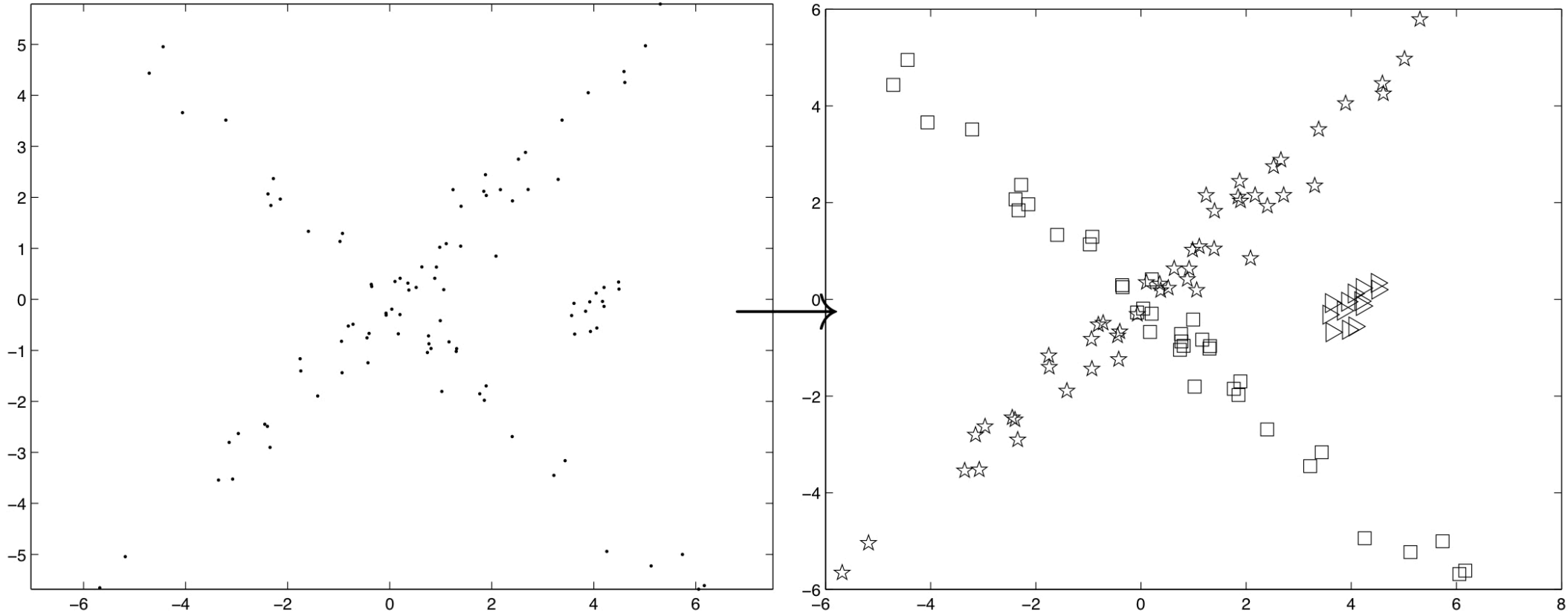


Dans le cadre de



Objectif 2 : comprendre (app. non supervisé)

Exemple : détecter des groupes de sportifs



Événement organisé par



En association avec

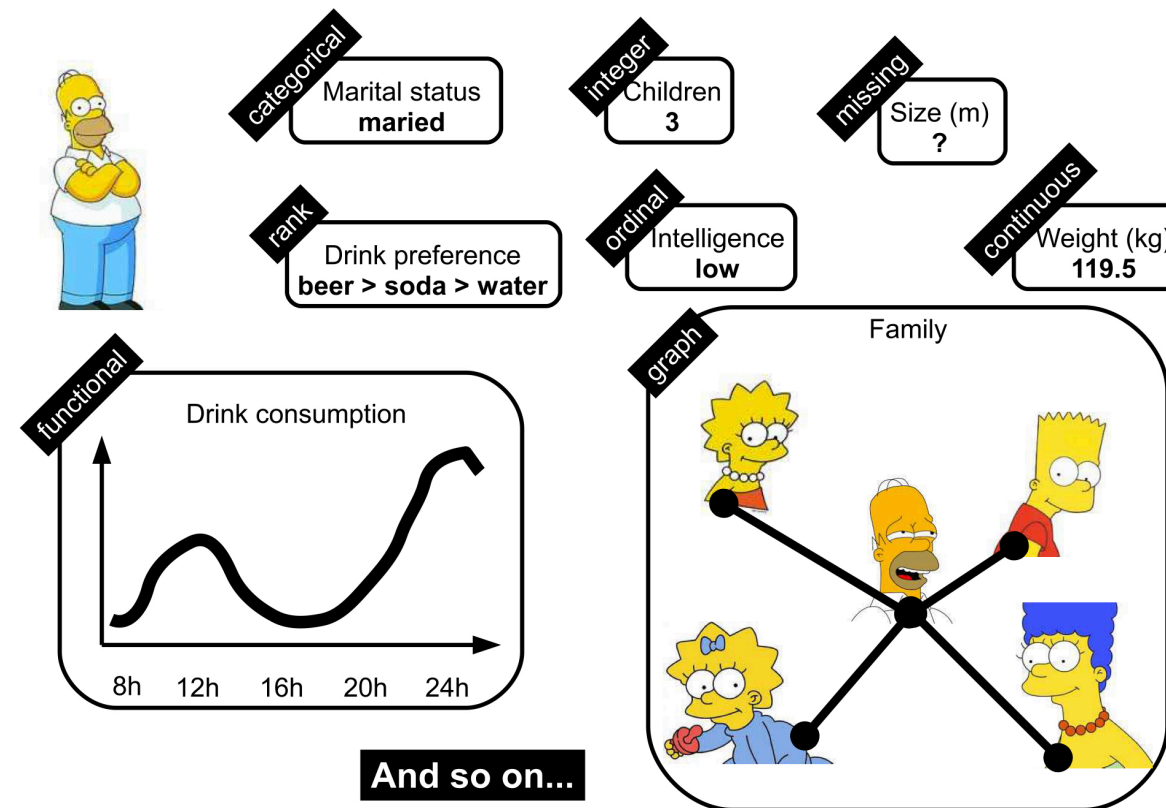


Dans le cadre de



Difficulté : complexité des données collectées

- **Valeur** : taille, poids
- **Catégorie** : gaucher/droitier
- **Ordinal** : ressenti
- **Angle(s)** : position du bras
- **Courbe** : rythme cardiaque
- **Graphe** : passes de ballon
- **Image, vidéo** : caméra de terrain
- **Manquant** : capteur en panne
- **Incertitude** : capteur imprécis
- **Haute dimension** : nbx capteurs



Traiter ces données est un **verrou recherche**

Illustrons sur plusieurs exemples et sur une application en sport...

Événement organisé par



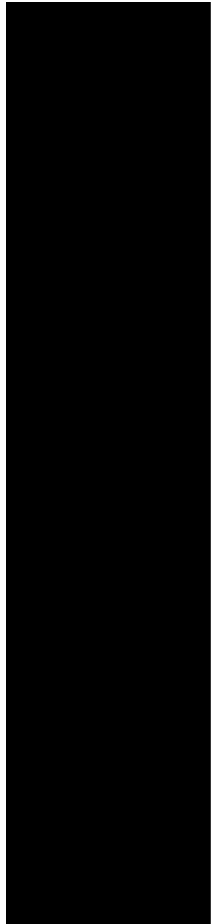
En association avec



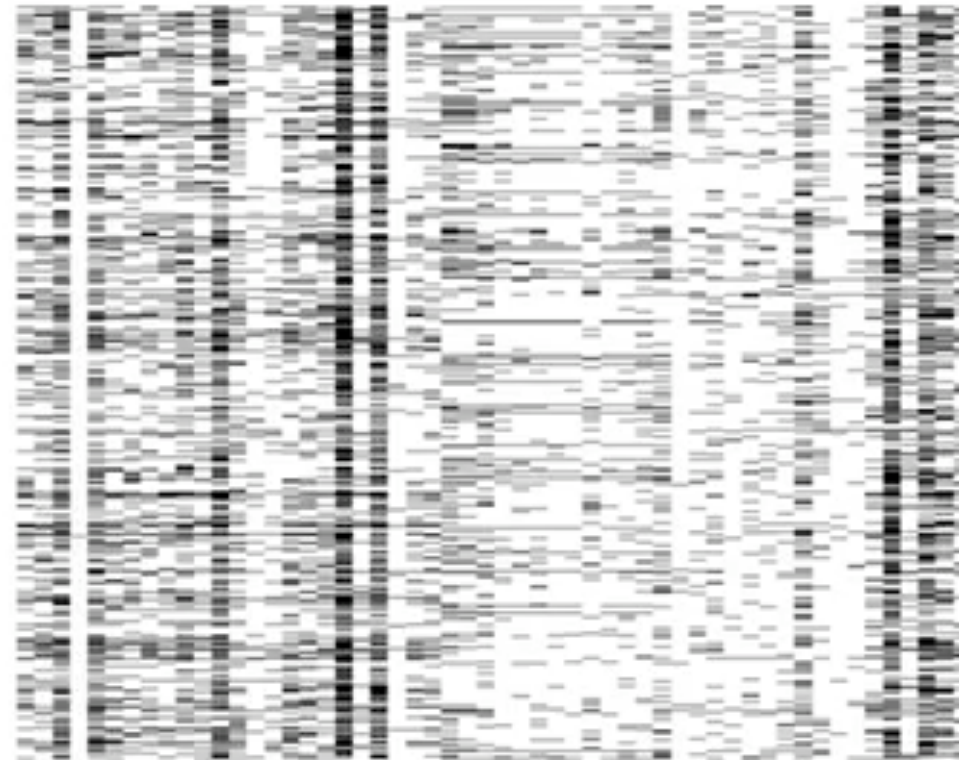
Dans le cadre de



Exemple pour des données binaires (emails)



Données initiales



Évènement organisé par



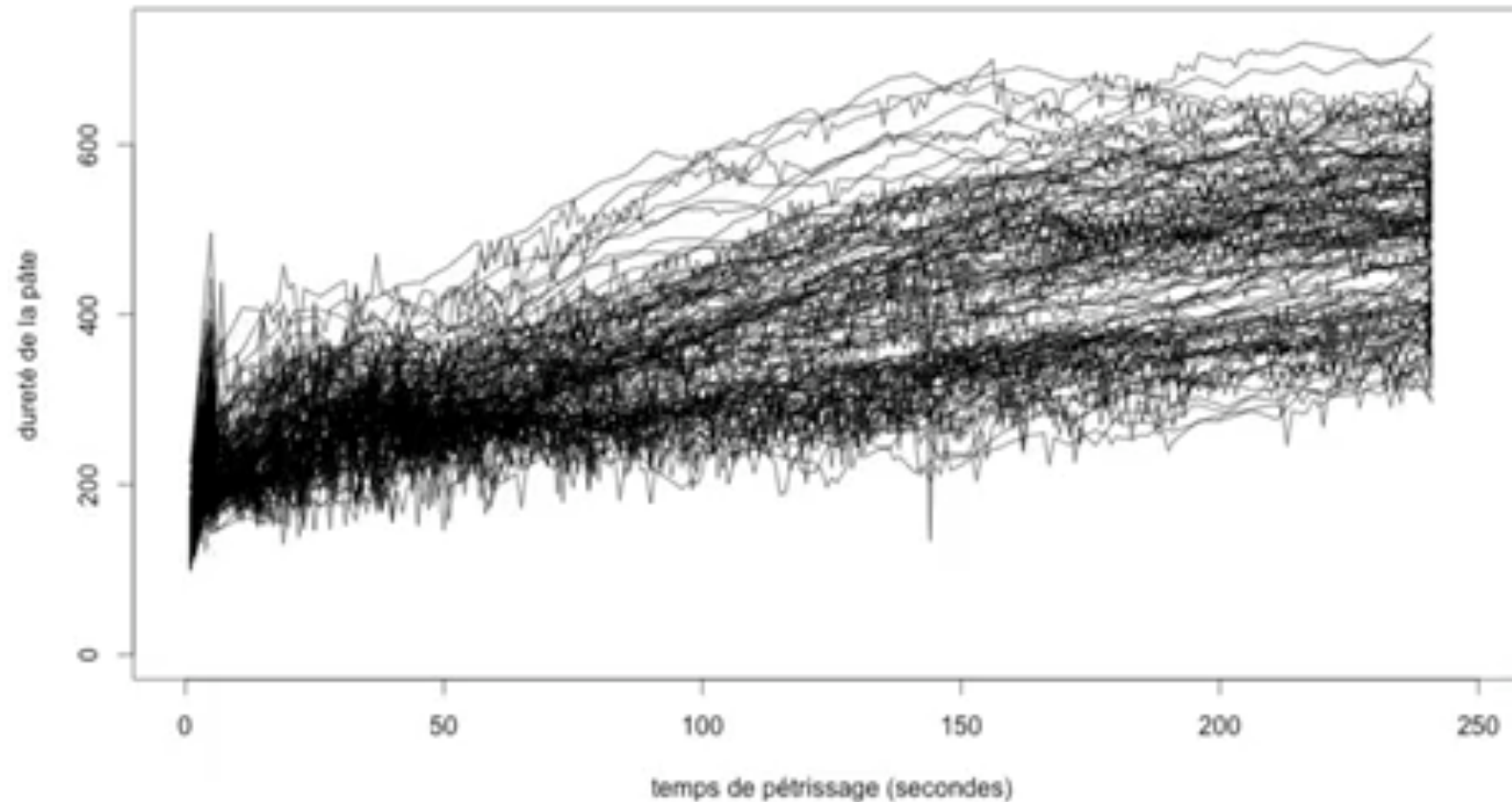
En association avec



Dans le cadre de



Exemple pour des courbes (cookies)



Événement organisé par



En association avec



Dans le cadre de



Détection précoce de jeunes nageurs prometteurs

Événement organisé par



En association avec



Dans le cadre de



Travail issu d'une thèse en mathématiques

THÈSE DE DOCTORAT

Spécialité : Mathématiques appliquées

par

ARTHUR LEROY

Apprentissage de données fonctionnelles par
modèles multi-tâches : application à la prédiction de
performances sportives

Événement organisé par



En association avec



Dans le cadre de



Problématique

- **Motivation** : des études récentes sur les carrières de nombreux sportifs suggèrent une faible relation entre le niveau de performance dans les jeunes années et celui à l'âge adulte
- Deux questions :
 - **Comprendre** : existe-t-il des profils typiques de progression pour les sportifs ?
 - **Prédire** : est-il possible d'utiliser ces similarités entre individus pour améliorer des prédictions de la performance future ?
- La réponse à ces deux questions permettra de **repérer précocement** les nageurs à fort potentiel
- Nécessité de mesurer les **risques de décision** (quantification probabiliste)

Complexité des données

- **Objet d'étude** : courbe de performance composée de résultats en compétition
- **Difficulté recherche** : ces séries temporelles sont observées irrégulièrement d'un individu à l'autre (parcimonie et irrégularité)

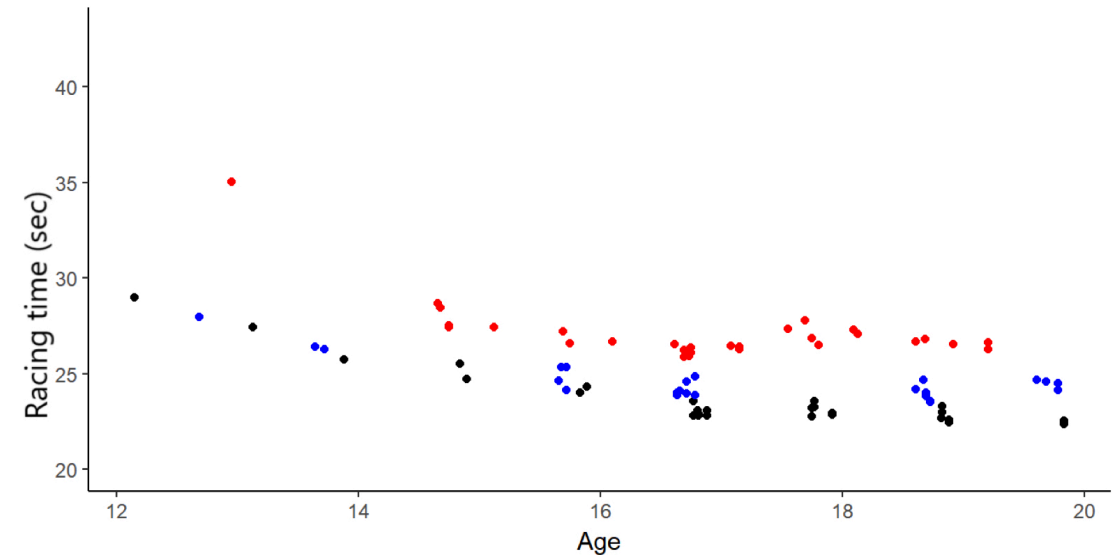
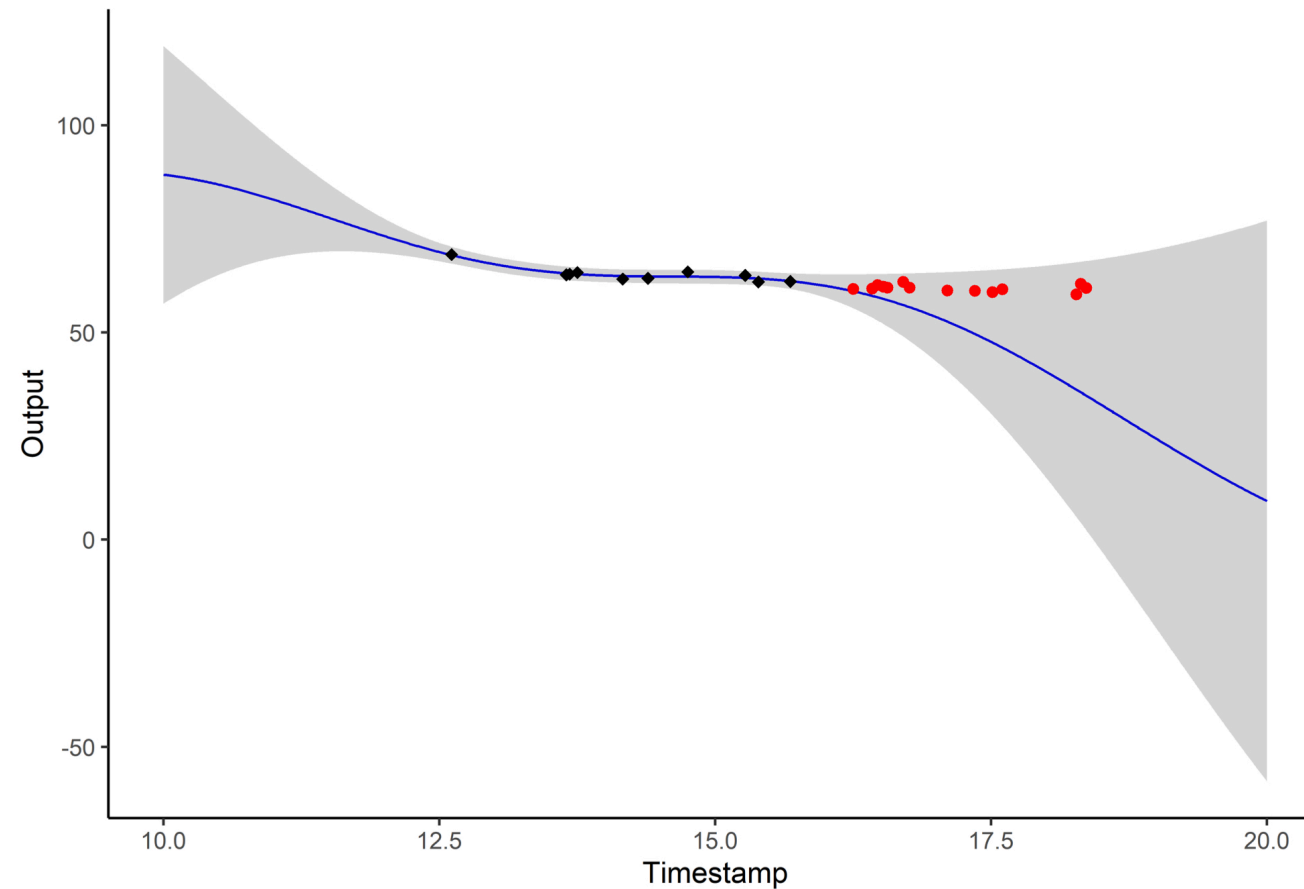
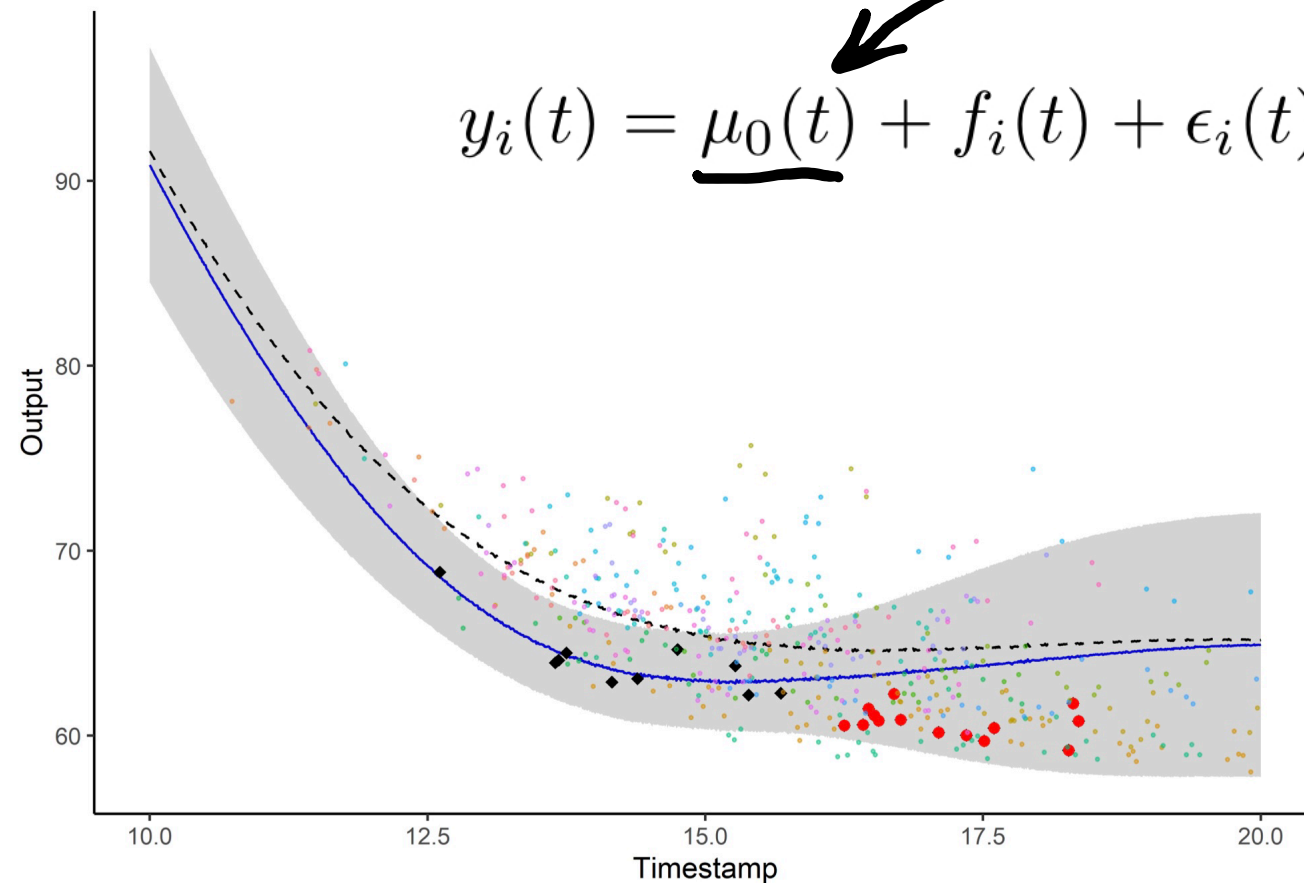


Figure 1 – Exemple de données observées pour 3 nageurs différents (respectivement en bleu, rouge et noir). L'âge des abscisses indique l'âge du nageur. L'axe des ordonnées indique la performance en course en secondes, ici pour des compétitions de 50m nage libre homme.

Méthode 1 : processus gaussiens individuels

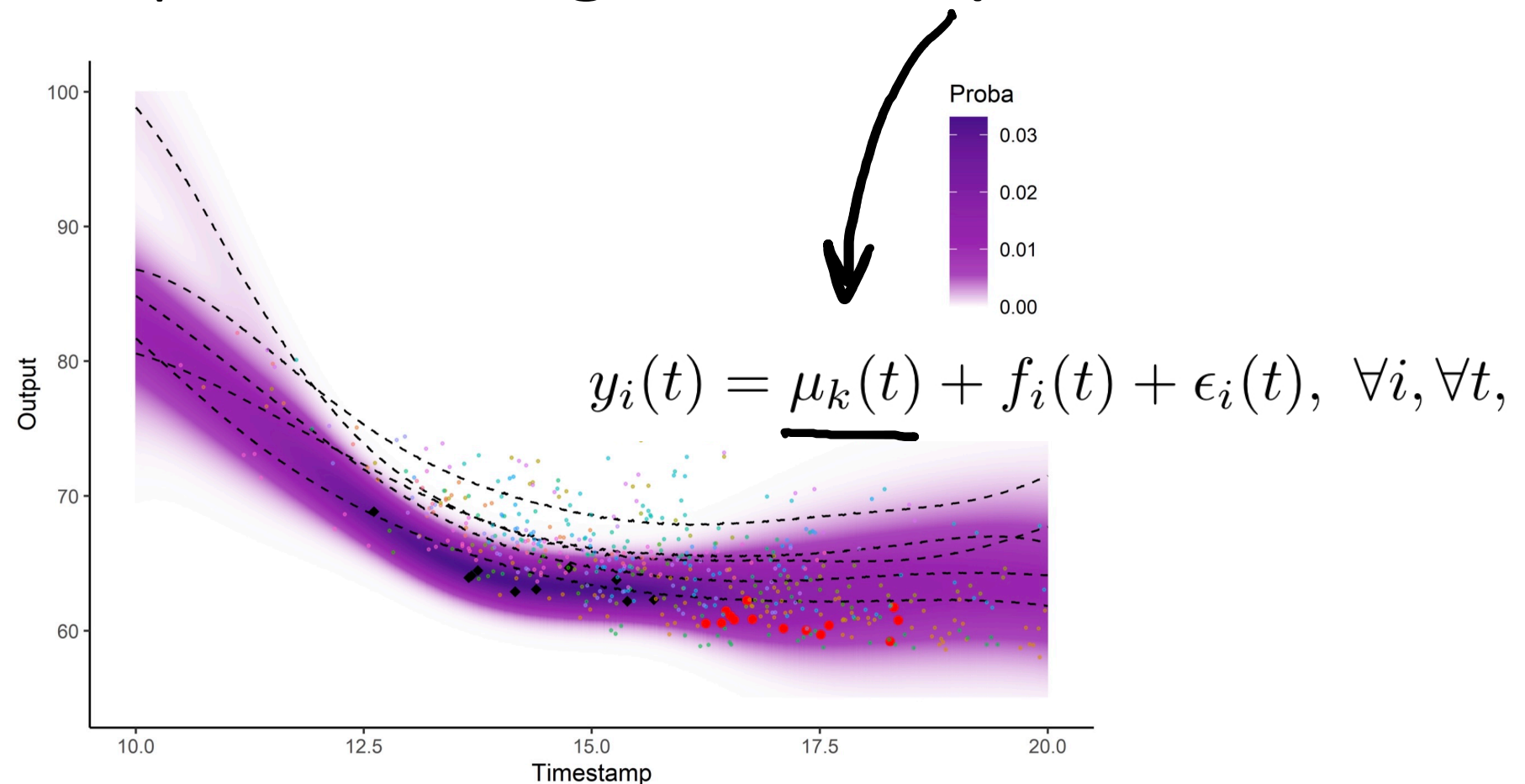


Méthode 2 : processus gaussiens collectifs



$$y_i(t) = \mu_0(t) + f_i(t) + \epsilon_i(t), \forall i, \forall t,$$

Méthode 3 : processus gaussiens par classes



Utilisation pratique

- Mise à disposition d'un **package informatique** utilisable par le coach
- **Interprétation** des classes de nageurs
- **Prévision de la classe** actuelle/future d'un nageur
- **Prévision de la performance** future (en s) d'un nageur

On a finalement associé prévision et compréhension !

Événement organisé par



En association avec



Dans le cadre de



Conclusion

Événement organisé par



En association avec



Dans le cadre de



- Le sport de haute performance s'appuie sur de nombreuses données
- Des problématiques de recherche en statistique en émergent
- Ces partenariats interdisciplinaires seront **mutuellement fructueux**

Événement organisé par



En association avec



Dans le cadre de



Bibliographie

- <https://www.usine-digitale.fr/article/comment-les-clubs-sportifs-capitalisent-sur-les-donnees-des-athletes-pour-augmenter-leurs-performances.N865315>
- <https://www.lesechos.fr/idees-debats/cercle/la-donnee-au-coeur-de-la-performance-dans-le-sport-134755>
- <https://www.powerslide.io/blog/data-et-sport>
- [Multi-task learning models for functional data and application to sports performances prediction A. Leroy - PhD Thesis - 2020](#)
- <https://www.inria.fr/fr/inria-prepare-ses-jeux-olympiques>
- https://www.sfds.asso.fr/fr/statistique_et_sport/618-statistique_et_sport/

Événement organisé par



En association avec



Dans le cadre de



Merci !

Événement organisé par



En association avec



Dans le cadre de

