



9èmes Rencontres de Statistique

Science des Données et Sport



26 et 27 novembre 2024

<https://www.lebesgue.fr/fr/rencontresstat24>

Université Bretagne Sud

Amphithéâtre du bâtiment Yves Coppens

Campus de Tohannic - 56000 Vannes

Comités d'organisation et scientifique

Présidents : Gilles Durrieu et François Septier

Baptiste Alglave, Chloé Friguet, Pierre Gloaguen, Evans Gouno, Ion Grama, Salim Lardjane, Calire Launay, Audrey Poterie

Conférenciers et intervenants invités

Simon Boisbluche, Préparateur physique du Rugby Club de Vannes

Antoine Bouvet, ENS-CREST, Rennes

Christian Derquenne, groupe Statistique et Sport, SFDS, Paris

Laetitia Fradet, UBS, Lab-STICC, Lorient

Emmanuel Frénod, UBS, LMBA, Vannes

Brigitte Gelein, ENSAI-IRMAR, Rennes

Léo Gerville-Réache, Université de Bordeaux, Bordeaux

Arthur Leroy, INRAE, Paris

Nolwenn Pinczon du Sel, Fédération Française de Rugby-IRMAR, Paris

Mardi 26 novembre 2024

— **9h15-09h30 : Accueil des participants et café** dans l'amphithéâtre, Bâtiment Yves Coppens, Université Bretagne Sud, Rue André Lwoff, 56000 Vannes.

— **9h45-10h : Introduction** par Gilles Durrieu, Gabel Hardy et François Septier.

— **10h00-11h00 : Christian Derquenne, SFDS, Paris**

“La Science Statistique au service des Jeux Paralympiques 2024.”

Cet exposé a pour objectif de montrer en quoi la Science Statistique peut répondre aux enjeux issus du Sport.

Nous illustrons cette dimension dans le domaine parasportif et plus particulièrement nous l'axons vers les Jeux Paralympiques de 2024 à Paris qui se sont déroulés dernièrement. Les travaux présentés entrent dans le cadre du projet Paraperf de l'INSEP et du Mécénat de compétences EDF appuyé par EDF R&D. Nous présenterons quatre disciplines : le tir sportif, le para-cyclisme, l'haltérophilie et le tir à l'arc. Pour chacune d'elles, nous expliquerons les enjeux et problématiques fixés par les fédérations sportives, les méthodes mises en œuvre et leurs applications à destination des coaches. Enfin, nous tenterons de faire le lien avec les résultats obtenus de nos para-athlètes Françaises et Français aux JP 2024.

— **11h00-12h00 : Nolwenn Pinczon du Sel, Fédération Française de Rugby-IRMAR, Paris**

Modélisation de couloirs de niveau non-linéaires : Application à l'effet de l'âge relatif dans l'évaluation de performance sportive

Les athlètes nés dans le premier quartile de l'année de sélection sont surreprésentés dans les compétitions d'élite chez les jeunes, un phénomène largement reconnu comme l'Effet de l'Âge Relatif (RAE). Le RAE survient lorsque les athlètes relativement plus âgés bénéficient d'avantages qui favorisent une performance accrue et une visibilité renforcée dans des contextes compétitifs, comparativement à leurs pairs plus jeunes. Nous proposons une méthode qui 1) prend en compte la progression non linéaire en fonction de l'âge et du niveau de performance, 2) fournit un nouvel indicateur de performance ajusté pour le RAE, et 3) permet le suivi de la progression dans un cadre de référence individuel.

— **12h00-14h00 : Pause Déjeuner** au Tableau (<https://www.au-tableau.com/>) pour les orateurs

et les organisateurs.

— **14h00-15h00 : Arthur Leroy, INRAE, Paris**

“Do sports problems require tailored methods or direct applications? On a swimming-oriented journey with a Bayesian roadmap”

When looking at a new problem for the first time, the initial question that comes to mind can generally be summarised as : has this problem, or a closely-related one, already been tackled before? Although being a rather recent scientific field of study, sports sciences often face the same pattern, despite the number of recognised ‘resolved problems’ remaining relatively low compared to more established disciplines. In particular, when it comes to statistical tools, there exists in the literature an abundance of methods to handle a variety of problems, such as image classification, missing data reconstruction, time series forecasting, dimensionality reduction, and data visualisation, among others. In this presentation, I will try to illustrate how, in the past 7 years, determining whether some sport-related problem required developing a novel tailored statistical tool or simply applying an established method on this particular dataset, was generally the most important step of the project to result in an adequate answer. From several articles using direct applications of Bayesian mixed models on morphological swimming datasets, to the 3-year long development of a novel machine learning framework dedicated to forecasting irregular time series of swimmers’ performances, let’s explore the variety of challenges coming from sports science applications, while accounting for their intrinsic complexity. In conclusion, I will try to emphasise how those new statistical models, originally tailored to handle sports-related problems, can often find natural applications in other fields like medicine, biology, or robotics (among others), thus contributing to the overall methodological toolbox shared by many applicative sciences.

— **15h00-16h00 : Laetitia Fradet, UBS-Lab-STICC, Lorient**

La Science Statistique au service de l’analyse du mouvement : du développement de la méthodologie à la compréhension des troubles moteurs.

L’analyse du mouvement permet de caractériser la performance motrice d’une population (sportive ou pathologique) : au quotidien pour permettre de suivre l’évolution de la performance sportive ; plus ponctuellement,

en laboratoire, pour optimiser la performance sportive ou mettre en évidence des troubles moteurs dans les populations pathologiques (personnes âgées chuteuses, enfants autistes etc.).

Le suivi quotidien de la performance motrice nécessite le recours à des méthodologies, dont le développement repose largement sur les sciences statistiques. De la même manière, l'analyse biomécanique cherchant à identifier les caractéristiques motrices spécifiques à une population ne peut se passer d'analyses statistiques avancées ce afin d'extraire les variables les plus pertinentes et révélatrices des schémas de motricité spécifique.

Cette présentation mettra en lumière ces deux dimensions. Dans une première partie, nous explorerons donc l'utilisation des statistiques dans l'élaboration des méthodologies de suivi de la performance. Ensuite, nous aborderons des études spécifiques visant à mieux comprendre les particularités motrices des enfants autistes, en montrant comment l'analyse statistique contribue à cette compréhension..

— **16h00-16h15 : Pause café**

— **16h15-17h15 : Simon Boisbluche, Rugby Club de Vannes**

L'interprétation des datas dans la prévention des blessures et la performance collective en rugby

Nous verrons à travers quelques exemples de terrain comment les datas peuvent nous aider à améliorer la performance en match à travers la diminution du nombre de blessure et l'augmentation des performances athlétiques de l'équipe.

— **17h15-18h15 : Témoignages d'anciens étudiants**

— **Aurélien Gaillard**, Data Analyst au Stade Rennais (diplômé en 2019)

— **Valentin Langrognat-Mulier**, Data Scientist chez STEF (diplômé en 2021)

— **Erwann Le Chenadec**, Data Scientist/Développeur Logiciel CGI (diplômé en 2020)

— **19h30 : Diner pour les orateurs, les organisateurs et les étudiants de Master 2ème année** au Piano barge (<http://www.pianobarge.com/Page/Accueil>).

Mercredi 27 novembre 2024

- **9h45-10h00 : Accueil des participants et café** dans l'amphithéâtre, Bâtiment Yves Coppens, Université Bretagne Sud, Rue André Lwoff, 56000 Vannes.

- **10h00-11h00 : Emmanuel Frénod UBS-LMBA, Vannes**

Entité connectée éloignée de son centre de contrôle : formalisation et modélisation de sa double boucle de pilotage.

Comment piloter de manière robuste une entité connectée (i.e. munie de capteurs et de leviers) dans un contexte où cette entité est très éloignée de son centre de contrôle (par exemple dans l'espace ou dans un lieu avec une connexion très dégradée) ? La réponse à cette question est : on utilise une double boucle de pilotage (i.e. assimilation de données - guidage). La première, située à proximité de l'entité, réalise un pilotage « au mieux ». La seconde apporte - avec retard - des instructions de guidage plus pertinentes et des évolutions à apporter à la première boucle pour la rendre plus adaptée. Cet exposé expliquera une formalisation de ce principe et une théorisation mathématique. Des premiers exemples d'applications seront montrés.

- **11h00-12h00 : Léo Gerville-Réache, Université de Bordeaux, Bordeaux**

Momentum en sport : approche statistique et application au tennis

Dans nombres de sports comme le tennis ou le volley par exemple, on observe au cours d'un match une série relativement longue de points gagnés ou perdus. Cette série est mathématiquement une simple suite (ou séquence) binaire de 1 et de 0 où le chiffre 1 traduit la victoire et le 0 la perte du point. Voici la série de victoire et de perte de points de Rafael Nadal lors de sa finale à Roland Garros en 2020 contre Novak Djokovic. Cette finale est remportée par Rafael Nadal sur le score de 6/0 – 6/2 – 7/5 et 183 points ont été joués :

0100111101110011011101101001010111000111101011101110111001001001001111100101011

01110011011010111111100001110011000001110110101000011011111110010100010001

011010101011000011111001111111...

Comment détecter sur cette séquence un ou plusieurs momentums ? Cette question nécessite, dans un premier temps, de construire un modèle mathématique de la probabilité de gain et de perte de chaque point de la

rencontre. Ensuite, un momentum pourrait être vu comme une série « atypique » de points au sein de la longue série constituant le match. Plus précisément, une série « anormalement » dense en 1 pourrait être vu comme un momentum en faveur de Rafael Nadal alors qu'une série « anormalement » dense en 0 serait un momentum en faveur de Novak Djokovic. Détecter un momentum consisterait donc à comparer la série binaire de la rencontre avec des probabilités « théoriques » associées à chaque point...

— **12h00-14h00 : Pause Déjeuner** au Tableau (<https://www.au-tableau.com/>) pour les orateurs et les organisateurs.

— **14h00-15h00 : Brigitte Gelein, ENSAI-IRMAR, Rennes**

“Modélisation de l'évolution des qualités athlétiques en fonction de l'âge et des blessures antérieures chez le jeune footballeur de niveau élite.”

Cette étude vise à modéliser l'évolution de certains paramètres athlétiques chez les jeunes footballeurs de niveau élite en utilisant des modèles mixtes (linéaires et non-linéaires). L'ensemble de l'académie d'un club de football professionnel a été suivi sur une période de quatre saisons consécutives en enregistrant certaines de leurs performances athlétiques telles que la vitesse, la force, la puissance ou encore l'endurance. Les modèles mixtes ont été utilisés afin de tenir compte du caractère longitudinal des données. Un apport important de ce travail consiste en la prise en compte non seulement de l'âge mais aussi des blessures antérieures. Plusieurs indicateurs du fardeau des blessures ont été construits et comparés pour obtenir les meilleurs modèles en prédiction. Les résultats indiquent l'obtention de modèles fiables et précis, avec de faibles marges d'erreurs pour la plupart des paramètres athlétiques. Ils mettent en lumière des trajectoires distinctes d'évolution des paramètres athlétiques en fonction des différents facteurs, offrant ainsi des informations précieuses pour la gestion de la performance et le développement des jeunes athlètes de football.

Auteurs : Arthur Guillotel (M2S, SRFC), Brigitte Gelein(IRMAR, Ensai), Anthony Sorel (M2S, Université Rennes 2)

— **15h00-16h00 : Antoine Bouvet ENS-CREST, Rennes**

Monitoring & Modélisation de la Performance en natation (SWIMO²P) : comment la science des données peut-elle catalyser l'accompagnement scientifique à la performance ?

A l'heure de la révolution digitale, la natation est une discipline encore sous-technologisée. Or, la densifi-

cation actuelle du haut-niveau contraint à toujours plus d'optimisation et d'individualisation du processus d'entraînement, ainsi que d'analyse fine des compétitions. L'apport d'approches data-driven, à travers l'utilisation de centrale inertielle en routine, fournit une perspective de suivi embarqué intéressante. Néanmoins, l'exploitation de ces données pour l'accompagnement scientifique à la performance pose de nombreux obstacles, notamment méthodologiques, afin d'extraire – automatiquement et de manière généralisable – des métriques pertinentes d'une série temporelle multivariée relative aux accélérations et vitesses angulaires 3D du nageur. Cette présentation détaillera un aperçu des travaux de thèse “SWIMO²P” comprenant, d'une part, des contributions méthodologiques en science des données basées apprentissage profond pour la reconnaissance d'activité humaine au service du monitoring biomécanique du nageur, et clustering fonctionnel pour le profilage in-situ des qualités techniques. D'autre part, des applications pratiques en sciences du sport, avec la Fédération Française de Natation et visant l'accompagnement scientifique à la performance, seront brièvement présentées. Ces outils et méthodes, issus du projet NePTUNE Sciences 2024, sont actuellement déployés dans le cadre de la préparation et du développement des équipes de France, aussi bien à l'entraînement quotidien qu'en compétition internationale bassin et eau libre. A la fois pour analyser et modéliser la performance, mais surtout les prérequis techniques sous-jacents, et ainsi individualiser les stratégies de gestion de course à haut-niveau et les axes d'entraînements associés.

— **16h00-17h00 : Guillaume Franchi, UBS-LMBA, Vannes**

Modélisation dynamique des données d'abondance en écologie

La modélisation de l'abondance jointe des espèces dans un écosystème est exhaustivement étudiée dans la littérature. Néanmoins, la dimension temporelle d'une telle abondance n'est que rarement prise en compte. Les modèles proposés ne permettent alors pas d'identifier clairement les impacts que les espèces peuvent avoir les unes sur les autres. Nous nous intéresserons dans cet exposé à la modélisation dynamique de deux principaux types d'abondance : l'abondance relative et l'absence-présence. La première, à valeurs dans le simplexe, consiste en un vecteur composé des proportions de chaque espèce dans l'écosystème tout entier. La seconde consiste en un vecteur binaire multivarié, où la i -ème composante est égale à 1 si l'espèce i est présente dans l'écosystème, 0 sinon. Nous présenterons des méthodes générales de construction de séries temporelles à

valeurs dans les espaces compacts considérés, dont les paramètres sont facilement interprétables d'un point de vue écologique. Nous nous pencherons tout particulièrement sur les garanties d'existence et d'unicité d'un processus satisfaisant nos modèles avec de bonnes propriétés statistiques, comme la stationnarité ou l'ergodicité. Nous discuterons également des stratégies d'estimation des paramètres des modèles présentés. Nous présenterons enfin, à titre illustratif, des applications de nos modèles à des données issues d'études écologiques.

Travaux de thèse réalisés sous la direction de Lionel Truquet (ENSAI), avec la collaboration de Marie-Pierre Etienne

— **Clôture des 9èmes Rencontres.**