



8èmes Rencontres de Statistique

Science des Données

Mer & Littoraux



30 novembre et 1er décembre 2023

<https://www.lebesgue.fr/fr/rencontresstat23>

Université Bretagne Sud

Amphithéâtre du bâtiment Coppens

Campus de Tohannic - F-56000 Vannes

Comités d'organisation et scientifique

Présidents : Gilles Durrieu et François Septier

Anne Cuzol, Chloé Friguët, Pierre Gloaguen, Evans Gouno, Ion Grama, Salim Lardjane, Audrey Poterie

Conférenciers et intervenants invités

Angelo Alcaraz, UBS, LMBA, Vannes

Baptiste Alglave, UBS, Lab-STICC, Vannes

Benjamin Dufée, UBS, LMBA, Vannes

Marie Pierre Etienne, Institut Agro, IRMAR, Rennes

Pierre Gloaguen, UBS, LMBA, Vannes

Juliette Legrand, UBO, LMBA, Brest

Mathilde Letard, Univ. Rennes, Géosciences Rennes, Rennes

Nicolas Raillard, IFREMER, Brest

Etienne Rivot, Institut Agro, DECOD, Rennes

Orlane Rossini, Univ Montpellier, IMAG, Montpellier

Benjamin Sautter, UBS, GeoOcean, Vannes

Jeudi 30 novembre 2023

— **10h30-11h00 : Accueil des participants et café** dans l’amphithéâtre, Bâtiment Yves Coppens, Université Bretagne Sud, Rue André Lwoff, 56000 Vannes.

— **11h00-12h00 : Etienne Rivot, Institut Agro, DECOD, Rennes**

“Counting fish is like counting trees, except they are invisible and they move”. Les modèles statistiques à variables latentes pour l’écologie halieutique.

“Counting fish is like counting trees, except they are invisible and they move”. Cette citation célèbre tirée d’une leçon de John Shepherd (Prof mérite, Univ Southampton, UK) nous raconte toute l’histoire d’amour entre l’écologie halieutique et la modélisation statistique. L’écologie halieutique vise à comprendre les mécanismes qui sous-tendent le fonctionnement des écosystèmes aquatiques, en particulier la dynamique des populations de poissons en réponses aux multiples facteurs de forçage et de pressions anthropiques. La dynamique de ces systèmes est caractérisée par une grande variabilité naturelle et par des difficultés d’observation. En bref, on ne peut pas vider la mer pour observer les poissons! On doit donc faire appel à des sources de données variées qui nous renseignent indirectement sur l’état du système. L’écologie halieutique fait donc largement appel aux modèles statistiques à variables latentes, qui permettent de faire de l’inférence sur des processus cachés non directement accessibles à l’observation, afin de mieux comprendre la dynamique, d’intégrer et de propager les multiples sources d’incertitudes dans les inférences et la prédiction, et de proposer des outils d’aide à la décision pour la gestion. Dans cet exposé, nous illustrerons ces questions au travers d’exemples de recherche qui portent sur la modélisation de la dynamique de population d’une espèce patrimoniale, le saumon atlantique. Les recherches visent à mieux comprendre la dynamique en lien avec les effets du changement climatique, et s’insèrent dans le cadre d’une recherche-action pour nourrir des outils d’expertise et d’aide à la décision pour la gestion de ces populations dans un contexte international.

— **12h00-14h00 : Déjeuner**

— **14h00-15h00 : Benjamin Sautter, UBS, GeoOcean, Vannes**

Influence of Geological Inheritance on Water Circulation and Storage in the Land-Sea Continuum

To understand the environmental implications of human activities along the land-sea continuum, it is critical to characterize the dynamics of water circulation in the heterogeneous geological subsurface. This research focuses on the often overlooked yet crucial influence of geological inheritance in shaping the hydrological characteristics of this continuum. Geological inheritance encompasses the structures that shaped the Earth's surface and subsurface such as lithological contacts, fault and fold systems, geomorphic and buried structures, and is investigated here as the primary driver influencing the distribution, flow paths, and residence times of water.

A key focus lies in elucidating how a geological discontinuity impacts ground-surface water interactions. Does it act as a driver, a barrier or has no impact? We aim to employ a comprehensive approach, integrating geological, geophysical and geochemical datasets in a GIS environment, to delineate the pathways of water movement, identify potential groundwater discharge zones, and assess the influence of geological inheritance on nutrient transport in coastal areas.

This research project emphasizes the importance of recognizing geological controls in developing sustainable water resource management strategies. Insights gained not only contribute to the fundamental understanding of the Land-Sea Continuum but also provide critical information for mitigating the impacts of anthropogenic activities and adapting to changing environmental conditions.

— **15h00-16h00 : Marie Pierre Etienne, Institut Agro, IRMAR, Rennes** *Comment l'étude du déplacement des individus nous informe sur l'écologie des espèces.*

Le déplacement d'un individu peut être envisagé comme la réponse à ses besoins internes, à son environnement, et potentiellement à ses relations aux autres. Le développement de système de suivi GPS léger et très autonome permet de suivre les déplacements des individus de manière précise et peu invasive. Ces données de trajectoires doivent être affectés par le changement dans les besoins internes et par l'environnement. Dans cet exposé, je présenterai plusieurs développements statistiques récents, reposant sur des approches à variables latentes, qui visent à extraire ce type d'information des données déplacement. Ces méthodes posent des questions intéressantes en terme de modélisation, d'inférence et de choix de modèles qui seront détaillées dans l'exposé.

— **16h00-16h15 : Pause café**

— **16h15-16h45 : Orlane Rossini, Univ Montpellier, IMAG, Montpellier** *Apprentissage bayésien par renforcement pour un processus semi-markovien déterministe par morceaux contrôlé : application à l'optimisation d'un traitement médical*

Le suivi et le traitement du cancer peuvent être considérés comme des problèmes de prise de décision séquentielle sous incertitude. Pour ce type d'application, l'objectif est de maximiser l'espérance de vie du.de la patient.e tout en minimisant les coûts de traitement et les effets secondaires. Nous faisons l'hypothèse d'une évolution déterministe de la maladie, entrecoupée de sauts stochastiques. Des résultats d'analyse sanguine permettent d'estimer l'état de santé global du.de la patient.e Pour modéliser l'évolution de la maladie, nous proposons un processus semi-Markov déterministe par morceaux partiellement observé (PDsMP). De plus, le processus évolue en temps et en espace continus et certains de ses paramètres sont inconnus. Par conséquent, aucune solution analytique ou numérique, ne peut être obtenue. On exprime le problème de contrôle comme un processus de décision markovien partiellement observé (POMDP) à temps discret. Pour apprendre les paramètres inconnus tout en contrôlant le.la patient.e, nous utilisons une approche basée sur l'apprentissage bayésien adaptatif.

— **16h45-17h45 : Témoignages d'anciens étudiants**

— **Luc Alapini**, Chief Technology Officer chez Bivariant (diplômé en 2019)

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

— **Orlane Rossini**, Doctorante à l'Univ Montpellier (diplômée en 2022)

— **19h30 : Dîner pour les orateurs, les organisateurs et les étudiants de Master 2ème année**

Vendredi 1er décembre 2023

- **8h50-9h00 : Accueil des participants et café** dans l'amphithéâtre, Bâtiment Yves Coppens, Université Bretagne Sud, Rue André Lwoff, 56000 Vannes.

- **9h00-10h00 : Mathilde Letard, Univ. Rennes, Géosciences Rennes, Rennes**

Etude de zones côtières avec lidar bathymétrique à retour d'ondes complètes

TBA

- **10h00-10h30 : Angelo Alcaraz, UBS, LMBA, Vannes**

Régression quantile avec contrôle de la variance : applications en écologie

Dans la communauté écologique, il est admis qu'une plus grande diversité alimentaire, appelée aussi diversité trophique, permet de mieux réagir à un changement d'environnement. Depuis une quarantaine d'années, une mesure permet d'avoir accès à des informations sur le régime alimentaire des animaux, et donc sur leur diversité trophique, à travers les signatures isotopiques en carbone et en azote. Afin d'établir un proxy de la diversité trophique, on définit la niche isotopique, qui correspond à la surface occupée par une espèce (ou une échelle taxonomique supérieure) dans le plan composé des signatures isotopiques du carbone et de l'azote. L'objectif est ici de réussir à développer un modèle permettant de prédire la taille de la niche isotopique en fonction de variables explicatives environnementales et biotopiques (dépendantes de l'espèce). La méthode proposée se situe à l'intersection de plusieurs formalisations de modélisation : données longitudinales, régression quantile, effets aléatoires, pénalisation ridge, etc.

- **10h30-10h45 : Pause café**

- **10h45-11h45 : Nicolas Raillard, IFREMER, Brest**

Méthodes statistiques pour l'océano-météo

Ces dernières années, les zones côtières ont fait l'objet d'une attention accrue, avec le développement accéléré des activités humaines d'une part, et le réchauffement climatique d'autre part, qui expose les populations vivant dans les zones côtières à des risques côtiers plus fréquents et plus intenses. Ces deux problématiques ont en commun la nécessité de bien connaître les conditions de mer à proximité des côtes, afin de qualifier

précisément les risques auxquels les personnes et les structures sont exposées, un objectif pour lequel la modélisation statistique est particulièrement bien adaptée. Dans cette présentation, nous nous concentrerons sur les structures marines, qui sont soumises aux effets combinés du vent, des vagues et des courants. Dans un premier temps, nous présenterons les données disponibles et les travaux menés à l'IFREMER pour acquérir de nouvelles données et connaissances. Dans un deuxième temps, nous présenterons les apports de la modélisation statistique pour préciser les conditions opérationnelles de ces structures, ainsi que les conditions extrêmes impactant leur conception.

— **11h45-12h15 : Baptiste Alglave, UBS, Lab-STICC, Vannes**

Statistiques, gestion des pêches et modèles spatialisés.

La gestion des populations halieutiques repose en grande partie sur les méthodes d'évaluation de stock qui permettent de quantifier le niveau d'exploitation des populations marines exploitées (ou stocks).

Cette présentation partira d'un modèle d'évaluation de stock spatio-temporel représentant l'une des approches les plus poussées dans le domaine de l'halieutique. En décrivant pas à pas le modèle et la méthode d'inférence des paramètres, j'illustrerai les progrès méthodologiques successifs qui ont permis le développement de ce type de modèle.

La base du modèle est un modèle structuré en taille. À l'origine, l'ajustement de ce type de modèle se faisait en réduisant les écarts entre les observations de captures et les captures prédites par le modèle. Le développement des modèles hiérarchiques (en particulier, des modèles à espace d'état) a permis (1) de donner un cadre formel à ce type d'approche et (2) d'intégrer des sources de données variées (*e.g.* données de capture, de classe de taille, de marquage-recapture) pour modéliser de façon conjointe différents types de processus (*e.g.* dynamique de population, mouvement des individus). Ces progrès ont notamment été rendus possibles par le développement de méthodes d'inférence efficaces pour ajuster ces modèles (*e.g.* les méthodes MCMC dans le cadre bayésien).

Plus récemment, la disponibilité de plus en plus de données spatialisées poussent les halieutes à développer des approches spatiales pour l'évaluation de stock. Les méthodes d'inférence classiques ne suffisent plus pour estimer les paramètres des modèles en un temps raisonnable. La dernière partie de l'exposé portera sur les avancées récentes des méthodes d'inférence qui permettent d'ajuster ces modèles de façon efficace. Elles reposent à la fois sur des approches d'approximation de la vraisemblance (*i.e.* approximation de Laplace), des

méthodes géostatistiques permettant de réduire la dimensionnalité des problèmes statistiques (*i.e.* approche SPDE) et des méthodes de dérivation efficaces (*i.e.* différentiation automatique).

Ces méthodes mises ensemble permettent d'estimer la dynamique de population dans l'espace et dans le temps en un temps compatible avec les besoins opérationnels (*i.e.* quelques heures); elles ouvrent ainsi la porte à l'implémentation opérationnelle de ces approches pour la gestion des ressources marines.

— **12h15-14h : Déjeuner**

— **14h00-15h : Juliette Legrand, UBO, LMBA, Brest**

Simulation jointe d'extrêmes de hauteurs significatives de vagues à la côte et au large.

La caractérisation des événements présents et futurs de vagues extrêmes est cruciale en raison de leurs multiples impacts (e.g. submersion, érosion, fiabilité et dimensionnement des structures). Dans ce travail, nous proposons et étudions un simulateur stochastique qui génère conjointement, en fonction de certaines conditions d'état de mer au large, des extrêmes de hauteur significative de vagues (H_s) au large et à la côte, quantité liée à l'énergie des vagues et donc à leur sévérité. Pour cela, nous nous appuyons sur l'approche par dépassements de seuils bivariés et nous développons un algorithme de simulation non-paramétrique de lois de Pareto généralisées bivariées. À partir de ce simulateur d'événements cooccurrents, nous dérivons un modèle de simulation conditionnel d'extrêmes. Les deux algorithmes de simulation (cooccurrent et conditionnel) sont appliqués aux H_s près des côtes bretonnes françaises. Un dernier développement est traité quant à la modélisation des lois marginales des H_s : afin de prendre en compte leur non-stationnarité, nous adaptons une extension de la loi de Pareto généralisée, en considérant l'effet de la période et de la direction pic sur ses paramètres.

— **15h00-15h30 : Benjamin Dufee, UBS, LMBA, Vannes**

Assimilation de données pour des modèles océaniques stochastiques.

L'assimilation de données consiste en l'incorporation de données d'observations partielles d'un système dans un modèle d'évolution numérique de ce système afin de corriger et d'enrichir sa dynamique. Cet exposé présentera une grande famille de techniques d'assimilation de données, appelés filtres de Kalman d'ensemble,

et ses applications à un modèle océanique stochastique. Ce dernier reposera sur une décomposition en semi-martingale du déplacement lagrangien. La nature stochastique du modèle sera également utilisée comme degré de liberté supplémentaire afin d'introduire dans le modèle un terme de guidage visant à mener les réalisations du modèle vers une région d'intérêt, proche des observations.

— **15h30-16h00 : Pierre Gloaguen, UBS, LMBA, Vannes**

Le modèle Poisson log-normal, un modèle statistique pour la description de communauté d'espèces

En écologie, l'analyse de distribution de la biodiversité passe souvent par le recensement des individus dans leur milieu naturel. Ceci aboutit à des données d'abondance (souvent des comptages) de plusieurs espèces sur plusieurs sites. Je parlerai du modèle Poisson log-normal qui permet d'expliquer la distribution jointe des espèces par des variables environnementales, mais aussi d'essayer d'inférer les interactions résiduelles entre espèces (cooccurrences ou évitements non expliqués par l'environnement). Je parlerai du problème d'inférence dans ce cas en montrant comment le cadre bayésien peut être utile pour l'estimation.

— **Clôture des 8èmes Rencontres.**