



10èmes Rencontres de Statistique Histoire & Territoires



27 et 28 novembre 2025

<https://www.lebesgue.fr/fr/rencontresstat25>

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, Lise Bellanger-Husi, Chloé Friguet, Pierre Gloaguen, Evans Gouno, Ion

Grama, Salim Lardjane, Claire Launay, Matthieu Marbac, Audrey Poterie

Jeudi 27 novembre 2025

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

— **9h45-10h : Introduction**

— **10h00-11h00 : Davide Adamo, Université Côte d'Azur**

“Archaeozoology & machine learning : a promising matching”

Understanding the relationship between past human societies and their environment relies on the ability to identify and compare faunal remains across space and time. Yet, in archaeozoology, morphological similarities between species and the subtle variability induced by domestication often challenge traditional visual and metric approaches. Recent advances in 3D scanning technologies now offer new quantitative tools to address these challenges.

In this talk, I will present two case studies illustrating how machine learning (ML) methods can support archaeological interpretation. We first focus on the automatic identification of wild and domestic ungulates using Topological Data Analysis and Multiple Kernel Learning, combining shape descriptors with probabilistic kernel selection via variational inference. We will then explore the use of Optimal Transport, introducing geometric-invariant Wasserstein barycenters, to track the morphological evolution of domestic animal bones over time. Together, these approaches demonstrate how statistical and ML models can contribute to reconstruct the evolutionary and cultural dynamics linked to animal domestication.

— **11h00-12h00 : Edith Darin, Leverhulme Centre for Demographic Science Oxford**

Estimation des populations infranationales à l'ère des données numériques : intégration de sources multiples, modélisation bayésienne et enjeux éthiques en contexte de crise

Dans de nombreux contextes à faibles ressources ou touchés par des crises, les données démographiques de base font souvent défaut, en raison de recensements obsolètes, incomplets, voire inexistant. Cette présentation examine des approches méthodologiques innovantes pour estimer les populations infranationales en l'absence de données traditionnelles fiables. Nous nous concentrons sur l'intégration de sources de données hétérogènes — registres administratifs partiels (état civil, santé, éducation), images satellites, et covariables géospatiales — au sein de cadres de modélisation bayésienne permettant une quantification rigoureuse de

l'incertitude. Nous abordons ensuite les défis liés au nowcasting des populations dans des contextes de crise (conflits, déplacements forcés, catastrophes naturelles), où des flux de données en temps quasi réel (données téléphoniques, imagerie satellitaire, réseaux sociaux) permettent de suivre les dynamiques démographiques avec une résolution spatio-temporelle inédite.

Enfin, nous soulignons les implications éthiques de ces approches, notamment en ce qui concerne la protection de la vie privée, les biais algorithmiques et la représentation équitable des populations vulnérables. Cette contribution met en lumière une nouvelle génération d'outils et de pratiques pour produire des estimations démographiques plus rapides, plus précises et plus responsables.

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

— **14h00-15h00 : Arthur Coulon, Université de Tours**

“Méthodes d’apprentissage statistique pour la périodisation et la datation en archéologie”

L’établissement de la chronologie, et donc de la périodisation des sites archéologiques en fonction des grandes transformations qui les affectent, constitue un préalable essentiel à tout discours historique. Les principales sources mobilisées pour déterminer cette périodisation — autrement dit, la temporalité des sites archéologiques - sont :

— la stratigraphie, qui correspond à la succession des niveaux anthropiques et fonde l’élaboration d’une chronologie relative,

— le mobilier archéologique, en particulier la céramique, omniprésente dans les fouilles et dont la typologie évolue rapidement au cours du temps. Ces deux éléments confèrent à la céramique une valeur chronologique particulièrement élevée.

Dans cette présentation, nous présenterons une approche interdisciplinaire originale associant archéologie et statistique, visant à produire une périodisation fondée conjointement sur la céramique et la stratigraphie.

La première étape consiste à définir des faciès céramiques stables et interprétables à l'aide d'une méthode de classification non supervisée reposant sur un compromis. Dans une seconde étape, les prévisions issues d'un modèle de classification supervisée permettent d'intégrer des assemblages céramiques moins fiables, mais néanmoins indispensables à la construction d'un modèle chronologique global. Cette approche a été

mise en œuvre sur des données issues de la cité mythique d'Angkor Thom (Cambodge), capitale de l'Empire khmer entre le IXe et le XVe siècle. Si le temps le permet, nous aborderons également, la mise en œuvre de modèles de régression pour répondre à la problématique de l'estimation de la datation de contextes archéologiques à l'aide de la céramique. Le projet Open CNRS ARCTOOLS (2025-2027) contribue à financer le développement du package R SPARTAAS (Statistical PArametric ReconsTruction for Archaeological Analysis and Stratigraphy), qui offre un cadre reproductible et automatisé pour le traitement de grands volumes de données archéologiques.

— **15h00-16h00 : Charlotte Brabant, LETG Rennes**

“Analyse statistique et spatiale de la réponse phénologique de la végétation urbaine à l’îlot de chaleur : le cas du bassin de Rennes”

L'étude des interactions entre climat urbain et végétation demande aujourd'hui des approches combinant statistiques inférentielles, analyses spatiales à haute résolution. Dans ce cadre, nous proposons une démarche statistique visant à quantifier la relation entre l'intensité de l'îlot de chaleur urbain (ICU) et la phénologie printanière d'espèces arborées dans le bassin rennais.

Notre approche s'appuie sur un réseau de mesures intra-urbaines de température (Rennes Urban Network – RUN) dans des villes de tailles différentes et sur des observations phénologiques réalisées sur sept espèces d'arbres au cours des printemps 2021 et 2022. Nous avons mobilisé des méthodes inférentielles pour tester l'effet des variables microclimatiques et du contexte urbain sur le départ végétatif. Enfin, une modélisation spatiale a permis d'explorer la variabilité intra-urbaine des réponses d'une espèce : le chêne pédonculé et de montrer les gradients phénologiques en lien avec l'intensité de l'ICU.

Ces résultats illustrent la complémentarité entre outils statistiques et modélisation spatiale pour comprendre les dynamiques écologiques en milieu urbain, et soulignent l'intérêt des réseaux de suivi microclimatique dans les villes moyennes et petites.

— **16h00-16h15 : Pause café**

— **16h15-17h15 : Nicolas Frerebeau, Université de Bordeaux Montaigne**

“Le projet tesselle : outils et méthodes pour une science des données archéologiques”

L'utilisation des langages de programmation en archéologique a connu une augmentation notable au cours

de la dernière décennie, notamment avec R, un langage polyvalent permettant de développer des modules spécialisés. Cette contribution sera l'occasion de présenter le projet tesselle (<https://www.tesselle.org/>), une collection de packages R destinés à la recherche en archéologie, développée et maintenue depuis 2018. Ces packages constituent une boîte à outils polyvalente facilitant l'exploration et l'analyse des types de données courants en archéologie, tels que les données de comptage, les données de composition ou les données chronologiques. Cette contribution propose ainsi un retour d'expérience sur le développement d'outils en science des données archéologiques. Il s'agira de discuter les usages, les facteurs limitants et d'explorer les évolutions possibles.

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

Vendredi 28 novembre 2025

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

— **10h00-11h00 : Ketsia Guichard, Université de Rennes**

“Des réseaux de mobilités comme un graphe attribué : partitionnement par des méthodes basées sur la distance”

Les données de mobilité peuvent être modélisées comme un graphe, qu'il s'agisse d'un réseau physique comme le réseau routier, ou d'un réseau défini par les connexions entre les points d'entrée et de sortie d'un système de transport. Le clustering de ces graphes permet d'identifier des communautés ou de segmenter des réseaux complexes. Après un aperçu de ces données de mobilité, cet exposé s'interessera aux graphes attribués, où chaque arête ou nœud est enrichi d'informations (courbes de flux, histogrammes de vitesse, etc.).

Pour tenir compte à la fois de la structure topologique du graphe et des attributs, nous nous concentrerons ici sur des méthodes de partitionnement basées sur les distances, en particulier les k-Fréchet means et des méthodes issues du transport optimal, utilisant la distance de Gromov–Wasserstein. Leurs performances seront évaluées à l'aide d'un cadre de simulation, analysant la sensibilité au bruit et l'influence du choix de la structure cible, puis illustrées sur des données réelles de trafic urbain et de vélo en libre-service.

— **11h00-12h00 : Mathieu Aubry, Ecole des Ponts ParisTech**

“Building Computer Vision tools for and with Historians”

Through a set of examples, this presentation will interrogate how historians and computer vision scientists can collaborate effectively.

The first part of the presentation will present our recent paper on "Segmenting France Across Four Centuries" (M. Lopez-Rauhut, H. Zhou, M. Aubry, L. Landrieu, ICDAR 2025). In this paper, we introduce a "computer-vision ready" benchmark for segmenting Cassini and Etat major maps across all of France, as well as several baselines. It illustrates indirect collaboration through public data sharing and annotations, and outlines the requirements of computer vision scientists to tackle a project.

The second part of the presentation, will outline two smaller scale and more direct collaborations, on

paleographic analysis and watermark recognition.

Finally, the end of the presentation will showcase our aikon-demo and aikon-platform web-based tools (<https://aikon-platform.github.io/>), designed respectively to enable Historians to test standard computer vision algorithms, and to enable teams of historians to collaborate together on large corpora, leverage computer vision tools, and interact with computer vision scientists.

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

— **14h00-14h30 : Baptiste Alglave, Université Bretagne Sud**

“Identifying urban profiles with regard to overheating factors. Application to French municipalities”

Urban areas are increasingly exposed to climate change, with rising urban temperature outpacing global rise. Buildings and land sealing exacerbate urban overheating through urban heat islands and prolonged thermal stress which significantly impacts populations health. This study aims to classify French municipalities based on their sensitivity to overheating by leveraging multivariate analysis methods, here the factorial analysis of mixed data. Indicators are derived from diverse data sources characterizing urban morphology, land use, climate, and topography. Five municipality profiles emerged, distinguished by factors such as building density, vegetation cover, and proximity to the coast. The classification was validated against urban heat islands data simulated using urban climate models. This typology provides a first tool to categorize urban territories with regards to their sensitivity to urban overheating. Such classification could be used to quantify population exposure and identify at-risk populations based on the type of municipality they live in.

— **14h30-15h30 : Florent Laroche, Ecole Centrale de Nantes**

“Patrimoines et Humanités numériques d’aujourd’hui et de demain”

Depuis deux décennies, des chercheurs en sciences du numérique (IRCCyN puis LS2N) et en sciences humaines (Centre François Viète) croisent leurs expertises autour du patrimoine et de la modélisation numérique. De la rétroconception d’objets techniques à la restitution monumentale par IA générative, ces travaux conjuguent rigueur scientifique et innovation culturelle. À travers plusieurs projets, thèses et dispo-

sitifs muséographiques réalisés, cette collaboration a façonné une véritable culture commune du numérique appliquée au patrimoine. Elle interroge aujourd’hui autant les méthodes que les finalités de la recherche interdisciplinaire, de la donnée à l’interprétation, de l’objet au territoire.

— **15h30-15h45 : Pause café**

— **15h45-16h15 : Matthieu Gousseff, Lab-STICC**

“Comparer des variables qualitatives spatialisées en l’absence de référence, application à la comparaison de cartes de zones climatiques locales”

La surchauffe urbaine et l’îlot de chaleur urbain sont des phénomènes de plus en plus préoccupants dans le cadre du dérèglement climatique. Dans ce contexte, la description du tissu urbain est un enjeu crucial, pour organiser des campagnes de mesure autant que pour la modélisation des phénomènes étudiés. Les Zones Climatiques Locales (LCZ) sont une façon de segmenter et de classer un territoire à partir des caractéristiques morphologiques de la ville et de l’occupation du sol, d’une manière pertinente pour l’étude des phénomènes sus-cités.

Les LCZ reposent sur des règles définies à dire d’expert, et appliquées en fonction de l’information disponible. Contrairement à une caractéristique physique unique, directement mesurable, il n’existe pas de vérité de terrain ni d’échantillon de référence auquel comparer les sorties d’un algorithme de classification d’un territoire en LCZ.

Cet exposé propose de comparer d’abord deux cartes produites par deux algorithmes différents, puis de généraliser l’étude à un plus grand nombre d’algorithmes de production de LCZ.

Le traitement des données a été fait avec le langage R et des paquets de référence ou développés pour l’étude.

— **15h45-16h15 : Aimi Okabayashi, Université Bretagne Sud**

“Modèles génératifs pour l’observation de la Terre : application à la super-résolution”

LL’imagerie satellite offre une ressource exceptionnelle pour observer la Terre, qu’il s’agisse de dynamiques environnementales, d’activités humaines ou de gestion des risques naturels. L’abondance et la diversité de ces données ouvrent de nouvelles perspectives, mais leur analyse reste complexe. Les modèles d’IA jouent un rôle clé pour automatiser des tâches telles que la classification des cultures agricoles, la détection d’infrastructures

ou la prédition des feux de forêts. Exploiter efficacement ces données nécessite des approches capables d'intégrer des observations multisources et multirésolutions. Les modèles génératifs permettent de pallier le manque de données et, en particulier, la super-résolution offre la possibilité de générer des images haute résolution à partir d'observations basse résolution, réduisant ainsi les coûts d'acquisition.

Clôture des 10èmes Rencontres.