

## Productions scientifiques

### Publications

- 2023 • **Multiple imputation of incomplete multilevel data using Heckman selection models**, *Múnoz, J., Egger, M., Efthimiou, O., Audigier, V., de Jong, V. M. T. & Debray, Thomas. P. A.*, *Statistics in Medicine*.
- 2022 • **Clustering with missing data: which equivalent for Rubin's rules?**, *Audigier, V., Niang, N.*, *Advances in Data Analysis and Classification*.
- **An ensemble learning method for variable selection: application to high dimensional data and missing values**, *Bar-Hen, A., Audigier, V.*, *Journal of Statistical Computation and Simulation*.
- 2021 • **Severe altered immune status after Burn injury is associated with bacterial infection and Septic Shock**, *Moins-Teisserenc, H., Jorge Cordeiro, D., Audigier, V., Ressaire, Q., Benyamina, M., Lambert, J., Maki, G., Homyrda, L., Toubert, A., Legrand, M.*, *Frontiers in Immunology*.
- 2020 • **Severe diffuse alveolar hemorrhage related to autoimmune disease: a multicenter study**, *Mirouse, A., Parrot, A., Audigier, V., Demoule, A., Mayaux, J. et al.*, *Critical Care*.
- 2018 • **Multiple imputation for multilevel data with continuous and binary variables**, *Audigier, V., White, I., Jolani, S., Debray, T., Quartagno, M., Carpenter, J., van Buuren, S. & Resche-Rigon, M.*, *Statistical Science*.
- 2017 • **MIMCA: Multiple imputation for categorical variables with multiple correspondence analysis**, *Audigier, V., Husson, F. & Josse, J.*, *Statistics and Computing*, 2017.
- 2016 • **Multiple imputation for continuous variables using a Bayesian principal component analysis**, *Audigier, V., Husson, F. & Josse, J.*, *Journal of Statistical Computation and Simulation*, 2016.
- **A principal components method to impute mixed data**, *Audigier, V., Husson, F. & Josse, J.*, *Advances in Data Analysis and Classification*, 2016.

### Pre-Publications

- **Clustering with missing data: which imputation model for which cluster analysis method?**, *Audigier, V., Niang, N., Resche-Rigon, M.*
- **A simple Bayesian model to estimate proportions and ratios from count data with a hierarchical error structure with an application to droplet digital PCR experiments data: which imputation model for which cluster analysis method?**, *Mouhou, E., Audigier, V., Noirel, J.*

### Chapitres de livres

- 2022 • **Gestion des données manquantes par imputation multiple**, *A. Gégout-Petit; M. Maumy; G. Saporta; C. Thomas-Agnan.*, *Données manquantes*, Editions TECHNIP ISBN 9782710811954.

- **Imputation multiple en grande dimension par analyse factorielle**, A. Gégout-Petit; M. Maumy; G. Saporta; C. Thomas-Agnan., Données manquantes, Editions TECHNIP ISBN 9782710811954.

## Conférences

- 2024 • **Subspace clustering sur données incomplètes**, Agliz, Y. & Audigier, V. & Niang, N., 55emes journées de la statistique, Bordeaux, 27-31 Mai.
- 2023 • **Clustering sur données incomplètes : méthodes directes ou imputation multiple ?**, Audigier, V. & Sadou Zouleya F., 54emes journées de la statistique, Bruxelles, 3-7 Juillet.
- **Multiple imputation for clustering on incomplete data**, Audigier, V. & N. Niang, CLADAG 2023, Salerno, 11-13 Septembre, (invité).
  - **Handling missing data in clustering using multiple imputation**, Audigier, V., ERCIM 2023, Berlin, 16-18 Décembre, (organized Invited Session).
- 2022 • **Clustering with missing data: which imputation model for which cluster analysis method?**, Audigier, V., Niang, N., Resche-Rigon, M., IFCS 2022, Porto, 19-23 Juillet.
- **Données manquantes et analyse de survie**, Audigier, V., Resche-Rigon, M., Bonneville, E., EPICLIN 2022, Paris, 18-20 Mai.
- 2021 • **SOM-based clusterwise regression**, Hassini, H., Niang, N., Audigier, V., DSSV 2021, Rotterdam, 7-9 Juillet.
- **Cluster analysis after multiple imputation**, Audigier, V., Niang, N., ASMDA 2021, Athènes, 1-4 Juin.
  - **Clustering sur données incomplètes : quel modèle d'imputation choisir ?**, Audigier, V., Niang, N., Resche-Rigon, M., EPICLIN 2021, Marseille, 8-11 Juin.
- 2019 • **Imputation multiple pour données mixtes par analyse factorielle**, Audigier, V., Husson, F. & Josse, J., Journées de la Statistique, Nancy, 3 - 7 juin.
- **Clustering with missing data: Pooling multiple imputation results with consensus clustering**, Faucheux, L., Resche-Rigon, M, Audigier, V., Curis, E., Soumelis, V. & Chevret, S., ISCB, Louvain, 14-18 Juillet.
  - **micemd: a smart multiple imputation R package for missing multilevel data**, Audigier, V., Resche-Rigon, M, User2019!, Toulouse, 9-12 Juillet.
- 2018 • **Multiple imputation for multilevel data with continuous and binary variables**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Chimiométrie, Paris, 30 et 31 Janvier.
- **Multiple imputation for multilevel data with continuous and binary variables**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Journée de rencontres scientifiques autour de la statistique pour la biologie et la médecine, Poitiers, 1er Février.
  - **Une méthode d'ensemble pour la sélection de variables : application à la grande dimension et aux données manquantes**, Audigier, V., Bar-Hen A., Journées de la Statistique, Saclay, 28 Mai - 1er juin.
- 2017 • **Multiple imputation for multilevel data with continuous and binary variables**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., SMB, Paris, 14 et 15 Septembre.
- 2016 • **Comparison of multiple imputation methods for systematically and sporadically missing multilevel data**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., \*ISCB\*, Birmingham, 21-25 Août.

- **Comparison of multiple imputation methods for systematically and sporadically missing multilevel data**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Journées GDR / SFB, Lyon, 27-28 juin.
- 2015
- **Multiple imputation with MCA**, Audigier, V., Husson, F. & Josse, J., Rencontres doctorales Lebesgue, Nantes, 28-30 Octobre.
  - **Multiple imputation with MCA**, Audigier, V., Husson, F. & Josse, J., \*CARME\*, Naples, 21-23 Septembre.
  - **Multiple imputation for categorical data using MCA**, Audigier, V., Husson, F. & Josse, J., \*MissData2015\*, Rennes, 17-19 Juin.
  - **Multiple imputation with MCA**, Audigier, V., Husson, F. & Josse, J., 47emes Journées de Statistique, Lille, 1-5 Juin.
- 2014
- **Multiple imputation with MCA**, Audigier, V., Husson, F. & Josse, J., Journées de Statistique de Rennes, Rennes, 23-24 Octobre.
  - **Imputation multiple pour variables quantitatives par Analyse en composantes principales Bayésienne**, Audigier, V., Husson, F. & Josse, J., 46emes Journées de Statistique, Rennes, 2-6 Juin.
- 2013
- **Imputation of mixed data: Random Forests versus PCA**, Josse, J., Audigier, V. & Husson, F., ERCIM, Londres, 14-16 Decembre.
  - **Imputation multiple à l'aide des méthodes d'analyse factorielle**, Audigier, V., Husson, F. & Josse, J., 45emes Journées de Statistique, Toulouse, 27-31 Mai.
- 2012
- **Missing values imputation for mixed data based on principal component methods**, Husson, F., Audigier, V. & Josse, J., COMPSTAT, Chypre, 27-31 Août.
  - **Imputation de données manquantes pour des données mixtes via les méthodes factorielles grâce à missMDA**, Audigier, V., Husson, F. & Josse, J., Premières rencontres R, Bordeaux, 2-3 Juillet.
- [Séminaires](#)
- 2023
- **Imputation multiple pour la gestion des données manquantes en clustering**, Audigier, V., Niang, N., Centre Borelli, Paris, 27 Septembre.
  - **Imputation multiple pour la gestion des données manquantes en clustering**, Audigier, V., Niang, N., Séminaire du laboratoire ERIC, Paris, 9 Octobre.
  - **Imputation multiple pour la gestion des données manquantes en clustering**, Audigier, V., Niang, N., Séminaire du laboratoire de Mathématiques de Bretagne Atlantique, Vannes, 20 Octobre.
- 2021
- **Clustering with missing data: which imputation model for which cluster analysis method?**, Audigier, V., Niang, N., Resche-Rigon, M., Séminaire MAP5, Paris, 4 Juin.
  - **Clustering with missing data: which imputation model for which cluster analysis method?**, Audigier, V., Niang, N., Resche-Rigon, M., Missing Data, Imputation & Analysis group, London School of Hygiene & Tropical Medicine, Londres, 18 juin.
- 2019
- **An ensemble learning method for variable selection: application to high dimensional data and missing values**, Audigier, V., Bar-Hen, A., Séminaire du SBIM, Paris, 5 Novembre.
  - **An ensemble learning method for variable selection**, Audigier, V., Bar-Hen, A., Sino-French meeting, Paris, 12 Octobre.
- 2018
- **An ensemble learning method for variable selection: application to high dimensional data and missing values**, Audigier, V., Bar-Hen, A., Missing Data, Imputation & Analysis group, London School of Hygiene & Tropical Medicine, Londres, 28 Novembre.

- 2017 • **Some contributions for handling missing values by multiple imputation**, Audigier, V., Séminaire de l'équipe MSDMA, Paris, 29 Septembre.
- **Multiple imputation with principal component methods**, Audigier, V., Josse, J. & Husson, F., Data Science Seminar de Telecom Paristech, Paris, 16 Novembre.
- **Multiple imputation for multilevel data with continuous and binary variables**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Séminaire de Statistique et Probabilités Appliquées du LJK, Grenoble, 14 Décembre.
- 2016 • **Comparison of multiple imputation methods for systematically and sporadically missing multilevel data**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Séminaire de l'équipe INRIA MODAL, Lille, 22 Novembre.
- **Comparison of multiple imputation methods for systematically and sporadically missing multilevel data**, Audigier, V., White, I., Jolani, S. Debray, T., Quartagno, M. van Buuren S. & Resche-Rigon, M., Missing Data, Imputation & Analysis group, London School of Hygiene & Tropical Medicine, Londres, 28 septembre.
- **Multiple imputation with principal component methods**, Audigier, V., Husson, F. & Josse, J., Institut de Recherche Mathématique Avancée, Strasbourg, 29 mars.
- **Multiple imputation with Multiple Correspondence Analysis**, Audigier, V., Husson, F. & Josse, J., Missing Data, Imputation & Analysis group, London School of Hygiene & Tropical Medicine, Londres, 21 mars.
- **Multiple imputation with principal component methods**, Audigier, V., Husson, F. & Josse, J., Institut de Santé Publique, d'Epidémiologie et de Développement, Bordeaux, 14 mars.
- 2014 • **Multiple imputation with Bayesian PCA**, Josse, J., Audigier, V. & Husson, F., Séminaire de l'institut de mathématiques de Bordeaux, 6 Mars.

## Enseignements

- 2022-2023 • **Fouille de Données**, *Formation à distance*, 78h, CNAM, Paris.  
**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)
- **Statistique exploratoire multidimensionnelle**, *Formation à distance et présentiel*, 180h, Cours, TD, CNAM, Paris.  
**Thématiques** Analyse univariée, bivariée, multivariée (ACP, AFC, ACM, AFDM) clustering (k-means, CAH), gestion des données manquantes
- **Analyse multivariée approfondie**, *Cours*, 4h, M1, CNAM, Paris.  
**Thématiques** Gestion des données manquantes dans modèle linéaire par imputation multiple
- 2021-2022 • **Gestion des données manquantes par imputation multiple**, *Cours/TP (Formation continue)*, 8h, Journées d'études en statistique (SFDS), Fréjus.  
**Thématiques** Imputation multiple, imputation multiple par analyse factorielle
- **Gestion des données manquantes**, *Cours/TP (Formation continue)*, 6h, Ateliers de la statistique (SFDS).  
**Thématiques** Données manquantes, Imputation multiple
- **Fouille de Données**, *Formation à distance*, 78h, CNAM, Paris.  
**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)

- **Statistique exploratoire multidimensionnelle**, *Formation à distance et présentiel*, 150h, Cours, TD, CNAM, Paris.  
**Thématiques** Analyse univariée, bivariée, multivariée (ACP, AFC, ACM, AFDM) clustering (k-means, CAH), gestion des données manquantes
  - **Analyse multivariée approfondie**, *Cours*, 4h, M1, CNAM, Paris.  
**Thématiques** Gestion des données manquantes dans modèle linéaire par imputation multiple
- 2020-2021
- **Fouille de Données**, *Formation à distance*, 55h, CNAM, Paris.  
**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)
  - **Statistique exploratoire multidimensionnelle**, 45h, Cours, CNAM, Paris.  
**Thématiques** Analyse univariée, bivariée, multivariée (ACP, AFC, ACM, AFDM) clustering (k-means, CAH), gestion des données manquantes
  - **Analyse multivariée approfondie**, *Cours*, 4h, M1, CNAM, Paris.  
**Thématiques** Gestion des données manquantes dans modèle linéaire par imputation multiple
- 2019-2020
- **Biostatistique**, *Préparation des cours, TD et évaluation des soutenances de M2*, 51h, Master 1 (Mastère spécialisé de santé publique), CNAM, Paris.  
**Thématiques** Statistique descriptive, probabilités élémentaires, variables aléatoires, tests paramétriques pour un ou deux groupes
  - **Fouille de Données**, *Formation à distance*, 55h, CNAM, Paris.  
**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)
  - **Statistique exploratoire multidimensionnelle**, *Formation à distance (165h) et présentiel (70h)*, Cours et TD, CNAM, Paris.  
**Thématiques** Analyse univariée, bivariée, multivariée (ACP, AFC, ACM, AFDM) clustering (k-means, CAH), gestion des données manquantes
  - **Analyse multivariée approfondie**, *Cours*, 4h, M1, CNAM, Paris.  
**Thématiques** Gestion des données manquantes dans modèle linéaire par imputation multiple
  - **Modélisation avec données manquantes**, 6h, Cours, Paris Dauphine.  
**Thématiques** Gestion des données manquantes par imputation multiple. Imputation multiple en grande dimension.
  - **Modélisation avec données manquantes**, 2h, Cours, ENSAE, Saclay.  
**Thématiques** Gestion des données manquantes par imputation multiple
- 2018-2019
- **Statistique mathématique**, *Cours et TD*, 9h, L3 (Formation continue), CNAM, Paris.  
**Thématiques** simulation de variables aléatoires, bootstrap
  - **Biostatistique**, *Préparation des cours, TD et évaluation des soutenances de M2*, 51h, Master 1 (Mastère spécialisé de santé publique), CNAM, Paris.  
**Thématiques** Statistique descriptive, probabilités élémentaires, variables aléatoires, tests paramétriques et non-paramétriques pour un ou deux groupes
  - **Fouille de Données**, *Formation à distance*, 45h, CNAM, Paris.  
**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)
  - **Analyse statistique**, *Cours et TD*, 18h, L3 (Formation ingénieur en partenariat), CNAM, Paris.  
**Thématiques** Statistique descriptive, ACP

- **Statistique exploratoire multidimensionnelle**, *Formation à distance et présentiel 165h*, Cours et TD, CNAM, Paris.

**Thématiques** Analyse univariée, bivariée, multivariée (ACP, AFC, ACM) clustering (k-means, CAH)

- **Analyse multivariée approfondie**, *Cours*, 4h, M1, CNAM, Paris.

**Thématiques** Gestion des données manquantes dans modèle linéaire par imputation multiple

- **Fouille de données**, *Cours/TP (Formation continue)*, 32h, Direction Générale des Impôts Maroc, Rabat.

**Thématiques** Analyse exploratoire, pré-traitement, clustering, recherche de règles d'association

- **Probabilités et Analyse statistique**, *Cours*, 18h, M1, CNAM/ENJMIN, Angoulême.

**Thématiques** Statistique descriptive, probabilités élémentaires, modélisation

- **Modélisation avec données manquantes**, *2h*, Cours, ENSAE, Saclay.

**Thématiques** Gestion des données manquantes par imputation multiple

- **Modélisation avec données manquantes**, *6h*, Cours, Paris Dauphine.

**Thématiques** Gestion des données manquantes par imputation multiple. Imputation multiple en grande dimension.

- 2017-2018
- **Evaluation de méthodes d'analyses en biosciences**, *Préparation et animation des TD*, 12h, Licence 1 (Première année commune aux études de santé), Université Paris-Diderot, Paris.

**Thématiques** Statistique descriptive, probabilités élémentaires, variables aléatoires

- **Statistique mathématique**, *Cours et TD*, 75h, L3 (Formation continue), CNAM, Paris.

**Thématiques** Statistique inférentielle, simulation de variables aléatoires, tests paramétriques et non paramétriques

- **Biostatistique**, *Préparation des cours, TD et évaluation des soutenances de M2*, 51h, Master 1 (Mastère spécialisé de santé publique), CNAM, Paris.

**Thématiques** Statistique descriptive, probabilités élémentaires, variables aléatoires, tests usuels pour un ou deux groupes

- **Fouille de Données**, *Formation à distance*, 50h, CNAM, Paris.

**Thématiques** Pré-traitement (réduction de la dimension, gestion des données manquantes, etc), Méthodes supervisées (arbres de décision, forêt aléatoires) et non-supervisées (cartes de Kohonen, règles d'association), Méthodes d'ensemble (Bagging, Boosting)

- **Probabilités et Analyse statistique**, *Cours et TD*, 36h, L3 (Formation ingénieur en partenariat), CNAM, Paris.

**Thématiques** Statistique descriptive, probabilités élémentaires, ACP, clustering

- 2016-2017
- **Evaluation de méthodes d'analyses en biosciences**, *Préparation et animation des TD, conception de sujet d'examen*, 64h, Licence 1 (Première année commune aux études de santé), Université Paris-Diderot, Paris.

**Thématiques** Statistique descriptive, probabilités élémentaires, variables aléatoires, tests statistiques usuels.

- **Biostatistique et logiciel R**, *Cours, préparation et animation des TD, contrôle des acquis*, 6h, Master 1 (Santé publique), Université Paris-Diderot, Paris.

**Thématiques** Inférence statistique, test de comparaison pour 3 groupes, test pour variables qualitatives.

- **Modélisation statistique**, *Cours/TD*, 2h, Master 2 (Santé publique), Université Paris-Diderot, Paris.

**Thématique** Gestion des données manquantes.

- 2012-2015 • **Statistique générale**, *Préparation et animation des TD, contrôle des acquis, suivi de projets*, 160h, Licence 2 et Master 1 (ingénieurs agroalimentaires, ingénieurs agronomes, apprentis ingénieurs en alternance, universitaires), Agrocampus-Ouest, Rennes.  
**Thématiques** regression multiple, analyse de la variance, plans d'expérience, ACP, classification, logiciel de statistique R.
- **Traitement de données d'enquêtes**, *Préparation et animation des TD, contrôle des acquis, suivi de projets*, 45h, Licence 3 (ingénieurs agroalimentaires, apprentis ingénieurs en alternance), Agrocampus-Ouest, Rennes.  
**Thématiques** Analyse des Correspondances Multiples, classification, logiciel de statistique R
- 2013 • **Sensométrie**, *Préparation et animation des TD, contrôle des acquis, suivi de projets*, 15h, Licence 3 (ingénieurs agroalimentaires), Agrocampus-Ouest, Rennes.  
**Thématiques** caractérisation de produits, performance d'un jury, cartographie des préférences, logiciel de statistique R

## Cursus

- sept 2017-Auj. • **Maître de conférences**, *Conservatoire National des Arts et Métiers, Equipe MSDMA, CEDRIC, Paris*.
- 2017 • **Assistant Hospitalo-Universitaire**, *Service biostatistique et information médicale de l'hôpital Saint Louis, Equipe ECSTRA, INSERM U1153, Paris*.  
**Activités** Recherche en statistique, enseignements de biostatistique, consulting clinique  
**Responsable** Matthieu Resche-Rigon (PU-PH, Equipe ECSTRA, INSERM U1153)  
**Mots-clés** imputation multiple, données mixtes, analyse factorielle, modèle à effets aléatoires, biostatistique
- 2016 • **Post-Doctorat en statistique**, *Service biostatistique et information médicale de l'hôpital Saint Louis, Equipe ECSTRA, INSERM U1153, Paris*.  
**Titre** Gestion des périodes à trous  
**Responsable** Matthieu Resche-Rigon (PU-PH, Equipe ECSTRA, INSERM U1153)  
**Mots-clés** Données manquantes, imputation multiple, données multiniveaux, périodes à trous, méta-analyse, équations chaînées.
- 2013 - 2015 (2 ans) • **Doctorat en statistique**, *Laboratoire de Mathématiques Appliquées d'Agrocampus-Ouest, IRMAR UMR 6625 CNRS, Rennes*.  
**Titre** Imputation multiple via l'utilisation des méthodes d'analyse factorielles : une nouvelle méthodologie pour traiter les données manquantes  
**Jury** Charles Bouveyron (PR, Université Paris Descartes, UMR 8145) [rapporteur], Iven van Michelen (PR, Université de Louvain) [rapporteur], Christophe Biernacki (Université de Lille 1, Laboratoire de Mathématiques, UMR 8524) [président], Matthieu Resche-Rigon (MCU-PH, Université Paris-Diderot, UMR 1153) [examinateur], François Husson (PR, Agrocampus-Ouest, IRMAR) [Directeur] Julie Josse (MC, Agrocampus-Ouest, IRMAR) [Directeur]  
**Mots-clés** Données manquantes, imputation simple, imputation multiple, analyse factorielle, ACP, ACM, AFDM, Bayésien, bootstrap données quantitatives, données qualitatives, données mixtes
- 2012 - 2013 (1 ans) • **Ingénieur d'études**, *Laboratoire de Mathématiques Appliquées d'Agrocampus-Ouest, IRMAR UMR 6625 CNRS, Rennes*.  
**Projet** Enseignements de statistique générale dans les filières agronomes et agroalimentaires d'Agrocampus-Ouest ; Recherche sur l'imputation par méthodes factorielles ; Appui logistique  
**Mots-clés** Enseignement, données manquantes, imputation par analyse factorielle
- 2010 - 2012 • **Master Modélisation Mathématiques et Analyse Statistique**, *Université de Poitiers*, Mention Très bien.  
**Mots-clés** : Statistique inférentielle, statistiques exploratoire, séries temporelles, processus à temps discrets et continus, algorithmes stochastiques, optimisation, théorie des graphes, imagerie, génomique, analyse de survie

## Expériences professionnelles

### Responsabilités scientifiques

2014-Auj. • **Relecture d'articles scientifiques**, Relecture de papiers en lien avec mes thématiques de recherche pour des revues internationales (Biometrika, Bioinformatics, CSDA, Neural Processing Letters, Journal of Chemometrics, ADAC, Sociological Methods and Research, Revstat, Plos One, Statistics and Probability Letters, Bernoulli Journal, Applied Soft Computing Journal).

**Mots-clés** génomique, chimiométrie, ACP, boosting, données mixtes, sélection de variables, données multi-niveaux, données manquantes

2015-2023 • **Organisation de congrès internationaux.**

**Symbolic Data Analysis Workshop 2023** CNAM, Paris, 2-4 Novembre. Membre du comité d'organisation. Tâches effectuées : organisation logistique, appui technique.

**ICML Artemiss 2020** on-line, 8 juin. Membre du comité de programmation. Tâches effectuées : relecture des propositions de communication.

**Chimiométrie 2018** CNAM, Paris, 30-31 janvier. Membre du comité d'organisation. Tâches effectuées : recherche de financement, organisation logistique.

**MissData2015** Agrocampus-Ouest, Rennes, 17-19 juin. Membre du comité scientifique et du comité d'organisation. Tâches effectuées : choix des invités, recherche de financements, sélection des résumés, organisation logistique, communication

2015 - • **Implication dans des instances administratives.**

**Membre du conseil de l'IRMAR** Rennes, 2015. Validation des dépenses allouées aux activités de recherche (déplacements d'invités, organisation de congrès, etc), et des politiques mises en œuvre. Transmission d'informations auprès de mon équipe de recherche.

**Membre du conseil du laboratoire CEDRIC** Paris, depuis 2019.

**Membre du conseil de l'équipe pédagogique nationale *mathématiques et statistique*** CNAM, Paris, depuis 2023.

2018 - • **Participation à des comités.**

**Membre de comités de sélection**

- Poste MCF en section 26 à l'Oniris (comité présidé par Mathieu Emily)
- Poste MCF en section 26 au CNAM (comité présidé par Crisitane Preda)
- Poste MCF en section 26 à l'Université de Versailles St-Quentin-en-Yvelines (comité présidé par E. Rio)
- Poste d'enseignant chercheur contractuel en IA (sections 26/27/61) à l'école d'ingénieurs Télécom Physique Strasbourg (comité présidé par Pierre Gançarski)

**Membre de comités de suivi de thèse**

2023 - 2026 K. Amezouwui, thèse dirigée par M. Marbac (ENSAI)

2021 - 2024 M. K. Benramdane, thèse dirigée par S. Bouzeffrane (CNAM)

2018 - 2020 N. Guard, thèse CIFRE dirigée par N. Niang (CNAM)

2019 W. Rao Fernandes, thèse dirigée par P. Cubaud (CNAM)

**Membre de comité technique**

2021 - 2024 H. Gadacha, thèse CIFRE dirigée par N. Niang (CNAM)

**Membre de comité de perfectionnement de diplômés**

2023 Master *Traitement de l'information et exploitation des données*

2015 - • **Autres responsabilités.**

**Co-organisateur du Forum Emploi Maths** Cité des Sciences et de l'Industrie, en 2018, 2019 et 2023. Choix du programme et des intervenants. Invitation d'une quarantaine d'intervenants et échanges avec eux sur les sujets abordés lors de l'évènement.

**Organisateur du séminaire de statistique de l'équipe MSDMA** CNAM, Paris, 2017-2022.

**Modérateur pour les évènements satellites des Journées de la statistique** Saclay, 2018

**Animateur du journal club** Hôpital Saint-Louis, Service de biostatistique et d'information médicale, Paris, 2017.

**Co-organisateur du séminaire des doctorants de statistique de l'IRMAR** Rennes, 2015.

### Vulgarisation

2023 **La Statistique dans les jeux vidéos (poster)**, *Salon Culture et Jeux Mathématiques 2023*, [https://www.sfds.asso.fr/fr/group/grand\\_public/728-posters/](https://www.sfds.asso.fr/fr/group/grand_public/728-posters/).

### Activités de conseil

2017-2018 • **Consulting clinique**, *Hôpital Saint-Louis, en cours.*

**Projet 1** intérêt du suivi de l'antigène galactomannane chez des patients d'hématologie

**Collaborateurs** Matthieu Resche-Rigon, Blandine DENIS, Hannah Moatti

**Mots-clés** biostatistique, survie

**Projet 2** variation de longueur après la pose d'une prothèse totale du genou

**Collaborateur** Guillaume Rigoulot

**Mots-clés** Modélisation, modèle linéaire

**Projet 3** Evolution des profils immunitaires chez les grands brûlés

**Collaborateur** Hélène Moins, Matthieu Legrand, Jérôme Lambert

**Mots-clés** Données multiblocs, données longitudinales

2015 (1 semaine) • **Consulting pour HydrOcean**, *Semaine d'études maths-entreprise, Université de Nantes, 18-22 Mai.*

**Projet** Estimation de la pression en un point d'un solide plongé dans l'eau

**Collaborateurs** Matthieu de Leffe (HydrOcean), Lise Bellanger (MC en statistique), Hélène Mathis (MC en analyse numérique), et 5 doctorants de thématiques mathématiques variées (géométrie, analyse d'EDP,...)

**Mots-clés** Estimation non-paramétrique, noyau, splines, transformée de Fourier

2013-2014 • **Enquête sur les pratiques agricoles au niveau des bassins versants bretons**, *Laboratoire de Mathématiques Appliquées d'Agrocampus-Ouest, IRMAR UMR 6625 CNRS, Rennes.*

**Projet** Déterminer l'évolution des pratiques agricoles par rapport à l'usage de pesticides et de fertilisants entre 2004 et 2011 (à partir de données déclaratives et de données de terrain)

**Collaborateurs** Jérôme Pagès (LMA2), Philippe Michel (DRAAF Bretagne), Yvan Hurvois (Agence de l'eau), Dikran Zakeossian (EPICES)

**Mots-clés** Analyse factorielle multiple, ACP

## Encadrement

- 2024 • **Encadrement stage de recherche niveau M2, CNAM, Equipe MSDMA, Paris.**  
(5 mois) **Titre**  
**Etudiant** Djafarou OULARE  
**Niveau** M2  
**Mots-clés** Données fonctionnelles, clustering, océanographie
- 2023 • **Encadrement stage de recherche niveau M2, CNAM, Equipe MSDMA, Paris.**  
(6 mois) **Titre** Imputation multiple par régression non-paramétrique  
**Etudiant** Koffi AMEZOUWUI  
**Niveau** M2  
**Mots-clés** Données manquantes, imputation multiple séquentielle, régression non-paramétrique
- (5 mois) • **Encadrement stage de recherche niveau M2, CNAM, Equipe MSDMA, Paris / INSERM, Equipe ECCSTRA, Paris.**  
**Titre** Sélection de variables pour les modèles d'imputation en grande dimension  
**Etudiant** Wilfrid Jakkie Nkodo  
**Niveau** M2  
**Mots-clés** Données manquantes, sélection de variables, imputation multiple séquentielle
- (en cours depuis sept. 2022) • **Co-direction de thèse 30%, Groupe Caisse des Dépôts / CNAM.**  
**Titre** Clustering and data embedding with missing data  
**Directrice** Ndèye Niang  
**Doctorant** Yasmine Agliz  
**Ecole doctorale** Sciences des Métiers de l'Ingénieur  
**Mots-clés** clustering, données manquantes
- 2020 - 2023 • **Co-direction de thèse 50%, CNAM, Equipe ROC/MSDMA, Paris.**  
**Titre** Chiffrement homomorphe et machine learning  
**Directrice** Samia Bouzefrane  
**Doctorant** Yulliwas Ameur  
**Ecole doctorale** Sciences des Métiers de l'Ingénieur  
**Mots-clés** Chiffrement homomorphe, données manquantes, machine learning
- 2022 • **Encadrement stage de recherche niveau M2, CNAM, Equipe MSDMA/Genomique**  
(5 mois) **Bioinformatique, Paris.**  
**Titre** Caractérisation de la relation génotype-phénotype des troubles mentaux par l'apprentissage statistique de matrices d'expression tout-tissu  
**Etudiant** Barbara Bricout  
**Niveau** ENSAI 3eme année  
**Mots-clés** génétique, grande dimension, données complexes
- (6 mois) • **Encadrement stage de recherche niveau M2, CNAM, Equipe MSDMA/ INSERM, Equipe ECCSTRA, Paris.**  
**Titre** Imputation multiple pour données biologiques hétérogènes  
**Etudiant** Arbnor Zenuni  
**Niveau** M2  
**Mots-clés** biologie, données manquantes, imputation multiple

- 2021 • **Encadrement stage de recherche niveau M2**, *CNAM, Equipe MSDMA, Paris.*  
 (6 mois) **Titre** Apport du consensus pondéré pour le clustering après imputation multiple  
**Etudiant** Lamine Ndao  
**Niveau** M2 TRIED  
**Mots-clés** données manquantes, consensus de partitions, imputation multiple, clustering
- **Encadrement stage de recherche niveau M1**, *CNAM, Equipe MSDMA, Paris.*  
 (3 mois) **Titre** Apport du consensus pondéré pour le clustering après imputation multiple  
**Etudiant** Fadela Sadou Zouleya  
**Niveau** ENSAE 2ème année  
**Mots-clés** données manquantes, clustering, imputation multiple
- 2020 • **Encadrement stage de recherche niveau M2**, *CNAM, Equipe MSDMA, Paris.*  
 (6 mois) **Titre** Sparse Clusterwise  
**Etudiant** Houda HASSINI  
**Niveau** M2 TRIED  
**Mots-clés** Données complexes, régression clusterwise, cartes de Kohonen, régularisation
- (6 mois) • **Encadrement stage de recherche niveau M2**, *CNAM, Equipe MSDMA, Paris.*  
**Titre** Régression clusterwise multiniveaux  
**Etudiant** Redha ALLICHE  
**Niveau** M2 TRIED  
**Mots-clés** Données complexes, régression clusterwise, modèles mixtes
- 2017 • **Encadrement stage niveau L2**, *Service biostatistique et information médicale de l'hôpital Saint Louis, Equipe ECSTRA, INSERM U1153, Paris.*  
 (3 mois) **Titre** Typologie du trouble bipolaire : une approche par analyse factorielle  
**Etudiant** Matthieu François  
**Niveau** DUT STID 2eme année  
**Mots-clés** Analyse factorielle, analyse de questionnaires, bipolarité

## Compétences informatiques

### Programmation

- **Statistique.**

Maîtrise du langage de programmation statistique R

auteur et mainteneur des packages *micemd* et *clusterMI*

contributions au package *missMDA* (apport de nouvelles fonctions)

calcul parallèle

Notions de programmation en SAS

- **Calcul formel.**

Bonnes bases en programmation sous Maple

- **Calcul numérique.**

Bonnes bases en programmation sous Scilab

- **Programmation orientée objet.**

Notions de programmation en langage Ada et C

### Système de gestion de base de données

- **Maîtrise du système ACCESS 2000.**

## Bureautique

- **Maîtrise des outils de bureautique (titulaire du certificat informatique et internet niveau 1).**

## Docimologie numérique

- **Certifié SIDE-S (Système informatique distribué d'évaluation en santé).**