



**Chayma DAAAYEB**  
Montpellier, 34090  
07 51 26 26 03  
<https://chaymadaayeb.github.io>  
[chayma.daayeb@umontpellier.fr](mailto:chayma.daayeb@umontpellier.fr)

## DOCTEUR EN MATHÉMATIQUES APPLIQUÉES (Biostatistique)

### EXPÉRIENCE PROFESSIONNELLE

**2021-2023 ATER à l'École polytechnique universitaire de Montpellier**

- **Analyse de données (Niveau M1) (CM/TP) 52h.**
  - Analyse de données (R et Python)
  - Data visualisation – Exploration
  - Régression et classification (régression simple et multiple, régression logistique, arbres de décision, forêts aléatoires, SVM, ...)
- **Machine Learning (Niveau M1) (CM/TP) 56h.**
  - Python
  - Scikit-learn
  - Numpy
  - Pandas
  - SQL
- **Analyse de données multidimensionnelles (Niveau M1) (TP) 20h.**
  - Rmarkdown
  - Visualisation de données
  - Shiny
- **Économétrie techniques de scoring et régression (Niveau M2) (CM/TP) 92h.**
  - Estimation : technique du maximum de vraisemblance
  - Test : Anova, Student, Khi2 et Fisher
  - Régression et classification
- **Probabilité et Statistique (Niveau L 3) (CM/TD) 18h.**
  - Probabilité
  - Statistiques descriptive
  - Inférence statistique
- **Maths pour l'Informatique (Niveau L3) (CM/TP/TD) 21h.**
  - Suites
- **Maths de la décision (Niveau L 3) (CM/TP/TD) 92h**
  - Optimisation

### FORMATIONS

**2018-2022 Doctorat de Mathématiques Appliquées et Biostatistique.**  
Université de Montpellier.

**Titre : Modèles linéaires fonctionnels avec des données partiellement observées.**

Mots-clés : Modèle linéaire fonctionnel, composantes principales fonctionnelles,

données partiellement observées, données manquantes, imputation par régression, Estimation, Prédiction.

- Calculer la vitesse de convergence pour l'erreur quadratique moyenne de prédiction, une fois que les courbes sont reconstruites.
- Présenter des simulations, des résultats numériques et des illustrations sur un jeu de données réelles.
- Développer un programme sur R pour les méthodes de reconstruction des courbes partiellement

observées, de l'imputation (simple et multiple) et de la prédiction de la réponse (réelle et fonctionnelle).

Directeurs de thèse :

Ali GANNOUN, IMAG - Institut Montpelliérain Alexander Grothendieck,

Mohamed MNIF, LAMSIN - Laboratoire de Modélisation Mathématique et Numérique dans les Sciences de l'Ingénieur.

Co-directeurs de thèse :

Christophe CRAMBES, IMAG - Institut Montpelliérain Alexander Grothendieck,

Yousri HENCHIRI, LAMSIN - Laboratoire de Modélisation Mathématique et Numérique dans les Sciences de l'Ingénieur.

**2017-2018 Mastère Recherche en Modélisation Mathématique et Calcul Scientifique (M2).**

Mémoire de recherche au sein du LAMSIN : Estimation non-paramétrique de la fonction de régression en présence des données manquantes : théorie et applications.

Directeur de stage : Yousri HENCHIRI.

ENIT, Ecole Nationale d'Ingénieurs de Tunis.

*Publié sous forme d'un livre par Éditions Universitaires Européennes.*

**2016-2017 Mastère Fondamentale en Mathématiques et Applications (M1).**

FST, Faculté des Sciences Mathématiques, Physiques et Naturelles de Tunis, Tunisie.

**2011-2015 Licence Fondamentale en Mathématiques et Applications.**

FST, Faculté des Sciences Mathématiques, Physiques et Naturelles de Tunis, Tunisie.

<b>PUBLICATIONS</b>
---------------------

- **Functional linear model with with partially observed covariate and missing values in the response, 2022** (*Journal of Nonparametric Statistics*, 1-26, doi: 10.1080/10485252.2022.2142222).

Les valeurs manquantes sont un des problèmes qui surviennent fréquemment dans le processus d'observation ou d'enregistrement des données. Dans ce travail, nous considérons le modèle de régression linéaire fonctionnelle, lorsque la variable d'intérêt, réelle, et la variable explicative, fonctionnelle, contiennent des valeurs manquantes. Nous utilisons un opérateur de reconstruction qui vise à reconstruire les parties manquantes dans les courbes, puis nous nous intéressons à la méthode d'imputation par régression des données manquantes sur la variable réponse, en utilisant la régression fonctionnelle sur composantes principales pour estimer le coefficient fonctionnel du modèle. Nous étudions le comportement asymptotique de l'erreur de prévision commise lorsque les valeurs manquantes sont remplacées par les valeurs imputées. Le comportement de la méthode est également étudié en pratique sur des données simulées et réelles.

- **Multiple imputation in the functional linear model with partially observed covariate and missing values in the response, 2022** (*TEST Statistical Journal*).

Les méthodes d'imputation multiple sont des outils populaires et naturels pour traiter les données manquantes, qui offrent une méthode calculable pour étudier un large éventail de problèmes sous une série de mécanismes d'absence de données en remplaçant chaque valeur manquante par un ensemble de valeurs raisonnables qui représentent l'incertitude sur les valeurs sous-jacentes. Cette méthode consiste à créer  $q \geq 2$  valeurs imputées pour remplacer les données manquantes par des valeurs estimées dans  $q$  fichiers de données complets. Nous étudions le modèle linéaire fonctionnel en présence des données manquantes dans la variable explicative fonctionnelle et la réponse réelle.

Une fois que les courbes sont bien reconstruites, nous utilisons l'imputation multiple. Nous étudions la vitesse de la convergence de l'erreur quadratique moyenne de la prévision.

- **Prediction in functional-on-functional linear model with partially observed functional covariate and response, 2023** (*Preprint*).

En analyse des données fonctionnelles, le modèle linéaire fonctionnel est certainement le modèle qui a été le plus exploré dans la recherche théorique et appliquée telle que la climatologie, la météorologie, l'économie, l'analyse d'images et bien d'autres domaines. En pratique, ce modèle peut présenter des données partiellement observées dans la covariable fonctionnelle ou dans la réponse ou dans les deux. Ces problèmes sont devenus critiques au cours de ces dernières années. Dans ce travail, nous proposons un modèle linéaire fonctionnel en présence des données manquantes sur la variable explicative et la réponse qui sont tous les deux des variables fonctionnelles. Nous cherchons à reconstruire les courbes par deux méthodes (opérateur de reconstruction et imputation déterministe). Une fois que les données fonctionnelles seront complètes, nous déterminons la vitesse de convergence de la prédiction ainsi le comportement des méthodes en pratique sur des données simulées et réelles.

## PRESENTATIONS

- 02/2023** **Journée des doctorant.es, IMAG**  
« Modèle linéaire fonctionnel avec des données partiellement observées »
- 11/2022** **Séminaire de Probabilités et Statistique, IMAG.**  
« Functional linear models with partially observed data »
- 11/2022** **SEmDoc 2022 - Séminaire des Doctorant·e·s IMAG.**  
« Functional linear models with partially observed data »
- 06/2022** **JDS 2022 - 53èmes Journées de Statistique de la Société Française de Statistique (Sfds).**  
« Imputation multiple dans le modèle linéaire fonctionnel avec une covariable partiellement observée et des valeurs manquantes dans la réponse. »
- 06/2021** **JDS 2021 - 52èmes Journées de Statistique de la Société Française de Statistique (Sfds).**  
« Préviation dans le modèle linaire fonctionnel en présence des données manquantes dans la réponse et covariable. »
- 03/2021** **READ'21 - Rencontre Annuelle des Doctorant.es, LAMSIN.**  
« Modèle linaire fonctionnel en présence des données manquantes dans la réponse et covariable. »
- 12/2019** **MYRPAM, Maghrebien Meeting of Young Researches in pure and Applied Mathematics, Tunisie.**  
« Estimation non-paramétrique de la fonction de régression en présence des données manquantes. »
- 12/2018** **READ'18 - Rencontre Annuelle des Doctorants, Tunisie.**  
« Estimation non-paramétrique de la fonction de régression en présence des données manquantes. »

## COMPETENCES INFORMATIQUES

- **Langages & Logiciels** : R, Python, Pack Office, Latex, MATLAB
- **Machine learning** : Pytorch, Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn
- **Base de données** : SQL, PostgreSQL
- **Systèmes d'exploitation**: Windows, Linux (Ubuntu)

## LANGUES

**Français** : Niveau courant

**Anglais** : Niveau courant : Diplôme – régime maîtrise (2013-2015)

## ACTIVITES DIVERSES

- 2018-2020** Participation à l'organisation du Séminaire doctorant LAMSIN-ENIT.
- 2019** Participation à l'organisation de l'ENIT Doctoral Day (Journée des Doctorants de l'ENIT).
- 2018** Participation à l'organisation du READ Rencontre Annuelle des Doctorants Hammamet.