Master's Internship Proposal – 2027

Development and Evaluation of Machine Learning Methods for the Analysis of Multidimensional Longitudinal Trajectories

Internship supervisors: C. Chouquet, S. Mercier, IMT
Contacts: cecile.chouquet@math.univ-toulouse.fr, sabine.mercier@math.univ-toulouse.fr

Background

The general background of this internship is to develop, adapt, and evaluate statistical and machine learning methods for the analysis of multidimensional longitudinal trajectories. Multidimensional longitudinal trajectories describe the joint evolution of several variables, each measured sequentially (e.g., over time) for the same statistical unit. The main objective is to investigate different approaches for identifying homogeneous trajectory profiles, while accounting for both the sequential evolution and the multidimensional nature of the data.

Several families of multidimensional clustering methods are already available, including partitioning clustering methods (K-means, hierarchical agglomerative clustering) [Den Teuling, 2026], as well as unsupervised machine learning approaches (Kohonen self-organizing maps, neural networks). While some of these methods have already been extended to the longitudinal setting (such as K-means), this is not the case for all of them, particularly Kohonen self-organizing maps [Kohonen, 2013]. Adapting these methods to longitudinal data still raises several methodological challenges. Investigating potential solutions to these challenges will be a central focus of this internship.

Objectives

1. **Literature review.** The intern will first become familiar with the various clustering methods through a comprehensive literature review. This review will aim to identify the assumptions, specific features, and limitations of the different approaches, as well as the methodological challenges associated with the longitudinal and multidimensional nature of the data.

2. **Mathematical feasibility and implementation.** The second part of the internship will focus on assessing the feasibility of adapting these methods to multidimensional longitudinal data, from both a mathematical and computational perspective. For example, the use of alternative distance or similarity measures may be investigated.

3. **Performance evaluation and applications.** Finally, the developed methods will first be implemented and tested on simulated data or benchmark datasets to evaluate their performance. Different scenarios will need to be considered in order to identify the conditions under which each approach is most relevant, as well as its potential limitations. In a second stage, the methods may be applied to real-world data from the EFEMERIS cohort (Evaluation of Medication Use and Associated Risks in Pregnant Women – http://www.efemeris.fr/), focusing on maternal exposure to psychotropic medications during pregnancy. This application is part of a collaboration between the Institut de Mathématiques de Toulouse (IMT) and the INSERM pharmacoepidemiology research group (CERPOP, SPHERE team).

All work will be implemented using well-documented, reusable R code.

Internship Details

- Supervisors: Cécile Chouquet and Sabine Mercier, Institut de Mathématiques de Toulouse.
- Location: Institut de Mathématiques de Toulouse, France.
- Duration: 4–6 months (preferred starting date: first quarter of 2027).
- Funding: MUMMY project, funded by CIMI. Compensation will be provided at the standard internship rate.

Candidate Profile

- Education: Master's student (M1/M2 level) in Applied Mathematics, Statistics, Artificial Intelligence, or Data Science.
- Technical skills: Solid background in statistical modeling and machine learning, strong analytical and abstract reasoning skills, and proficiency in R programming.
- Personal qualities: Rigorous and methodical approach to work, with the ability to identify, analyze, and clearly communicate challenges encountered during the project.

References

- Den Teuling, N. G. P., Pauws, S. C., & van den Heuvel, E. R. (2026). Clustering of longitudinal data: a tutorial on a variety of approaches. Statistical Analysis and Data Mining: An ASA Data Science Journal, 19(1), e70052. https://doi.org/10.1002/sam.70052
- Kohonen, T. (2013). Essentials of the self-organizing map. Neural Networks, 37, 52–65.