



Atelier statistique
Bonnes pratiques et développements récents
en modélisation non linéaire à effets mixtes et
conjointe de données longitudinales et de survie
Applications en amélioration variétale

25 au 27 novembre 2025, Saint-Rémy-lès-Chevreuse

Cet atelier pour objectif de former un public interdisciplinaire de scientifiques issus de la biologie et des statistiques à la modélisation statistique des variabilités de mesures répétées, par exemple dynamiques longitudinales, et de mesures de temps d'intérêt observées au sein d'une population et de leur donner l'opportunité de partager leurs expériences respectives.

L'atelier débutera par une introduction générale aux modèles à effets mixtes linéaires et non linéaires. La modélisation statistique des différents niveaux de variabilité sera détaillée et illustrée. Elle pourra plus largement être hybridée à un modèle mécaniste d'un phénomène d'intérêt, comme par exemple un modèle éco-physiologique. Des tests des composantes de la variance permettant d'identifier les paramètres du modèle variant au sein de la population seront présentés. Les variabilités individuelles des paramètres du modèle pourront également être expliquées via des variables descriptives de grande dimension, comme des marqueurs moléculaires, parmi lesquelles on cherchera à sélectionner les plus pertinentes pour expliquer la variabilité observée. Les modèles conjoints pour des données longitudinales et de temps d'intérêt seront présentés et illustrés par des exemples. Les méthodes seront mises en oeuvre sur deux cas pratiques pour analyser des données réelles de biologie. Chaque thème sera exploré à travers des cours suivis de sujets pratiques réalisés sous R ou sous Python. Une bonne maîtrise des notions élémentaires de statistique est requise ; il n'est en revanche pas requis de maîtriser R ou Python : les expériences pratiques seront présentées via diaporama et pourront être reproduites par les participants intéressés lors de temps d'échange libre.

Comité d'organisation :

Charlotte Baey (Université de Lille, laboratoire Painlevé)
Maud Delattre (INRAE, MaIAGE)
Estelle Kuhn (INRAE, MaIAGE)
Jean-Benoist Leger (UTC, laboratoire Heudiasyc)
Sarah Lemler (CentraleSupélec, MICS)
Céline Richard-Molard (INRAE, ECOSYS)
Renaud Rincet (INRAE, GQE Le Moulon)
Laure Sansonnet (Sorbonne Université, LPSM)

Lieu : [Domaine de Saint Paul](#), 78470 Saint-Rémy-lès-Chevreuse, Yvelines.
Accès : station RER B Saint-Rémy-lès-Chevreuse, puis 20 minutes à pied.

Frais : Les frais d'inscription s'élèvent à 150 euro. Les frais de séjour en chambre simple du 25 novembre matin au 27 novembre après-midi sont pris en charge pour les participants par le projet ANR [Stat4Plant](#). Les frais de transport sont à la charge des participants.

Le nombre de places est limité à 25 afin d'assurer un bon fonctionnement. Le processus d'inscription se fera en deux étapes : une [pré-inscription en ligne ici](#) à effectuer avant le 28 juillet 2025, suivi d'une validation par le comité d'organisation par mail avant le 31 juillet 2025. Les frais d'inscription devront être réglés avant le 15 septembre par bon de commande pour confirmer l'inscription.

Programme prévisionnel

Mardi 25 novembre 2025

09:00 - 9:30 : introduction et tour de table

09:30 - 10:30 : modélisation à effets mixtes linéaires et non linéaires, notion d'approche populationnelle (notion de mesures répétées, variabilités intra et inter individuelles, notion d'effets mixtes, d'effets fixes, d'effets aléatoires, composantes de la variance, aspect hiérarchique des modèles à effets mixtes).

10:30 - 11:00 : pause

11:00 - 12:30 : estimation et prédiction individuelle dans les modèles mixtes, applications numériques en simulation et sur données réelles (estimation par maximum de vraisemblance, information de Fisher, algorithmes d'estimation EM et gradient stochastique, région de confiance, prédiction)

12:30 - 13:30 : déjeuner

14:00 - 15:30 : test des effets fixes et des composantes de la variance des effets aléatoires, utilisation du package VarTestNlme (notion d'hypothèses nulle et alternative, problématiques des tests de variances nulles et des paramètres de nuisances)

15:30 - 16:00 : pause

16.00 - 17.30 : cas pratique d’une approche populationnelle à partir du modèle mécaniste ARNICA et d’un modèle à effets mixtes : ARNICA est un modèle de plante entière mécaniste qui simule les flux de carbone et d’azote entre compartiment aérien et racinaire chez *Arabidopsis thaliana* et prédit dynamiquement les biomasses aérienne et racinaire, la surface foliaire ainsi que l’absorption et le stockage d’azote durant la phase végétative. L’objectif est d’identifier sur quels paramètres du modèle repose la variabilité génétique de la réponse de cette espèce à la disponibilité en azote chez une quarantaine d’écotypes soumis à deux niveaux d’azote en développant une méthode d’ajustement globale des paramètres. Cette approche permet aussi de réduire la charge expérimentale dans les études d’interaction GxN utilisant un modèle mécaniste.

17.30-18.30 temps d’échange libre avant dîner

Mercredi 26 novembre 2025

09:00 - 10:30 : modélisation à effets mixtes intégrant des variables de grande dimension, sélection de variables par LASSO et bayésienne (variables descriptives de grande dimension, cas des marqueurs moléculaires, problématique en estimation, méthode régularisée, choix du paramètre de régularisation, intérêt de l’approche populationnelle pour la sélection de variables)

10:30 - 11:00 : pause

11:00 - 12:30 : cas pratique de sélection de variables dans le modèle mixte en grande dimension sur données réelles de dynamique de senescence : l’objectif est d’identifier dans une population de blé tendre des marqueurs moléculaires associés à la variabilité génétique de la dynamique de senescence. L’innovation méthodologique consiste à faire la sélection de marqueurs et la modélisation de la dynamique non-linéaire dans un même modèle statistique pour réduire la perte d’information et ainsi obtenir une meilleure puissance de détection. Les méthodes proposées sont directement applicables pour tout processus non-linéaire (modèle de culture, dynamique de croissance, etc) avec sélection de variables en grande dimensions (SNPs et autres omiques).

12:30 -13.30 : déjeuner

14:00 – 15.30 : bonnes pratiques de code et de reproductibilité (présentation d’un ensemble de pratiques et de méthodes permettant le maintien en fonctionnement du code, sa reproduction, son évolution, sa reprise dans d'autres activités et la capitalisation sur celui-ci)

15.30 - 16:00 : pause

16.00 - 17.30 : introduction à la différentiation automatique (présentation de la méthode qui permet l'évaluation exacte des dérivées d’une fonction non linéaire complexe sans l'écriture de sa forme analytique. Cette méthode de calcul est très rapide et constitue une brique fondamentale des méthodes de Machine Learning modernes. Elle est utilisée dans le code du cas pratique du modèle mécaniste ARNICA).

17.30-18.30 temps d’échange libre avant dîner

Jeudi 27 novembre 2025

09:00 - 10:30 : modélisation jointe de données longitudinales et de données de survie en grande dimension (introduction aux données de survie, notions de fonction de survie, de risque instantané, modèle de Cox, modélisation jointe causale via une fonction de lien combinant un modèle à effets mixtes et un modèle de survie)

10:30 - 11:00 : pause

11:00 - 12:30 : application numérique sur l’estimation et la sélection de variables dans le modèle joint (estimation des paramètres du modèle joint causal, intégration de variables descriptives de grande dimension, sélection de variables)

12:30 -13.30 : déjeuner

14:00 - 15.00 : prédiction dynamique dans le modèle joint (utilisation du modèle joint pour obtenir des prédictions dynamiques individuelles sous forme d’intervalle de confiance uniforme)

15:00 - 15:30 conclusion et fin de l’atelier