

## Proposition de stage de M2

### Méthodes de statistique en grande dimension pour l'étude de cinétiques de données transcriptomiques

#### Description

Ce stage se déroulera à AgroParisTech au sein de l'équipe Statistique & Génome de l'UMR MIA-Paris en relation avec l'Institut des plantes de Paris Saclay (IPS2).

Il s'agit d'un travail de modélisation statistique, à la fois théorique et pratique, qui est motivé par l'étude de données d'expression de gènes de type RNA-Seq sur la plante modèle *Arabidopsis Thaliana*. La technologie du RNA-Seq est une méthode d'analyse du transcriptome permettant de quantifier les ARN d'un tissu à un instant donné. Grâce à ce type de données, on peut déterminer le degré de transcription d'un gène et donc étudier la variation de cette intensité au cours du temps ou en réponse à des stimuli externes.

L'objectif du stage est de i) proposer un modèle adapté à la nature particulière des données considérées et aux questions biologiques posées, ii) développer une méthode d'inférence et en étudier les propriétés statistiques, iii) réaliser des simulations pour valider les résultats, et iv) l'appliquer à des données réelles.

Les particularités des données considérées sont leur nature discrète et temporelle, ainsi que leur taille conséquente. Les biologistes souhaitent savoir quels gènes dépendent les uns des autres, à quels instants et sous quelles conditions. Ce sont ces aspects de comptage, de temporalité et de sélection de variables que le stagiaire devra combiner dans l'élaboration d'un modèle innovant pour l'étude des données de type RNA-Seq.

Il est à noter d'une part qu'aucune connaissance en sciences du vivant n'est exigée dans ce stage et d'autre part que celui-ci pourra éventuellement déboucher sur une thèse.

#### Laboratoire d'accueil

UMR MIA-Paris, INRA, AgroParisTech, Université Paris-Saclay  
16, Rue Claude Bernard, 75005 Paris

Site web du laboratoire : [www6.inra.fr/mia-paris](http://www6.inra.fr/mia-paris)

L'UMR MIA-Paris est une unité de recherche en statistique et informatique orientée vers les applications en sciences du vivant. Son objectif essentiel est de développer et diffuser des méthodes d'inférence statistique (modèles complexes, modèles à variables latentes, inférence bayésienne, apprentissage, sélection de modèles, ...) et algorithmiques. Les activités de l'unité s'appuient sur une bonne culture dans les disciplines destinatrices : écologie, environnement, biologie moléculaire et biologie des systèmes.

#### Encadrants :

Céline Lévy-Leduc : [celine.levy-leduc@agroparistech.fr](mailto:celine.levy-leduc@agroparistech.fr)

Sarah Ouadah : [sarah.ouadah@agroparistech.fr](mailto:sarah.ouadah@agroparistech.fr)

Laure Sansonnet : [laure.sansonnet@agroparistech.fr](mailto:laure.sansonnet@agroparistech.fr)

Thomas Blein : [thomas.blein@ips2.universite-paris-saclay.fr](mailto:thomas.blein@ips2.universite-paris-saclay.fr)

**Durée :** 6 mois.

**Gratification :** 554,40 euros par mois.

### Références

**Dunsmuir, W. T. and Scott, D. J.** The glarma package for observation driven time series regression of counts. *Journal of Statistical Software*, Volume 67, Issue 7, p. 1–36, 2015.

**Weiss, C. H.** An Introduction to Discrete-valued Time Series. John Wiley & Sons, 2018.