

# Post-doctorat de 18 mois :

## Adaptation de méthodes d'analyse d'image par transformée en ondelettes la comparaison de séquences ADN, pour la cartographie du risque et la classification de données satellites

**Contexte et objectifs du projet post-doctoral :** L'objectif est d'étudier l'adaptation de méthodes d'analyse d'images par transformée en ondelettes pour différentes applications en épidémiologie et en génomique. Les ondelettes font partie des méthodes usuelles en analyses d'images. Nous souhaitons les utiliser pour des analyses qui s'apparentent à de l'analyse d'image, mais ne sont généralement pas assimilées à de l'analyse d'images, plus précisément en génomique et en cartographie du risque. Le post-doctorant travaillera tout d'abord dans le contexte de la génomique. Il s'intéressera à la recherche de similarité entre des séquences ADN stockées au sein de larges banques de données, afin par exemple de déterminer l'espèce dont provient une séquence d'origine inconnue. Les séquences ADN seront enregistrées sous forme d'images en codant chaque nucléotide (c'est-à-dire chaque base A, T, C ou G) en niveaux de gris. Puis des techniques d'analyse d'images seront appliquées pour l'étude de ces séquences. La recherche de similarités s'appuie sur des méthodes de hachage perceptuel. Il s'agit de produire un "résumé" appelé clé de hachage de chaque image. Ces clés de hachage réduisent significativement les données à comparer et ont pour objectif d'accélérer la recherche de séquences similaires. Une distance faible entre deux clés indique un fort taux de similarité entre deux images, c'est-à-dire dans notre cas deux séquences ADN. Ceci permet d'effectuer un premier tri rapide dans de grandes bases de données et de sélectionner rapidement un ensemble de séquences probablement proches qui seront ensuite à nouveau testées par des méthodes de comparaison plus exactes, mais aussi plus lentes. La construction de ces clés de hachage est possible par différentes techniques. Au cours de ce post-doc une transformée en ondelettes sera utilisée. Le post-doctorant sera chargé des adaptations d'un point de vue mathématique, puis apportera une expertise pour l'implémentation.

Dans un deuxième temps, le post-doctorant travaillera en cartographie du risque. Les cartes de risque sont des outils largement utilisés pour l'étude des maladies non contagieuses humaines (par exemple en cancérologie). Elles permettent l'étude de la répartition spatiale, et plus récemment l'évolution spatio-temporelle, de ces maladies, ainsi que la mise en évidence de facteurs épidémiologiques. Elles ont été récemment introduites en épidémiologie animale, notamment par les travaux de l'unité d'Epidémiologie Animale. Dans ce contexte, la cartographie du risque a montré un potentiel intéressant en termes de compréhension de la dynamique des épidémies et d'aide à la décision. Elle peut être un support pour la mise en œuvre de mesures de protections spatialement délimitées comme la vaccination, la restriction de circulation ou abattage du cheptel. Le post-doctorant étudiera la possibilité d'adapter les méthodes d'analyse d'images basées sur des transformées en ondelettes à la cartographie du risque. En effet, la production d'une carte de risque lissée à partir de données de cas et de population par unités géographiques correspond au lissage d'une image bruitée. Les méthodes classiques d'analyse d'images, en particulier les transformées en ondelettes, pourront être adaptées à ce contexte particulier. Des indicateurs comme l'intensité et l'énergie

d'une transformée en ondelettes continue pourront être utiles pour capturer et représenter les niveaux de risque dans différentes régions. Le post-doctorant pourra en particulier se concentrer sur l'aspect spatio-temporel, ainsi que sur la modélisation de maladie contagieuses, deux points qui intéressent particulièrement l'unité d'Epidémiologie Animale.

Enfin, de nombreuses variables explicatives peuvent être intégrées dans les modèles épidémiologiques. Celles-ci peuvent être collectées sur le terrain en même temps que les données. Mais il arrive que des photos aériennes ou satellites soient utilisées pour obtenir des données spatiales sur un vaste territoire. Elles permettent par exemple de produire des informations sur le type de végétation ou la présence d'eau. Une classification doit être effectuée sur ces photos afin de déterminer les zones de présence des catégories qui nous intéressent. Le post-doctorant sera donc amené à travailler sur l'utilisation de transformées en ondelettes pour l'analyse d'images satellite et en particulier la classification.

Bien que les domaines d'application soient différents, les trois parties de ce post-doctorat (analyse d'images satellite, cartes de risque et comparaison de séquences ADN) présentent les mêmes questionnements généraux. Il s'agit d'extraire l'information essentielle des données, d'analyser et/ou de classer différents "motifs" ou "signatures", et enfin de quantifier l'intensité de certains phénomènes. Nous sommes donc dans un contexte général d'analyse et reconnaissance de motifs (pattern analysis and recognition). Concernant plus précisément la méthodologie mathématique, le post-doctorant évaluera l'apport potentiel des ondelettes (transformées discrètes ou continues), par rapport aux outils statistiques traditionnellement utilisés dans ces contextes. Il pourra également envisager d'autres approches mathématiques alternatives et/ou complémentaires.

**Étapes du projet :** Le post-doctorant débutera son étude par la comparaison de séquences ADN qui est une suite de travaux réalisés durant une thèse en cours dans l'unité. En parallèle, il travaillera sur l'analyse de photos aériennes ou satellites en vue de déterminer des variables écologiques à entrer dans les modèles épidémiologiques. Dans les deux cas, de précédents travaux serviront de base à son travail. Nous envisageons de consacrer 9 à 12 mois à cette partie du projet post-doctoral. Dans un second temps, le post-doctorant travaillera sur la cartographie du risque qui sera plus exploratoire.

**Compétences recherchées :** Docteur en statistiques avec une expérience sur les transformées en ondelettes. Avoir déjà travaillé sur des applications en biologie sera un plus.

**Conditions matérielles :** Rémunération selon la grille de la fonction publique en fonction de l'expérience du candidat.

Le post-doctorant sera rattaché à l'unité d'Epidémiologie Animale (Clermont-Ferrand). Il réalisera son travail dans cette unité et collaborera avec les statisticiens, informaticiens et génomiciens de l'unité. La partie génomique du projet post-doctoral étant une suite d'une thèse menée en collaboration avec le LIMOS, le post-doctorant aura des contacts dans ce laboratoire d'informatique. Le post-doctorant pourra s'intégrer dans les collaborations existantes avec le laboratoire de Mathématiques de Clermont-Ferrand ou en créer de nouvelles. Il participera au groupe de travail de statistique commun à l'unité d'Epidémiologie Animale et au laboratoire de Mathématiques de Clermont Ferrand ainsi qu'au séminaire de mathématiques appliquées organisé par ce laboratoire auquel il sera également amené à présenter ses travaux.

**Contacts :** Vos candidatures (CV + lettre de motivation) sont à faire parvenir à Myriam CHARRAS-GARRIDO (myriam.charras-garrido@clermont.inra.fr).