

Analyses statistiques multivariées de données géochimiques : Discrimination des sources de matières colorantes préhistoriques

Responsable du stage : Amandine BLIN

Laboratoire d'accueil : UAR 2700 2AD – Acquisition et Analyse de Données pour l'Histoire naturelle. Muséum National d'Histoire Naturelle - CNRS

Durée du stage : 4 à 6 mois selon le profil

Profil recherché : Etudiant de Master 2, école d'ingénieur ou équivalent (BAC+5)

Année universitaire : 2024-2025

Date de début souhaitée : à partir du 06/01/2025

Date limite pour postuler : 07/10/2024

Compétences requises :

L'étudiant-e devra avoir de solides compétences en statistiques et en analyse de données, en particulier en statistique multivariée. Une maîtrise du langage de programmation R est requise. Il/elle devra être inscrit-e à une formation de statistiques/chimométrie au niveau Master ou école d'ingénieur dans un établissement d'enseignement supérieur, ou avoir une expérience notable dans le domaine. Des compétences dans au moins un des domaines suivants seraient appréciées : géochimie/biochimie, analyses physico-chimiques, archéométrie. Des connaissances en apprentissage automatique seront un plus. Le/la candidat-e devra être intéressé-e par la recherche, disposé-e à apprendre de nouvelles techniques et savoir analyser et interpréter les résultats. Il/elle évoluera dans un environnement multidisciplinaire et devra être capable d'interagir avec les différents acteurs.

Contexte scientifique :

Les minerais de fer et de manganèse se retrouvent dans des contextes géologiques très variés qui ont été largement exploités aux périodes historiques, même pour les plus petits d'entre eux. Mais les premières traces d'utilisation de sources de minerais de fer et de manganèse se retrouvent dès la préhistoire ; de nombreux sites préhistoriques ont ainsi livré des matériaux jaunes, rouges (hydroxydes et oxydes de fer) ou noirs (oxydes de manganèses). Ces matériaux ont été utilisés comme pigments pour la réalisation d'œuvres pariétales au paléolithique supérieur. Ils ont aussi été exploités dès le paléolithique moyen et ont été impliqués dans des activités aussi variées que le tannage des peaux, la protection de la peau contre des insectes ou les UV du soleil, ou encore la confection de colles et résines permettant d'emmancher des lames de silex sur des supports. Afin de comprendre le rôle de ces matériaux dans les activités techniques, il est nécessaire d'en déterminer la nature et d'identifier leur origine géologique. Pour cela, leur nature pétrographique et leur composition minéralogique et élémentaire sont analysées puis comparées à des sources géologiques collectées à des échelles régionales voire nationales. Cela nécessite l'utilisation de nombreuses techniques d'analyses fournissant des données de nature variées, ou présentant des sensibilités et précisions différentes. Devant la complexité et l'hétérogénéité de ces données, il est nécessaire de recourir à des analyses statistiques multivariées permettant de discriminer de manière robuste les sources géologiques de natures différentes et d'établir les correspondances entre sources géologiques et matériel archéologique.

Objectifs du stage :

Les objectifs de ce stage sont d'évaluer la pertinence et la robustesse des différents protocoles de pré-traitement des données élémentaires, combinés à différentes méthodes d'analyses multivariées et d'apprentissage automatique afin de déterminer les plus performantes. Plusieurs jeux de données géochimiques seront utilisés pour appliquer différentes méthodes de transformation des données (transformation additive, transformation centrée, transformation isométrique) en pré-traitement propres aux données de composition. Puis différentes analyses multivariées exploratoires (analyse en composantes principales (ACP)) ou encore des méthodes d'apprentissage automatique supervisées (analyses discriminantes linéaires (LDA), PLS-DA (Partial Least-Squares Discriminant Analysis), Support Vector Machine (SVM), Random Forest (RF)...) seront appliquées à chacune de ces transformations. Chacun des modèles sera évalué par l'utilisation de différentes mesures de performance. Afin de limiter les biais, le risque de sur-ajustement sera évalué par les techniques de validation croisée.

Contexte de travail :

Ce stage se déroulera au sein de l'Unité d'Appui et de Recherche «2AD – Acquisition et Analyse de Données pour l'Histoire naturelle » (CNRS/Muséum National d'Histoire Naturelle, Paris) en partenariat avec l'UMR « HNHP - Histoire naturelle de l'Homme Préhistorique » (CNRS/Muséum National d'Histoire Naturelle, Paris) dans le cadre de l'ANR Color-Sources.

Durée : 4 à 6 mois selon le profil

Date de prise de fonction : à partir du 06/01/2025

Rémunération : Gratification selon la réglementation en vigueur

Pour postuler :

Veillez envoyer votre CV et votre lettre de motivation à Amandine Blin (amandine.blin@mnhn.fr) et Matthieu Lebon (matthieu.lebon@mnhn.fr) avec pour objet du mail M2_NOM_2025 avant le **7 octobre 2024**. Un entretien individuel sera programmé pour les candidat-e-s sélectionné-e-s.