

Analyses de co-évolution dans les données paléoenvironnementales pour la reconstruction des dynamiques environnementales passées

direction: Lionel Tabourier (LIP6 - SU/CNRS)
co-encadrement: Jerry Lonlac-Konlac (CERI/SN - IMT Nord-Europe)
et Yannick Miras (IPH - MNHN/UPVD/CNRS)

contact: lionel.tabourier@lip6.fr

1 Contexte

Ce projet est situé à l'intersection de l'informatique et de la paléoécologie, c'est-à-dire la science qui étudie les environnements passés par les traces que laissent les espèces vivantes sous forme de fossiles. Un pan entier de ce domaine se fonde sur l'analyse des grains de pollens et des microfossiles non-polliniques fongiques, qui sont extraits par carottage des sédiments de milieux spécifiques (lacs, tourbières). L'énumération et la datation des espèces, ou plus exactement des taxons, présents permet d'identifier les changements environnementaux et climatiques passés, ainsi que l'impact des sociétés passées et leur adaptabilité à ces changements.

L'utilisation du dénombrement de motifs de co-évolution a été proposée en palynologie dans le cadre du projet Mobipaléo (2015-2018) dans le but d'améliorer la reconstruction des dynamiques des écosystèmes passés, ainsi que de discerner des signaux faibles traduisant des changements environnementaux dans les données palynologiques. Le principe en est de repérer les co-évolutions de différents indicateurs polliniques au cours du temps (sur plusieurs millénaires) qui caractérisent de modifications de l'environnement végétal passé [LMB⁺18] (cf. Figure 1). De ce projet est né la plateforme <https://mobipaleo.limos.fr/> gérée par le LIMOS (Université Clermont-Auvergne) à destination des palynologues.

Cette méthode a permis de mettre en évidence certains taxons comme possibles taxons d'intérêt pour identifier des traces d'anthropisation, c'est-à-dire de modification de l'environnement sous l'effet de l'activité humaine (pastoralisme, agriculture) [MKB⁺21]. Malgré ces premiers résultats prometteurs, celle-ci n'a été jusqu'à présent éprouvée que dans des contextes encore relativement restreints et nécessite des approfondissements méthodologiques et techniques pour pouvoir être diffusée plus largement dans la communauté paléoécologique.

2 Résumé du projet

L'utilisation du dénombrement de motifs de co-évolution a été proposé en palynologie pour **améliorer la reconstruction des dynamiques passées des écosystèmes** en permettant de caractériser la variabilité naturelle de la végétation, l'intensité, la vitesse et la récurrence des changements et des types de réponses aux forçages naturels (fluctuations climatiques) et anthropiques (e.g., pastoralisme). L'ambition est d'**identifier les signaux d'alerte aux changements de la biodiversité floristique** au cours du temps [LMB⁺18]. Cette méthode a permis de mettre en évidence certains taxons comme possibles taxons d'intérêt pour identifier des traces d'anthropisation (pastoralisme, agriculture) [MKB⁺21]. L'objet de ce projet de thèse est de **rendre la méthode opérationnelle et robuste à grande échelle** et de **démontrer son intérêt dans une variété de contextes palynologiques**. Pour ce faire, on propose des tests de significativité réalisés par comparaison des mesures sur les données polliniques à des modèles aléatoires

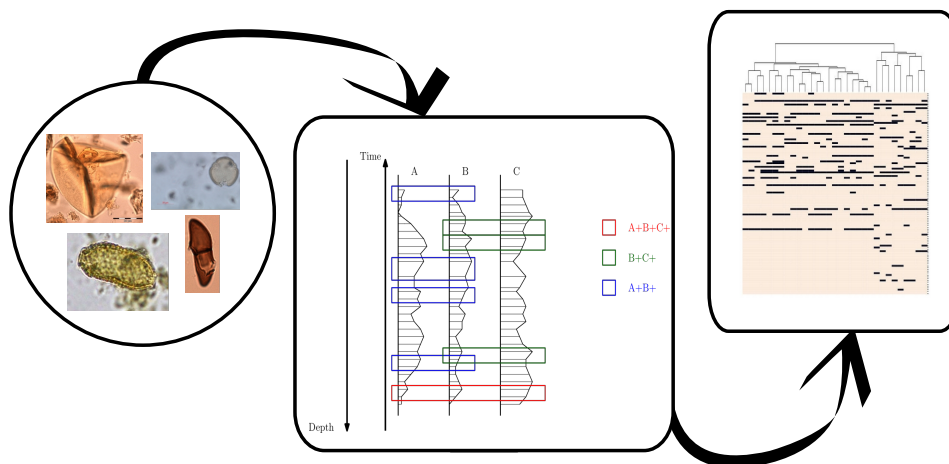


FIGURE 1 – Résumé schématique de la méthode Mobipaléo : les palynologues dénombrent les grains de pollens et microfossiles fongiles dans les sédiments, l’outil dénombre les motifs de co-évolution des taxons (espèces) et fait apparaître des groupes de motifs potentiellement significatifs pour la reconstitution des environnements passés.

réalistes [GG96]. On doit également démontrer la robustesse de la méthode au bruit, en évaluant l’effet des petites fluctuations sur les motifs détectés [CBLT⁺23].

3 Description détaillée du projet de thèse

L’objet de ce projet de thèse est de rendre la méthode Mobipaléo opérationnelle et robuste à grande échelle et de démontrer son intérêt dans une variété de contextes palynologiques. Pour ce faire, nous proposons un projet de thèse en trois volets.

Le premier est **méthodologique** : les méthodes statistiques développées demandent à être consolidées théoriquement par la proposition de tests de significativité et de robustesse. De premières propositions ont été faites dans ce sens : la significativité des motifs est évaluée par comparaison des mesures sur les données polliniques à des modèles aléatoires réalistes [GG96]. Les mesures exploratoires réalisées par Lionel Tabourier indiquent que ce travail est nécessaire à une bonne interprétation des résultats produits par la méthode. Par ailleurs, elle doit également être prouvée robuste en évaluant la variabilité de ces résultats au bruit des données (incertitudes de mesures liées au comptage et à la datation des échantillons). De premières propositions ont été faites pour prendre en compte l’effet des petites fluctuations sur les motifs détectés [CBLT⁺23], celles-ci doivent maintenant être mises en œuvre dans un cadre théorique complet. La première étape du travail de thèse consiste à systématiser la méthodologie, en sélectionnant les modèles de référence et à les tester empiriquement sur les données palynologiques, grâce à l’expertise de Yannick Miras et son équipe, qui peuvent évaluer la pertinence des motifs détectés par leur connaissance des états de référence pré- et post-changement.

Le second volet est **algorithmique** et suit chronologiquement le premier : les tests proposés doivent être conçus de façon à traiter en temps réel de grands volumes de données. En effet, l’identification des motifs de co-évolution dans les données palynologiques est confrontée à une explosion combinatoire. Pour que les méthodes soient utilisables par les praticiens, il est nécessaire de développer des méthodes de test efficaces pour pouvoir évaluer la significativité d’un motif « à la volée ». Ce volet comporte aussi un travail de développement dont le but est de mettre à disposition une nouvelle application mettant en œuvre les tests de robustesse, qui soit utilisable

par la communauté des paléoenvironmentalistes et bioarchéologues. Jerry Lonlac-Konlac, qui était à l'initiative de la méthode et le principal développeur de la plateforme initiale fait partie de l'équipe d'encadrement et pourra apporter une expertise essentielle au déroulement de cet aspect du projet.

Le troisième est **applicatif** : au cours des deux premiers volets, on fera usage des données paléocéologiques disponibles grâce à différentes collaborations impliquant l'unité de Yannick Miras (HNHP). Trois contextes biogéographiques et chronologiques distincts ont été identifiés comme des cadres privilégiés et complémentaires pour l'évaluation de la méthodologie, en raison de la richesse des données disponibles et des expertises établies dans ces domaines. La mise en place de démarches comparatives dans ces différents contextes complémentaires permettront ainsi de tester l'aptitude de la méthode à analyser, à différentes échelles spatiales et temporelles, les similitudes et les variabilités des réponses de la végétation à différents aléas environnementaux (notamment des fluctuations climatiques d'intensités très différentes) et à des typologies variées d'activités humaines. Ces zones sont :

- la région Auvergne (biome tempéré, carottes continentales lacustres et tourbeuses, période holocène soit les 11,7 derniers millénaires), dont Yannick Miras est expert,
- la Californie, l'Espagne et la Tunisie (biomes méditerranéens à semi-arides, carottes continentales et marines, période holocène), qui relève également du champ de compétence de Yannick Miras, notamment par sa participation au projet ANR MeSCAL (dir. A. Ejarque, ISEM, CNRS),
- les carottes marines de l'Ocean Drilling Project (principalement en mer méditerranée, période pléistocène soient le dernier million d'années), projet dans lequel est fortement impliquée HNHP (deux thèses et un post-doctorat en cours).

Ces données pourront potentiellement être enrichies grâce aux bases de données internationales récemment mises en place dans les domaines (par exemple, European Pollen Database, Neotoma). En démontrant la pertinence de l'approche dans une variété de contextes, nous serons en mesure d'apporter des arguments probants en faveur de la reconnaissance de cette approche comme méthodologie de référence dans le domaine.

4 Verrous méthodologiques

Les tests de significativité et de robustesse nécessaires doivent être mis en œuvre dans un cadre théorique complet. Un premier verrou de la thèse consiste alors à systématiser la méthodologie, en sélectionnant les modèles de référence et à les tester empiriquement sur les données palynologiques pour vérifier leur pertinence, ceci sera fait dans une variété de contextes géographiques et chronologiques identifiés comme des cadres privilégiés pour l'évaluation de la méthodologie. Par ailleurs, les tests proposés doivent être conçus de façon à traiter en temps réel de grands volumes de données, il est nécessaire de développer des algorithmes efficaces pour pouvoir évaluer la significativité d'un motif « à la volée », ce qui constitue le second verrou de la thèse, celui-ci est aussi associé à un travail de développement pour la réalisation d'une application utilisable par les palynologues.

5 Équipe encadrante et profil recherché

Les approches méthodologiques décrites précédemment relèvent de l'informatique, et la réalisation d'algorithmes efficaces pour les mettre en œuvre dans une application demande une expertise propre à ce domaine. Cependant, la confrontation aux données expérimentales et l'évaluation des approches mises en œuvre ne peuvent être faites qu'en collaboration avec des spécialistes de la palynologie, qui peuvent apprécier la pertinence des motifs détectés par leur connaissance des états de référence pré- et post-changement. Les contextes géographiques de travail ont été choisis en raison de la richesse des données disponibles et des expertises de HNHP et de ses collaborateurs sur ces contextes.

La thèse serait encadrée par Lionel Tabourier, Professeur d'Informatique au LIP6. Une partie importante de son travail de recherche consiste à développer des outils de fouilles de données pour

des questions de recherche interdisciplinaires (principalement pour l'analyse de réseaux sociaux). Le projet de thèse a été défini au cours de sa délégation CNRS au sein de l'unité Histoire Naturelle et Humanités Préhistoriques (HNHP), au premier semestre 2025. Yannick Miras est Ingénieur de Recherche CNRS (HDR) à HNHP, spécialiste de palynologie, il est un des initiateurs du premier projet Mobipaléo et apportera à l'équipe son expertise de la reconstitution des paléoenvironnements par l'analyse des pollens et fossiles micropolliniques. Enfin, Jerry Lonlac-Konlac, Maître de Conférences en informatique au CERI-SN (IMT Nord-Europe), est le principal architecte de la première plateforme Mobipaléo et est donc expert de la conception et de la mise en œuvre du dénombrement de motifs de co-évolution pour la palynologie.

Nous recherchons des étudiants dont les principaux centres d'intérêt sont l'algorithmique et la programmation pour la fouille de données. Il est également essentiel d'avoir un goût pour les questions interdisciplinaires.

Références

- [CBLT⁺23] Michael Chirmeni Boujike, Jerry Lonlac, Norbert Tsopze, Engelbert Mephu Nguifo, and Laure Pauline Fotso. Grapgt : Gradual patterns with gradualness threshold. *International Journal of General Systems*, 52(5) :525–545, 2023.
- [GG96] Nicholas J Gotelli and Gary R Graves. *Null models in ecology*. Smithsonian Institution Press, 1996.
- [LMB⁺18] Jerry Lonlac, Yannick Miras, Aude Beauger, Vincent Mazonod, Jean-Luc Peiry, and Engelbert Mephu Nguifo. An approach for extracting frequent (closed) gradual patterns under temporal constraint. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2018.
- [MKB⁺21] Yannick Miras, Jerry Garvin Konlac, Aude Beauger, Benjamin Legrand, Delphine Latour, Paul M Ledger, Vincent Mazonod, Engelbert Mephu-Nguifo, et al. Tracking plant, fungal and algal diversity through a data mining approach : towards an improved analysis of holocene lake aydat (puy-de-dôme, france) dynamics and ecological legacies./trajectoires des biodiversités végétale, fongique et algale par une approche de fouilles de données : vers une meilleure caractérisation de la dynamique holocène du lac d'aydat (puy-de-dôme, france) et des héritages écologiques. *Revue des sciences naturelles d'Auvergne*, 85(1) :83–104, 2021.