

ATELIER BIODIVERSITE OSUG : LUNDI 27 NOVEMBRE Campus Universitaire Gières, salle de conférences André Rassat, Bat E de Chimie, 470 rue de la Chimie 38400 Saint-Martin-d'Hères

Les Bases de données en Biodiversité à l'ère du big data : exemples de projets de recherche et nouveaux outils d'analyses (IA...)

Accueil - 9h30

Introduction de la journée : Bases de données en biodiversité – quels sont les enjeux ? Comment compiler les différentes sources d'information ? Comment les utiliser pour mieux évaluer la dynamique de la biodiversité (méta-analyses, monitoring temporel / spatial, intelligence artificielle) ?

9h45 - 10h15

1/ Présentation du PNDB : Olivier Norvez

titre :

Le Pôle National de Données de Biodiversité, une infrastructure de recherche et centre de référence thématique pour et par les communautés.

résumé :

Le Pôle National de Données de Biodiversité - [PNDB](#) - est une infrastructure de recherche numérique donnant accès aux données de biodiversité, de la molécule aux anthroposystèmes et sur le temps long, ainsi qu'à un certain nombre d'outils de traitement des données et autres services à la recherche. Le PNDB joue le rôle de centre de référence thématique national pour accompagner et faciliter la compréhension, le partage, et l'utilisation des {méta}données de biodiversité. Le PNDB est le fruit d'un partenariat entre 18 établissements partenaires avec le soutien de la Fondation pour la Recherche sur la Biodiversité (FRB), et dont la coordination technique est assurée par le Muséum national d'Histoire naturelle via PatriNat, afin d'assurer une complémentarité d'outils et de services entre les différents systèmes d'informations sur la biodiversité (SINP/INPN, SIB, point nodal français GBIF).

2/ Outils d'analyse des données : méta-analyse, intelligence artificielle

10h15-11h

Nicolas Le Guillaume, LECA

Titre : Intégration de données et graphe de connaissance pour la reconstruction de réseaux multitrophiques

Résumé

Lancé en 2016, l'observatoire ORCHAMP est un outil de suivi de la dynamique spatio-temporelle de la biodiversité au sein d'un réseau de gradients d'altitude répartis sur l'ensemble de la partie française des massifs alpin et pyrénéen. Les données collectées sont notamment utilisées pour créer des modèles prédictifs de la biodiversité en milieu alpin. Dans cet exposé, nous nous intéresserons à la modélisation prédictive de la biodiversité du sol à l'échelle des alpes françaises. Le sol est un écosystème multi-trophique complexe. La modélisation de sa biodiversité nécessite de prendre en compte les interactions entre une grande diversité d'organismes qui couvre la majeure partie de l'arbre de la vie, organismes dont la taille varie de plus de six ordres de grandeur et qui présentent un large éventail de préférences en termes de (micro)habitats et de ressources alimentaires. Au cours des dernières années, l'accumulation d'informations sur les habitudes alimentaires des organismes du sol s'est accélérée. Néanmoins, le manque de standardisation entraîne une dispersion de ces informations au sein de multiples bases de données au format et à la structure hétérogènes, ce qui les rend difficilement exploitables pour la reconstruction de réseaux multitrophiques. Nous évoquerons notre expérience en matière d'intégration de données et de construction semi-automatique d'un graphe de connaissances sur l'écologie trophique des organismes du sol. Ce graphe de connaissance fournit un point d'accès unique à une multitude de bases de données harmonisées et consolidées, ce qui facilite la reconstruction de réseaux multitrophiques qui peuvent ensuite être intégrés au sein de modèles prédictifs de la biodiversité.

11h-11h45

Sara Si Moussi et Yuqing Hu, LECA

Titre : Contribution de l'IA à la modélisation spatiale de la biodiversité

Vue d'ensemble sur l'utilisation de l'IA dans les modèles prédictifs de la biodiversité :

- 1) Utilisation de l'IA pour intégrer des données de paysage complexes dans les modèles prédictifs de biodiversité. Exemple: imagerie aérienne et satellite.
 - Deux cas d'étude rapides: Habitats (échelle Européenne, EUNIS), Végétation (échelle des alpes, CBNA)
- 2) Un des aspects qui alimentent la réticence à l'usage de l'IA est l'aspect boîte noire des algos d'IA. Les outils d'Interpretable machine learning (IML) présentent une solution potentielle pour extraire des connaissances à partir de modèles opaques comme les réseaux de neurones ou les forêts aléatoires.
 - Cas d'étude : CBNA.
- 3) Intégrer des connaissances expertes (réseaux d'interaction) dans les modèles d'apprentissage automatique de la distribution des espèces (SDM). Cette partie fait le lien entre graphes de connaissances et modèles de distribution.
 - Cas d'étude : données ORCHAMP.

11h45-12h15

**3/ Les bases de données de biodiversité en forêt
Laurent Bergès, INRAE LESSEM**

Titre: Importance de la stratégie d'échantillonnage et de la répartition spatiale des observations pour analyser les patrons de distribution de la biodiversité forestière. Le cas de l'Inventaire forestier national.

Résumé

Le dispositif national de suivi des ressources forestières de l'Inventaire forestier national a deux principaux avantages par rapport à la plupart des autres bases de données de biodiversité : elles suivent un protocole standardisé (flore) et sont bien réparties sur l'ensemble du territoire national. Nous illustrons par deux exemples l'utilisation de cette base de données pour aborder des questions d'écologie du paysage à l'échelle de plusieurs régions françaises. Nous montrons que les patrons de réponse de la biodiversité à la fragmentation forestière n'auraient jamais pu être mis en évidence sans travailler à une large échelle spatiale (répétitions de paysages différents), et sans disposer d'un échantillon de taille conséquente et bien réparti spatialement. D'autres travaux ont mis en évidence que les jeux de données de biodiversité forestière (SILENE, SINP) peuvent présenter une répartition spatiale fortement biaisée (diminution très rapide de la fréquence des observations en fonction de la distance à des routes ou des chemins, densité des observations très variable au sein d'un territoire). Tout en rappelant l'importance de suivre une stratégie d'échantillonnage pour bâtir les dispositifs d'inventaire et de suivis de biodiversité, nous proposons des pistes d'amélioration pour les dispositifs actuels.

12h30 - 14h

Buffet – échanges

14h-16h

5/ Sciences participatives : projets du CREA (Colin Van Reeth)

titre : Les défis du CREA Mont-Blanc pour compiler et analyser les données de biodiversité, notamment issues de sciences participatives.

résumé :

Le CREA Mont-Blanc est une association qui comme beaucoup d'organismes doit faire face à de multiples enjeux dans la gestion et le traitement de la donnée naturaliste:

- la compilation de l'information naturaliste pour un partage ouvert de la donnée (état, INPN, partenaires, etc.)
 - l'évaluation et le traitement de la qualité de données, notamment en sciences participatives
 - la comparaison des performances de différentes méthodes d'acquisition de la donnée (exemple: sciences participatives et intelligence artificielle) et identification de leur complémentarité
- Cette intervention abordera ces différents enjeux avec des cas concrets.

6/ Table ronde : les BD biodiversité à l'OSUG : quels liens / quelles spécificités par rapport aux BD nationales ?

Intervenantes : Véronique Chaffard référente d'un futur axe "gestion des données scientifiques" au sein de l'OSUG et Geneviève Michaud, nouvelle directrice technique OSUG, membre du conseil d'administration de DataCite