UM4MRM55	ANALYSES MULTIVARIÉES ET APPRENTISSAGE MACHINE POUR L'ÉCOLOGIE	
6 ECTS	*Mots clefs*	écologie numérique, analyse de données, ordination, classification, apprentissage machine
École d'été	*Responsables*	Jean-Olivier IRISSON, Stéphane GASPARINI (LOV, Villefranche)
Villefranche	*Intervenants*	Moniteur·trice ou vacataire pour les TD

Descriptif

Format de l'UE

Modalités d'enseignement

30h de cours théoriques (dont 9h de formation aux logiciels), 30h de TD dont une partie d'application des cours et une partie de projet personnel (traitement de jeux de données multivariés en autonomie).

Modalités d'évaluation

Les cours théoriques sont évalués par un examen écrit (2h). Le projet est évalué par une présentation orale (~10 min).

Résumé

Cette unité d'enseignement forme aux méthodes de traitement numérique des données multivariées et des séries temporelles en océanographie et plus généralement en sciences de l'environnement. Cette formation présente la conduite à suivre pour choisir et associer les méthodes appropriées, les bases théoriques de ces différentes méthodes et leur mise en œuvre pratique via l'utilisation de logiciels spécialisés et de la programmation sous R.

Les thèmes abordés sont : les mesures de distances pertinentes en écologie (Euclidienne, Bray-Curtis, etc.), les méthodes d'ordination en espace réduit (ACP, AFC, MDS), les méthodes de régression multiple (modèle linéaire, RDA, CCA), les méthodes de groupement non supervisé (hiérarchique, k-means), les méthodes de classification et régression supervisées (i.e. apprentissage machine : LDA, arbres de régression et de classification, RandomForest, perceptron multicouche, réseaux convolutifs).

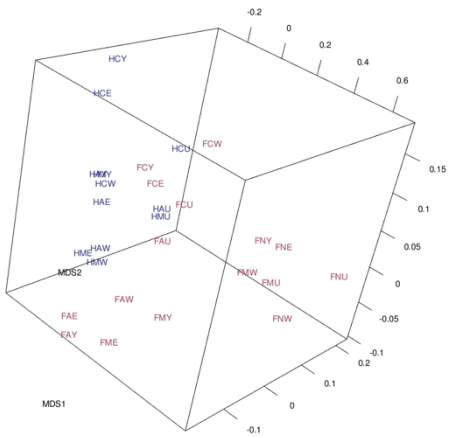

Fig: Analyse MDS, projection en 3 dimensions

Objectifs d'apprentissage

Au terme de cette UE, vous saurez :

1. choisir une méthode d'analyse appropriée à la question posée et à la nature des données;
2. mettre en œuvre des analyses multivariées et interpréter leur résultat;
3. conduire une analyse complète d'un jeu de données déjà préparé.
4. programmer des analyses sous R et utiliser des logiciels spécialisés;

Prérequis

Ce module aborde les analyses multivariées et suppose une connaissance préalable des statistiques inférentielles classiques (notion de variance, ANOVA, régression simple).

Bibliographie

Numerical ecology. Legendre, Legendre, 2012. 3dr English edition. Elsevier.
Numerical ecology with R. Borcard, Gillet, Legendre, 2011. Springer
An Introduction to Applied Multivariate Analysis with R. Everitt, Hothorn, 2011. Springer.
Analyses factorielles simples et multiples. Escoffier, Pagès, 1998. Dunod.
The elements of statistical learning. Hastie et al. 2nd Ed. 2009
An introduction to statistical learning. James et al. 2013

Fonctionnement

Les bases théoriques de chaque méthode sont détaillées sur une demi-journée de cours, en se focalisant sur les conditions d'application, le choix des méthodes et l'interprétation des résultats. Le fonctionnement mathématique des méthodes est abordé pour permettre de bien les comprendre mais sans exiger un niveau de connaissances poussé.

Ensuite, chaque méthode fait l'objet d'une séance de travaux dirigés d'une demi-journée, permettant de l'appliquer sur un ou plusieurs jeux de données qui mettent bien en évidence ses caractéristiques. Ces TDs sont tous basés sur des logiciels libres et gratuits ; certaines méthodes sont présentées avec des logiciels disposant d'une interface graphique (Tanagra et Past), d'autres via de la programmation sous R. Une formation à ces logiciels et à la programmation en R est inclue au début du module.

Enfin, les deux ou trois derniers jours sont consacrés à des projets personnels, au cours desquels les étudiants analysent un petit jeu de données fourni ou qu'ils auront amené avec eux. Chaque projet fait l'objet d'une présentation orale.

De façon pratique :

- Ce module peut, en général, être validé comme une formation doctorale, finalisée par un examen. Au minimum un certificat de participation pourra être délivré.
- Le module est ouvert aux étudiants français et étrangers de niveau Master 1, Master 2 ou en thèse, maîtrisant les prérequis mentionnés ci-dessus. Si des étudiants étrangers s'inscrivent, les cours seront donnés en Anglais.
- La station de Villefranche offre des possibilités d'hébergement et de restauration.
- Les étudiants sont attendus le samedi ou le dimanche précédant le début du module. Les cours débutent le premier jour à 9h et finissent le dernier jour à 12h.
- L'inscription se fait par la plateforme e-candidat.

NB : Ce document est indicatif. Les détails du contenu et de la forme des enseignements et des évaluations peuvent évoluer d'une année à l'autre.