

Dr. Daniel Borcard
Département de sciences biologiques
Université de Montréal
C.P. 6128, succursale Centre ville
Montréal QC H3C 3J7
Canada

Plan de cours

Cours condensé d'écologie numérique

Université Laval, Québec, février 2006

L'ensemble du cours prend 7 jours, qui comprennent des périodes de théorie couvrant les grands domaines de l'écologie numérique, des démonstrations de programmes d'ordinateur et des applications. Ces dernières se feront d'une part sur des jeux de données de démonstration communs à tous les étudiants (période d'apprentissage), et d'autre part sur des jeux de données apportés par les participants (période d'application).

Théorie

5 matinées de 9h à 12h

Matière couverte:

1. Introduction: matrices de données brutes et matrices d'association; indices d'association (similarité, distance, dépendance).
2. Groupement hiérarchique (groupements à liens, groupements moyens) et non hiérarchique (K-means).
3. Ordination en espace réduit (analyse en composantes principales, analyse factorielle des correspondances, analyse en coordonnées principales).
4. Ordination canonique (analyse de redondance RDA, analyse canonique des correspondances CCA); tests multivariés (tests sur les ordinations canoniques, test de Mantel). Partitionnement de la variation. ANOVA multivarié par le biais de la RDA.
5. Analyse spatiale des données écologiques multivariés (partitionnement en contexte spatial, coordonnées principales de matrices de voisinage).

Pratique

La pratique sur ordinateur sera axée principalement sur l'apprentissage de l'utilisation du langage statistique R (<http://cran.r-project.org/>), qui allie souplesse et universalité. Ce langage est en train de devenir le standard mondial en statistique, et est adopté par un très grand nombre d'institutions universitaires. Il est gratuit et existe pour toutes les plateformes principales (PC, Mac OSX, Linux...). Le langage R comprend des bibliothèques permettant de réaliser la quasi-totalité des analyses

démontrées au cours. Son apprentissage ne nécessite aucune connaissance préalable en programmation.

Un autre programme démontré sera CANOCO, qui est un standard en ordination simple et canonique, et permet notamment l'utilisation de méthodes sophistiquées de permutations très utiles pour analyser les résultats de plans expérimentaux.

Afin d'assurer un encadrement de qualité aux étudiants, le professeur sera secondé par un démonstrateur qualifié, connaissant bien le langage R.

Temps nécessaire: 5 après-midis + 2 jours.

Phase d'apprentissage: 5 après-midis correspondant aux 5 matinées de théorie.

1. Présentation du langage R; manipulation de données; statistiques de base en R.
2. Matrices d'association et groupement en R.
3. Ordination simple en R.
4. Ordination canonique en R; présentation du logiciel CANOCO.
5. Analyse spatiale en R.

Le 6^e jour est consacré à des présentations préparées par chaque étudiant. En 10 minutes, l'étudiant doit présenter brièvement sa problématique, ses hypothèses, ses données, et les analyses qu'il envisage de réaliser. Pour 20 étudiants, l'ensemble des présentations se ferait en 4 périodes de 50 minutes. Le reste de la journée sera consacré à la mise en place technique des analyses des données personnelles, avec discussion avec le professeur et le démonstrateur.

Le dernier jour sera consacré aux analyses de données personnelles. En fin de cours, l'étudiant devra remettre un court rapport d'analyse présentant sa problématique, ses hypothèses et ses données (une page), suivies des analyses qu'il a réalisées durant la journée (nombre de pages dépendant des analyses).

La formule proposée nécessite un travail de préparation de l'étudiant, à la fois pour la présentation (un excellent exercice) et pour les données. Des instructions lui seront fournies à l'avance au sujet du format des données. Cette option a le mérite de permettre à l'étudiant d'appliquer à ses propres données les méthodes vues au cours, sans court-circuiter les nécessaires étapes d'apprentissage des programmes.

En support au cours et aux travaux pratiques, le professeur montera une page web fournissant un document de notes de cours, les données, les scripts d'analyse en langage R, et tout autre matériel (p.ex. tirés-à-part en format pdf) jugé utile.

Dates

14, 15, 16 et 17 février 2006

22, 23 et 24 février 2006

Heures: 8h30 à 11h30 (cours théorique)
13h30 à 17h00 (exercices)