
Fédération de Recherche ECCOREV n° 3098

CNRS/Université Paul Cézanne Aix-Marseille

Europôle Méditerranéen de l'Arbois
Bâtiment du CEREGE BP 80
13545 Aix en Provence cedex 4

Direction : Joël Guiot
Tél : 04 42 97 15 32
guiot@eccorev.fr

Administration : Joëlle Cavaliéri
Tél : 04 42 97 15 21 Fax : 04 42 97 15 47
cavaliéri@eccorev.fr

Site internet : <http://www.eccorev.fr/>



**Journées d'analyse statistique des données sur R
Edition 2013
CEREGE, 2e étage porte ouest, salle 301 (géomatique)
27-28-29-30 mai 2013**

**Inscriptions : Joëlle Cavaliéri, FR ECCOREV - mel :
cavaliéri@eccorev.fr**

**Responsable : Franck Torre, IMBE, OSU Institut Pythéas
d'Aix-Marseille Université - mel : franck.torre@imbe.fr**

Ces journées se dérouleront sur ordinateur, avec le logiciel R. Le conférencier présentera les techniques en les illustrant avec des exemples reproductibles par les étudiants immédiatement sur leur ordinateur. Il y aura donc à la fois vidéo-projection et TP en parallèle.

L'inscription se fera par module. Les participants pourront sélectionner les modules qui les intéressent, ce qui permettra d'optimiser les présences.

On dispose au CEREGE de 20 ordinateurs en réseau WIFI. La salle étant relativement petite, on ne pourra accueillir qu'une **trentaine d'étudiants** par module. Le public visé sera en priorité: les thésards, post-docs et chercheurs des institutions membres d'ECCOREV, selon la place disponible. L'objectif est de rendre l'utilisateur suffisamment autonome dans les domaines les plus

« populaires » de la statistique, de manière à ce qu'il soit capable de pratiquer directement les cas standards et d'être capable de personnaliser son approche.

J 1 – Lundi 27 mai 2013 – Prise en main du logiciel R – Session 1

Packages requis : base !

J1 – Lundi 27 mai 2013 – 9h30-12h30

Introduction à R

Par Franck Torre, IMBE-AMU

Le but de cette session est de permettre à l'utilisateur novice de naviguer au sein l'environnement et d'utiliser les outils offerts par R pour l'analyse de données. Quelques applications graphiques et statistiques simples seront travaillées.

J1 – Lundi 27 mai 2013 – 13h30-16h30

Notions avancées

Par David Nérini & Séverine Martini, MIO-AMU

Ce module s'adresse à des utilisateurs de R ayant déjà pratiqués les bases du logiciel. Au travers d'exemples pris en sciences de l'environnement, nous montrerons l'utilisation de certains packages ainsi que la création de fonctions sous R. Nous montrerons les possibilités du logiciel en termes de lecture de données (lecture conditionnelle de tableau, ...) ainsi que sur l'utilisation de représentations graphiques interactives. Toutes ces notions seront abordées en utilisant des méthodes statistiques classiques (régression multiple, ACP, ...).

J 2 – Mardi 28 mai 2013 – Analyse multivariée en environnement – Session 2

Packages requis : ade4, adegenet, vegan ...

J2 – Mardi 28 mai 2013 – 9h30-12h30

Analyse multivariée

Par Franck Torre, IMBE-AMU

L'analyse de données permet de mettre en évidence l'information contenue dans un tableau de données multivariées. En fonction de la nature de ces variables, différentes méthodes ont été proposées et leur présentation est au programme de cette séance : analyse en composantes principales normées ou centrées, analyse factorielle des correspondances, analyse des correspondances multiples, analyses de données mixtes. Des exemples provenant d'échantillonnage en écologie serviront d'illustration.

J2 – Mardi 28 mai 2013 – 13h30-16h30

Analyse multivariée

Par Alex Baumel & Marine Pouget, IMBE-AMU

Dans le cadre des analyses de diversité (moléculaire, spécifique, biologique etc ...) les analyses multivariées ont l'avantage d'être économes en postulats et de pouvoir traiter des volumes de données importants. Durant cette séance nous présenterons les analyses du package adegenet visant à révéler les groupes génétiques avec ou sans a priori et nous terminerons par la SPCA, analyse multivariée et spatialisée de la diversité.

J3 matin – Mercredi 29 mai 2013 – Régression, optimisation – Session 3

J3 – Mercredi 29 mai 2013 – 9h30-12h30

Régression et optimisation

Par David Nérini & Séverine MARTINI, MIO-AMU

Dans le domaine des sciences de l'environnement, beaucoup de données peuvent être traitées comme des courbes (profil de résistivité, profils de température, évolution d'une population au cours du temps, ...). On s'intéresse dans ce module à un ensemble de méthodes d'optimisation permettant de reconstituer ces courbes à partir d'échantillons de données ponctuelles. Il s'agira d'aborder des méthodes de régression non-paramétriques comme les splines ou des méthodes de régression à noyaux mais également les méthodes nécessaires à l'ajustement d'un modèle d'équations différentielles ordinaires à des données échantillonnées. Les travaux seront illustrés à partir de données océanographiques de température dans l'Océan austral mais aussi à partir de données expérimentales de suivi de populations zooplanctoniques en chémostat.

J3 ap midi – Mercredi 29 mai 2013 – Bioindicateurs et application à la paléo – Session 4

J3 – Mercredi 29 mai 2013 – 13h30-16h30

Méthodes de calcul de bioindicateurs environnementaux, application aux assemblages (paléo-)biologiques

Par Joël Guiot, CEREGE-AMU

La recherche d'indicateurs environnementaux à partir de données biologiques implique l'utilisation de méthodes particulières. Quelques-uns des outils utiles ont été rassemblés dans le package BIOINDIC. D'autres packages sont également disponibles. La paléoclimatologie et la dendroclimatologie sont particulièrement visées.

1) méthodes de synthétisation des assemblages (ACP) et standardisation des données
2) fonctions de réponse de la végétation (de la croissance des arbres) au climat : régression linéaire, réseaux de neurones

3) fonctions de transfert permettant de reconstruire le climat (et tout paramètre environnemental) à partir d'assemblages paléoécologiques sur les continents (pollen, insectes, mollusques, diatomées) et les océans (foraminifères, coccolithes, dinocystes, diatomées). Seront abordées les méthodes de calibration (régression, PLS, WA-PLS, GAM, réseaux de neurones) et les méthodes de similarité (meilleurs analogues, surfaces de réponse).

J4 – Jeudi 30 mai 2013 – Analyse spatiale – Session 5

J4 – Jeudi 30 mai 2013 – 9h30-12h30

Analyse spatiale

Par Pascal Monestiez, BIOSP-INRA Avignon

Après une présentation rapide des types de questions et de données auxquelles s'appliquent les méthodes des statistiques spatiales (processus ponctuels, analyses sur réseaux et sur grille, géostatistique), la demi-journée sera consacrée à une introduction des concepts et méthodes de la Géostatistique au travers d'exemples et de petits programmes sous R. Visualisation et description de données spatiales. Hypothèses générales et modèles utilisés en géostatistique (utilisation de méthodes

de simulations pour visualiser le potentiel et les limites du cadre théorique). Outils d'analyse de la variabilité spatiale: variogramme expérimental, fonction de covariance spatiale, choix de modèles et ajustement (présentation autour d'exemples). Méthodes d'interpolation par Krigeage (ordinaire et universel) dans des cas simples et univariés. Influence du choix du modèle et réflexion sur les types d'échantillonnage.

J4 – Jeudi 30 mai 2013 – 13h30-16h30

Régression spatiale

Par Ghislain GENIAUX, Ecodéveloppement-INRA Avignon

Ce module traitera des régressions spatiales en présence de 1) dépendance spatiale et/ou 2) d'hétérogénéité spatiale, avec un focus sur les problèmes liés à l'estimation sur gros échantillons. 1) Estimation et tests de spécification des modèles avec dépendance spatiale (SAR, SEM, SDM et SARAR) à partir des packages spdep et sphet par maximum de vraisemblance et par la méthode des moments généralisés (GMM). La création et l'utilisation (estimation par GMM linéarisée) de matrices de voisinage spatial sur larges échantillons seront traités avec le package sield (bientôt disponible); 2) Prise en compte de l'hétérogénéité spatiale non observée et de modèles à coefficient variable spatialement à partir de modèles géo-additifs (mgcv) ou de modèles localement pondérés (sield).+ Diverses techniques d'import de données spatiales (rgdal, RPostgreSQL, ...).