

---

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@cerege.fr](mailto:guiot@cerege.fr)

Administration : Joëlle Cavaliéri  
Tél : 04 42 97 15 21 Fax : 04 42 97 15 47  
[cavaliéri@cerege.fr](mailto:cavaliéri@cerege.fr)

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

---



## Journées d'analyse statistique des données sur R

Quand ? 17, 18 et 19 mai 2010

Où ? CEREGE, 2<sup>e</sup> étage porte ouest, salle 301 (géomatique)

<p><b>Inscription : Joëlle Cavaliéri, FR ECCOREV, tel 04 42 97 15 21</b> <b>mel : <a href="mailto:cavaliéri@cerege.fr">cavaliéri@cerege.fr</a></b></p>
------------------------------------------------------------------------------------------------------------------------------------------------------------

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 en trois modules.** Ces derniers étant indépendants (à l'exception de l'introduction à R), les étudiants pourront sélectionner les modules qui les intéressent, ce qui permettra d'accueillir un nombre total d'étudiants plus élevé et ne travailler qu'avec ceux qui sont réellement intéressés par le module.

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 standard et d'être capable de personnaliser son approche.

## PROGRAMME

**Lundi 17 mai 9h-12h30 : introduction à R par J. Guiot et Y. Gally (CEREGE)**  
(éventuellement répétition l'après-midi si le nombre d'inscrits l'impose)

**MODULE 1:** Le but de cette session est de permettre à l'utilisateur novice de naviguer au sein de l'environnement et d'utiliser les outils offerts par R pour l'analyse de données. Quelques applications graphiques et statistiques simples seront travaillées.

**Mardi 18 mai 9h00-12h30: Plans d'expérience et ANOVA par F. Torre (IMEP)**

**MODULE 2a:** L'analyse de variance permet de quantifier et de tester les effets de facteurs expérimentaux dans un grand nombre de situations (facteurs contrôlés ou décrits, fixes ou aléatoires, plans équilibrés ou déséquilibrés, split-plot, split-split-plot, blocs ou mesures répétées...). L'accent sera notamment mis sur le retour aux données de base et les graphiques à réaliser quand un effet est significatif. Une comparaison entre tests post-hoc de comparaisons multiples de moyennes et techniques des contrastes sera proposée. L'ensemble de ces approches sont utiles au traitement des données expérimentales de toute nature et des exemples en écologie expérimentale seront traités plus particulièrement.

**Mardi 18 mai 13h30-17h : Couplage de tableaux par S. Dray (Labo. BBE, Univ. C. Bernard - Lyon I)**

**MODULE 2b:** Les méthodes de couplage de tableaux permettent d'étudier le lien entre deux tableaux. On présentera l'analyse de co-inertie qui permet d'étudier la structure commune à deux tableaux contenant différents descripteurs sur les mêmes individus. On présentera également les analyses multivariées explicatives type analyse de redondances (RDA) ou analyse des correspondances sous contrainte (CCA).

**Mercredi 19 mai 9h-12h30 : statistiques spatiales, krigeage par P. Monestiez (INRA Avignon) et V. Garreta (CEREGE)**

**MODULE 3a:** Présentation des types de problèmes et de la nature des données auxquelles s'appliquent les méthodes d'analyses spatiales: analyses de processus ponctuels, analyses sur réseaux et sur grilles, et géostatistique (exemples, visualisations simples de jeux de données). Présentation des hypothèses générales et des modèles utilisés en géostatistique (utilisation de programmes de simulations pour visualiser le potentiel et les limites du cadre théorique). Présentation des 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 (krigeages ordinaire et universel) dans des cas simples et univariés, influence du choix du modèle et réflexion sur les types d'échantillonnage.

**Mercredi 19 mai 13h30-17h : Apport des méthodes géostatistiques dans l'évaluation du risque environnemental et industriel par Jean Baccou, IRSN, Cadarache**

**MODULE 3b:** L'objectif de ce module est de montrer comment les outils géostatistiques peuvent être utilisés en analyse de risque. Dans une méthodologie d'évaluation d'un risque, ils interviennent à plusieurs niveaux, notamment pour la caractérisation d'un milieu aléatoire (grâce à l'analyse variographique) ou pour la construction du krigeage. Dans ce dernier cas, il est aussi possible de se servir du phénomène interpolé (et de l'information fournie par le krigeage) pour planifier de nouvelles expériences afin de mieux le décrire. Plusieurs applications en analyse de risque environnemental et industriel seront présentées pour illustrer ces différents types d'utilisation des approches géostatistiques.