

Science des données III : cours 4b



Séries spatio-temporelles (partie 2)

Philippe Grosjean & Guyliann Engels

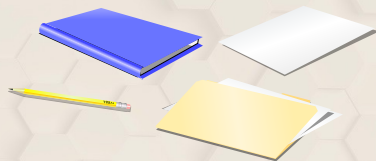
Université de Mons, Belgique
Laboratoire d'Écologie numérique des Milieux aquatiques



<http://biodatascience-course.sciviews.org>
sdd@sciviews.org

Objectifs du cours

- Prendre conscience des caractéristiques importantes des séries spatio-temporelles
- Savoir détecter une tendance globale ou locale
- Pouvoir détecter des cycles à l'aide de l'analyse spectrale



Les composantes d'un signal spatio-temporel

- Une série spatio-temporelle peut être considérée comme la **superposition de plusieurs signaux plus simples** :
 - Une ou plusieurs **tendances générales** (augmentation ou diminution progressive sur le long terme),
 - Des **tendances locales**, responsables de la variation *locale* de la valeur moyenne des valeurs,
 - Des **cycles** qui correspondent à des variations périodiques (cycle circadien, effet saisonnier, lunaire ou des marées, ...)
 - Un **bruit blanc** représenté par une variation purement aléatoire (analogue aux **résidus** d'un modèle linéaire)
- Ces différentes composantes peuvent se combiner de **deux façons différentes** :
 - De manière **additive** (modèle additif)
 - De manière **multiplicative** (modèle multiplicatif issu d'un effet à caractère exponentiel)

Illustration

Présentation au tableau de la façon dont divers signaux peuvent se combiner.

Tendance générale

- La tendance générale s'exprime par, soit une **augmentation**, soit une **diminution** progressive dans le temps (sur le long terme) du signal étudié, à l'exclusion de signaux cycliques.
- Elle est généralement visible, mais *est-elle significative* ?
- Un test existe pour déterminer si une tendance générale de forme quelconque est significative : le **test de tendance par bootstrap du coefficient de Spearman**

Concept du test

Il doit exister une corrélation significative entre les observations et le temps en présence d'une tendance. Mais comme les données **ne sont pas indépendantes** entre elles, on ne peut pas se référer à une distribution théorique. La distribution de référence est simulée en randomisant un grand nombre de fois les observations dans le temps (= technique du *bootstrap*).

Tendance locale

- La recherche de **tendances locales** vise à détecter les périodes dans le temps où les valeurs moyennes des observations changent
- La meilleure façon de détecter ces changements consiste à transformer le signal en **sommes cumulées**, et d'analyser le signal ainsi transformé :

$$S_q = \sum_{i=1}^q x_i - q.r$$

ou :

- q est le nombre d'observations déjà cumulées à l'observation actuelle x_i
- r est une constante de référence (généralement, la moyenne de toute la série)
- Les sommes cumulées forment alors une **ligne brisée**. Chaque segment représente une période de valeur moyenne constante donnée par la pente du segment additionnée de r .

Analyse spectrale

- Analyse de cycles : décomposition de la série en une série de signaux sinusoïdaux et cosinusoidaux de fréquences différentes. Détermination des composantes qui sont significatives.
- L'ensemble est représenté sur un graphique appelé **périodogramme**.
- Démonstration du concept à : <https://go.sciviews.org/spectral>
- Exemple sur `nottem`