Science des données III : cours 6

Séries spatio-temporelles (partie 5)

Philippe Grosjean & Guyliann Engels

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

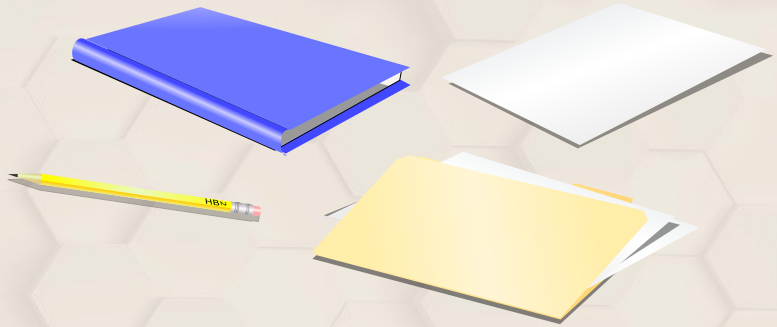
http://biodatascience-course.sciviews.org
sdd@sciviews.org

UMONS

Régularisation de séries temporelles

UMONS

Objectifs du cours



- Pouvoir régulariser une série régulière
- Décider du meilleur pas de temps
- Choisir entre plusieurs méthodes de régularisation

UMONS

Régularisation

- **Série à trous** = série régulière mais avec des valeurs manquantes. Dans ce cas, le pas de temps est connu
- **Série irrégulière**. Il faut décider du meilleur pas de temps (celui pour lequel on doit interpoler moins de valeurs)
- Utilisation des propriétés d'autocorrelation pour **interpoler** afin de "bricoler" une série régulière
- Méthode acceptable lorsque l'interpolation ne concerne pas trop de données

UMONS

Choix du pas de temps pour la régularisation

- Fonctions `regul.screen()` et `regul.adj()` du package **pastecs**
- Test de combinaisons de divers pas de temps et de diverses dates de départ avec `regul.screen()`
- Garder la combinaison qui permet d'interpoler le moins de valeurs possibles
- Choix de la fenêtre de tolérance (pas d'interpolation, mais utilisation directe de la valeur mesurée dans la fenêtre) à l'aide de `regul.adj()`

Démonstration

Détermination des paramètres de régularisation pour le jeu de données `releve` (phytoplancton mesuré en une station), exemple de la série `Melosul`

UMONS

Méthode de régularisation par valeurs constantes

- La méthode par **valeurs constantes** est la plus simple: elle reporte simplement la valeurs la plus proche rencontrée à gauche ($f = 1$), à droite ($f = 0$) , ou combinaison des deux ($0 > f < 1$) :

$$X_j = x_i.f + x_{i-1}.(1 - f)$$

- Méthode impliquant le moins d'hypothèse sur la forme du signal
- Utile lorsque le signal est plutôt stable

En pratique...

Utiliser `regconst()` ou `regul(method = "c")`. Illustration sur `relevel$Melosul`. Objet `regul` avec méthode `plot()` pour le diagnostic.

UMONS

Méthode de régularisation linéaire

- **Interpolation linéaire** entre le point avant et le point après :

$$X_j = \frac{(t_i - T_j).x_{i-1} + (T_j - t_{i-1}).x_i}{t_i - t_{i-1}}$$

- Méthode impliquant une variation linéaire ou quasi-linéaire d'un point à l'autre
- Technique la plus universelle (hypothèse raisonnable dans beaucoup de cas)

En pratique...

Utiliser `reglin()` ou `regul(method = "l")`. Illustration sur `relevel$Melosul`.

UMONS

Méthode de régularisation par courbes splines

- **Interpolation polynômiale**, en utilisant deux ou plus de points avant et après la valeur à interpoler
- Formulation mathématique complexe (voir syllabus)
- Prend en compte la variation du signal autour du point à interpoler en utilisant plus d'information que les deux méthodes précédentes
- Méthode efficace, mais **attention aux pics et creux artificiels possibles (surtout si des valeurs négatives sont à éviter)**

En pratique...

Utiliser `regspline()` ou `regul(method = "s")`. Illustration sur `relevel$Melosul`.

UMONS

Méthode de régularisation par les aires

- Utilisation de l'information présente dans une **fenêtre glissante de largeur fixe**.
- Moyenne des observations présentes dans le fenêtre pondérée par rapport à leur "aire d'influence" (voir schéma dans le syllabus)
- Prend en compte la densité variable d'information disponible localement
- Attention: **méthode efficace pour des données physico-chimiques, mais un lissage est également effectué**. Donc, le signal n'est plus brut!

En pratique...

Utiliser `regarea()` ou `regul(method = "a")`. Illustration sur `relevel$Melosul`.

UMONS

Prédiction (démonstration)

- Les **modèles autorégressifs** (AR, ARMA, ARIMA, etc.) permettent de modéliser un signal en prenant en compte l'autocorrelation. Ils peuvent être ensuite utilisés à des fins de prédiction
- Souvent, ces prédictions sont peu fiables, surtout pour des données en biologie contenant du bruit important
- Récemment, des techniques plus sophistiquées issues de l'intelligence artificielle apparaisse.

Démonstration

La méthode **prophet** développée par facebook fait partie de ces techniques plus efficaces, … mais nécessitant beaucoup de données (milliers d'observations, voire bien plus)! Voir https://facebook.github.io/prophet/

UMONS