

Traitement de séries temporelles

Jordi Inglada



Novembre 2012

Sommaire

- 1 Introduction
- 2 Modèles ARMA
- 3 Ajustement de fonctions de base
- 4 Séparation de composantes
- 5 Échantillonnage irrégulier et distorsions temporelles

Sommaire

- 1 Introduction**
Introduction
- 2 Modèles ARMA**
- 3 Ajustement de fonctions de base**
- 4 Séparation de composantes**
- 5 Échantillonnage irrégulier et distorsions temporelles**

Introduction

- ▶ Traitement ou analyse de séries temporelles ?
 - ▶ Transformation du signal
 - ▶ Filtrage, détection, etc.
 - ▶ TF, ondelettes
 - ▶ Modélisation du signal
 - ▶ Caractérisation, interprétation, prédiction
 - ▶ ARMA, ajustement de modèles, composantes
 - ▶ Outils privilégiant la parcimonie des descriptions obtenues

Ce que nous ne verrons pas mais qui est très utile et intéressant

- ▶ Filtrage de Kalman
- ▶ Filtrage particulaire
- ▶ Bispectres et statistiques d'ordres supérieurs
- ▶ Séries multi-variées

Les séries temporelles

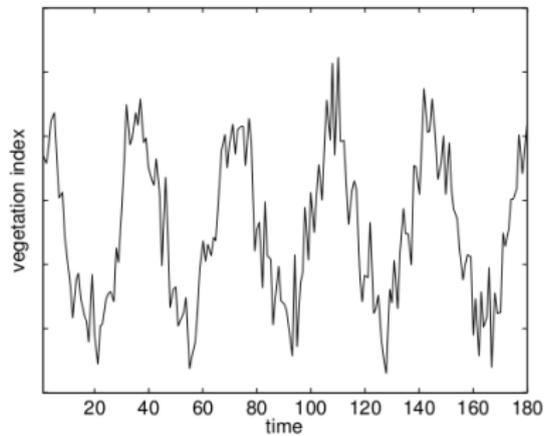
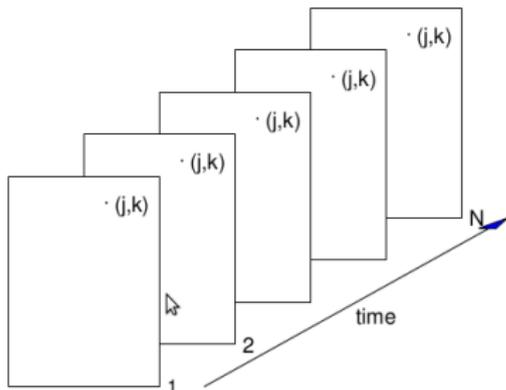
- ▶ Des mesures (liées à des processus) dont l'ordre temporel est structurant
- ▶ Structuration des données
 - ▶ autocorrélation
 - ▶ tendances (trends)
 - ▶ variations saisonnières
- ▶ Caractérisation au premier ordre
 - ▶ mode d'échantillonnage
 - ▶ longueur de la série
 - ▶ stationnarité
 - ▶ niveau de bruit

Objectifs et applications

- ▶ Obtenir une compréhension des processus générateurs des séries observées
- ▶ Ajuster des modèles pour faire
 - ▶ du suivi
 - ▶ de la prédiction
 - ▶ du contrôle
- ▶ Domaines d'application
 - ▶ Économétrie
 - ▶ Rendements de production
 - ▶ Processus industriels
 - ▶ Dynamique des populations
 - ▶ Consommation d'énergie
 - ▶ Météorologie

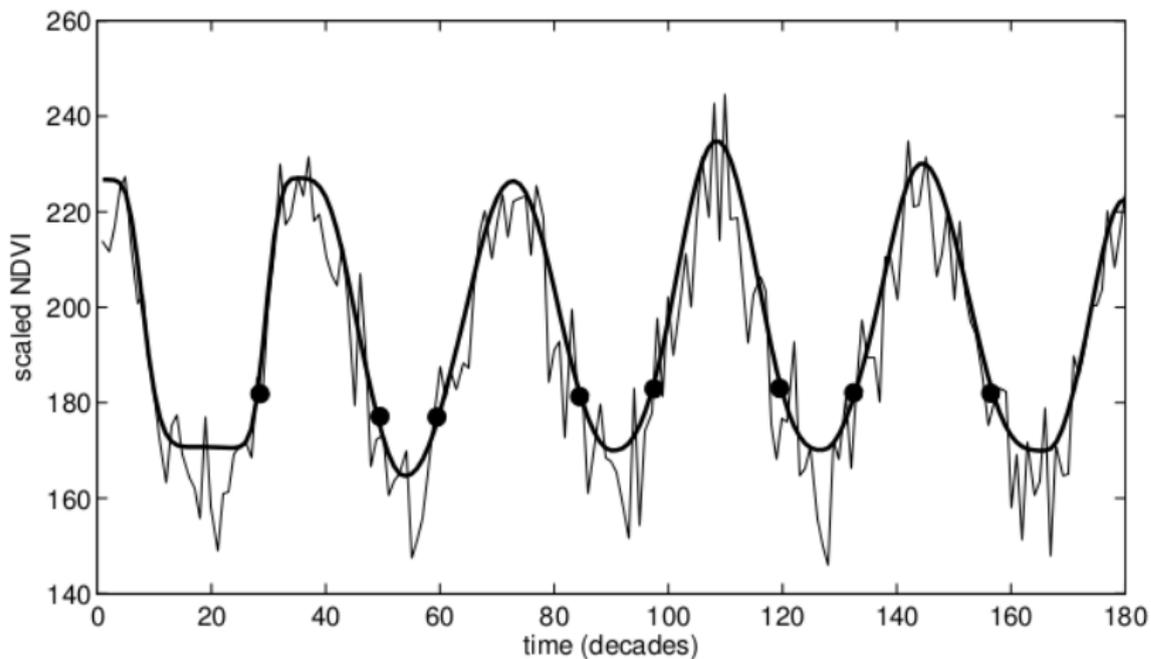
Exemple

Série temporelle d'images satellites



Exemple

Série temporelle d'images satellites



Sommaire

1 Introduction

2 Modèles ARMA

Moyennes

ARMA

Modèle autorégressif

Modèle moyenne mobile

Modèle autorégressif et moyenne-mobile

3 Ajustement de fonctions de base

4 Séparation de composantes

5 Échantillonnage irrégulier et distorsions temporelles

Moyennes et moyennes glissantes

- ▶ Moyenne de la série
- ▶ Moyenne glissante
 - ▶ lissage (boxcar)
- ▶ Lissage exponentiel
 - ▶ Simple
 - ▶ $S_t = y_{t-1} + (1 - \alpha)S_{t-1}; \alpha \leq 1; t > 2$
 - ▶ Double (permet de prendre en compte des tendances)
 - ▶ $S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + b_{t-1})$
 - ▶ $b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$

Sommaire

1 Introduction

2 Modèles ARMA

Moyennes

ARMA

Modèle autorégressif

Modèle moyenne mobile

Modèle autorégressif et moyenne-mobile

3 Ajustement de fonctions de base

4 Séparation de composantes

5 Échantillonnage irrégulier et distorsions temporelles

Modèles ARMA

- ▶ En statistiques, les modèles ARMA (modèles autorégressifs et moyenne mobile), ou aussi modèle de Box-Jenkins, sont les principaux modèles de séries temporelles.
- ▶ Étant donné une série temporelle X_t , le modèle ARMA est un outil pour comprendre et prédire, éventuellement, les valeurs futures de cette série.
- ▶ Le modèle est composé de deux parties : une part autorégressive (AR) et une part moyenne-mobile (MA).
- ▶ Le modèle est généralement noté ARMA(p,q), où p est l'ordre de la partie AR et q l'ordre de la partie MA.

Sommaire

1 Introduction

2 Modèles ARMA

Moyennes

ARMA

Modèle autorégressif

Modèle moyenne mobile

Modèle autorégressif et moyenne-mobile

3 Ajustement de fonctions de base

4 Séparation de composantes

5 Échantillonnage irrégulier et distorsions temporelles

Modèle autorégressif

- ▶ La notation $AR(p)$ réfère au modèle autorégressif d'ordre p . Le modèle $AR(p)$ se note

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t.$$

- ▶ où ϕ_1, \dots, ϕ_p sont les paramètres du modèle, c est une constante et ϵ_t un bruit blanc.
- ▶ La constante est bien souvent omise dans la littérature.
- ▶ Des contraintes supplémentaires sur les paramètres sont nécessaires pour garantir la stationnarité.
- ▶ Par exemple, pour le modèle $AR(1)$, les processus tels que $|\phi_1| \geq 1$ ne sont pas stationnaires.

Exemple : un processus AR(1)

- ▶ Un modèle AR(1) est donné par :

$$X_t = c + \phi X_{t-1} + \epsilon_t,$$

- ▶ où ϵ_t est un bruit blanc, de moyenne nulle et de variance σ^2 .
- ▶ Le modèle est stationnaire en variance si $|\phi| < 1$.
- ▶ Si $\phi = 1$, alors le processus exhibe une racine unitaire, ce qui signifie qu'il est une marche aléatoire, et n'est pas stationnaire en variance.
- ▶ Supposons donc $|\phi| < 1$, et en notant la moyenne μ , on obtient

$$E(X_t) = E(c) + \phi E(X_{t-1}) + E(\epsilon_t) \Rightarrow \mu = c + \phi\mu + 0.$$

Exemple : un processus AR(1)

- ▶ Ainsi

$$\mu = \frac{c}{1 - \phi}.$$

- ▶ En particulier, prendre $c = 0$ revient à avoir une moyenne nulle.
- ▶ La variance vaut

$$\text{var}(X_t) = E(X_t^2) - \mu^2 = \frac{\sigma^2}{1 - \phi^2}.$$

- ▶ La fonction d'autocovariance se donne par

$$B_n = E(X_{t+n}X_t) - \mu^2 = \frac{\sigma^2}{1 - \phi^2} \phi^{|n|}.$$

Exemple : un processus AR(1)

- ▶ On peut voir que la fonction d'autocovariance décroît avec un taux de $\tau = -1/\ln(\phi)$.
- ▶ La densité spectrale de puissance est la transformée de Fourier de la fonction d'autocovariance. Dans le cas discret, cela s'écrit :

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} B_n e^{-i\omega n} = \frac{1}{\sqrt{2\pi}} \left(\frac{\sigma^2}{1 + \phi^2 - 2\phi \cos(\omega)} \right).$$

- ▶ Ce développement est périodique dû à la présence du terme en cosinus au dénominateur.

Exemple : un processus AR(1)

- ▶ En supposant que le temps d'échantillonnage ($\Delta t = 1$) est plus petit que la constante de temps (τ), alors on peut utiliser une approximation continue de B_n :

$$B(t) \approx \frac{\sigma^2}{1 - \phi^2} \phi^{|t|}$$

- ▶ qui présente une densité spectrale :

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}} \frac{\sigma^2}{1 - \phi^2} \frac{\gamma}{\pi(\gamma^2 + \omega^2)}$$

- ▶ où $\gamma = 1/\tau$ est la fréquence angulaire associée à τ .

Exemple : un processus AR(1)

- ▶ Une expression alternative pour X_t peut être dérivée en substituant X_{t-1} par $c + \phi X_{t-2} + \epsilon_{t-1}$ dans l'équation définissante.
- ▶ En continuant cette manipulation N fois fournit

$$X_t = c \sum_{k=0}^{N-1} \phi^k + \phi^N X_{t-N} + \sum_{k=0}^{N-1} \phi^k \epsilon_{t-k}.$$

Exemple : un processus AR(1)

- ▶ Pour N devenant très grand, ϕ^N s'approche de 0 et :

$$X_t = \frac{c}{1 - \phi} + \sum_{k=0}^{\infty} \phi^k \epsilon_{t-k}.$$

- ▶ On peut voir que X_t est le bruit blanc convolé avec le noyau ϕ^k plus une moyenne constante.
- ▶ Si le bruit blanc est gaussien, alors X_t est aussi un processus normal.
- ▶ Dans les autres cas, le théorème central limite indique que X_t sera approximativement normal lorsque ϕ est proche de l'unité.

Sommaire

1 Introduction

2 Modèles ARMA

Moyennes

ARMA

Modèle autorégressif

Modèle moyenne mobile

Modèle autorégressif et moyenne-mobile

3 Ajustement de fonctions de base

4 Séparation de composantes

5 Échantillonnage irrégulier et distorsions temporelles

Modèle moyenne mobile

- ▶ La notation $MA(q)$ réfère au modèle moyenne-mobile d'ordre q :

$$X_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i},$$

- ▶ où les $\theta_1, \dots, \theta_q$ sont les paramètres du modèle
- ▶ et $\epsilon_t, \epsilon_{t-i}, \dots$ sont encore une fois des termes d'erreur.

Sommaire

1 Introduction

2 Modèles ARMA

Moyennes

ARMA

Modèle autorégressif

Modèle moyenne mobile

Modèle autorégressif et moyenne-mobile

3 Ajustement de fonctions de base

4 Séparation de composantes

5 Échantillonnage irrégulier et distorsions temporelles

Modèle autorégressif et moyenne-mobile

- ▶ La notation $ARMA(p, q)$ réfère le modèle avec p termes autorégressifs et q termes moyenne-mobile.
- ▶ Ce modèle contient à la fois les modèles $AR(p)$ et $MA(q)$:

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}.$$

- ▶ Une note sur les termes d'erreur
 - ▶ Les termes d'erreur ϵ_t sont généralement supposés indépendants et identiquement distribués (i.i.d.) selon une loi normale de moyenne nulle.
 - ▶ Ces hypothèses peuvent être assouplies mais ceci changerait les propriétés du modèle, comme par exemple supposer le simple caractère i.i.d.

Spécification en termes de l'opérateur de retard

- ▶ Les modèles ARMA peuvent s'écrire en termes de L , qui est l'opérateur retard.
- ▶ Le modèle autorégressif $AR(p)$ s'écrit

$$\epsilon_t = \left(1 - \sum_{i=1}^p \phi_i L^i \right) X_t = \phi X_t$$

- ▶ où ϕ représente le polynôme

$$\phi = 1 - \sum_{i=1}^p \phi_i L^i.$$

Spécification en termes de l'opérateur de retard

- ▶ Pour le modèle moyenne mobile $MA(q)$, on a

$$X_t = \left(1 + \sum_{i=1}^q \theta_i L^i \right) \epsilon_t = \theta \epsilon_t$$

- ▶ où θ représente le polynôme

$$\theta = 1 + \sum_{i=1}^q \theta_i L^i.$$

Spécification en termes de l'opérateur de retard

- ▶ Finalement, en combinant les deux aspects, on en tire l'écriture du modèle $ARMA(p, q)$:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t$$

- ▶ où plus court :

$$\phi X_t = \theta \epsilon_t.$$

Modèle d'ajustement

- ▶ Les modèles ARMA, une fois choisis les ordres p et q , peuvent être ajustés sur des données par la méthode des moindres carrés :
 - ▶ on recherche les paramètres qui minimisent la somme des carrés des résidus.
- ▶ Prendre des valeurs de p et q les plus petites est généralement vu comme une bonne pratique (principe de parcimonie).
- ▶ Pour un modèle AR pur, les équations de Yule-Walker permettent de réaliser l'ajustement.

Obtention des équations de Yule-Walker

- ▶ L'équation définissant le processus AR est

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t.$$

- ▶ En multipliant les deux membres par X_{t-m} et en prenant l'espérance, on obtient

$$E[X_t X_{t-m}] = E \left[\sum_{i=1}^p \phi_i X_{t-i} X_{t-m} \right] + E[\epsilon_t X_{t-m}].$$

- ▶ Or, il se trouve que $E[X_t X_{t-m}] = \gamma_m$ par définition de la fonction d'autocovariance.

Obtention des équations de Yule-Walker

- ▶ Les termes du bruit blanc sont indépendants les uns des autres et, de plus, X_{t-m} est indépendant de ϵ_t où m est plus grand que zéro.
- ▶ Pour $m > 0$, $E[\epsilon_t X_{t-m}] = 0$.
- ▶ Pour $m = 0$,

$$E[\epsilon_t X_t] = E \left[\epsilon_t \left(\sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \right) \right] = \sum_{i=1}^p \phi_i E[\epsilon_t X_{t-i}] + E[\epsilon_t^2] = 0 + \sigma_\epsilon^2,$$

Obtention des équations de Yule-Walker

- ▶ Maintenant, on a pour $m \geq 0$,

$$\gamma_m = E \left[\sum_{i=1}^p \phi_i X_{t-i} X_{t-m} \right] + \sigma_\epsilon^2 \delta_m.$$

- ▶ Par ailleurs,

$$E \left[\sum_{i=1}^p \phi_i X_{t-i} X_{t-m} \right] = \sum_{i=1}^p \phi_i E[X_t X_{t-m+i}] = \sum_{i=1}^p \phi_i \gamma_{m-i},$$

Obtention des équations de Yule-Walker

- ▶ Et finalement :

$$\gamma_m = \sum_{i=1}^p \phi_i \gamma_{m-i} + \sigma_\epsilon^2 \delta_m; m \geq 0$$

$$\gamma_m = \gamma_{-m} = \sum_{i=1}^p \phi_i \gamma_{|m|-i} + \sigma_\epsilon^2 \delta_m, m < 0$$

- ▶ Les paramètres peuvent donc être estimés à partir de la fonction d'autocorrélation !

Quelques remarques

- ▶ On suppose que la série temporelle est stationnaire
- ▶ Dans le cas d'une série non-stationnaire on travaille sur ses dérivées
 - ▶ Modèles ARIMA
- ▶ Souvent, on enlève la moyenne et on normalise la variance
- ▶ L'utilisation de ce type de modèles demande des séries relativement longues (> 100 échantillons).

Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

4 Séparation de composantes



Méthode

- ▶ Nous avons une série temporelle de type (t_i, y_i) , $i = 1, 2, \dots, N$
- ▶ et un modèle de type $f(t) = c_1\phi_1(t) + c_2\phi_2(t) + \dots + c_M\phi_M(t)$
- ▶ $\phi_i(t)$ sont des fonctions de base connues.
- ▶ Les valeurs optimales – au sens des moindres carrés – pour les coefficients c_i sont obtenues en posant
 - ▶ **$\mathbf{A}^T \mathbf{A} \mathbf{c} = \mathbf{A}^T \mathbf{b}$**
 - ▶ avec $A_{ij} = \omega_i \phi_j(t_i)$ et $b_i = \omega_i y_i$
 - ▶ ω_i est le poids de la date i et supposé connu.

Comment choisir les poids

- ▶ Connaissance à priori
 - ▶ Connaissance générale sur les données
 - ▶ plages de valeurs attendues
 - ▶ lissage
 - ▶ Données exogènes
 - ▶ évaluation de l'erreur
 - ▶ probabilité de mesure fausse
- ▶ Estimation itérative

Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

4 Séparation de composantes



Suppression de points aberrants

- ▶ En utilisant des données exogènes comme indiqué précédemment
- ▶ Filtrage médian
 - ▶ suppose des points aberrants isolés
- ▶ Décomposition STL (Seasonal Trend decomposition based on Loess)
- ▶ Combinaison de la décomposition STL avec des données exogènes

Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

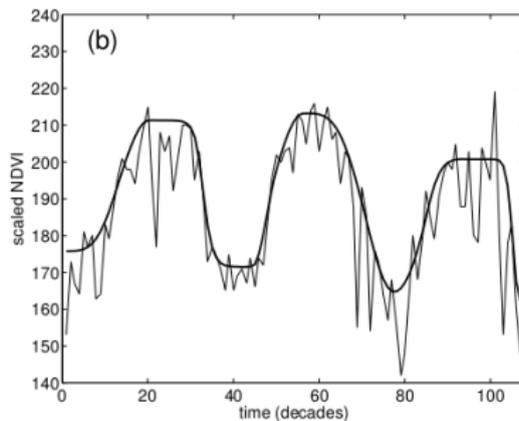
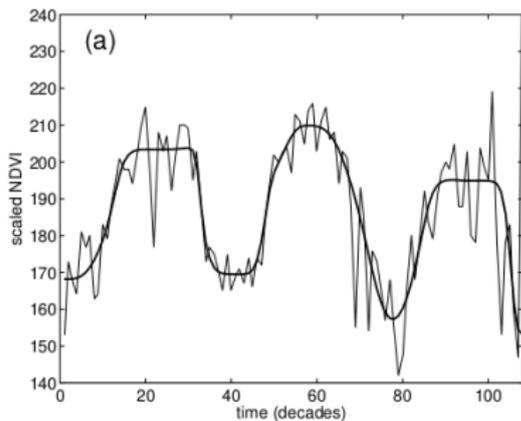
4 Séparation de composantes



Estimation de l'enveloppe d'une série

- ▶ Certaines séries peuvent être biaisées
 - ▶ bruit/points aberrants toujours de même signe
- ▶ Procédure d'approximation itérative
 - 1 Approximation par moindres carrés classique
 - 2 Estimation des poids en fonction de la position de chaque point par rapport à l'approximation précédente
 - 3 Nouvelle approximation en utilisant les nouveaux poids

Exemple de Savitzky-Golay enveloppe supérieure



Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

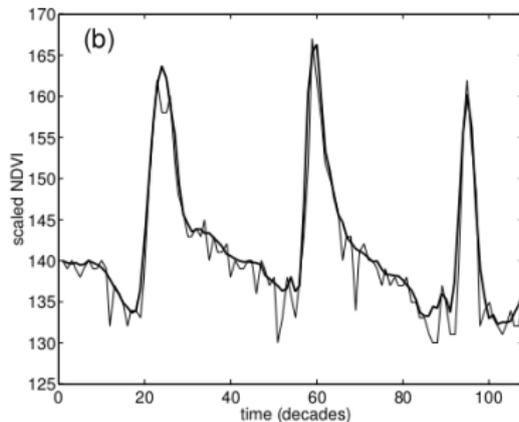
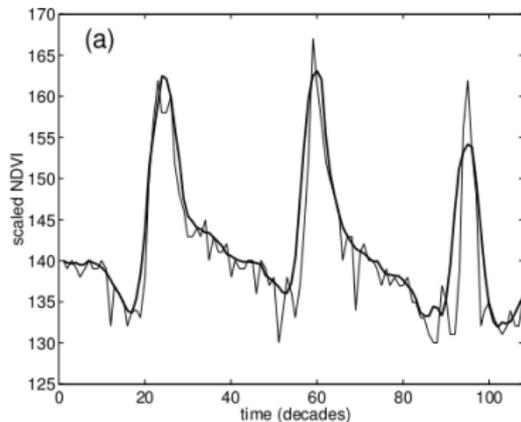
4 Séparation de composantes



Lissage de Savitzky-Golay

- ▶ Remplacer chaque valeur y_i par une combinaison linéaire des valeurs voisines
 - ▶ $\sum_{j=-n}^n c_j y_{i+j}$
 - ▶ la moyenne glissante correspond à $c_j = 1/(2n + 1)$
- ▶ Savitzky-Golay ajuste un polynôme de degré 2 aux valeurs dans la fenêtre
 - ▶ $f(t) = c_1 + c_2 t + c_3 t^2$
 - ▶ utilisation des moindres carrés
 - ▶ et on ne remplace que la valeur centrale
- ▶ Peut s'utiliser de façon itérative pour trouver l'enveloppe
- ▶ Comment trouver la *bonne* : taille de fenêtre n ?

Illustration



- ▶ (a) fenêtre trop large ; (b) fenêtre adaptée.

Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

4 Séparation de composantes



Approche générale

- ▶ On ajuste des fonctions paramétriques autour des maxima de la série
 - ▶ $f(t) \equiv f(t; \mathbf{c}, \mathbf{x}) = c_1 + c_2 g(t; \mathbf{x})$
 - ▶ $\mathbf{c} = (c_1, c_2)$ déterminent le plancher et l'amplitude du signal
 - ▶ $\mathbf{x} = (x_1, x_2, \dots, x_p)$ déterminent la forme de la fonction de base $g(t; \mathbf{x})$
- ▶ Fonction de coût : $\sum_{i=n_1}^{n_2} [\omega_i (f(t; \mathbf{c}, \mathbf{x}) - y_i)]^2$
 - ▶ Moindres carrés non-linéaires (i.e. Levenberg-Marquardt)

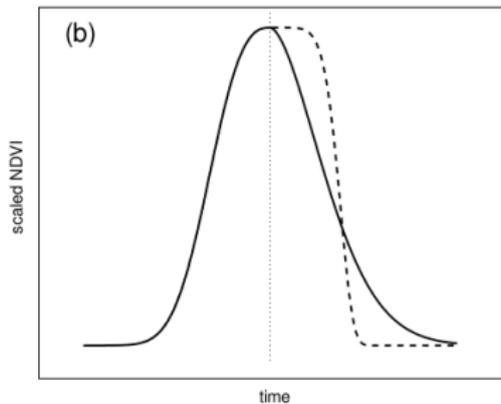
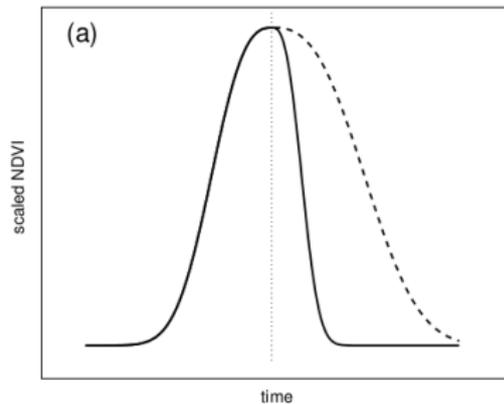
Gaussiennes asymétriques

$$g(t; x_1, \dots, x_5) = \begin{cases} \exp \left[- \left(\frac{t-x_1}{x_2} \right)^{x_3} \right] & t > x_1 \\ \exp \left[- \left(\frac{t-x_1}{x_4} \right)^{x_5} \right] & t < x_1 \end{cases} \quad (1)$$

- ▶ Les valeurs des paramètres sont restreintes
 - ▶ x_3 et x_5 sont supposés > 2

Gaussiennes asymétriques

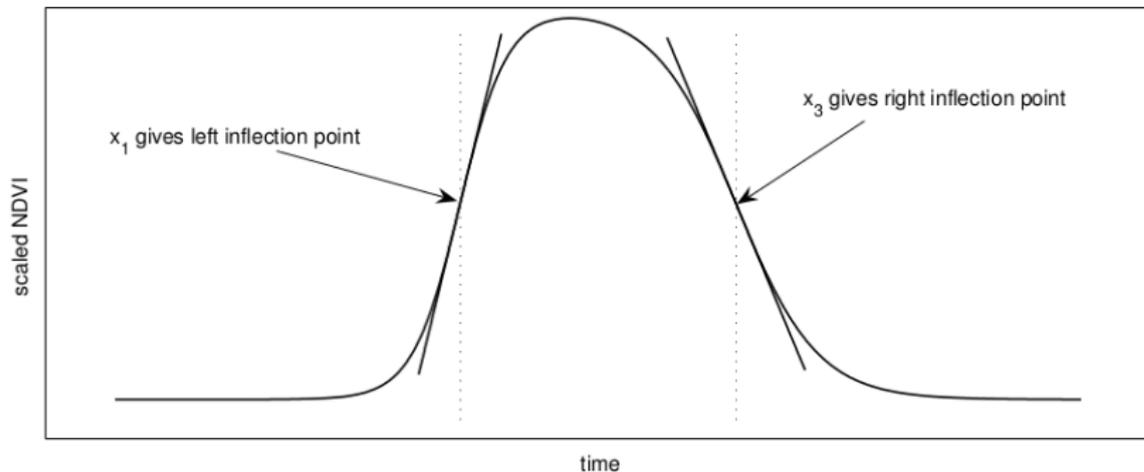
Illustration



- ▶ (a) effet du paramètre x_2
- ▶ (b) effet du paramètre x_3

Doubles logistiques

$$g(t; x_1 \cdots x_4) = \frac{1}{1 + \exp\left(\frac{x_1 - t}{x_2}\right)} - \frac{1}{1 + \exp\left(\frac{x_3 - t}{x_4}\right)}$$



Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

Ajustement par moindres carrés

Suppression de points aberrants

Enveloppe

Savitzky-Golay

Gaussiennes asymétriques et doubles logistiques

LOESS

4 Séparation de composantes



LOESS

- ▶ Similaire à Savitzky-Golay
- ▶ Ajustement local par moindres carrés de polynômes de degré faible
 - ▶ choix des poids $w(x) = (1 - |x|^3)^3$ [$|x| < 1$]
- ▶ La *localité* de l'ajustement est déterminée par le paramètre α avec
 - ▶ $(\lambda + 1) / n \leq \alpha \leq 1$
 - ▶ λ est le degré du polynôme
 - ▶ le nombre de points est donné par l'entier supérieur le plus proche de $n\alpha$

Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

4 Séparation de composantes

Estimation du nombre de périodes

STL

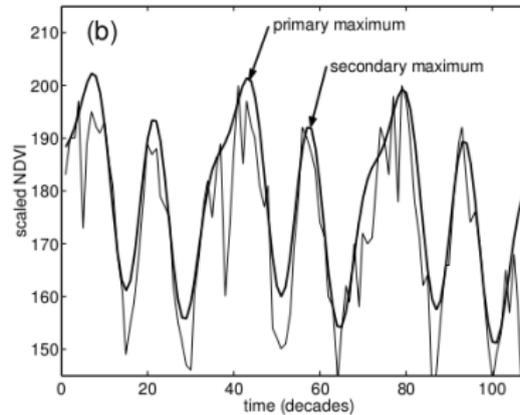
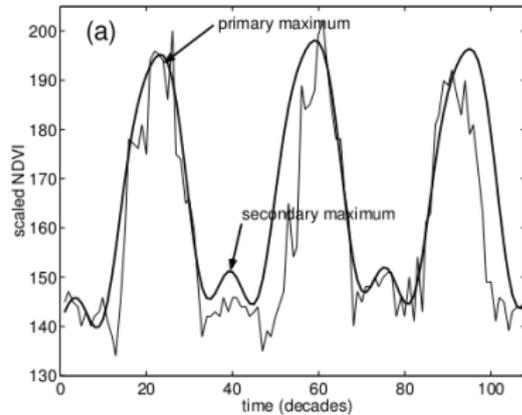
BFAST

5 Échantillonnage irrégulier et distorsions temporelles

Estimer les saisons

- ▶ Dans le cas général, recherche de pics dans l'espace de Fourier
- ▶ Exemple des saisons de la végétation
 - ▶ On suppose que la végétation peut avoir 1 ou 2 saisons
 - ▶ Une année contient un nombre entier de saisons
 - ▶ Modèle de type
$$f(t) = c_1 + c_2 \sin(\omega t) + c_3 \cos(\omega t) + c_4 \sin(2\omega t) + c_5 \cos(2\omega t)$$
 - ▶ avec $\omega = 3 \times \frac{2\pi}{N}$
 - ▶ et N est le nombre d'échantillons par an
 - ▶ Cette procédure trouvera 1 ou 2 maxima

Illustration



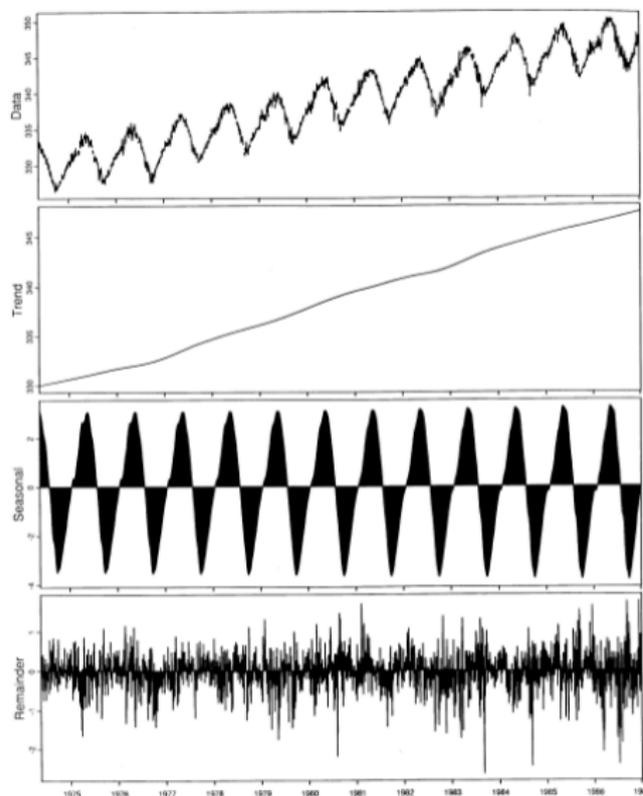
Sommaire

- 1 Introduction**
- 2 Modèles ARMA**
- 3 Ajustement de fonctions de base**
- 4 Séparation de composantes**
 - Estimation du nombre de périodes
 - STL
 - BFAST
- 5 Échantillonnage irrégulier et distorsions temporelles**

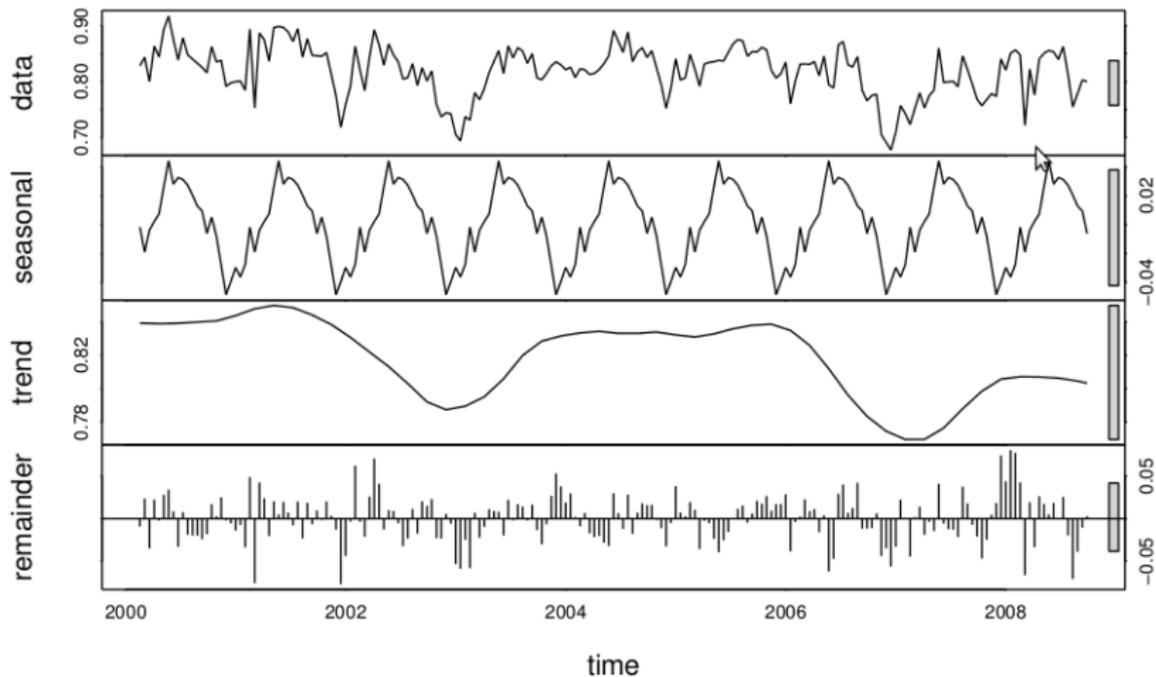
STL

- ▶ Seasonal Trend Decomposition based on LOESS
- ▶ $X_t = S_t + T_t + R_t$
- ▶ Application séquentielle de plusieurs lissages de type LOESS
 - 1 Estimation de la tendance
 - 2 Suppression de la tendance
 - 3 Interpolation de la composante saisonnière
 - 4 Lissage de la composante saisonnière
 - 5 Suppression de la composante saisonnière
 - 6 Lissage de la tendance
 - 7 Aller à 2

STL



STL avec tendance non-linéaire



Sommaire

1 Introduction

2 Modèles ARMA

3 Ajustement de fonctions de base

4 Séparation de composantes

Estimation du nombre de périodes

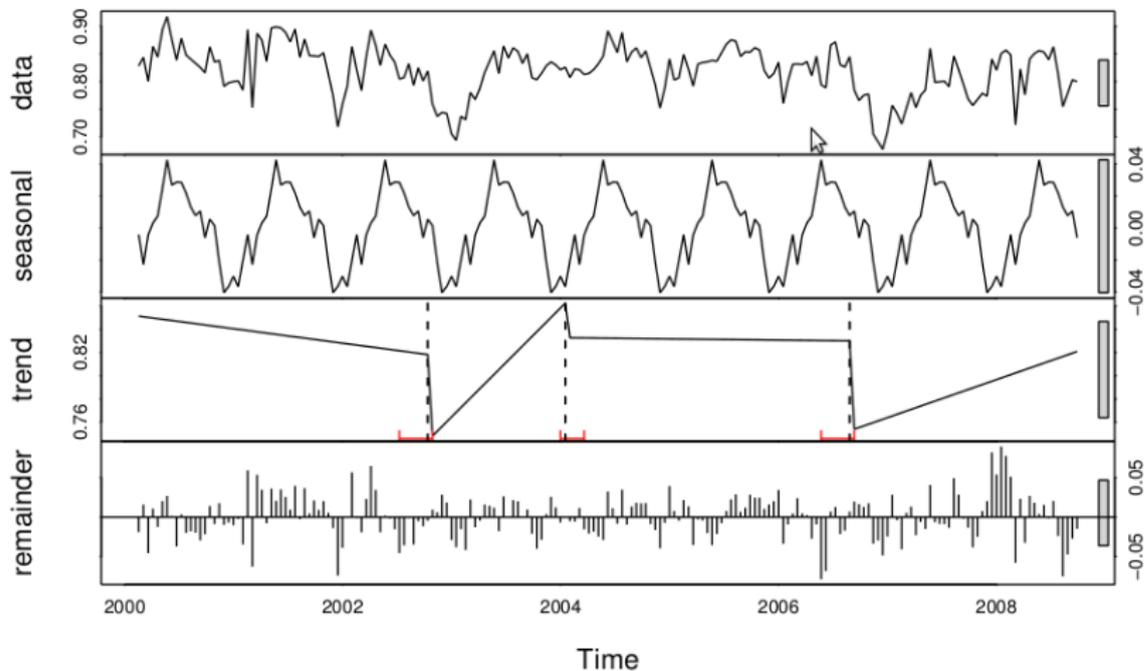
STL

BFAST

5 Échantillonnage irrégulier et distorsions temporelles

BFAST

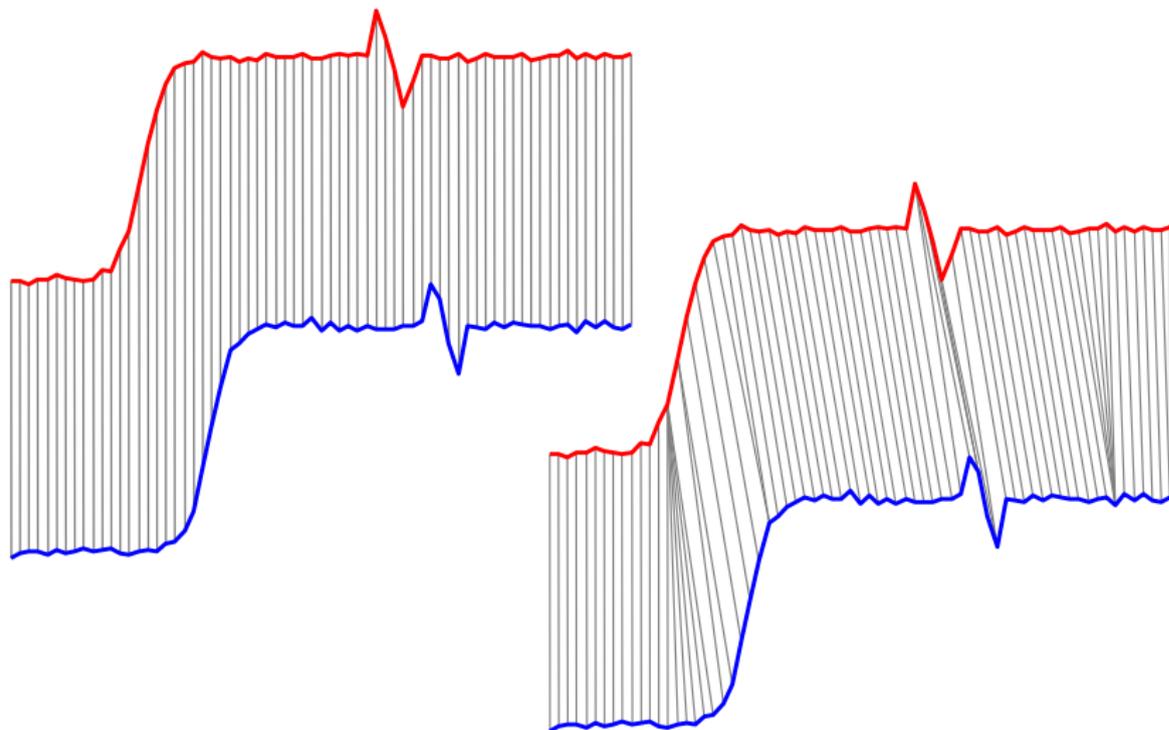
- ▶ Comme STL, mais la tendance est linéaire par morceaux



Sommaire

- 1 Introduction
- 2 Modèles ARMA
- 3 Ajustement de fonctions de base
- 4 Séparation de composantes
- 5 Échantillonnage irrégulier et distorsions temporelles
Dynamic Time Warping

DTW



Principe de DTW

- ▶ DTW a pour objectif de positionner les associations entre les éléments des séquences de façon à minimiser le coût de l'alignement.
- ▶ Calcul

$$DTW(A_{1\dots i}, B_{1\dots j}) \\ = \delta(a_i, b_j) + \min \left\{ \begin{array}{l} DTW(A_{1\dots i-1}, B_{1\dots j-1}) \\ DTW(A_{1\dots i}, B_{1\dots j-1}) \\ DTW(A_{1\dots i-1}, B_{1\dots j}) \end{array} \right.$$

- ▶ Robuste aux distorsions et aux échantillons manquants

Exemple

