# Intelligence Artificielle & Machine Learning pour la modélisation de séries temporelles et de signaux

## Séance 3 : Modèles pour les séries temporelles

Laurent Oudre
laurent.oudre@ens-paris-saclay.fr

Diplôme ARIA
ENS Paris Saclay
2024-2025

# Modèles et représentations

- Nous avons vu un certain nombre d'outils pour l'exploration des données temporelles. Ces outils sont issus de la physique ou des statistiques
- Nous allons dans cette séance passer à l'étape suivante, c'est à dire à la modélisation des phénomènes. Nous verrons en particulier comment apprendre les paramètres de ces modèles à partir des données.

# Pourquoi utiliser des modèles ?

- Afin de prédire, classifier ou extraire de l'information de séries temporelles, il est nécessaire de les comprendre
- Cette compréhension passe par la définition de modèles (déterministes ou statistiques)
- Ces modèles permettent à la fois de résumer l'information contenue dans le signal sous la forme d'un petit nombre de paramètres, mais aussi d'avoir une idée de son comportement dans le futur
- En revanche ils ne sont que des approximations et il faut garder à l'idée que dans *la vraie vie*, la plupart des hypothèses sous jacentes ne sont pas validées...

# Programme de la séance

- Connaître les techniques pour modéliser les séries temporelles et inférer les paramètres associés
- Modèles courants physiques et statistiques

Session 3 : Time series models

# Plan du cours

# Plan du cours

# Capteurs et mesures

- Lorsque l'on mesure un phénomène physique (température, tension électrique, son...), il est impossible d'accéder exactement aux quantités pertinentes que l'on cherche à quantifier.
- On observe l'apparition de signaux parasites qui viennent se superposer au signal dit utile (i.e l'information que l'on souhaite récupérer). Ces signaux sont une gêne pour la compréhension de l'information que le signal transporte.
- Ces signaux parasites sont nécessairement aléatoires car non connus : on appelle ce type de perturbation un **bruit**

# Modèle de signal et bruit

- Le modèle le plus courant pour exprimer ce phénomène est le suivant

$$\boxed{y[n] = x[n] + b[n]}$$

  - $x[n]$ est le signal utile, celui que l'on veut mesurer
  - $b[n]$ est le bruit de mesure, aléatoire, que l'on connaît pas
  - $y[n]$ est le signal mesuré
  - $x[n]$ et $b[n]$ sont décorrélés
- Bien souvent, le but est de retrouver $x[n]$ à partir de l'observation bruitée $y[n]$

# Pourquoi parler de bruit ?

$$y[n] = x[n] + b[n]$$

Exemple historique de la parole dans la téléphonie

- On cherche à transmettre un signal de voix, des phrases et des mots. Il s'agit du signal utile qui contient de l'information
- Tout le reste (grésillements, bruit ambiant...) est considéré comme du bruit, qui parasite la communication
- La qualité de la communication va dépendre des niveaux relatifs du signal utile (la voix de l'interlocuteur) et du bruit (tous les autres sons)

# Rapport signal sur bruit

- Etant donné un signal $x[n]$ déterministe corrompu par un bruit $b[n]$, le signal bruité $y[n]$ s'écrit :
$$y[n] = x[n] + b[n]$$
- On appelle le rapport signal sur bruit, la quantité :

$$\boxed{SNR = \frac{P_x}{P_b}}$$

Plus cette quantité est élevée, moins le bruit n'a d'importance

- On exprime souvent cette quantité en décibels :

$$\boxed{SNR_{dB} = 10 \log_{10}\left(\frac{P_x}{P_b}\right)}$$

# Notion de bruit blanc

On appelle **bruit blanc** un signal aléatoire $b[n]$ tel que

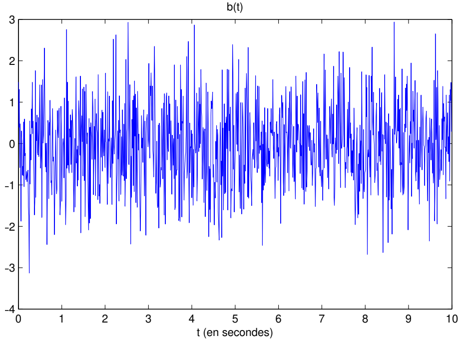
1. Le signal est stationnaire au sens large
2. Sa moyenne statistique est nulle

$$\forall n, \quad \mathbb{E}\left[b[n]\right] = 0$$

3. Deux valeurs $b[n_1]$ et $b[n_2]$ avec $n_1 \neq n_2$ sont décorrélées

$$\forall n_1, n_2 \text{ tels que } n_1 \neq n_2 \qquad \mathrm{cov}\left(b[n_1], b[n_2]\right) = 0$$

# Notion de bruit blanc




Un bruit blanc $b[n]$ est un signal aléatoire qui prend des valeurs au hasard à chaque instant $n$

- Les valeurs prises à chaque instant sont toutes décorrélées les unes des autres
- La moyenne de ce signal est égale à 0
- La variance du signal est égale à $\sigma^2$
- Plus $\sigma^2$ est élevé, plus le bruit blanc a tendance à prendre des valeurs éloignées de 0 et plus les amplitudes sont élevées (en valeur absolue)
- Très utilisé pour modéliser des perturbations dont on ne connaît que l'ordre de grandeur (paramétré par $\sigma^2$)

# Fonction d'autocorrélation

- Comme un bruit blanc est stationnaire au sens large et de moyenne nulle on a

$$\gamma_b[m] = \mathbb{E}\left[b[n]b[n+m]\right] = \mathrm{cov}\left(b[n], b[n+m]\right)$$

- Donc la fonction d'autocorrélation est nulle pour tout $m \neq 0$, et pour $m = 0$ on a

$$\gamma_b[0] = \mathbb{E}\left[b[n]^2\right] = \mathrm{var}\left[b[n]\right] = \sigma^2$$

- Pour un bruit blanc on a donc

$$\boxed{\gamma_b[m] = \begin{cases} \sigma^2 & \text{si } m = 0 \\ 0 & \text{sinon} \end{cases}}$$

# Densité spectrale de puissance

- Par définition de la densité spectrale de puissance, on a

$$\Gamma_b[k] = \mathcal{TFD}\left\{\gamma_b[m]\right\}$$

- En faisant le calcul on peut démontrer que la densité spectrale de puissance (DSP) d'un bruit blanc est

$$\boxed{\Gamma_b[k] = \sigma^2}$$

# Densité spectrale de puissance

- Dans le cas d'un bruit blanc, comme tout est aléatoire et décorrélé, il n'y a pas de raison qu'une fréquence ait plus d'importance que les autres. La densité spectrale de puissance est donc constante.
- Dans un bruit blanc, toutes les fréquences sont donc présentes avec des contributions égales. On l'appelle bruit blanc par analogie avec la lumière blanche qui mélange toutes les fréquences lumineuses.

# Plan du cours

1. Modèle signal + bruit
2. Modèle sinusoïdal
3. Modèle tendance et saisonnalité
4. Modèle autorégressif
5. Évaluation des modèles

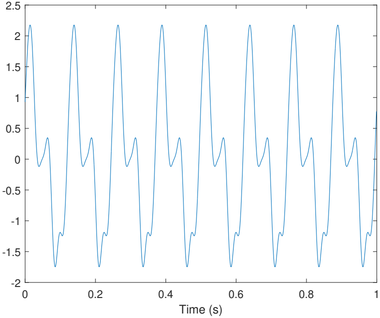
# Modèle sinusoïdal

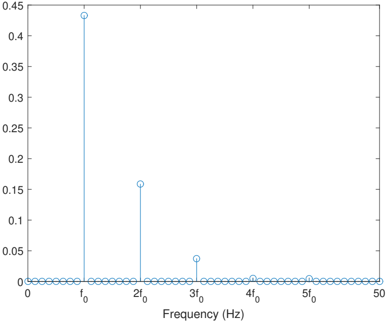
Pour tous les phénomènes répétitifs périodiques (notamment son ou phénomènes vibratoires), les fréquences présentes sont toutes multiples d'une seule et même fréquence

$$x[n] = \sum_{i=1}^{N_h} a_i \sin\left(2\pi(if_0)\frac{n}{F_e} + \phi_i\right) + b[n]$$

- $f_0$ : fréquence fondamentale (en Hz)
- $i \times f_0$ : $i$ ème harmonique
- $a_i$ et $\phi_i$ : amplitudes et phases de l'harmonique $i$
- $N_h$ : nombre d'harmoniques
- $b[n]$ : bruit blanc décorrélé

# Exemple




$$F_e = 1000 \text{ Hz}, f_0 = 8 \text{ Hz}$$

# Modèle sinusoïdal

$$x[n] = \sum_{i=1}^{N_h} a_i \sin\left(2\pi(if_0)\frac{n}{F_e} + \phi_i\right) + b[n]$$

- La connaissance de la fréquence fondamentale $f_0$ et des amplitudes relatives des harmoniques $a_i$ permet de caractériser le signal
- Utilité pour l'indexation, la description et la prédiction
- On stocke en général les amplitudes renormalisées par celle de la fondamentale

$$\frac{a_2}{a_1}, \frac{a_3}{a_1}, \frac{a_4}{a_1}, \ldots$$

que l'on nomme également *timbre*, par analogie avec les instruments de musique

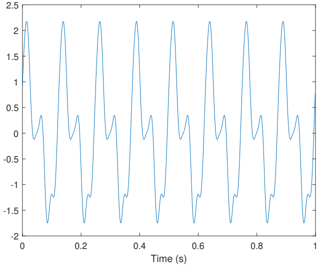
# Estimation des paramètres

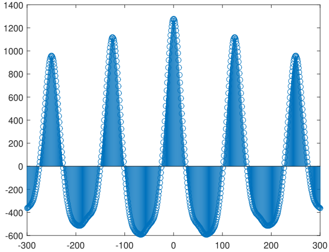
Deux paramètres doivent être estimés

- **La fréquence fondamentale $f_0$ :**
  - Recherche d'un pic sur le spectre et détermination de la fréquence associée
  - Souvent plus robuste : utilisation de la fonction d'autocorrélation (voir slide suivant)
- **Paramètres des harmoniques $a_i$ et $\phi_i$ :**
  - Etude de la TFD pour les fréquences multiples de $f_0$ (ou autour, si elles ne sont pas observables)
  - Régression sur une base de fonctions $\sin\left(2\pi(if_0)\frac{n}{F_e}\right)$ et $\cos\left(2\pi(if_0)\frac{n}{F_e}\right)$ (voir modèle suivant)

# Estimation de $f_0$

1. Tracer la fonction d'autocorrélation
2. Le premier pic pour les lags positifs correspond à la période (en échantillons) du signal $m_0$
3. $f_0$ est estimée grâce à $\hat{f}_0 = \frac{F_e}{m_0}$




$F_e$ = 1000 Hz, pic visible pour $m_0 = 125$ donc $\hat{f}_0 = \frac{1000}{125} = 8$ Hz

# Plan du cours

# Extension à d'autres fonctions

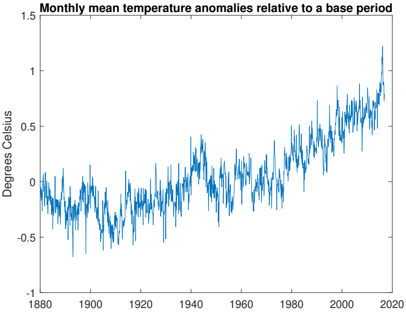
- Dans le modèle sinusoïdal on a utilisé la fonction $\beta(t) = \sin(2\pi f_0 t + \phi)$ pour modéliser la composante déterministe du signal
- Dans un cas plus général on peut considérer un grand nombre de fonctions pour modéliser une série temporelle
  - Certains fonctions à variations lentes, que l'on appellera **tendance**

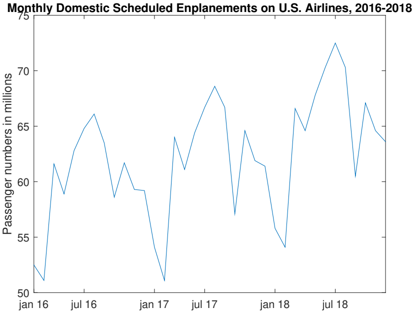
$$\beta(t) = cts \text{ (constante)}, \beta(t) = t \text{ (linéaire)}, \beta(t) = t^2 \text{ (quadratique)}...$$

  - D'autres à variations plus rapides et périodiques, que l'on appellera **saisonnalité**

$$\beta(t) = \cos(2\pi f t), \beta(t) = \sin(2\pi f t)...$$

# Exemples


Monthly mean temperature anomalies relative to a base period
Degrees Celsius
1.5
1
0.5
0
-0.5
-1
1880
1900
1920
1940
1960
1980
2000
2020


Monthly Domestic Scheduled Enplanements on U.S. Airlines, 2016-2018
Passenger numbers in millions
75
70
65
60
55
50
jan 16
jul 16
jan 17
jul 17
jan 18
jul 18

# Modèle tendance et saisonnalité

$$x[n] = \underbrace{\alpha_1\beta_1\left(nT_e\right) + \ldots + \alpha_j\beta_j\left(nT_e\right)}_{\text{tendance}} + \underbrace{\alpha_{j+1}\beta_{j+1}\left(nT_e\right) + \ldots + \alpha_d\beta_d\left(nT_e\right)}_{\text{saisonnalité}} + b[n]$$

- $\beta_1(t), \ldots \beta_d(t)$ : fonctions supposées connues
- $\alpha_1, \ldots \alpha_d$ : paramètres réels à estimer
- L'estimation du modèle est ainsi assimilable à une tâche de regression, c'est à dire une approximation de la série temporelle comme une combinaison linéaire de fonctions connues

# Estimation des paramètres

- Pour apprendre le modèle associé à $x$, il suffit d'apprendre les paramètres $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^T$
- Utilisation d'une estimation au sens des moindres carrés

$$\mathbf{x} = (x[0], \ldots, x[N-1])^T$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1(0) & \cdots & \beta_d(0) \\ \beta_1(T_e) & \cdots & \beta_d(T_e) \\ \vdots & \ddots & \vdots \\ \beta_1((N-1)T_e) & \cdots & \beta_d((N-1)T_e) \end{pmatrix}$$

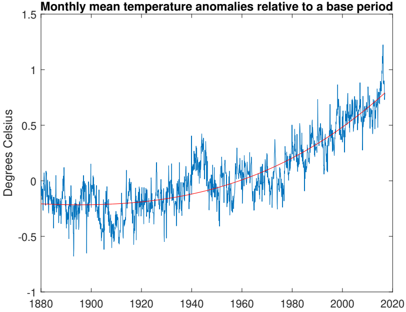
- On souhaite minimiser

$$\|\mathbf{x} - \boldsymbol{\beta}\boldsymbol{\alpha}\|_2$$

- Solution explicite

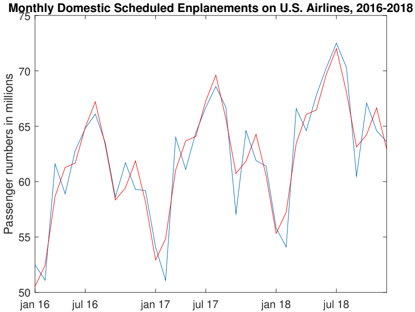
$$\boxed{\hat{\boldsymbol{\alpha}} = \left(\boldsymbol{\beta}^T\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}^T\mathbf{x}}$$

# Exemple : météo




$$\beta_1(t) = cst, \beta_2(t) = t, \beta_3(t) = t^2, \beta_4(t) = t^3$$

# Exemple : avion




$$\beta_1(t) = cst, \beta_2(t) = t$$

$$\beta_3(t) = \cos(2\pi f_0 t), \beta_4(t) = \sin(2\pi f_0 t) \text{ avec } f_0 = \frac{1}{12}$$

$$\beta_5(t) = \cos(4\pi f_0 t), \beta_6(t) = \sin(4\pi f_0 t), \beta_7(t) = \cos(6\pi f_0 t), \beta_8(t) = \sin(6\pi f_0 t)$$

# Plan du cours

# Liens temporels

- Lorsque l'on considère une série temporelle, il est logique de considérer que ce que l'on observe à l'échantillon $x[n]$ dépend des valeurs prises auparavant par le signal $x[n-1], \ldots, x[n-p]$
- Plusieurs modèles statistiques permettent de prendre en compte ces interactions et corrélations temporelles
- Le modèle le plus populaire est appelé **autorégressif** (AR) : au lieu d'utiliser une regression grâce à des fonctions extérieures, on va utiliser les valeurs passées de la série temporelle

# Modèle autorégressif $AR(p)$

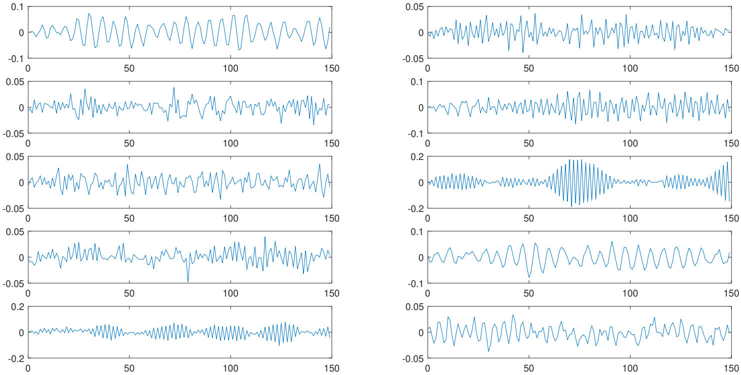
$$x[n] = -\sum_{i=1}^{p} a_i x[n-i] + b[n]$$

- $p$ : ordre du modèle
- $a_i, \ldots, a_p$ : coefficients du modèle
- $b[n]$ : bruit blanc

# Propriétés

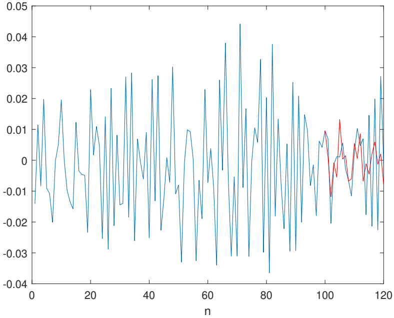
- Un signal $AR(p)$ peut être stationnaire pour certaines configurations de coefficients $a_i$ : on se placera généralement sous cette hypothèse
- Les coefficients $a_i$ peuvent être utilisés pour l'indexation et la prédiction.
- Ce modèle est également lié à un modèle physique dit source/filtre où l'on suppose que l'on excite un filtre linéaire grâce à un bruit blanc : il est très utilisé pour modéliser des signaux de parole
- Ces coefficients permettent également de fournir une approximation de la DSP du signal

# Exemples




Plusieurs signaux $AR(5)$

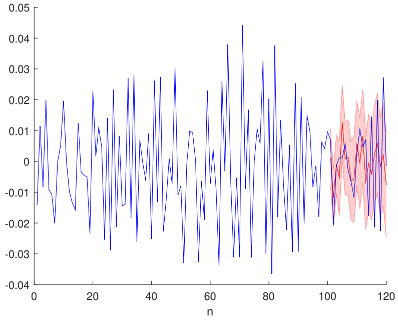
# Prédiction

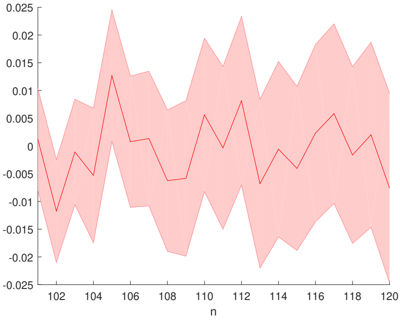



Estimation des paramètres AR sur les 100 premiers échantillons et prédiction des 20 suivants par récursivité

$$\hat{x}[n] = -\sum_{i=1}^{p} \hat{a}_i x[n-i]$$

# Prédiction




Estimation des paramètres AR sur les 100 premiers échantillons et prédiction des 20 suivants par récursivité

$$\hat{x}[n] = -\sum_{i=1}^{p} \hat{a}_i x[n-i] + b[n]$$

avec 1000 réalisations différentes de $b[n]$

# Estimation des paramètres

- Pour estimer les paramètres du modèle, il faut revenir à l'équation de base

$$x[n] + a_1 x[n-1] + \ldots + a_p x[n-p] = b[n]$$

- En multipliant par $x[n-1]$ on obtient

$$x[n]x[n-1] + a_1 x[n-1]^2 + \ldots + a_p x[n-p]x[n-1] = b[n]x[n-1]$$

- En prenant l'espérance on obtient

$$\mathbb{E}\left[x[n]x[n-1]\right] + a_1 \mathbb{E}\left[x[n-1]^2\right] + \ldots + a_p \mathbb{E}\left[x[n-p]x[n-1]\right] = \mathbb{E}\left[b[n]x[n-1]\right]$$

- Comme $b[n]$ et $x[n-1]$ sont décorrélés et que $b[n]$ est un bruit blanc

$$\mathbb{E}\left[b[n]x[n-1]\right] = \mathbb{E}\left[b[n]\right]\mathbb{E}\left[x[n-1]\right] = 0$$

- Si l'on suppose que $x[n]$ est stationnaire on a

$$\mathbb{E}\left[x[n_1]x[n_2]\right] = \gamma_x[|n_1 - n_2|]$$

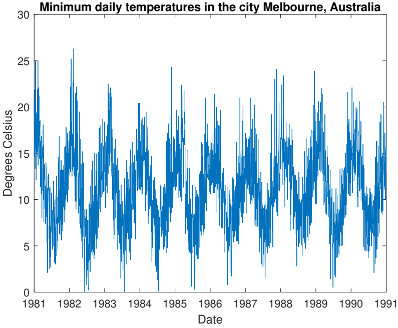
# Estimation des paramètres

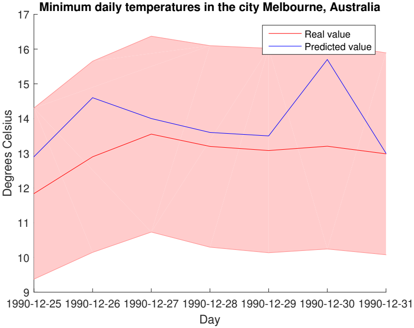
- Au final on a :
$$\gamma_x[1] + a_1\gamma_x[0] + \ldots + a_p\gamma_x[p-1] = 0$$
- Le même procédé peut être réitéré en multipliant l'équation de base par $x[n-2], x[n-3], \ldots$. On arrive alors à un système d'équations :

$$\begin{bmatrix} \gamma_x[0] & \gamma_x[1] & \cdots & \gamma_x[p-1] \\ \gamma_x[1] & \gamma_x[0] & \cdots & \gamma_x[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_x[p-1] & \gamma_x[p-2] & \cdots & \gamma_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \gamma_x[1] \\ \gamma_x[2] \\ \vdots \\ \gamma_x[p] \end{bmatrix}$$

- La connaissance de la fonction d'autocorrélation du signal $\gamma_x[m]$ pour $m = 0 \ldots p$ est suffisante pour estimer les paramètres
- On peut utiliser un estimateur empirique lorsque la fonction d'autocorrélation n'est pas connue

# Exemple




Prédiction avec un modèle $AR(29)$

# Plan du cours

# Évaluation des modèles

- Afin d'évaluer la capacité d'un modèle à rendre compte du signal, on commence par estimer les paramètres du modèle puis on calcule le signal ainsi prédit

$$\hat{x}[n] = \sum_{i=1}^{\hat{N}_h} \hat{a}_i \sin\left(2\pi i \hat{f}_0 \frac{n}{Fe} + \hat{\phi}_i\right)$$

$$\hat{x}[n] = \underbrace{\hat{\alpha}_1 \beta_1 (nT_e) + \ldots + \hat{\alpha}_j \beta_j (nT_e)}_{\text{tendance}} + \underbrace{\hat{\alpha_{j+1}} \beta_{j+1} (nT_e) + \ldots + \hat{\alpha}_d \beta_d (nT_e)}_{\text{saisonnalité}}$$

$$\hat{x}[n] = -\sum_{i=1}^{p} \hat{a}_i x[n-i]$$

# Évaluation des modèles

- On évalue ensuite l'adéquation du modèle grâce à la Root Mean Square Error (RMSE) définie par

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n}\left(x[n] - \hat{x}[n]\right)^2}$$

- Intuitivement plus on augmente le nombre de paramètres, meilleure sera la modélisation : d'autres critères (BIC, AIC) existent pour réaliser le meilleur compromis entre erreur de modélisation et nombre de paramètres à apprendre

# Références


- D. Chafaï, C. Lévy-Leduc, A. Roche. Introduction aux séries temporelles. Master 1 Dauphine Université Paris
  https://www.ceremade.dauphine.fr/~roche/Enseignement/Series_temp/cours.pdf
- A. Charpentier. Modèles de prévision - Séries temporelles. UQAM, ACT6420.
  http://freakonometrics.free.fr/uqamts.pdf
- V. Monbet. Modélisation de séries temporelles, Université de Rennes
  https://perso.univ-rennes1.fr/valerie.monbet/Cours_ST/CoursST2010.pdf
- G. Peeters. Traitement du signal musical. ENSEA 3e.
  http://recherche.ircam.fr/anasyn/peeters/pub/cours/20171017_Peeters_20172018_ENSEA_3SYM_Cours_ModeleEstimationDescripteursApprentissage.pdf
- J.Y. Dauxois. Introduction à l'étude des séries temporelles, INSA Toulouse
  https://perso.math.univ-toulouse.fr/jydauxoi/files/2017/04/poly_eleves.pdf
- R.J. Hyndman, G. Athanasopoulos. Forecasting : Principles and Practice. Book-online, 2018.