Les Séries Temporelles

Emmanuel César & Bruno Richard
Université de Versailles Saint-Quentin-en-Yvelines

Module XML et Data Mining - Mars 2006

Table des matières

Introduction

Les séries temporelles constituent une branche de l'économétrie dont l'objet est l'étude des variables au cours du temps. Parmi ses principaux objectifs figurent la détermination de tendances au sein de ces séries ainsi que la stabilité des valeurs (et de leur variation) au cours du temps. On distingue notamment les modèles linéaires (principalement AR et MA, pour Auto-Regressive et Moving Average) des modèles conditionnels (notamment ARCH, pour Auto-Regressive Conditional Heteroskedasticity).

L'analyse de ces séries touche énormément de domaines de la vie professionnelle, et plus précisément celui de l'informatique décisionnelle. L'image que l'on pourrait se faire de cette analyse ressemblerait à un homme très âgé avec beaucoup d'expérience et une sagesse assez grande pour tirer des événements passés des indications sur le futur, une sorte d'oracle. En informatique, ce serait plutôt une structure fondée sur les bases de données, fournissant ainsi le volume nécessaire d'information permettant de dresser une chronique historique des événements passés. Dessus viendrait se greffer un protocole d'extraction des données, intégré suivant un modèle judicieusement adapté à l'analyse que l'on voudrait faire. Enfin, au sommet de cette pyramide, la réponse à la question posé au départ, qui sera la prévision.

Afin de pouvoir bien appréhender les séries temporelles, l'article débutera par une approche assez générale (Partie2 *Définitions&Explications*), puis s'attardera sur les notions mathématiques indispensables à la compréhension de celles-ci (Partie3 *Concepts Mathématiques pour aborder les séries temporelles*). On s'intéressera ensuite aux méthodes courantes (Partie4 *Quelques méthodes courantes*), pour poursuivre par la présentation de quelques modèles (Partie5 *Les Algorithmes*), et terminer sur une conclusion.

2. Définitions & Explications

2.1 Qu'appelle-t-on série temporelle ?

Contrairement à l'économétrie traditionnelle, le but de l'analyse des séries temporelles n'est pas de relier des variables entre elles, mais de s'intéresser à la « dynamique » d'une variable. Cette dernière est en effet essentielle pour deux raisons : les avancées de l'économétrie ont montré qu'on ne peut relier que des variables qui présentent des propriétés similaires, en particulier une même stabilité ou instabilité ; les propriétés mathématiques des modèles permettant d'estimer le lien entre deux variables dépend de leur dynamique.

Définition (Série Temporelle) *La suite d'observations* $(y_t, t \in T)$ *d'une variable y à différentes dates t est appelée série temporelle. Habituellement, T est dénombrable, de sorte que* $t=1,\dots T$.

Une série temporelle est donc toute suite d'observations correspondant à la m^me variable : il peut s'agir de données macroéconomiques (le PIB d'un pays, l'inflation, les exportations…), microéconomiques (les ventes d'une entreprise donnée, son nombre d'employés, le revenu d'un individu, le nombre d'enfants d'une femme…), financières (le CAC40, le prix d'une option d'achat ou de vente, le cours d'une action), météorologiques (la pluviosité, le nombre de jours de soleil par an…), politiques (le nombre de votants, de voix reçues par un candidat…), démographiques (la taille moyenne des habitants, leur âge…). En pratique, tout ce qui est chiffrable et varie en fonction du temps. La dimension temporelle est ici importante car il s'agit de l'analyse d'une chronique historique : des variations d'une même variable au cours du temps, afin de pouvoir comprendre la dynamique. La périodicité de la série n'importe en revanche pas : il peut s'agir de mesures quotidiennes, mensuelles, trimestrielles, annuelles… voire même sans périodicité.

On représente en général les séries temporelles sur des graphiques de valeurs (ordonnées) en fonction du temps (abscisses). Lorsqu'une série est stable autour de sa moyenne, on parle de **série stationnaire**. Inversement, on trouve aussi des **séries non stationnaires**. Lorsqu'une série croît sur l'ensemble de l'échantillon et donc possède une moyenne qui n'est pas constante, on parle de **tendance**. Enfin lorsqu'on observe des phénomènes qui se reproduisent à des périodes régulières, on parle de **phénomène saisonnier**.

2.2 Quels sont les buts de cette analyse ?

Parmi les multiples applications de l'analyse des séries temporelles, il est possible d'en distinguer neuf principales.

2.2.1 Prévoir

La fonction première pour laquelle il est intéressant d'observer l'historique d'une variable vise à en découvrir certaines régularités afin de pouvoir établir une prévision. Il s'agit ici de supposer que les mêmes causes produisent les mêmes effets. Avec une analyse fine, il est même possible d'établir des prévisions "robustes" vis-à-vis de ruptures brusques et de changements non anticipables.

2.2.2 Relier les variables

Il s'agit ici de créer des liens entre des variables, afin d'établir des comparaisons ainsi que des corrélations. Ainsi, on va pouvoir écarter certaines relations qui ne présentent aucun

sens avec la série, ou au contraire associer d'autres relations qui interagissent avec la série observée.

2.2.3 Déterminer la causalité

Pour qu'un mouvement un provoque un autre, il est nécessaire qu'il le précède. Ainsi deux évènements similaires révèlent l'existence probable d'une source commune. L'utilisation de retards d'une variable, va permettre a partir des valeurs aux périodes précédentes de deviner la durée de transmission entre une source et son effet.

2.2.4 Etudier des anticipations des agents

L'idée que l'on se fait de l'avenir peut intervenir dans certaines équations. Il faut donc dans certaines équations faire intervenir des valeurs avancées des variables, en utilisant la manière dont elles ont été formées dans le passé.

2.2.5 Repérer les tendances et cycles

Grâce aux tendances et aux cycles, il est ainsi possible d'analyser les interactions entres diverses variables, afin d'atteindre un équilibre.

2.2.6 Corriger des variations saisonnières

En comparant le niveau saisonnier entre deux années par exemple, on va pouvoir en déduire un comportement. Celui-ci apportera des informations supplémentaires indispensable afin d'affiner les valeurs saisonnières, et appréhender leurs évolutions.

2.2.7 Détecter les chocs structurels

Un choc structurel est défini comme une modification permanente ou temporaire de la façon dont est générée une variable. Ils sont fréquents, souvent non anticipables et difficiles à mesurer. Cependant il est nécessaire de savoir qu'une rupture a eu lieu, car sa présence change les interactions et les équilibres.

2.2.8 Contrôler les processus

Il est indispensable de dresser une carte des variables ayant une forte influence sur les reste de l'économie, afin d'anticiper les évolutions possibles.

2.3 En quoi cette démarche consiste-t-elle ?

2.3.1 But

Le but poursuivi est la formulation d'un modèle statistique qui soit une représentation congruente du processus stochastique (inconnu) qui a généré la série observée ? Tout comme un probabilités/statistiques, il faut bien comprendre la différence entre le processus qui génère des données, sa réalisation telle qu'on l'observe sur les échantillons historiques à notre disposition, les futures réalisations et le modèle qu'on construit afin de tâcher de le représenter. Par représentation congruente, on entend un modèle qui soit conforme aux données sous tous les angles mesurables et testables.

2.3.2 Approche

Il est en pratique impossible de connaître la distribution d'une série temporelle $\{y_t\}_{t\geq 0}$, on s'intéresse par conséquent à la modélisation de la distribution conditionnelle (à priori constante dans le temps) de $\{y_t\}$ via sa densité :

$$f(y_t|Y_{t-1})$$

Conditionnée sur l'historique du processus : $Y_{t-1} = (y_{t-1}, y_{t-2},\ldots,y_0)$. Il s'agit donc d'exprimer y_t en fonction de son passé.

2.3.3 Résultat

L'approche conditionnelle fournit une Décomposition Prévision Erreur, selon laquelle :

$$y_t = E[y_t|Y_{t-1}] + \varepsilon_t$$

Où $E[y_t|Y_{t-1}]$ est la composante de y_t qui peut donner lieu à une prévision, quand l'historique du processus, Y_{t-1} est connu.

Et ε_t représente les informations imprévisibles.

Exemple (Modèles de séries temporelles)

1. *Processus autorégressifs d'ordre 1, AR(1) :*

$$y_t = ay_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim WN(0,\sigma^2) \text{ (bruit blanc)}$$

La valeur y_t ne dépend que de son prédécesseur. Ses propriétés sont fonctions de α qui est un facteur d'inertie :

- *quand $\alpha = 0$: y_t est imprévisible et ne dépend pas de sont passé, on parle de bruit blanc*
- *si $\alpha \in]-1,1]$: y_t est stable autour de zéro*
- *si $|\alpha| = 1$: y_t est instable et ses variations $y_t - y_{t-1}$ sont imprévisibles*
- *si $|\alpha| < 1$: y_t est explosif*

2. *Séries multivariées :*

$$y_t = Ay_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim WN(0,\sum)$$

3. *Processus autorégressif vectoriel, VAR(1) :*

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},$$
$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim \text{WN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2{}_1 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right).$$

4. *Modèle autorégressif à retards distribués, ADL :*

$$y_{1,t} = a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + \varepsilon_{1t}$$
$$\varepsilon_{1t} \sim WN(0,\sigma_1^2)$$

3. Concepts mathématiques pour aborder les séries temporelles

3.1 Variables aléatoires

Soit (Ω,M,P) un espace de probabilité, où Ω est l'espace des évènements, M est une tribu adaptée à Ω (c'est l'ensemble qui contient les combinaisons possibles d'évènements) et P est une mesure de probabilité définie sur M.

Définition Une variable aléatoire réelle (v.a.r) est une fonction à valeurs réelles $y : \Omega \rightarrow R$ *telle que pour tout réel c,* $A_c = \{\omega \in \Omega | y(\omega) \leq c\} \in M$.

En d'autres termes, A_c est un événement dont la probabilité est définie en termes de P. La fonction $F : R \rightarrow [0,1]$ définie par $F(c) = P(A_c)$ est la *fonction de distribution* de y.

3.2 Processus stochastiques

Soit T en ensemble d'indexation dénombrable contenu dans l'ensemble des entiers naturels ou dans celui des entiers relatifs.

Définition *Un processus stochastique (discret) est une fonction à valeurs réelles*

$$y : T \times \Omega \rightarrow R$$

Telle que pour tout $t \in T$ *donné,* $y_t(.)$ *soit une variable aléatoire.*

En d'autres termes, un processus stochastique est une suite ordonnée de variables aléatoires $\{y_t(\omega),\omega \in \Omega,t \in T\}$, telle que pour tout $t \in T$, y_t soit une variable aléatoire sur Ω et que pour tout $\omega \in \Omega$, $y_t(\omega)$ soit une réalisation du processus stochastique sur l'ensemble d'indexation T.

Définition *Une série temporelle* $\{y_t\}_{t=1}^{T}$ *est une réalisation d'un processus stochastique* $\{y_t\}$.

3.3 Stationnarité

Définition *Le processus* $\{y_t\}$ *est dit* ***stationnaire au sens faible****, ou* ***stationnaire au second ordre*** *si les premier (moyenne ou espérence mathématique) et second (variance et autocovariances) moments du processus existent et sont indépendants de t.*

La stationnarité est une propriété de stabilité, la distribution de y_t est identique à celle de y_{t-1}. La série oscille autour de sa moyenne avec une variance constante. Le lien entre y_t et y_{t-h} ne dépend alors que de l'intervalle h et non de la date t.

Définition *Le processus* $\{y_t\}$ *est dit* ***strictement ou fortement stationnaire*** *si pour tous* $h_1,\ldots,h_n$*, la distribution jointe de* $(y_t,y_{t+h},...,y_{t+hn})$ *dépend uniquement des intervalles* $h_1,\ldots,h_n$ *et non de t.*

$$f(y_t,y_{t+h1},\ldots,y_{t+hn}) = f(y_T,y_{T+h1},\ldots,y_{T+hn})$$

La stationnarité stricte implique que tous les moments soient indépendants du temps.

3.4 Quelques processus courant

Définition *Un* ***bruit blanc*** *(white noise) est un processus stationnaire au sens faible de moyenne zéro et qui est dynamiquement non corrélé.*

$$u_t \sim WN(0,\sigma^2)$$

Ainsi $\{u_t\}$ *est un bruit blanc si pour tout* $t \in T$: $E[u_t] = 0$, $E[u_t^2] = \sigma^2 < \infty$, *avec* u_t *et* u_{t-h} *indépendants si* $h \neq 0$, t *et* $(t-h) \in T$.

Définition *Si un bruit blanc* $\{u_t\}$ *est distribué Normalement, on parle de* ***bruit blanc Gaussien*** *:*

$$u_t \sim NID(0,\sigma^2)$$

L'hypothèse d'indépendance est alors équivalente à celle de non corrélation :
$E[u_t u_{t-h}] = 0$ *si* $h \neq 0$, t *et* $(t-h) \in T$.

4. Les méthodes courantes

4.1 Extrapolation déterministe des séries

Les modèles sont dits déterministes lorsque leurs valeurs futures sont connues avec certitude à tout instant. Elles ne font donc pas référence aux sources d'incertitudes des processus stochastiques.

Si on dispose d'un échantillon de T observations d'une série : $y_1,y_2,\ldots,y_{T-1},y_T$, il existe un polynôme de degré n = T-1 qui passe par tous les points y_t :

$$f(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n$$

4.1.1 Tendances Linéaires

Un caractéristique simple de y_t est sa tendance de long terme : si on pense qu'une tendance à la hausse existe et va perdurer, il est possible de construire un modèle simple qui va permettre de prévoir y_t. Le plus simple consiste en une **tendance linéaire** selon laquelle la série va s'accroître du même montant à chaque période :

$$y_t = a + bt$$
$$\Delta y_t = y_t - y_{t-1} = b$$
$$y_{T+h} = a + b(T+h)$$

Il peut sembler plus réaliste de penser que y_t va s'accroître du même pourcentage à chaque période, auquel cas une tendance exponentielle s'impose :

$$y_t = Ae^{rt}$$

4.1.2 Tendances autorégressives

Ici la valeur de t dépend de la valeur précédente :

$$y_t = a + by_{t-1}$$

Selon les valeurs de b et a, le comportement de la série diffère. Si a = 0 et $|b| \neq 1$, b est le taux de croissance de la série, et si b = 1, y_t suit une tendance déterministe.

4.2 Moyennes Mobiles

Il existe deux types de moyenne mobile, l'un correspond au **modèle MA** qui sera étudié plus loin et l'autre est davantage une **méthode ad hoc** permettant de donner une estimation d'une série. On suppose alors que la variable sera proche de sa moyenne récente. Une moyenne mobile est alors simplement une moyenne sur une fenêtre glissante d'observations :

$$\overline{y}_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} y_{t+k-i}$$

où k est librement fixé selon les besoins du modélisateur, pour une prévision, il est nécessaire que $k \leq 0$.

Mais il peut paraître peu réaliste que la prochaine valeur y_{T+1} puisse être proche d'une simple moyenne des dernières observations. Si on souhaite accorder plus de poids aux observations les plus récentes, on peut alors utiliser le modèle EWMA (Exponentially Weighted Moving Average) selon lequel :

$$\begin{aligned}\widehat{y}_{T+1} &= \alpha y_T + \alpha(1-\alpha)\, y_{T-1} + \alpha(1-\alpha)^2 y_{T-2}\ldots. \\ &= \alpha\sum_{i=0}^{\infty}(1-\alpha)^i y_{T-i},\end{aligned}$$

où α est compris entre 0 et 1 et indique l'importance accordée aux observations les plus récentes. Si α = 1 :

$$\widehat{y}_{T+1} = y_T$$

Notons qu'il s'agit bien d'une moyenne puisque la somme des coefficients est unitaire :

$$\alpha \sum_{i=0}^{\infty} (1-\alpha)^i = 1$$

Le modèle EWMA se prête mal aux variables présentant une tendance de fond à la hausse ou à la baisse, car il va dans ces cas sous- ou sur-prédire. Il est en revanche possible de l'appliquer à une série dont on a ôté la tendance.

Pour une prévision à horizon h > 1, il semble logique d'étendre

$$\widehat{y}_{T+h} = \alpha \sum_{i=1}^{h-1} (1-\alpha)^{i-1} \widehat{y}_{T+h-i} + \alpha \sum_{i=0}^{\infty} (1-\alpha)^{h-1+i} \widehat{y}_{T-i}$$

ce qui donne

$$\widehat{y}_{T+h} = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i y_{T-i}$$

et ainsi le modèle EWMA fournit la même prévision à tous horizons.

4.3 Lissage

Les méthodes de lissage ont pour but de retirer ou de réduire les fluctuations (cycliques ou non) de court terme des séries. Les deux méthodes les plus employées pour lisser une série sont les moyennes mobiles et le lissage exponentiel.

4.3.1 Moyennes mobiles

Les moyennes mobiles présentées précédemment permettent aussi d'obtenir des séries lissées : par exemple en utilisant une moyenne mobile d'ordre n données par :

$$\widetilde{y}_t = \frac{1}{n}\sum_{i=0}^{n-1} y_{t-i}.$$

Plus n est élevé, plus la série sera lissée. Le problème est de n'utiliser que les valeurs passées et présentes. Pour y remédier, on peut faire appel à une moyenne mobile centrée :

$$\widetilde{y}_t = \frac{1}{2k+1}\sum_{i=-k}^{k} y_{t+i}$$

4.3.2 Lissage exponentiel

Le lissage exponentiel fait appel aux modèles EWMA :

$$\widetilde{y}_t = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + ...\alpha(1-\alpha)^{t-1} y_1.$$

En pratique, il est plus facile d'écrire :

$$(1-\alpha)\widetilde{y}_{t-1} = \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + ...\alpha(1-\alpha)^{t-1} y_1$$

En soustrayant ces 2 équations on obtient la formule de récurrence du lissage exponentiel simple :

$$\widetilde{y}_t = \alpha y_t + (1-\alpha)\widetilde{y}_{t-1}$$

Plus α est proche de zéro, plus la série est lissée. En pratique toutefois, on peut souhaiter effectuer un lissage important mais sans donner trop de poids aux observations lointaines. On applique pour ce faire un lissage exponentiel double pour obtenir :

$$\widetilde{\widetilde{y}}_t = \alpha\widetilde{y}_t + (1-\alpha)\widetilde{\widetilde{y}}_{t-1}$$

avec une valeur plus élevée de α.

Enfin il est possible d'appliquer cette formule aux changements moyens de la tendance de long terme de la série en utilisant la formule de lissage exponentiel à deux paramètres de Holt-Winters :

$$\begin{aligned} \widetilde{y}_t &= \alpha y_t + (1-\alpha)(\widetilde{y}_{t-1} + r_{t-1}) \\ r_t &= \gamma(\widetilde{y}_t - \widetilde{y}_{t-1}) + (1-\gamma) r_{t-1}, \end{aligned}$$

où r_t est la série lissée représentant la tendance, i.e. le taux moyen de croissance.

Cette tendance est ajoutée lors du lissage afin d'éviter que le lissage exponentiel de y_t ne s''eloigne trop des valeurs récentes de la série originale y_t. Une prévision à horizon h peut être obtenue en posant

$$\widehat{y}_{T+h} = \widetilde{y}_T + hr_T$$

4.4 Ajustements saisonniers

Il existe diverses méthodes de correction des variations saisonnières. Elles fonctionnent pour la plupart sur une décomposition entre tendance et variations saisonnières de la forme :

$$Yt = L \times S \times C \times I$$

Avec L la valeur de long terme, S le coefficient saisonnier, C le cycle saisonnier, et I une composante irrégulière. Le but est d'isoler S x I.

4.5 Les équations de Yule-Walker

En l'absence de composante MA (q = 0 dans ARMA(p,q) la méthode à utiliser correspond aux moindres carrés ordinaires ou résolution des équations de Yule-Walker :

$$\lambda 1,\ldots,\lambda_p \quad \begin{cases} \lambda_1 = \alpha_1 + \alpha_2\lambda_1 + \ldots + \alpha_p\lambda_{p-1} \\ \lambda_2 = \alpha_1\lambda_1 + \alpha_2 + \ldots + \alpha_p\lambda_{p-2} \\ \ldots \\ \lambda_p = \alpha_1\lambda_{p-1} + \alpha_{p-2} + \ldots + \alpha_p \end{cases}$$

$$\lambda_k = \alpha_1\lambda_{k-1} + \alpha_{k-2} + \ldots + \alpha_{k-p}\text{, pour } k>p$$

en remplaçant les auto corrélations théoriques par leurs estimateurs

5 Les Algorithmes

5.1 Présentation générale des modèles usuels

Voici une liste non exhaustive des modèles couramment utilisés dans les séries temporelles :

- ARIMA(Box&Jenkins) and Autocorrelations
- Interrupted Time Series ARIMA
- Exponential Smoothing
- Seasonal Decomposition (Census1)
- X-11 Census method II seasonal adjustement
- Distributed Lags Analysis
- Single Spectrum (Fourier) Analysis
- Cross Spectrum Analysis
- Spectrum Analysis
- Fast Fourier Transformations

5.2 Fonctionnement de l'algorithme intégré dans SQL Server 2005

L'algorithme est en fait une version hybride d'autorégression et des techniques des arbres de décision.

5.2.1 Autorégression

Une des étapes clés de l'algorithme ART (Auto Regression Tree) est la transformation des cases simples d'une série temporelle en plusieurs cases interne :

Mois	Lait	Pain
Jan-2005	**5000**	**4500**
Fev-2005	**5200**	**4600**
Mar-2005	**5240**	**5130**
Avr-2005	**6390**	**6280**
Mai-2005	**6750**	**6160**
Jui-2005	**6280**	**6560**
Juy-2005	**7680**	**7200**
…		

Tableau avant Case Transform

CaseId	Lait(t-2)	Lait(t-1)	Lait(t0)	Pain(t-2)	Pain(t-1)	Pain(t0)
1	**5000**	**5200**	**5240**	**4500**	**4600**	**5130**
2	**5200**	**5240**	**6390**	**4600**	**5130**	**6280**
3	**5240**	**6390**	**6750**	**5130**	**6280**	**6160**
4	**6390**	**6750**	**6280**	**6280**	**6160**	**6560**
5	**6750**	**6280**	**7680**	**6160**	**6560**	**7200**
…						

Tableau après Case Transform

Dans l'ART, la méthode « Case Transform » utilise par défaut les 8 valeurs précédentes. Le principal avantage de cette méthode c'est qu'elle regroupe dans une même table toutes les séries temporelles utilisant le même modèle (ici typiquement le Lait et le Pain sont les variables).

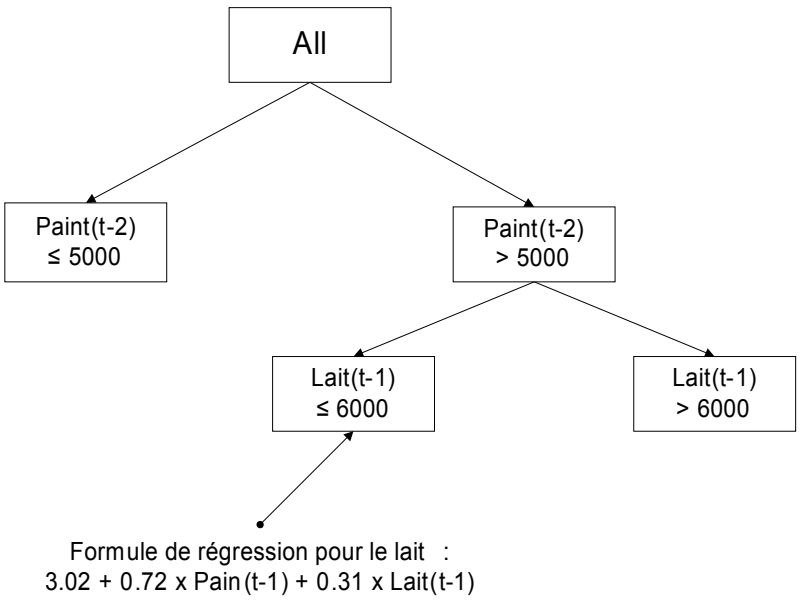
Ensuite, le but va être de trouver la fonction f, fonction linéaire possédant des cœfficients d'autorégression, et établi comme dans les modèles précédents, en fonction de son passé. Le processus va consister en un système d'équations linéaires, qui va être résolu grâce aux équations de Yule-Walker. Cela va nous permettre de calculer les coefficients d'autorégression, grâce à la matrice de covariance ainsi obtenue.

$$\begin{pmatrix} 1 & r_1 & r_2 & r_3 & r_4 & \dots & r_{n-1} \\ r_1 & 1 & r_1 & r_2 & r_3 & \dots & r_{n-2} \\ r_2 & r_1 & 1 & r_1 & r_2 & \dots & r_{n-3} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ r_{n-1} & r_{n-2} & r_{n-3} & r_{n-4} & r_{n-5} & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ . \\ . \\ . \\ r_n \end{pmatrix}$$

Un autre avantage de l'ART, et qui n'est pas anodin, c'est qu'il reconnaît les séries croisées. Ainsi il va pouvoir « Relier les variables », grâce aux concepts mathématiques et aux méthodes courantes déjà cités.

5.2.2 Arbre d'autorégression

La fonction f défini dans la partie 5.2.1, correspond a un arbre de régression. Une représentation que l'on pourrait en avoir serait la suivante :



Comme dans les arbres de décisions, il va falloir choisir par diverses méthodes le nombre de nœud maximum possible, ainsi que la hauteur à ne pas dépasser. En descendant ensuite dans cet arbre, on va ainsi atteindre la feuille la plus adaptée, qui va nous permettre d'approcher la valeur recherchée. Après un calcul par une formule de régression, celle-ci sera la valeur prédite que l'on remplacera dans la table.

5.2.3 Saisonnalité

Pour traiter ce phénomène dans l'ART, SQL Server 2005 va utiliser un paramètre de saisonnalité appelé **Periodicity_Hint**. Ainsi, pendant l'étape « Case Transform », l'algorithme va ajouter des points de donnée basé sur ce paramètre (défini par l'utilisateur). Par exemple, si la période de saisonnalité est de 12mois pour le Lait et le Pain, l'algorithme va ajouter dans la table les valeurs Lait(t-8x12) … Lait(t-24), Lait(t-12), Pain(t-8x12) … Pain(t-24), Lait(t-12).

De plus, il est possible de spécifier plusieurs Periodicity_Hint, et si on ne spécifie aucune périodicité, l'algorithme va essayer de détecter automatiquement la saisonnalité (algorithme basé sur la méthode Fast Fourier Transform)

5.3 Fonctionnement de la méthode ARIMA

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que **les données sont une fonction du temps** ($y = f(t)$). Cette catégorie de modèle peut être ajustée par la méthode des moindres carrés, ou d'autres méthodes itératives. L'analyse des modèles par transformée de Fourier est une version sophistiquée de ce type de modèle.

Une seconde catégorie de modèles cherche à **déterminer chaque valeur de la série en fonction des valeurs qui la précède** ($y_t = f(y_{t-1}, y_{t-2}, \ldots)$). **C'est le cas des modèles ARIMA** ("Auto - Regressive – Integrated – Moving Average"). Cette catégorie de modèles a été popularisée et formalisée par Box et Jenkins (1976).

5.3.1 Définitions

Les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents, plus un terme aléatoire d'erreur.

Le processus d'intégration suppose que chaque point présente une différence constante avec le
point précédent.

Les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédant, plus sa propre erreur.

5.3.2 Typologie du modèle

Un modèle ARIMA est étiqueté comme modèle ARIMA (p,d,q), dans lequel:
p est le nombre de termes auto-régressifs
d est le nombre de différences
q est le nombre de moyennes mobiles.

5.3.2 Analyse du modèle

5.3.2.1. Premier critère : la différenciation.

L'estimation des modèles ARIMA suppose que l'on travaille sur une série stationnaire. Ceci signifie que la moyenne de la série est constante dans le temps, ainsi que la variance. La meilleure méthode pour éliminer toute tendance est de différencier, c'est-à-dire de remplacer la série originale par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme *Integrated*).

La correction d'une non-stationnarité en termes de variance peut être réalisée par des transformations de type logarithmique (si la variance croît avec le temps) ou à l'inverse exponentielle. Ces transformations doivent être réalisées avant la différenciation. Une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de y est constante.

$$y_t - y_{t-1} = \mu + \varepsilon_t$$

μ est la constante du modèle, et représente la différence moyenne en y. Un tel modèle est un ARIMA(0,1,0). Il peut être représenté comme un accroissement linéaire en fonction du temps. Si μ est égal à 0, la série est stationnaire.

Les modèles d'ordre 2 travaillent non plus sur les différences brutes, mais sur les différences de différence. La seconde différence de y au moment t est égale à $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, c'est-à dire à $y_t - 2y_{t-1} + y_{t-2}$.

Un modèle ARIMA(0,2,0) obéira à l'équation de prédiction suivante :

$$y_t - 2y_{t-1} + y_{t-2} = \mu + \varepsilon_t \text{ ou encore } y_t = \mu + 2y_{t-1} - y_{t-2} + \varepsilon_t$$

5.3.2.2. Deuxième critère : l'auto-régression

Les modèles autorégressifs supposent que yt est une fonction linéaire des valeurs précédentes.

$$y_t = \mu + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \varphi_3 y_{(t-3)} + \varepsilon_t$$

Littérairement, chaque observation est constituée d'une composante aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des observations précédentes. φ_1, φ_2 et φ_3 dans cette équation sont les coefficients d'auto-régression

A noter que cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA(1,1,0) on aura :

$$y_t - y_{t-1} = \mu + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t$$

Notez qu'un processus autorégressif ne sera stable que si les paramètres sont compris dans un certain intervalle ; par exemple, s'il n'y a qu'un paramètre autorégressif, il doit se trouver dans l'intervalle $-1<\varphi_1<+1$.

Dans les autres cas, les effets passés s'accumuleraient et les valeurs successives des x_t se déplaceraient infiniment vers l'avant, ce qui signifie que la série ne serait pas stationnaire.

S'il y a plus d'un paramètre autorégressif, des restrictions similaires (générales) sur les valeurs des paramètres peuvent être posées (par exemple, voir Box et Jenkins, 1976 ; Montgomery, 1990).

5.3.2.3. Troisième critère : la moyenne mobile

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne. On considère alors que la meilleure estimation est représentée par la moyenne pondérée d'un certain nombre de valeurs antérieures (ce qui est le principe des procédures de moyennes mobiles utilisées pour le lissage des données). Ceci revient en fait à considérer que l'estimation est égale à la moyenne vraie, auquel on ajoute une somme pondérée des erreurs ayant entaché les valeurs précédentes :

$$y_t = \mu - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \theta_3 \varepsilon_{(t-3)}. + \varepsilon_t.$$

Littérairement, chaque observation est composée d'une composante d'erreur aléatoire (choc aléatoire, ε) et d'une combinaison linéaire des erreurs aléatoires passées. θ_1, θ_2 et θ_3 sont les coefficients de moyenne mobile du modèle.

Comme précédemment cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA(0,1,1) on aura :

$$y_t - y_{t-1} = \mu - \theta \varepsilon_{t-1} + \varepsilon_t$$

Ce qui peut également être écrit:

$$y_t = \mu + y_{t-1} - \theta \varepsilon_{t-1} + \varepsilon_t$$

Un modèle de moyenne mobile correspond à des séries avec des fluctuations aléatoires autour d'une moyenne variant lentement. Plutôt que de prendre comme précédemment la valeur précédente comme prédicateur, on utilise une moyenne de quelques observations précédentes, de manière à éliminer le bruit, et estimer plus précisément la moyenne locale.

Cette logique correspond au **lissage exponentiel simple**, qui considère chaque observation comme la résultante d'une constante (b) et d'un terme d'erreur ε, soit :

$y_t = b + \varepsilon_t$.

La constante b est relativement stable sur chaque segment de la série, mais peut se modifier lentement au cours du temps.

Si ce modèle est approprié, l'une des manières d'isoler la réelle valeur de b, et donc la partie systématique ou prévisible de la série, consiste à calculer une sorte de moyenne mobile, ou les observations courantes et immédiatement précédentes ("les plus récentes") ont une pondération plus forte que les observations plus anciennes.

C'est exactement ce que fait un lissage exponentiel simple, où les pondérations les plus faibles sont affectées exponentiellement aux observations les plus anciennes. La formule spécifique de lissage exponentiel simple est :

$$y_t = \alpha\hat{y}_t - (1-\alpha)\, y_{t-1}$$

Lorsqu'on l'applique de façon récurrente à chaque observation successive de la série, chaque nouvelle valeur prédite est calculée comme la moyenne pondérée de l'observation courante et de l'observation précédente prédite ; la précédente observation prédite était elle-même calculée à partir de la valeur (précédente) observée et de la valeur prédite avant cette valeur (précédente), et ainsi de suite.

Par conséquent, chaque valeur prédite est une moyenne pondérée des observations précédentes, où les poids décroissent exponentiellement selon la valeur des paramètres α. Si α est égal à 1 les observations précédentes sont complètement ignorées ;

si α est égal à 0, l'observation courante est totalement ignorée, et la valeur prédite ne porte que sur les valeurs prédites précédentes (qui est calculée à partir de l'observation lissée qui lui précède, et ainsi de suite ; c'est pourquoi toutes les valeurs prédites auront la même valeur que la valeur initiale $\hat{y}_0$). Les valeurs intermédiaires de α produiront des résultats intermédiaires (noter que la valeur 1-α correspond au θ des équations précédentes).

On peut également envisager des modèles mixtes: par exemple un modèle ARIMA(1,1,1) aura l'équation de prédiction suivante:

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) - \theta_1\varepsilon_{t-1} + \varepsilon_t$$

Néanmoins on préfère généralement utiliser de manière exclusive les termes AR ou MA.

5.3.3 Signification des paramètres des modèles ARIMA

L'objectif essentiel des modèles ARIMA est de permettre une prédiction de l'évolution future d'un phénomène. Son développement dans le domaine de l'économétrie est basé sur ce principe.

Un autre intérêt, peut-être plus essentiel en ce qui concerne la recherche scientifique, est de comprendre la signification théorique de ces différents processus.

Il est clair cependant que cette interprétation dépend de la nature du phénomène étudié, et des modèles dont le chercheur dispose pour en rendre compte.

- Un **processus non différencié à bruit blanc** (ARIMA(0,0,0) suggère des fluctuations aléatoires autour d'une valeur de référence. Cette valeur de référence peut être considérée comme une caractéristique stable du système étudié (trait de personnalité, mémoire, capacité stabilisée, etc..)
- Un **processus de moyenne mobile** suggère que la valeur de référence évolue d'une mesure à l'autre. Plus précisément, la valeur de référence est fonction de la valeur de référence précédente et de l'erreur ayant entaché la mesure précédente.
- Un **processus autorégressif** suggère que le phénomène étudié n'est pas déterminé par une valeur de référence. C'est la performance précédente (ou les performances précédentes) qui déterminent entièrement la performance présente.

Par exemple, Spray et Newell (1986) analysent des données tirées d'une expérimentation portant sur le rôle de la connaissance des résultats dans l'apprentissage. Les sujets réalisent 77 essais dans une tâche manuelle. Le protocole comprenait plusieurs groupes, différenciés par des combinaisons spécifiques d'essais avec ou sans connaissance des résultats. Notamment, certains sujets disposaient de connaissance des résultats tout au long des 77 essais, pour d'autre la connaissance des résultats était supprimée au-delà de 17, 32 ou 52 essais. Un groupe n'avait pas du tout connaissance des résultats.

Les résultats de la modélisation montrent que les séries avec connaissance des résultats (ou les portions de séries avec connaissance des résultats) peuvent être représentée par des processus à bruit blanc du type:

$$y_t = \mu + \varepsilon_t$$

C'est-à-dire un modèle ARIMA (0,0,0). Cette équation suggère donc que les performances successives oscillent de manière aléatoire autour d'une valeur moyenne, sorte de référence interne construite par la connaissance des résultats.

Les séries sans connaissance des résultats (ou les portions de série sans connaissance des résultats) sont quant à elles modélisées selon un ARIMA(0,1,1) selon la formule:

$$y_t = \mu - \theta_1 \varepsilon_{(t-1)} + \varepsilon_t \text{ ou } y_t = r_t + \varepsilon_t$$

r_t représentant la valeur de référence, qui cette fois change à chaque essai. On peut dériver du modèle que :

$$r_t = r_{t-1} - \theta_1 \varepsilon_{(t-1)}$$

C'est-à-dire que la référence est une combinaison de la référence précédente et de l'erreur ayant entaché l'essai précédent. Ce modèle indique clairement que l'essai en cours est influencé par l'essai précédent, ce qui n'était pas le cas dans les essais avec connaissance des résultats.

Ce modèle peut également être écrit sous la forme d'une interpolation pondérée entre la performance au temps t et la référence au temps t-1:

$$r_t = -\theta_1 y_t + (1+\theta_1) r_{t-1}$$

L'analyse des données de Diggles (1977) suggère que la référence précédente est plus importante que la performance actuelle.

On peut noter que pour les sujets ayant bénéficié de la connaissance des résultats durant 52 essais sur 77, la série demeure stationnaire et à bruit blanc jusqu'à la fin de l'expérimentation.

5.3.4 Les différentes étapes

5.3.4.1. Détermination de l'ordre de différenciation.

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'auto corrélation décline rapidement vers zéro. Si une série présente des auto-corrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée. La différenciation tend à introduire des auto-corrélations négatives.

Si l'auto-corrélation de décalage 1 est égale à 0 ou négative, la série n'a pas besoin d'être différenciée. Si l'auto-corrélation de décalage 1 est inférieure à –0.5, la série est sur différenciée.

L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de sur différenciation.

Un troisième symptôme de sur-différenciation est un changement systématique de signe d'une observation à l'autre.

Un modèle sans différenciation suppose que la série originale est stationnaire. Un modèle avec une différenciation d'ordre 1 suppose que la série originale présente une tendance constante. Un modèle avec une différenciation d'ordre 2 suppose que la série originale présente une tendance variant dans le temps.

Les modèles ARIMA peuvent inclure une constante ou non (sans constante signifie que la constante est égale à 0). L'interprétation d'une constante (significativité statistique) dépend du modèle.

- **Un modèle sans différenciation** possède généralement une constante (qui représente dans ce cas la moyenne de la série).
- Si la série est différenciée, la constante représente la moyenne ou l'ordonnée à l'origine de la série différenciée ; par exemple, si la **série est différenciée une fois**, et qu'il n'y a pas de paramètre autorégressif dans le modèle, la constante représentera la moyenne de la série différenciée, et donc la pente du trend linéaire de la série non différenciée.
- Dans le cas **des modèles avec un ordre de différenciation de 2**, la constante représente la tendance moyenne de la tendance.

Dans la mesure où en général on ne suppose pas l'existence de telles tendances, la constante est généralement omise.

- S'il n'y a **pas de paramètre autorégressif** dans le modèle, l'espérance mathématique de la constante est m, la moyenne de la série ;
- S'il y a des **paramètres autorégressifs dans la série**, la constante représente l'ordonnée à l'origine.

A noter que la moyenne, dans les modèles ARIMA, renvoie à la moyenne des séries différenciées, alors que la constante est un facteur qui apparaît dans la partie droite des équations de prédiction. Moyenne et constante sont liées par l'équation suivante:

μ = moyenne x (1 - ΣAR(p))

La constante est égale à la moyenne, multipliée par 1 moins la somme des coefficients des termes autorégressifs.

5.3.4.2. Identification des termes AR.

Après que la série ait été stationnarisée, l'étape suivante consiste à identifier les termes AR et MA nécessaires pour corriger les auto- corrélations résiduelles. Cette analyse est basée sur l'examen des fonctions d'auto-corrélation et d'auto-corrélation partielle. Rappelons que l'auto corrélation est la corrélation d'une série avec elle-même, selon un décalage défini.

L'auto-corrélation de décalage 0 est par définition égale à 1. La fonction d'auto-corrélation fait correspondre à chaque décalage l'auto-corrélation correspondante.

D'une manière générale, une corrélation partielle entre deux variables est la quantité de corrélations qui n'est pas expliquée par les relations de ces variables avec un ensemble spécifié d'autres variables. Supposons par exemple que l'on réalise la régression de Y sur trois variables X1, X2 et X3. La corrélation partielle entre Y et X3 contrôlant X1 et X2 est la quantité de corrélation entre Y et X3 qui n'est pas expliqué par leurs relations communes avec X1et X2. Elle peut être calculée comme la racine carrée du gain de variance expliquée obtenu en ajoutant X3 à la régression de Y sur X1 et X2.

Dans le cas des séries temporelles, la corrélation partielle de décalage k est la corrélation entre y_t et y_{t-k}, contrôlant l'influence des k-1 valeurs interposées.

L'auto corrélation de décalage 1 est la corrélation entre y_t et y_{t-1} . On suppose que c'est également la corrélation entre y_{t-1} et y_{t-2} Si y_t et y_{t-1} sont corrélés, et que y_{t-1} et y_{t-2} le sont également, on peut supposer qu'une corrélation sera présente entre y et y_{t-2}. C'est-à-dire que la corrélation de décalage 1 se propage au décalage 2 et sans doute aux décalages d'ordre supérieurs. Plus précisément, la corrélation attendue au décalage 2 est la carré de la corrélation observée au décalage 1.

L'auto-corrélation *partielle* de décalage 2 est donc la différence entre l'auto-corrélation de décalage 2 et la corrélation attendue due à la propagation de la corrélation de décalage 1.

Si l'on revient à la fonction d'auto-corrélation de l'exemple précédent (avant différenciation), on peut supposer que la présence d'auto-corrélations fortes pour un grand nombre de décalages successifs est lié à ce phénomène de propagation. Ceci est confirmé par l'examen de la fonction d'auto-corrélation partielle, qui n'a qu'un valeur significative au décalage 1 (notons que l'auto-corrélation partielle de décalage 1 est égale à l'auto-corrélation correspondante, aucune valeur n'étant intercalée).

5.3.4.3.. Identification des termes MA.

La fonction d'auto-corrélation joue pour les processus de moyenne mobile le même rôle que la fonction d'auto-corrélation partielle pour les processus autorégressifs. Si l'auto-corrélation est significative au décalage k mais plus au décalage k+1, ceci indique que k termes de moyenne mobile doivent être ajoutés au modèle.

A noter que si les coefficients AR peuvent être estimés par une analyse en régression multiple, une telle démarche est impossible pour les coefficients MA. D'une part, parce que l'équation de prédiction est non-linéaire, et d'autre part les erreurs ne peuvent être spécifiées en tant que variable indépendantes. Les erreurs doivent être calculées pas à pas en fonction des estimations courantes des paramètres.

Une signature MA est généralement associée à une auto-corrélation négative au décalage 1, signe que la série est sur-différenciée. Une légère sur-différenciation peut donc être compensée par l'ajout d'un terme de moyenne mobile.

5.3.5 Conclusion

Ces deux modèles peuvent ajuster de manière alternative la série de départ. Sachant que les termes AR peuvent compenser une légère sous différenciation, et les termes MA une légère sous-différenciation, il est courant que deux modèles alternatifs soient possibles: un premier avec 0 ou 1 ordre de différenciation combiné avec des termes AR, et un autre avec le niveau de différenciation supérieur, combiné à des termes MA. Le choix d'un ou l'autre modèle peut reposer sur des présupposé théoriques liés au phénomène observé.

Les outils principaux utilisés lors de la phase d'identification sont donc les tracés de la série, les corrélogrammes d'auto corrélation (FAC), et d'auto corrélation partielle (FACP). La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres ARIMA).

Toutefois, les composantes des séries chronologiques empiriques peuvent souvent être assez bien approchées en utilisant l'un des 5 modèles de base suivants, identifiables par la forme de l'autocorrélogramme (FAC) et de l'autocorrélogramme partiel (FACP). Puisque le

nombre de paramètres (à estimer) de chaque type dépasse rarement 2, il est souvent judicieux d'essayer des modèles alternatifs sur les mêmes données.

(1) Un paramètre autorégressif (p) : FAC - décomposition exponentielle ; FACP - pic à la période 1, pas de corrélation pour les autres périodes.
(2) Deux paramètres autorégressifs (p) : FAC - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles ; FACP - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes.
(3) Un paramètre de moyenne mobile (q) : FAC - pic à la période 1, aucune corrélation pour les autres périodes ; FACP - exponentielle amortie.
(4) Deux paramètres de moyenne mobile (q) : FAC - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes ; FACP - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles.
(5) Un paramètre autorégressif (p) et un de moyenne mobile (q) : FAC – décomposition exponentielle commençant à la période 1 ; FACP - décomposition exponentielle commençant à la période 1.

6 Conclusion : l'intérêt des séries temporelles

De façon générale, il est d'usage de considérer l'intérêt des séries temporelles selon trois perspectives : descriptive, explicative et prévisionnelle.

Description

• L'analyse temporelle permet de connaître la structure de la série de données étudiée ;

• Elle peut être utilisée pour comparer une série à d'autres séries (varicelle et oreillons, par exemple) ;

Explication

• Les variations d'une série peuvent être expliquées par une autre série (exposition météorologique, pollution atmosphérique, etc.) ;

• Il est possible de modéliser une intervention externe grâce à l'analyse de séries temporelles ;

• Ces analyses permettent de réaliser des scénarios pour la période contemporaine : en agissant sur une variable explicative, il est possible d'observer le comportement de la variable expliquée ;

Prévision

La prévision *a priori* permet la planification ;

La prévision *a posteriori* permet d'estimer l'impact d'une perturbation (dépistage, par exemple) sur la variable expliquée ;

Des scénarios pour le futur, enfin, peuvent être réalisés.

Dans le domaine environnemental, le grand avantage des études de séries temporelles est d'analyser des données facilement accessibles en général car mesurées en routine (données de mortalité en population, données d'hospitalisation, données d'exposition, etc.). D'autre part, les analyses de séries temporelles, bénéficiant souvent de longues périodes de données, voient leur puissance statistique être tout à fait honorable.

Bibliographie

Des données à la connaissance (Daniel T. Larose)

Datamining et scoring (StéphaneTufféry)