

Dur, dur!! Les séries chronologiques!!

Bernard Rapacchi
Centre Interuniversitaire de Calcul de Grenoble

18 août 1993

Dans cette note, nous nous intéressons aux séries chronologiques sous le point de vue de l'Analyse des modèles Box-Jenkins (ARMA, ARIMA, SARIMA, ...), ou d'Analyse Spectrale. Nous supposons que vous avez déjà lu une petite note [8] introduisant les séries chronologiques et leurs aspects sous le point de vue Analyse Exploratoire et Méthodes Robustes. On aurait pu s'arrêter là, mais pourquoi faire simple quand on peut faire compliqué ...

Table des matières

1	INTRODUCTION AUX MODELES BOX-JENKINS	1
1.1	Introduction	1
1.2	Série Stationnaire	1
1.2.1	Définition	1
1.3	Comment rendre une série stationnaire	2
1.3.1	Pourquoi des séries sont non stationnaires	2
1.3.2	Le lissage de la série	2
1.3.3	Les opérateurs de Box-Jenkins	2
1.4	Fonction d'autocorrélation	3
2	LE MODELE LINEAIRE DE BOX-JENKINS	5
2.1	Hypothèse	5
2.2	Filtre	5
3	PROCESSUS A MOYENNE MOBILE	7
3.1	MA d'ordre 1	7
3.2	MA d'ordre q	8
3.3	En résumé	9
4	PROCESSUS AUTOREGRESSIF	10
4.1	AR d'ordre 1	10
4.2	AR d'ordre p	11
4.3	La Fonction d'Autocorrélation Partielle	12
4.3.1	ACF et PACF d'un AR(1)	12
4.3.2	PACF d'un AR(p)	14
4.4	Dualité	14
5	PROCESSUS ARMA, ARIMA ET SARIMA	15
5.1	Les processus ARMA	15
5.1.1	Définition	15
5.1.2	Processus ARMA(1,1)	15
5.2	Les processus ARIMA	16
5.3	Les processus SARIMA	16
6	UN EXEMPLE	17
6.1	Données de Pollution Atmosphérique	17
6.2	Les logiciels	17
6.3	Processus de modélisation	17

6.4	Modélisation	18
7	ETUDE MULTIVARIEE DES SERIES CHRONOLOGIQUES	25
7.1	Fonction de corrélation croisée	25
7.1.1	Définition	25
7.1.2	L'exemple de la pollution	25
7.2	Modèle de Fonction de Transfert	26
7.3	L'influence de l'inversion de température sur la pollution	27
7.4	Les influences de l'inversion et de la température	34
7.5	Prévisions	36

Chapitre 1

INTRODUCTION AUX MODELES BOX-JENKINS

1.1 Introduction

Nous supposons donc que vous savez ce qu'est une série chronologique (sinon voir [8]) et quels sont les pièges pouvant survenir face à l'étude de celles-ci. Dans la suite, nous aurons une série notée:

$$X_t \text{ avec } t = 1, \dots, n$$

C'est-à-dire que nous avons une série de n observations séquentielles ordonnées suivant leurs indices.

1.2 Série Stationnaire

1.2.1 Définition

La notion de série stationnaire est indispensable pour la suite de l'analyse des séries. Une série stationnaire X_t est une série dont les propriétés sont inchangées par changement d'origine des temps, c'est-à-dire:

1. L'espérance de X_t est constante dans le temps: on peut penser "*espérance*" comme espérance mathématique, mais il faut garder en tête que c'est la valeur espérée pour l'observation au temps t .
2. Pour tout h fixé, la covariance entre X_t et X_{t+h} est invariante dans le temps. Ceci implique, entre autres, que la série est stable en dispersion. Cette notion de décalage de la série est très importante pour la suite. On peut faire le nuage de points de la série avec les valeurs de la série décalées d'un décalage h . La stationnarité implique que, si on prend un certain nombre de points au début ou à la fin de la série on doit toujours retrouver le même nuage de points.

1.3 Comment rendre une série stationnaire

1.3.1 Pourquoi des séries sont non stationnaires

1. La première série non stationnaire à laquelle on pense est celle qui est croissante dans le temps. Par exemple la série “*Airline Data*” (cf [8]) est croissante, qu’elle soit prise telle quelle, ou transformée par Logarithme.
2. Une série saisonnière est également non stationnaire puisque la valeur espérée dépend du temps dans la période de la saison. Par exemple la série “*Airline Data*” n’est pas stationnaire également par le fait de saisonnalité.
3. Une série dont la dispersion varie dans le temps n’est pas stationnaire. Par exemple, la série des valeurs brutes “*Airline Data*” n’est pas stationnaire du fait que sa dispersion grandit dans le temps. La transformation par le Logarithme permet d’éviter ce problème.

En fait, en prenant une série brutalement, on a fort peu de chances pour qu’elle soit stationnaire.

1.3.2 Le lissage de la série

On peut d’abord lisser les données (cf [8]). On regardera alors le rugueux et celui-ci est stationnaire. Nous ne nous attarderons pas sur cet aspect de l’analyse.

1.3.3 Les opérateurs de Box-Jenkins

L’opérateur de recul B

L’opérateur de recul B est défini comme agissant sur la série. A un instant t on fait correspondre la valeur de la série à l’instant $t - 1$. On définit ainsi une nouvelle série $B X$ comme:

$$B X_t = X_{t-1}$$

On peut appliquer plusieurs fois cet opérateur, on définit ainsi des nouvelles séries:

$$B^2 X_t = B (B X_t) = B X_{t-1} = X_{t-2}$$

et

$$B^m X_t = X_{t-m}$$

L’opérateur de différenciation ∇

L’opérateur ∇ (prononcer “*nabla*”) est défini par:

$$\nabla X_t = X_t - X_{t-1} \tag{1.1}$$

Nous entrons maintenant dans des considérations de notations. En effet, par écriture purement formelle on peut écrire:

$$\nabla X_t = X_t - B X_t = (1 - B) X_t \tag{1.2}$$

On peut écrire ∇ sous la forme d’un polynôme en B avec:

$$\nabla = 1 - B$$

Ce mode d'écriture sous forme de polynôme en B est en fait très pratique mais totalement formel. Il ne faut pas oublier que, quand on écrit $(1 - B)X$, on définit à partir d'une série X , une nouvelle série qui à t , fait correspondre la différence entre la valeur de la série observée à l'instant t et celle observée à l'instant $t - 1$.

L'opérateur de désaisonnalisation ∇_s

L'opérateur ∇_s est défini par:

$$\nabla_s X_t = X_t - X_{t-s}$$

En d'autres termes:

$$\nabla_s = (1 - B^s)$$

Les effets des opérateurs de Box-Jenkins

L'opérateur ∇

1. permet d'éliminer la tendance de la série.
2. peut être répété plusieurs fois si la tendance n'est pas linéaire. Par exemple:

$$\nabla^2 X_t = (1 - B)^2 X_t = (1 - 2B + B^2) X_t$$

permet d'éliminer une tendance quadratique. Le nombre de fois où on applique ∇ est appelé *ordre de différenciation*.

L'opérateur ∇_s

1. permet d'éliminer la saisonnalité de période s .
2. On peut également l'appliquer plusieurs fois.

$$\nabla_s^2 X_t = \nabla_s (X_t - X_{t-s})$$

Le nombre de fois où on applique ∇_s est appelé *ordre de désaisonnalisation*.

1.4 Fonction d'autocorrélation

- L'autocovariance au décalage k , $\gamma(k)$ est définie par:

$$\gamma(k) = 1/n \sum_{j=1}^{n-k} (X_{j+k} - \bar{X})(X_j - \bar{X}) \text{ avec } \bar{X} = 1/n \sum_{i=1}^n X_i \quad (1.3)$$

Il faut remarquer que la somme fait intervenir $n - k$ termes mais on divise par n et non pas par $n - k$ comme on le fait dans un calcul de covariance habituel. Ceci est fait dans le but de donner à la matrice de corrélation de bonnes propriétés.

- L'autocorrélation au décalage k $\rho(k)$ est définie par:

$$\rho(k) = \gamma(k)/\gamma(0) \tag{1.4}$$

- La fonction d'autocorrélation est la fonction qui à k associe la valeur $\rho(k)$. On notera cette fonction comme étant **l'ACF** de la série X .

Chapitre 2

LE MODELE LINEAIRE DE BOX-JENKINS

2.1 Hypothèse

On suppose que notre série X_t est générée à partir d'une autre série a_t qui suit une loi de Gauss de moyenne 0 et d'écart-type σ_a , sous le forme:

$$X_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

a_t vérifie donc:

$$\left\{ \begin{array}{lcl} a & \rightsquigarrow & N(0, \sigma_a^2) \\ E(a_t) & = & 0 \quad \forall t \\ \gamma(k) & = & Cov(a_t, a_{t+k}) = \begin{cases} \sigma_a^2 & \text{si } k=0 \\ 0 & \text{si } k \neq 0 \end{cases} \end{array} \right.$$

La série a_t est un “*bruit blanc*”. C’est l’exemple parfait d’une série stationnaire puisque son espérance est toujours nulle, que sa dispersion est stable et qu’il y a indépendance entre les observations.

La valeur μ est la moyenne de la série X_t . **Nous supposons dorénavant, que la série X_t est centrée et donc que $\mu = 0$.**

L’Analyse des séries sous l’aspect Box-Jenkins consiste à chercher la série a_t et les coefficients ψ_i qui permettent de passer de la série a_t à X_t .

2.2 Filtre

De l’équation précédente, on peut écrire:

$$\begin{aligned} X_t &= a_t + \psi_1 B a_t + \psi_2 B^2 a_t + \dots \\ X_t &= (1 + \psi_1 B + \psi_2 B^2 + \dots) a_t \\ X_t &= \Psi(B) a_t \end{aligned} \tag{2.1}$$

$\Psi(B)$ est un polynôme en B dont les coefficients sont les ψ_i . Si on revient à la notion de filtre [8], la série a_t est en entrée d’un filtre dont la série X_t est en sortie et le polynôme $\Psi(B)$ est la “*fonction de transfert*” du filtre.

On peut remarquer que formellement le polynôme $\Psi(B)$ a une infinité de termes mais dans la pratique on ne dispose que de n observations et on ne peut remonter que jusqu'à la valeur $t = 1$.

Chapitre 3

PROCESSUS A MOYENNE MOBILE

En anglais, moyenne mobile se dit “*Moving Average*” c’est pourquoi nous dirons que ces séries sont des **MA**.

3.1 MA d’ordre 1

On dit que la série X_t suit un processus de moyenne mobile d’ordre 1 (**MA(1)**) si elle est générée par un bruit blanc a_t sous la forme:

$$\begin{aligned}X_t &= a_t - \theta a_{t-1} \\X_t &= (1 - \theta B)a_t\end{aligned}$$

La fonction de transfert du filtre se réduit à un seul terme.

Variance de X Par calculs simples, on peut montrer que

$$\begin{aligned}Var(X) &= (1 + \theta^2)\sigma_a^2 \\Var(X) &> \sigma_a^2\end{aligned}$$

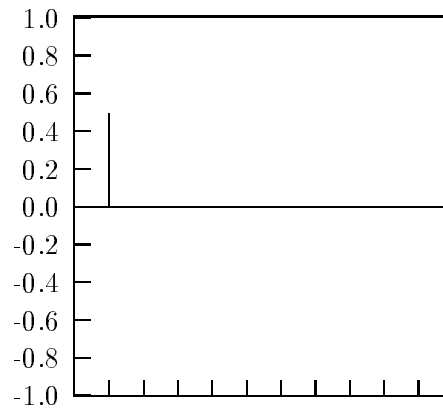
C’est-à-dire qu’en modélisant, on diminue la variance du phénomène ce qui est, par nature, le propre de toute modélisation!

Autocorrélation de X De même, on peut montrer que

$$\rho(k) = \begin{cases} \frac{-\theta}{1+\theta^2} & \text{Si } k = 1 \\ 0 & \text{Si } k \geq 2 \end{cases}$$

On peut dessiner la fonction d’autocorrélation (ACF). Ce dessin est d’ailleurs appelé au-

toautocorrélogramme.



3.2 MA d'ordre q

On suppose que la série X est générée par un bruit blanc a sous la forme:

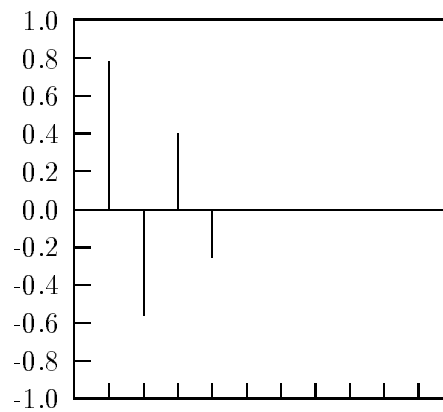
$$\begin{aligned} X_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \\ X_t &= \Theta(B)a_t \end{aligned}$$

On suppose bien évidemment que q est inférieur au nombre d'observations.

Autocorrélation de X On peut montrer que

$$\rho(k) \begin{cases} \neq 0 & \text{Si } k \leq q \\ = 0 & \text{Si } k > q \end{cases}$$

L'autocorrélogramme de X devient:



3.3 En résumé

Pour reconnaître qu'une série X est un processus moyenne mobile d'ordre q , on va dessiner l'autocorrélogramme de cette série. Pour chacune des valeurs d'autocorrélation, on aura un intervalle de confiance sur cette valeur. Si on peut décider que jusqu'à un décalage q , ces valeurs sont différentes de 0, et qu'ensuite elles ne sont pas statistiquement différentes de 0, alors on pourra dire que la série suit un processus $MA(q)$.

On peut remarquer aussi qu'une telle série est stationnaire.

Chapitre 4

PROCESSUS AUTOREGRESSIF

De tels processus seront notés AR.

4.1 AR d'ordre 1

On dit que la série X_t suit un processus autorégressif d'ordre 1 (**AR(1)**) si on peut écrire:

$$\begin{aligned}X_t &= \phi X_{t-1} + a_t \\X_t - \phi X_{t-1} &= a_t \\(1 - \phi B)X_t &= a_t\end{aligned}$$

où la série a_t est un bruit blanc.

On peut remarquer qu'on fait une régression de la série décalée de 1 sur la série elle-même et les résidus forment un bruit blanc.

Fonction de transfert On peut écrire:

$$\begin{aligned}X_t &= a_t + \phi X_{t-1} \\X_t &= a_t + \phi a_{t-1} + \phi^2 X_{t-2} \\X_t &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots + \phi^k a_{t-k} + \dots \\X_t &= \sum_{i=0}^{\infty} \phi^i a_{t-i} \\X_t &= \sum_{i=0}^{\infty} \phi^i B^i a_t = \Psi(B)a_t\end{aligned}$$

La fonction de transfert a donc une infinité de termes. Si on revient à notre écriture formelle sous forme de polynôme, on remarque qu'on a calculé l'inverse du polynôme $(1 - \phi B)$ et que

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

Variance de X Par calculs simples, on peut montrer que

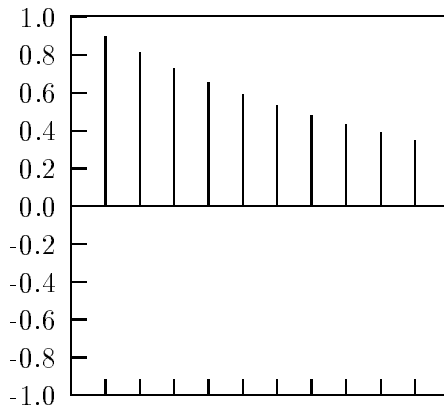
$$Var(X) = \frac{\sigma_a^2}{1 - \phi^2}$$

Ici aussi, on diminue la variance en modélisant à condition que ϕ soit en valeur absolue inférieure à 1.

Autocorrélation de X De même, on peut montrer que

$$\rho(k) = \phi^k \text{ pour } k \geq 1$$

On peut obtenir deux sortes de corrélogramme suivant si ϕ est positif ou négatif.



Stationnarité Un processus AR(1) est stationnaire si et seulement si la valeur absolue de ϕ est inférieure à 1.

$$|\phi| < 1$$

4.2 AR d'ordre p

Une série X suit un processus autorégressif d'ordre p (**AR(p)**) si on peut écrire:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \\ X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} &= a_t \\ (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t &= a_t \\ \Phi(B) X_t &= a_t \end{aligned}$$

où la série a_t est un bruit blanc.

Ici encore, on fait une régression des p séries décalées sur la série elle-même.

Autocorrélation Pour un processus $AR(p)$, on ne peut rien dire de sa fonction ACF, si ce n'est que:

$$\rho(k) \neq 0 \text{ pour tout } k$$

C'est pour pouvoir reconnaître un processus autorégressif qu'on a introduit la Fonction d'Autocorrélation Partielle.

4.3 La Fonction d'Autocorrélation Partielle

L'Autocorrélation Partielle au décalage k (**PACF(k)**) est définie comme étant la corrélation entre:

- Le résidu de la régression de la série X_{t+k} par les séries $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$

et

- Le résidu de la régression de la série X_t par les séries $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$

En d'autres termes:

$$\begin{aligned} X_{t+k} &= \alpha_1 X_{t+1} + \alpha_2 X_{t+2} + \dots + \alpha_{k-1} X_{t+k-1} + U \\ X_t &= \beta_1 X_{t+1} + \beta_2 X_{t+2} + \dots + \beta_{k-1} X_{t+k-1} + V \end{aligned}$$

et

$$PACF(k) = Corr(U, V)$$

Il faut comprendre que l'autocorrélation partielle est la corrélation entre X_t et X_{t+k} , une fois qu'on a expliqué ceux-ci par les valeurs entre eux deux, $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

4.3.1 ACF et PACF d'un AR(1)

Prenons une série X_t suivant un processus autorégressif d'ordre 1. Par exemple:

$$X_t = 0.9X_{t-1} + a_t$$

ACF(1) La corrélation entre X_t et X_{t+1} peut se calculer facilement

$$\begin{aligned} ACF(1) &= Corr(X_{t+1}, X_t) \\ &= Corr(0.9X_t + a_{t+1}, X_t) \\ &= 0.9Corr(X_t, X_t) + Corr(a_{t+1}, X_t) \\ &= 0.9 + 0 = 0.9 \end{aligned}$$

Ceci est vrai car X_t et a_{t+1} sont indépendants.

ACF(2) En écrivant la relation entre une valeur à un instant et la valeur précédente à l'ordre $t - 1$, on a :

$$\begin{aligned} X_{t-1} &= 0.9X_{t-2} + a_{t-1} \\ \text{d'où} \\ X_t &= 0.9(0.9X_{t-2} + a_{t-1}) + a_t \\ X_t &= 0.81X_{t-2} + 0.9a_{t-1} + a_t \end{aligned}$$

Comme X_{t-2} est indépendant à la fois de a_{t-1} et a_t , on en déduit facilement que :

$$Corr(X_t, X_{t-2}) = 0.81$$

ACF(p) En continuant ce type de raisonnement, on montre que :

$$Corr(X_t, X_{t-k}) = 0.9^k$$

On voit ici où se situe le problème de la fonction d'autocorrélation des processus autorégressifs : il semblerait, par la valeur de cette fonction, qu'il y ait une relation entre X_t et X_{t-2} . Or cette relation n'est qu'indirecte puisqu'elle n'existe que par X_{t-1} et par les relations de celui-ci avec X_t et X_{t-2} .

La fonction d'autocorrélation partielle existe justement pour connaître jusqu'à quel niveau de décalage, il existe une relation **directe** entre X_t et les valeurs précédentes.

PACF(1) L'autocorrélation partielle au décalage 1 est :

$$\begin{aligned} PACF(1) &= Corr(X_{t+1}, X_t) \\ &= 0.9 \end{aligned}$$

Rappel Dans une régression :

$$Y_i = \alpha X_i + \epsilon_i$$

le coefficient α peut être calculé comme :

$$\alpha = \frac{Cov(X, Y)}{Var(X)} = Corr(X, Y) \frac{\sigma(X)\sigma(Y)}{Var(X)} = Corr(X, Y) \frac{\sigma(Y)}{\sigma(X)}$$

C'est-à-dire que, si les deux variables X et Y ont même écart-type, que l'on fasse la régression de X sur Y ou la régression de Y sur X on retrouve le même coefficient α :

$$\alpha = Corr(X, Y)$$

Pour résumer, on peut dire que si on dispose de 2 variables X et Y de moyennes nulles et de même écart-type, si on fait la régression de X sur Y :

$$Y = \alpha X + \epsilon$$

alors, la régression de Y sur X est :

$$X = \alpha Y + \delta$$

et non, comme on aurait pu le penser :

$$X = 1/\alpha Y + \delta$$

PACF(2) Il faut, pour calculer PACF(2), faire la régression de X_{t+1} sur X_{t+2} et sur X_t . On sait déjà que:

$$X_{t+2} = 0.9X_{t+1} + a_{t+2}$$

Du rappel précédent, on en déduit que, si on fait la régression de X_{t+1} sur X_t on obtient:

$$X_t = 0.9X_{t+1} + V_t$$

Pour le calcul de PACF(2), on calcule la corrélation entre les 2 résidus de ces régressions:

$$\begin{aligned} PACF(2) &= Corr(X_{t+2} - 0.9X_{t+1}, X_t - 0.9X_{t+1}) \\ &= Corr(a_{t+2}, X_t - 0.9X_{t+1}) \\ &= Corr(a_{t+2}, X_t) - 0.9Corr(a_{t+2}, X_{t+1}) \\ PACF(2) &= 0 \end{aligned}$$

PACF(k) Par de calculs similaires on peut montrer que:

$$PACF(k) = 0 \quad \text{Pour tout } k > 1$$

4.3.2 PACF d'un AR(p)

Nous ne nous attarderons pas sur cette fonction d'autocorrélation partielle mais on peut montrer que:

$$PACF(k) = 0 \quad \text{Pour tout } k > p$$

4.4 Dualité

Pour introduire un tableau de dualité entre les processus MA et AR:

	AR(p)	MA(q)
ACF		= 0 Pour tout k > q
PACF	= 0 Pour tout k > p	

On reconnaît qu'une série suit un processus MA(q) si sa fonction d'autocorrélation ACF s'annule à partir d'un décalage q, ou qu'elle suit un processus AR(p) si sa fonction d'autocorrélation partielle PACF s'annule à partir d'un décalage p.

Chapitre 5

PROCESSUS ARMA, ARIMA ET SARIMA

5.1 Les processus ARMA

5.1.1 Définition

On dit qu'une série X_t suit un processus ARMA d'ordre (p,q) (noté **ARMA(p,q)**), si on peut écrire:

$$\begin{aligned}X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} &= a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \\ \Phi(B)X_t &= \Theta(B)a_t\end{aligned}$$

où la série a_t est un bruit blanc.

5.1.2 Processus ARMA(1,1)

En prenant p=1 et q=1, la série X_t suit un processus ARMA(1,1) si on peut écrire:

$$(1 - \phi B)X_t = (1 - \theta B)a_t$$

Les propriétés d'un ARMA(1,1) sont:

- Pour que la série X_t soit stationnaire et inversible (c'est-à-dire qu'on puisse calculer a_t en fonction de X_t il faut que:

$$|\phi| < 1 \text{ et } |\theta| < 1$$

- L'ACF d'un ARMA(1,1) décroît exponentiellement à partir de k=1, alors que l'ACF d'un AR(1) décroît exponentiellement à partir de k=0.
- La PACF d'un ARMA(1,1) ressemble à celle d'un MA(1) à partir de k=2.

On peut dire, en gros, que si on ne se retrouve pas d'une façon évidente en face d'un processus AR ou MA, on a de fortes chances de se trouver en face d'un processus ARMA.

5.2 Les processus ARIMA

Une série X_t suit un processus ARIMA (*AutoRegressive Integrated Moving Average*) d'ordre (p,d,q) si elle suit un processus ARMA d'ordre (p+d,q):

$$\Xi(B)X_t = \Theta(B)a_t$$

où la valeur $B=1$ est racine d'ordre d du polynôme $\Xi(B)$. On modélise alors la série sous la forme:

$$\begin{aligned}\Phi(B)(1-B)^d X_t &= \Theta(B)a_t \\ \Phi(B)\nabla^d X_t &= \Theta(B)a_t\end{aligned}$$

où le polynôme $\Phi(B)$ est de degré p et le polynôme $\Theta(B)$ est de degré q. On écrit que la série X_t suit un processus **ARIMA(p,d,q)**.

On peut remarquer que la série X_t suivant un processus ARIMA n'est pas stationnaire puisqu'il faut lui appliquer l'opérateur de différenciation pour avoir un bruit blanc la générant.

5.3 Les processus SARIMA

Une série X_t suit un processus SARIMA (*Seasonnal AutoRegressive Integrated Moving Average*) d'ordre $(p, d, q) * (P, D, Q)_s$ si cette série a une saisonnalité de période s et qu'on peut écrire:

$$\Phi_1(B)\Phi_2(B^s)(1-B)^d(1-B^s)^D X_t = \Theta_1(B)\Theta_2(B^s)a_t$$

où Φ_1 est un polynôme de degré p , Φ_2 est de degré P , Θ_1 est de degré q et Θ_2 est de degré Q .

Chapitre 6

UN EXEMPLE

6.1 Données de Pollution Atmosphérique

Les données dont nous disposons nous ont été fournies par l'ASCOPARG (Association pour le Contrôle de la Pollution Atmosphérique dans la Région Grenobloise) par Mme Marie-Blanche Personnaz.

Cette Association mesure, en divers points de la cuvette grenobloise, chaque quart d'heure plusieurs polluants et plusieurs paramètres météorologiques. Son but est non seulement de contrôler la pollution atmosphérique mais également de prévenir les personnes compétentes en cas de pollution aigüe.

Nous nous intéresserons ici seulement à quelques paramètres: le dioxyde de soufre (SO_2) mesuré à la Villeneuve de Grenoble, la température au Peuil-de-Claix (situé environ à 800m d'altitude) l'inversion de température, c'est-à-dire à la différence de température entre le Peuil-de-Claix et le Pont-de-Claix (situé dans la vallée). La situation encaissée de la vallée grenobloise permet effectivement de mesurer facilement cette inversion de température, mais elle apporte beaucoup de problème de pollution. Des études antérieures ont montré l'influence de ces paramètres sur la pollution.

6.2 Les logiciels

Plusieurs logiciels permettent d'estimer des modèles Box-Jenkins. La pauvreté de la commande "*Box-Jenkins*" du logiciel SPSSx [10] ne nous permet pas de nous étendre sur celle-ci. Nous avons ajouté au logiciel EDA [5] plusieurs commandes qui ne sont en fait que des interfaces entre ce logiciel et la bibliothèque NAG. Ces commandes, regroupées sous le vocable de TSEDA, ont été écrites pour EDA car, comme nous allons le voir, l'étude de la modélisation suit l'idée du logiciel d'interactivité, d'itérativité et de graphique. Mais nous préférons ici, dans un souci d'universalité, prendre notre exemple à l'aide du logiciel BMDP: nous utiliserons donc le programme BMDP2T de ce logiciel [4].

6.3 Processus de modélisation

La modélisation d'une série sous la forme Box-Jenkins se fait en 3 étapes:

1. Identification du modèle.
2. Estimation des paramètres du modèle.

3. Verification du modèle.

4. Si la vérification ne permet pas de conclure à un bon modèle on revient alors en 1.

On voit bien dans ce schéma l'itérativité du processus. L'identification du modèle se fait à l'aide des dessins des ACF et PACF; ici on voit apparaître l'aspect graphique du travail.

6.4 Modélisation

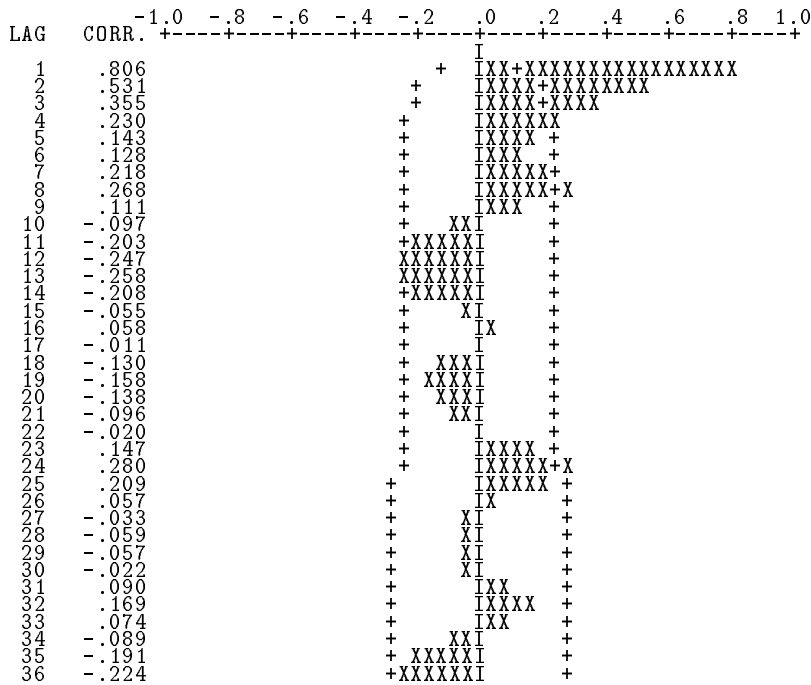
Le fichier **DATA2** de données contient 4 variables: la température à Pont-de-Claix, celle au Peuil-de-Claix, le SO_2 à la Villeneuve et la vitesse du vent. Ces mesures sont ramenées par moyenne sur 3 heures à partir de 0heure de chaque jour. Nous avons pris 240 mesures, soit 30 jours, situées en période d'hiver, à cheval sur les mois de janvier et février 1986.

Les ordres de BMDP2T permettant de faire les auto-corrélogrammes ACF et PACF de l'inversion sont les suivants:

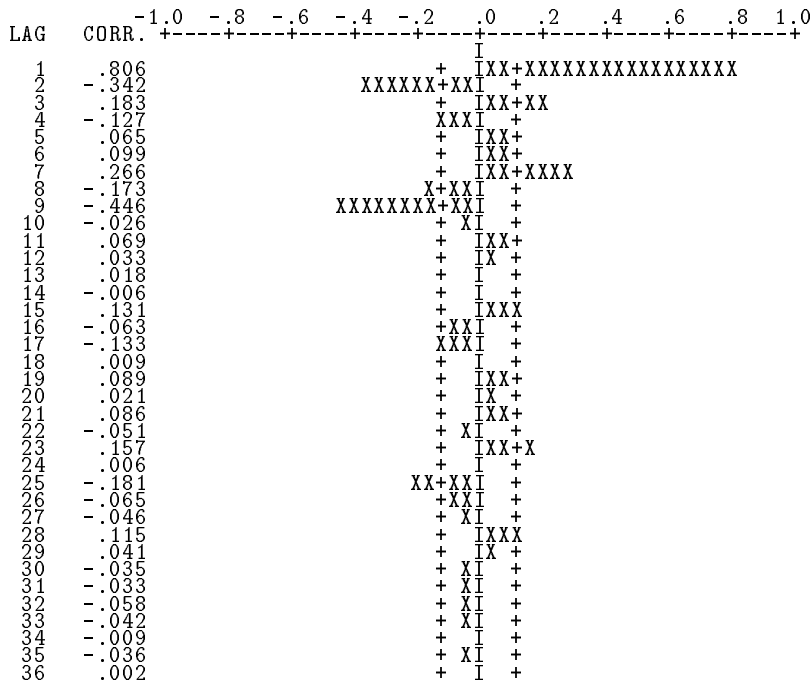
```
/PROBLEM TITLE IS 'MODELISATION DE LA POLLUTION'.  
/INPUT  VARIABLES ARE 4.  
        FORMAT IS '(4F4.0)'.  
        FILE IS DATA2 .  
/VARIABLE ADDED IS 1.  
        NAMES ARE TPNTCLAI,TPEUIL,SO2VILLE,VARITH,INVERS.  
/TRANSFORM INVERS=TPEUIL-TPNTCLAI .  
/END  
ACF VAR=5./  
PACF VAR=5./  
END/
```

Le résultat est le suivant:

PLOT OF AUTOCORRELATIONS



PLOT OF PARTIAL AUTOCORRELATIONS



Laissons de côté pour le moment l'aspect saisonnier de cette série; saison d'ailleurs qui est quotidienne puisque la période est de 8 fois 3 heures.

L'aspect des auto-corrélogrammes laisserait supposer un modèle ARMA(1,1). Faisons une estimation de ces paramètres et dessinons les corrélogrammes des résidus. Les ordres sont:

```

ARIMA VAR=5. CENTERED. AROR='(1)'.MAOR='(1)'. /
ESTIMATION RESIDUALS=RES5. /
ACF VAR=RES5. /
PACF VAR=RES5. /

```

BMDP a un avantage: il permet de faire l'estimation des paramètres par deux méthodes différentes ce qui permet d'avoir une qualité du modèle choisi; en effet, si les paramètres estimés par les deux méthodes sont très différents on peut douter de la justesse de celui-ci.

ESTIMATION BY CONDITIONAL LEAST SQUARES METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
INVERS	RANDOM	REMOVED	1-	240	

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	INVERS	MA	1	1	-.1214	.0741	-1.64
2	INVERS	AR	1	1	.8366	.0404	20.73

ESTIMATION BY BACKCASTING METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
INVERS	RANDOM	REMOVED	1-	240	

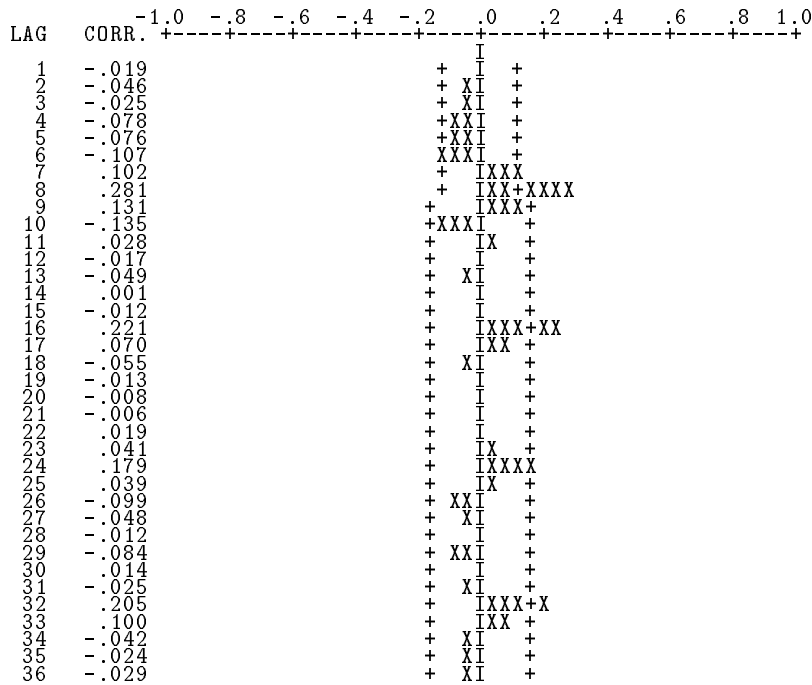
PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	INVERS	MA	1	1	-.1293	.0729	-1.77
2	INVERS	AR	1	1	.8441	.0389	21.69

On peut remarquer deux choses:

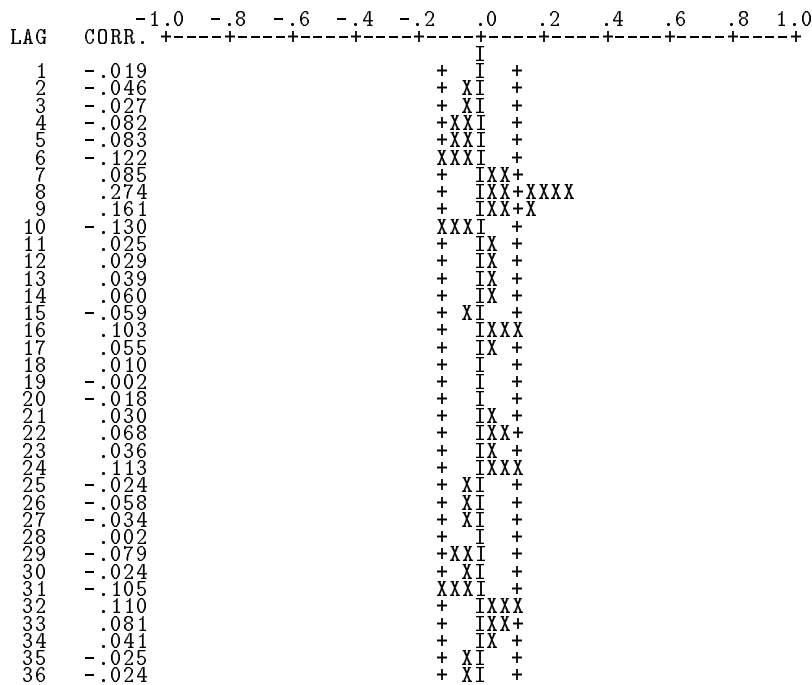
1. les paramètres sont stables: ce qui tendrait à prouver qu'on ne s'est pas trop éloigné du vrai modèle.
2. le paramètre MA(1) a un "*T-RATIO*" inférieur à 2., ce qui ne nous permet pas de conclure qu'il est différent de zéro: nous le supprimerons donc.
3. le paramètre AR(1) est différent de 1., ce qui nous conduit à dire qu'il n'y a pas lieu de différencier la série.

L'aspect "*Moving Average*" de la série vient en fait de la saisonnalité de celle-ci comme on peut le constater sur les corrélogrammes. Pour bâtir un modèle saisonnier, il faut s'intéresser aux graphiques seulement aux décalages 8, 16, 24, 32, ...

PLOT OF AUTOCORRELATIONS



PLOT OF PARTIAL AUTOCORRELATIONS



En regardant les graphiques précédents, on peut en déduire un ajout au modèle comme étant un ARMA(1,1) sur une période de 8. Les ordres pour cela sont:

```
ARIMA VAR=5. CENTERED. AROR='(1),(8)'.MAOR='(8)'./
ESTIMATION RESIDUALS=RES5./
ACF VAR=RES5./
PACF VAR=RES5./
```

Le résultat en est:

ESTIMATION BY CONDITIONAL LEAST SQUARES METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
----------	------	------	------	------	-------------

INVERS	RANDOM	REMOVED	1-	240	
--------	--------	---------	----	-----	--

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	INVERS	MA	1	8	.7542	.0656	11.49
2	INVERS	AR	1	1	.8359	.0375	22.30
3	INVERS	AR	2	8	.8437	.0337	25.04

RESIDUAL SUM OF SQUARES = 59632.947267

DEGREES OF FREEDOM = 228

RESIDUAL MEAN SQUARE = 261.548014

ESTIMATION BY BACKCASTING METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
----------	------	------	------	------	-------------

INVERS	RANDOM	REMOVED	1-	240	
--------	--------	---------	----	-----	--

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	INVERS	MA	1	8	.6793	.0613	11.08
2	INVERS	AR	1	1	.8373	.0358	23.38
3	INVERS	AR	2	8	.9332	.0206	45.38

RESIDUAL SUM OF SQUARES = 64231.664197

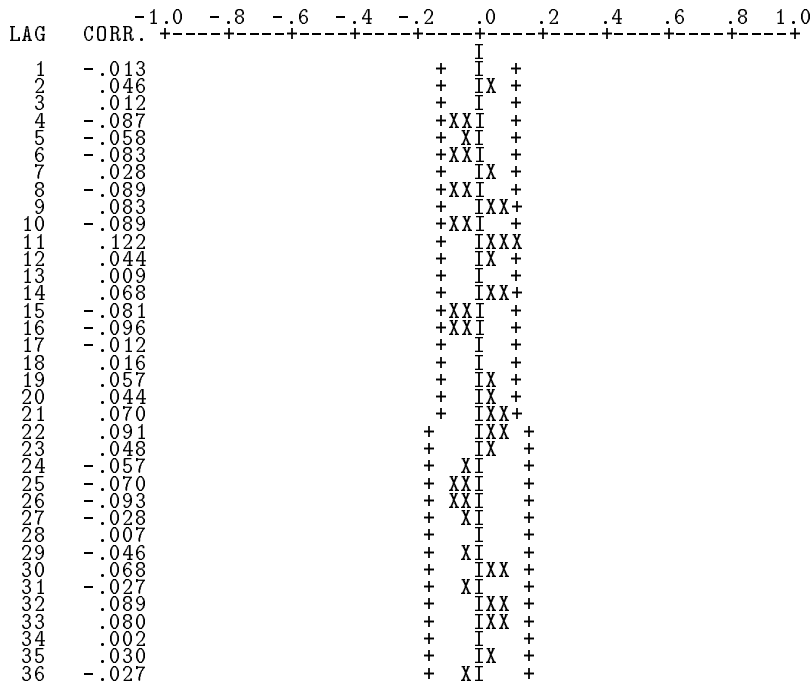
DEGREES OF FREEDOM = 228

RESIDUAL MEAN SQUARE = 281.717825

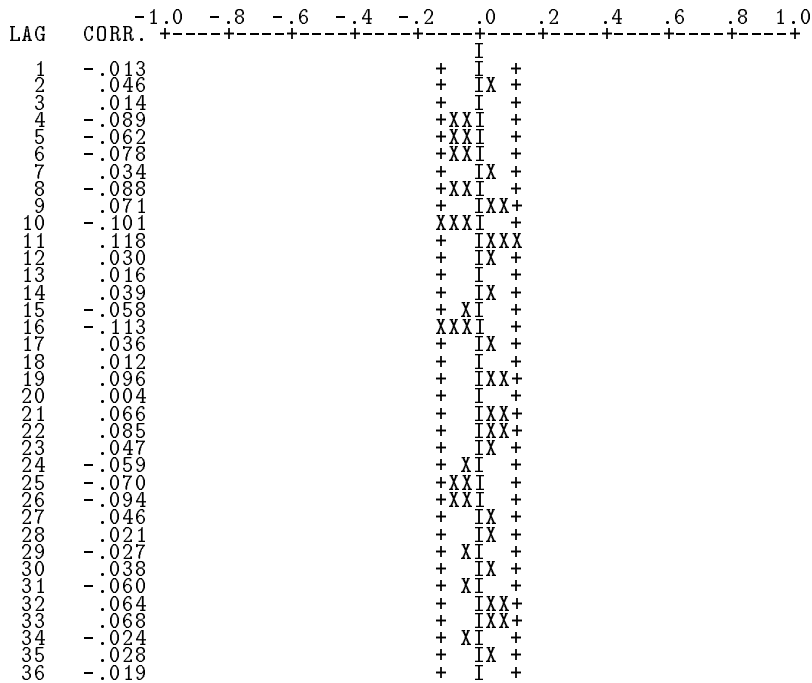
(BACKCASTS EXCLUDED)

La moyenne de la variance des résidus vaut environ 270, celle de la série de départ étant de 1450, on déduit que notre modèle permet d'expliquer environ 81% de la variance de l'inversion de température. Les corrélogrammes des résidus sont les suivants:

PLOT OF AUTOCORRELATIONS



PLOT OF PARTIAL AUTOCORRELATIONS



On peut constater sur ces dessins que les résidus forment bien un bruit blanc.

L'inversion de température suit donc un modèle $ARMA(1,0)*(1,1)_8$. Pour la température au Peuil-de-Claix, on peut trouver par la même technique, qu'elle suit un modèle $ARIMA(1,0,1)*(0,1,1)_8$ dont l'estimation des paramètres est la suivante:

ESTIMATION BY CONDITIONAL LEAST SQUARES METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
					8
TPEUIL		RANDOM		1- 240	(1-B)

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	TPEUIL	MA	1	1	-.4365	.0614	-7.11
2	TPEUIL	MA	2	8	.8245	.0385	21.41
3	TPEUIL	AR	1	1	.9346	.0267	34.94

RESIDUAL SUM OF SQUARES	=	25531.598432
DEGREES OF FREEDOM	=	228
RESIDUAL MEAN SQUARE	=	111.980695

ESTIMATION BY BACKCASTING METHOD

VARIABLE	VAR.	TYPE	MEAN	TIME	DIFFERENCES
					8
TPEUIL		RANDOM		1- 240	(1-B)

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	TPEUIL	MA	1	1	-.4311	.0599	-7.20
2	TPEUIL	MA	2	8	.9490	.0114	82.91
3	TPEUIL	AR	1	1	.9390	.0232	40.48

RESIDUAL SUM OF SQUARES	=	21095.997980
DEGREES OF FREEDOM	=	228
RESIDUAL MEAN SQUARE	=	92.526307

La moyenne de la variance des résidus vaut environ 100, celle de la série de départ étant d'environ 1520, on peut en déduire que notre modèle explique environ 93% de la variance de la température au Peuil-de-Claix.

Chapitre 7

ETUDE MULTIVARIEE DES SERIES CHRONOLOGIQUES

7.1 Fonction de corrélation croisée

7.1.1 Définition

- La covariance croisée au décalage k entre une série X_t et une série Y_t , $\gamma_{X,Y}(k)$ est définie par:

$$\gamma_{X,Y}(k) = 1/n \sum_{j=1}^{n-k} (Y_{j+k} - \bar{Y})(X_j - \bar{X}) \text{ avec } \bar{X} = 1/n \sum_{i=1}^n X_i \text{ et } \bar{Y} = 1/n \sum_{i=1}^n Y_i \quad (7.1)$$

Il faut remarquer que, ici encore, la somme ne fait intervenir que $n - k$ termes mais on divise par n et non pas par $n - k$ comme on le fait dans un calcul de covariance habituel. Il faut aussi dire que dans ce calcul de covariance on essaie de connaître l'influence de la série X sur la série Y avec un certain décalage dans le temps.

- La corrélation croisée au décalage k entre deux séries X_t et Y_t , $\rho_{X,Y}(k)$ est définie par:

$$\rho_{X,Y}(k) = \gamma_{X,Y}(k) / \gamma_{X,Y}(0) \quad (7.2)$$

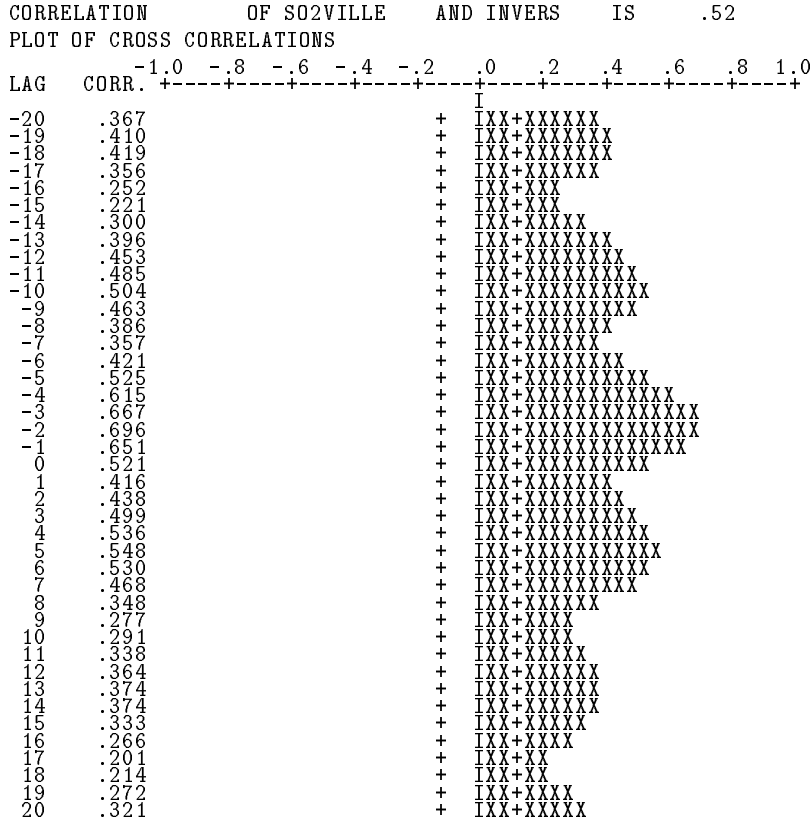
- La fonction de corrélation croisée est la fonction qui à k associe la valeur $\rho_{X,Y}(k)$. On notera cette fonction comme étant **la CCF** de la série X sur la série Y.

7.1.2 L'exemple de la pollution

Regardons la corrélation croisée entre l'inversion de température et le niveau de SO_2 mesuré à la Villeneuve de Grenoble. Le programme que nous utilisons permet d'étudier en même temps l'influence des 2 variables l'une sur l'autre. Les ordres de BMDP sont les suivants:

CCF VAR=3,5./

Le résultat est, en partie, le suivant:



Comme on a mis en premier l'indice de pollution, les premières valeurs pour des décalages négatifs concernent l'influence de l'inversion sur la pollution et celles des décalages positifs concernent l'influence de la pollution sur l'inversion de température.

Comme on le constate, la corrélation croisée entre deux séries qui évoluent en même temps (ici elles sont périodiques de même période) ne veut rien dire. C'est dans ce problème qu'on voit l'aspect nécessaire de la modélisation univariée des séries.

Constatons toutefois que dans ce corrélogramme, il semblerait que l'effet de l'inversion de température se fasse avec un décalage entre 6 et 9 heures de retard.

7.2 Modèle de Fonction de Transfert

On peut modéliser l'influence d'une série X_t sur une autre série Y_t sous une forme linéaire par:

$$\begin{aligned}
Y_t - \delta_1 Y_{t-1} - \dots - \delta_r Y_{t-r} &= \omega_0 X_{t-b} - \omega_1 X_{t-b-1} - \dots - \omega_s X_{t-b-s} + N_t \\
\delta(B)Y_t &= \omega(B)X_{t-b} + N_t \\
Y_t &= \delta^{-1}(B)\omega(B)X_{t-b} + n_t \\
Y_t &= v(B)X_t + n_t
\end{aligned}$$

où N_t est un bruit. On retrouve une formulation d'un filtre linéaire où la série en entrée est X_t , la sortie est la série Y_t , la fonction de transfert est décrite dans le polynôme $v(B)$.

Tout le problème réside alors dans l'estimation de ce polynôme $v(B)$.

7.3 L'influence de l'inversion de température sur la pollution

Revenons à notre exemple pour plus de simplicité. Pour essayer d'identifier notre modèle, le principe est:

1. Modéliser suivant un modèle SARIMA, l'inversion de température et récupérer les résidus.
2. Filtrer la pollution suivant le **même** modèle et récupérer les résidus.
3. Calculer la corrélation croisée entre ces résidus.

Les ordres du programme BMDP sont les suivants:

```
ARIMA VAR=5. CENTERED. AROR='(1)',(8)'. MAOR='(8)'./  
ESTIMATION RESIDUALS=RES5./  
FILTER VAR=3. RESIDUALS=RES3./  
CCF VAR=RES5,RES3. MAXLAG=10./
```

Le résultat est le suivant:

CORRELATION OF RES5 AND RES3 IS .07

CROSS CORRELATIONS OF RES5 (I) AND RES3 (I+K)

1- 10	.16	.06	.04	.07	.11	.03	.05	-.12	-.13	.05
ST.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

CROSS CORRELATIONS OF RES3 (I) AND RES5 (I+K)

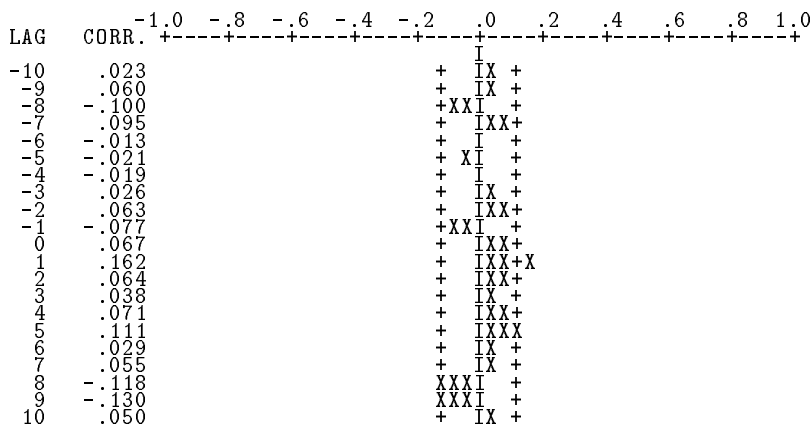
1- 10	-.08	.06	.03	-.02	-.02	-.01	.10	-.10	.06	.02
ST.E.	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

TRANSFER FUNCTION WEIGHTS

LAG	SCCF(X(I),Y(I+K))		SCCF(Y(I),X(I+K))	
	*SY/SX	*SX/SY	*SY/SX	*SX/SY
0	.11126	.04062	.11126	.04062
1	.26738	.09763	-.12688	-.04633
2	.10517	.03840	.10410	.03801
3	.06226	.02273	.04290	.01566
4	.11743	.04288	-.03200	-.01168
5	.18321	.06689	-.03534	-.01290
6	.04845	.01769	-.02228	-.00813
7	.09044	.03302	.15734	.05745
8	-.19448	-.07101	-.16512	-.06029
9	-.21450	-.07832	.09859	.03600
10	.08346	.03047	.03751	.01370

WHERE X(I) IS THE FIRST SERIES, Y(I) THE SECOND
SERIES, SX THE STANDARD ERROR OF X(I), AND SY
THE STANDARD ERROR OF Y(I)

PLOT OF CROSS CORRELATIONS



Remarquons qu'ici on a bien mis la variable de météorologie en premier. Si on regarde le dessin de la fonction de corrélation croisée, on constate que la seule corrélation vraiment significative est au décalage 1 (soit 3 heures de décalage). On remarquera ainsi que l'hypothèse d'un retard entre 6 et 9 heures que nous avions faite précédemment était fausse.

Ici on peut proposer deux modèles:

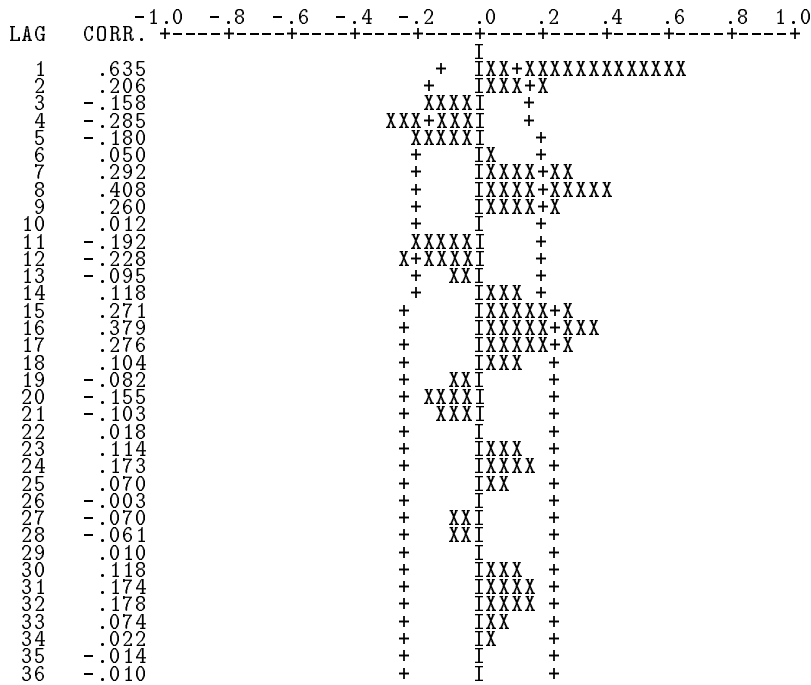
1. On remarque une certaine forme de courbe en exponentielle décroissante sur les décalages 1, 2, 3 pour la CCF. On propose alors un modèle où le polynôme $\delta(B)$ est du deuxième degré. Mais dans ce cas l'évaluation des estimateurs du polynôme $v(B)$ devient relativement compliqué. Nous vous proposons de vous reporter au livre de Box et Jenkins pour cela [1].
2. On suppose sinon que le polynôme $\delta(b)$ se réduit à l'unité. Alors une première estimation du polynôme est donnée directement par le programme; il faut pour cela regarder dans le paragraphe **TRANSFERT FUNCTION WEIGHTS**, à l'intersection de la ligne **LAG 0** et de la première colonne. On obtient ainsi une valeur initiale du paramètre: $v_1 = 0.27$, et $b = 1$ avec $\omega_0 = 0.27$. Cette valeur va être estimée à nouveau quand on soumettra le modèle final.

Dans le programme BMDP, on va soumettre la variable de pollution à un modèle où elle est influencée par l'inversion à un décalage de 3 heures avec un coefficient de 0.27, puis on va modéliser par un modèle SARIMA les résidus résultants.

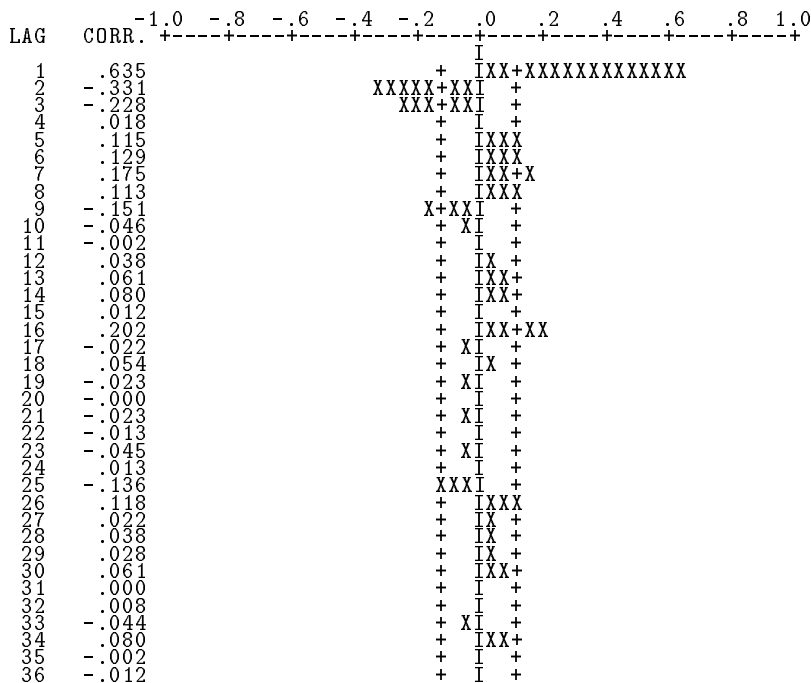
```
ARIMA VAR=3. CENTERED. /
INDEP VAR=5. CENTERED.
      UPORDERS='(1)'. UPVALUES=0.27./
ESTIMATION
      RESIDUALS=RES53./
ACF   VAR=RES53./
PACF  VAR=RES53./
```

Le résultat des fonctions d'autocorrélation et d'autocorrélation partielle sont les suivants:

PLOT OF AUTOCORRELATIONS



PLOT OF PARTIAL AUTOCORRELATIONS



En regardant la PACF, on peut supposer qu'à part le problème dû à la saisonnalité elle n'est pas nulle seulement aux décalages 1, 2, 3. On modéliserait ces résidus comme un SARIMA $(3, 0, 0) * (1, 0, 0)_8$. Mais l'estimation montre que le troisième paramètre ne peut pas être jugé significativement différent de 0. On modélise donc ces résidus comme un SARIMA $(2, 0, 0) * (1, 0, 0)_8$.

ARIMA VAR=3. CENTERED. /
INDEP VAR=5. CENTERED.

```
UPORDERS='(1)'. UPVALUES=0.27./  
ARIMA VAR=3. CENTERED. AROR='(1,2),(8)'./  
ESTIMATION  
RESIDUALS=RES53./
```

Le résultat est le suivant:

ESTIMATION BY CONDITIONAL LEAST SQUARES METHOD

SUMMARY OF THE MODEL

OUTPUT VARIABLE -- S02VILLE

INPUT VARIABLES -- NOISE INVERS

VARIABLE	VAR. TYPE	MEAN	TIME	DIFFERENCES
----------	-----------	------	------	-------------

S02VILLE	RANDOM	REMOVED	1- 240	
----------	--------	---------	--------	--

INVERS	RANDOM	REMOVED	1- 240	
--------	--------	---------	--------	--

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	S02VILLE	AR	1	1	.8749	.0653	13.41
2	S02VILLE	AR	1	2	-.2430	.0636	-3.82
3	S02VILLE	AR	2	8	.4203	.0593	7.09
4	INVERS	UP	1	1	.3910	.1013	3.86

RESIDUAL SUM OF SQUARES = 166052.610876

DEGREES OF FREEDOM = 225

RESIDUAL MEAN SQUARE = 738.011604

ESTIMATION BY BACKCASTING METHOD

SUMMARY OF THE MODEL

OUTPUT VARIABLE -- S02VILLE

INPUT VARIABLES -- NOISE INVERS

VARIABLE	VAR. TYPE	MEAN	TIME	DIFFERENCES
----------	-----------	------	------	-------------

S02VILLE	RANDOM	REMOVED	1- 240	
----------	--------	---------	--------	--

INVERS	RANDOM	REMOVED	1- 240	
--------	--------	---------	--------	--

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	S02VILLE	AR	1	1	.8501	.0604	14.09
2	S02VILLE	AR	1	2	-.2298	.0594	-3.87
3	S02VILLE	AR	2	8	.4812	.0533	9.02
4	INVERS	UP	1	1	.4267	.0968	4.41

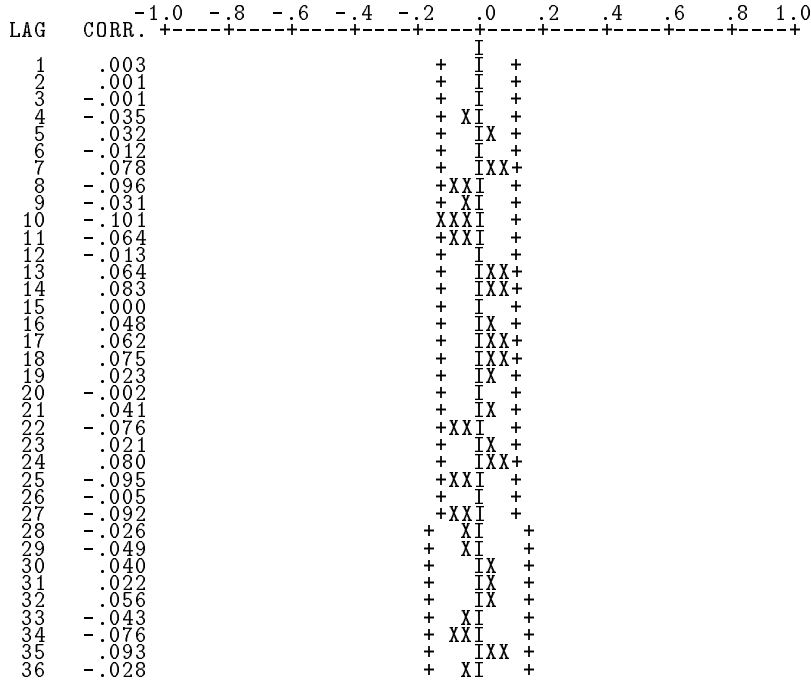
RESIDUAL SUM OF SQUARES = 167033.000869

DEGREES OF FREEDOM = 225

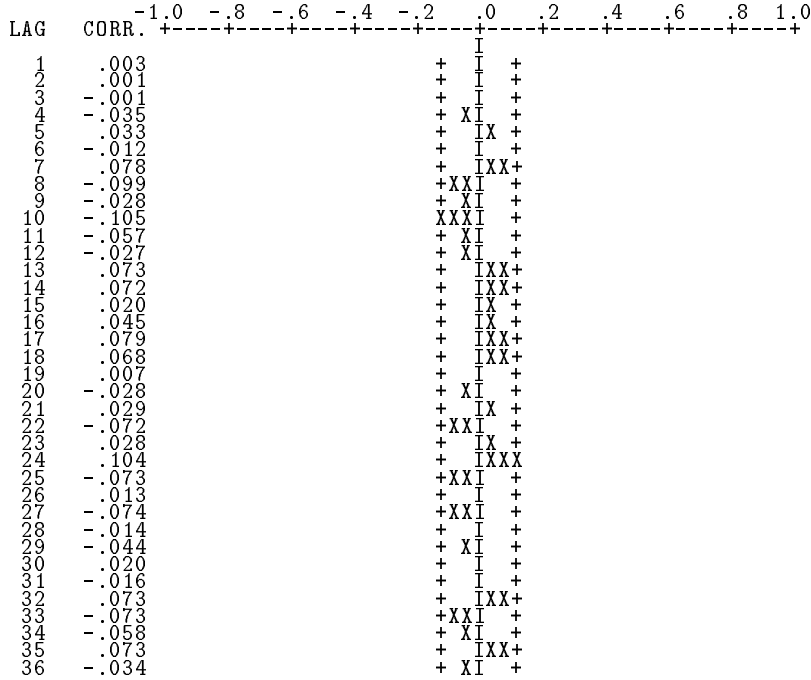
RESIDUAL MEAN SQUARE = 742.368893

(BACKCASTS EXCLUDED)

PLOT OF AUTOCORRELATIONS



PLOT OF PARTIAL AUTOCORRELATIONS



Si on note S_t le dioxyde de soufre mesuré à la Villeneuve et I_t l'inversion de température à l'instant t , on peut écrire le modèle:

$$(1 - 0.86B + 0.23B^2)(1 - 0.45B^8)S_t = 0.41I_{t-1} + a_t$$

7.4 Les influences de l'inversion et de la température

Par la même démarche, on peut étudier l'influence de la température mesurée au Peul-de-Claix sur la pollution. On remarquerait alors que la corrélation est positive et est à effet immédiat, et qu'une première estimation du paramètre ω_0 est 0.67.

Le programme qui permet d'analyser les influences de ces deux paramètres en même temps est:

```
ARIMA    VAR=3. CENTERED./
INDEP    VAR=5. CENTERED.
          UPORDERS='(1)'. UPVALUES=0.27./
INDEP    VAR=2. CENTERED.
          UPORDERS='(0)'. UPVALUES=0.67./
ARIMA    VAR=3. CENTERED. AROR='(1,2),(8)'./
ESTIMATION RESIDUALS=RES523./
ACF      VAR=RES523./
PACF     VAR=RES523./
```

Le résultat en est:

ESTIMATION BY CONDITIONAL LEAST SQUARES METHOD
SUMMARY OF THE MODEL

OUTPUT VARIABLE -- SO2VILLE

INPUT VARIABLES -- NOISE INVERS TPEUIL

VARIABLE	VAR. TYPE	MEAN	TIME	DIFFERENCES
----------	-----------	------	------	-------------

SO2VILLE	RANDOM	REMOVED	1- 240	
INVERS	RANDOM	REMOVED	1- 240	
TPEUIL	RANDOM	REMOVED	1- 240	

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	SO2VILLE	AR	1	1	.8350	.0648	12.88
2	SO2VILLE	AR	1	2	-.2590	.0637	-4.07
3	SO2VILLE	AR	2	8	.3734	.0607	6.15
4	INVERS	UP	1	1	.3832	.0939	4.08
5	TPEUIL	UP	1	0	.5313	.0980	5.42

RESIDUAL SUM OF SQUARES = 147714.352336
DEGREES OF FREEDOM = 224
RESIDUAL MEAN SQUARE = 659.439073

ESTIMATION BY BACKCASTING METHOD
SUMMARY OF THE MODEL

OUTPUT VARIABLE -- SO2VILLE

INPUT VARIABLES -- NOISE INVERS TPEUIL

VARIABLE	VAR. TYPE	MEAN	TIME	DIFFERENCES
----------	-----------	------	------	-------------

SO2VILLE	RANDOM	REMOVED	1- 240	
INVERS	RANDOM	REMOVED	1- 240	
TPEUIL	RANDOM	REMOVED	1- 240	

PARAMETER	VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	ST. ERR.	T-RATIO
1	SO2VILLE	AR	1	1	.8126	.0600	13.55
2	SO2VILLE	AR	1	2	-.2387	.0597	-4.00
3	SO2VILLE	AR	2	8	.4312	.0552	7.81
4	INVERS	UP	1	1	.4063	.0897	4.53
5	TPEUIL	UP	1	0	.5733	.0980	5.85

RESIDUAL SUM OF SQUARES = 148664.011513
DEGREES OF FREEDOM = 224
RESIDUAL MEAN SQUARE = 663.678623

Si on note S_t le dioxyde de soufre mesuré à la Villeneuve, I_t l'inversion de température et T_t la température au Peuil-de-Claix à l'instant t , on peut modéliser les influences des deux

paramètres météorologiques sous la forme:

$$(1 - 0.82B + 0.24B^2)(1 - 0.40B^8)S_t = 0.40I_{t-1} + 0.55T_t + a_t$$

Le pourcentage de variance expliquée par ce modèle est d'environ 80%.

Il est bien évident que dans une étude plus approfondie nous utiliserions d'autres paramètres pour connaître leurs influences sur la pollution, par exemple on peut supposer que la vitesse du vent doit intervenir pour "chasser" la pollution ainsi que sa direction suivant que ce vent amène ou non les émissions de pollution.

7.5 Prévisions

Enfin, pour terminer la justification d'un modèle, il faut vérifier si les prévisions que nous pouvons faire, à partir de ce modèle, sont compatibles avec ce qui est observé.

On peut utiliser 2 manières de faire:

1. Faire une prévision sur plusieurs temps (par exemple, le dernier jour observé) à partir des données mesurées sur la période précédant ceux-ci.
2. Faire une prévision réactualisée à chaque fois par ce qui a été réellement observé.

Notre programme BMDP se terminera donc par:

```
FORECAST CASES=8. START=233./
FORECAST CASES=1. START=234./
FORECAST CASES=1. START=235./
FORECAST CASES=1. START=236./
FORECAST CASES=1. START=237./
FORECAST CASES=1. START=238./
FORECAST CASES=1. START=239./
FORECAST CASES=1. START=240./
```

Le résultat sera le suivant:

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
233	63.28706	25.14692	63.00000	.28706
234	82.49854	32.23587	140.00000	-57.50146
235	130.99423	33.77615	189.00000	-58.00577
236	155.05760	33.92746	129.00000	26.05760
237	135.17631	33.92769	120.00000	15.17631
238	101.64390	33.93442	61.00000	40.64390
239	108.05569	33.93925	62.00000	46.05569
240	80.80442	33.94053	64.00000	16.80442

STANDARD ERROR = 25.1469 BY CONDITIONAL METHOD

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
234	82.26831	25.08865	140.00000	-57.73169

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
235	177.18237	25.33572	189.00000	-11.81763

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
236	187.64872	25.29021	129.00000	58.64872

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
237	100.22288	25.54174	120.00000	-19.77712

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
238	95.78532	25.51849	61.00000	34.78532

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
239	79.13462	25.56798	62.00000	17.13462

PERIOD	FORECASTS	ST. ERR.	ACTUAL	RESIDUAL
240	53.71337	25.53624	64.00000	-10.28663

Comme on peut le constater, notre modèle serait à améliorer puisqu'il aurait une certaine tendance à estimer avec un peu de retard la pointe de pollution; ceci est peut-être dû également au pas de temps choisi de 3 heures qui est sans doute un peu grand.

Bibliographie

- [1] G.E.P. Box et G.M. Jenkins. *Time Series Analysis: forecasting and control*. Holden-Day (1976).
- [2] P.J. Brockwell et R.A. Davis. *Time Series: Theory and Method*. Springer-Verlag (1987).
- [3] P.F. Coutrot et F. Droesbeke. *Les méthodes de prévision*. Que sais-je? n° 2157 (1984).
- [4] W.J. Dixon et coll. *BMDP Statistical Software*. University of California Press (1985).
- [5] E. Horber. *Exploratory Data Analysis: User's Manual. Version 2.1/1*. Université de Genève. (1989).
- [6] C.W. Ostrow Jr. *Time Series Analysis: Regression Techniques*. Série: Quantitative Applications in the Social Sciences. Sage Publications (1982).
- [7] B. Rapacchi. *TSEDA. User's Manual*. Centre InterUniversitaire de Calcul de Grenoble (1989).
- [8] B. Rapacchi. *Lisons les données!!*. Centre InterUniversitaire de Calcul de Grenoble (1989).
- [9] D.H. Sanders, A.F. Murph, R.J. Eng. *Les statistiques: une approche nouvelle*. Mac-Graw Hill (1984).
- [10] SPSS Inc. *SPSSx. User's Manual*. MacGraw-Hill Book Company (1986).