

NOTES DE STATISTIQUE ET D'INFORMATIQUE

2009/1

ETUDE DES SERIES CHRONOLOGIQUES
PAR LES METHODES DE DECOMPOSITION

R. PALM et Y. BROSTAUX

Faculté universitaire des Sciences agronomiques
Unité de Statistique, Informatique et Mathématique appliquées

GEMBLOUX

(Belgique)

ÉTUDE DES SÉRIES CHRONOLOGIQUES PAR LES MÉTHODES DE DÉCOMPOSITION

R. PALM* et Y. BROSTAUX†

RÉSUMÉ

L'objectif de cette note est de présenter brièvement les méthodes de décomposition les plus utilisées lors de l'analyse des séries chronologiques.

Ces méthodes sont ensuite illustrées par un exemple concret, traité à l'aide du logiciel Minitab.

SUMMARY

This note briefly describes the main methods of analyzing time series by a decomposition into a trend, a seasonal variation, and residuals.

The methods are illustrated by an example analysed with Minitab software.

1. INTRODUCTION

Parmi les nombreuses méthodes d'étude des séries chronologiques, les méthodes de décomposition constituent sans doute l'approche la plus ancienne. La simplicité du principe qui est à la base de ces méthodes et le faible volume de calculs qu'elles exigent expliquent l'intérêt qu'on leur accorde encore actuellement. L'objectif de cette note est de présenter brièvement les méthodes de décomposition les plus courantes et de les illustrer par un exemple réel.

*Professeur à la Faculté universitaire des Sciences agronomiques de Gembloux.

†Chef de Travaux et Chargé de cours à la Faculté des Sciences agronomiques de Gembloux.

Le but poursuivi par ces méthodes est de décomposer la série chronologique en ses principales caractéristiques. Nous définirons d'abord ces différentes composantes, ainsi que leur mode de combinaison (paragraphe 2). Ensuite, nous tenterons d'analyser la tendance générale (paragraphe 3) et la composante saisonnière (paragraphe 4). Nous donnerons alors quelques indications relatives à l'étude des résidus (paragraphe 5). Enfin, nous terminerons par quelques conclusions (paragraphe 6).

Des informations complémentaires concernant les méthodes de décomposition sont données notamment par CALOT [1965] et par MORICE et CHARTIER [1954]. Pour les autres méthodes d'analyse (méthodes fréquentielles, prévisions par lissage, procédure de BOX et JENKINS et méthodes multivariées, notamment), des informations sont données par de nombreux auteurs, parmi lesquels on peut citer BOX et *al.* [1994], CHATFIELD [2003], COUTROT et DROESBEKE [1984], CRYER [1986] et GOURIEROUX et MONTFORT [1990]. Les méthodes de lissage et la méthode de BOX et JENKINS ont également fait l'objet de notes publiées dans la présente collection (PALM, 1987 et 2007).

2. TYPES DE VARIATIONS ET MODÈLES

On distingue généralement quatre composantes dans une série chronologique : la tendance, la composante cyclique, la composante saisonnière et la composante irrégulière.

La *tendance* représente l'évolution moyenne à long terme de la variable étudiée.

La *composante cyclique* présente théoriquement un comportement périodique, dont la période est suffisamment importante pour être qualifiée de mouvement à moyen ou à long terme. En pratique, elle est cependant difficile à dissocier de la tendance, car on ne dispose en général pas de séries stables d'une longueur correspondant à plusieurs cycles. De plus, l'amplitude et la période du phénomène cyclique ne sont pas nécessairement constantes. Par la suite, nous ne chercherons pas à dissocier la composante cyclique de la tendance proprement dite.

La *composante saisonnière* est également une composante périodique, dont la période est généralement dépendante de notre environnement. Cette liaison peut être naturelle (influence des saisons ou de l'alternance du jour et de la nuit, par exemple) ou due au mode d'organisation de notre société (impact des vacances, des fêtes de fin d'année, des week-ends, etc.). Plusieurs composantes saisonnières peuvent éventuellement être présentes dans une même série. Ce serait par exemple le cas lors de l'étude du chiffre de vente des grands magasins à partir de données journalières relatives à plusieurs années d'observations. La première composante prendrait en compte les variations liées aux jours de la semaine, et la seconde composante les variations liées aux mois de l'année.

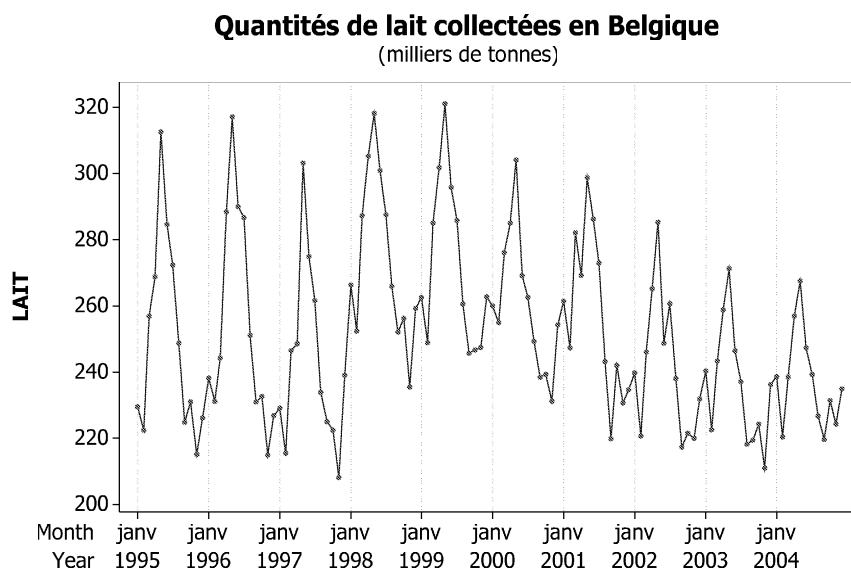


Figure 1 – Quantités de lait collectées en Belgique de janvier 1995 à décembre 2004

La *composante irrégulière* regroupe tout ce qui n'est pas pris en compte par les composantes précédentes. On considère qu'il s'agit d'une composante aléatoire, qui correspond à un résidu.

Pour certaines séries chronologiques, l'une ou l'autre des composantes citées ci-dessus peut éventuellement être absente de la série, ce qui en simplifie l'étude.

A titre d'illustration, considérons la série chronologique des quantités de lait de vache collectées en Belgique de janvier 1995 à décembre 2004 (en milliers de tonnes).¹

La figure 1 donne une représentation graphique de la série.

L'examen de cette figure fait clairement apparaître un effet saisonnier. On observe en effet systématiquement des quantités plus élevées durant les mois de printemps et des quantités plus faibles en hiver. Par ailleurs, on note également une tendance générale à la baisse à partir de 1999, précédée d'une hausse importante vers 1997-1998.

L'association des diverses composantes peut se faire de plusieurs façons. Un modèle classique est le schéma additif :

$$y_{ij} = m_{ij} + s_j + e_{ij} \quad (i = 1, \dots, l; j = 1, \dots, k).$$

¹Données EUROSTAT au 23/03/06 (<http://epp.eurostat.cec.eu.int/>).

Dans cette relation, et pour des données mensuelles par exemple ($k = 12$), y_{ij} est l'observation relative au j ème mois de la i ème année, m_{ij} représente la valeur de la tendance pour le j ème mois de la i ème année, s_j est la composante saisonnière du j ème mois, et e_{ij} est le résidu du j ème mois de la i ème année.

Un autre modèle courant est le modèle multiplicatif :

$$y_{ij} = m_{ij}s_je_{ij},$$

qui peut être ramené à un modèle additif par transformation logarithmique :

$$\log y_{ij} = \log m_{ij} + \log s_j + \log e_{ij}.$$

Enfin, on peut encore envisager des modèles mixtes, dont le plus courant est :

$$y_{ij} = m_{ij}s_j + e_{ij}.$$

Le choix du modèle dépend de la série à décomposer. Le schéma additif correspond à un mouvement saisonnier d'amplitude constante, alors que les deux autres schémas correspondent à un mouvement saisonnier d'amplitude proportionnelle au niveau de la série. De même, le modèle additif et le modèle mixte correspondent aux situations où l'écart-type de la composante irrégulière est constant, tandis que le modèle multiplicatif correspond aux situations où le coefficient de variation de la composante irrégulière est constant.

Pour les quantités de lait collectées, il semble bien que l'amplitude du phénomène saisonnier diminue avec les années, sans être toutefois clairement lié à la moyenne de l'année (figure 1). Un modèle multiplicatif ou mixte sera probablement plus indiqué qu'un modèle complètement additif.

3. ANALYSE DE LA TENDANCE GÉNÉRALE

La tendance générale, qui englobe le plus souvent aussi l'éventuelle composante cyclique, est désignée par m_{ij} pour indiquer qu'il s'agit du niveau moyen de la série au temps ij .

L'étude de la tendance générale se base uniquement sur le lien qui existe entre les valeurs observées y_{ij} et leur rang dans la série chronologique. Pour simplifier les notations, on peut donc remplacer le double indice ij par l'indice simple t :

$$m_{ij} = m_t,$$

avec :

$$t = k(i - 1) + j \quad (t = 1, \dots, n).$$

L'analyse de la tendance peut se faire par l'ajustement d'un modèle. Cette approche repose sur l'hypothèse que la tendance peut être représentée par une équation :

$$\hat{m}_t = f(t),$$

qu'on ajuste à la série par la méthode des moindres carrés, en considérant provisoirement que la composante saisonnière fait partie du résidu.

Le cas le plus simple est la tendance linéaire :

$$\hat{m}_t = b_0 + b_1 t,$$

mais des modèles plus complexes peuvent être envisagés si la tendance est non linéaire. Parmi ces modèles, les plus courants sont les suivants :

$$\hat{m}_t = b_0 + b_1 t + \dots + b_p t^p, \quad (1)$$

$$\hat{m}_t = b_0 e^{b_1 t}, \quad (2)$$

$$\hat{m}_t = b_0 t^{b_1}, \quad (3)$$

$$\hat{m}_t = b_0 / (1 + b_1 b_2^t), \quad (4)$$

$$\hat{m}_t = \exp(b_0 + b_1 b_2^t). \quad (5)$$

L'estimation des paramètres de l'équation polynomiale (équation 1) ne soulève pas de difficultés, car elle se fait par la méthode classique des moindres carrés, la non-linéarité ne portant pas sur les paramètres. Pour la fonction exponentielle (équation 2) et pour la fonction puissance (équation 3), on effectue une transformation logarithmique :

$$\log \hat{m}_t = \log b_0 + b_1 t \quad \text{ou} \quad \hat{m}'_t = b'_0 + b_1 t,$$

et

$$\log \hat{m}_t = \log b_0 + b_1 \log t \quad \text{ou} \quad \hat{m}'_t = b'_0 + b_1 t',$$

qui linéarise les modèles [DAGNELIE 2007]. Par contre, pour la fonction logistique (équation 4) et la fonction de GOMPERTZ (équation 5), le recours à la méthode des moindres carrés non linéaires s'impose, la non-linéarité portant sur les paramètres [DRAPER et SMITH, 1988].

On notera que la réalisation des tests statistiques permettant, par exemple, de vérifier le caractère significatif des coefficients de régression ne présente aucun intérêt à ce stade de l'étude, puisque la composante saisonnière, qui sera analysée ultérieurement, est provisoirement intégrée aux résidus.

Pour l'exemple du lait, la tendance étant, à première vue croissante au début de la série et ensuite décroissante, un modèle quadratique peut être envisagé. La figure 2 donne la représentation du modèle ajusté. Les paramètres de précision donnés dans la figure seront définis au paragraphe 5. Signalons simplement que le troisième paramètre, noté MSD, correspond à la variance des écarts par rapport au modèle quadratique. L'écart-type de ces écarts est donc égal à 24,7.

La tendance peut également être estimée par la méthode de la moyenne mobile. En particulier, si la série présente une composante saisonnière, l'ordre de la moyenne mobile est le plus souvent égal à la période k du phénomène saisonnier.

Si k est impair, on calcule la moyenne mobile symétrique simple d'ordre k :

$$\hat{m}_t = \frac{1}{k} (y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}),$$

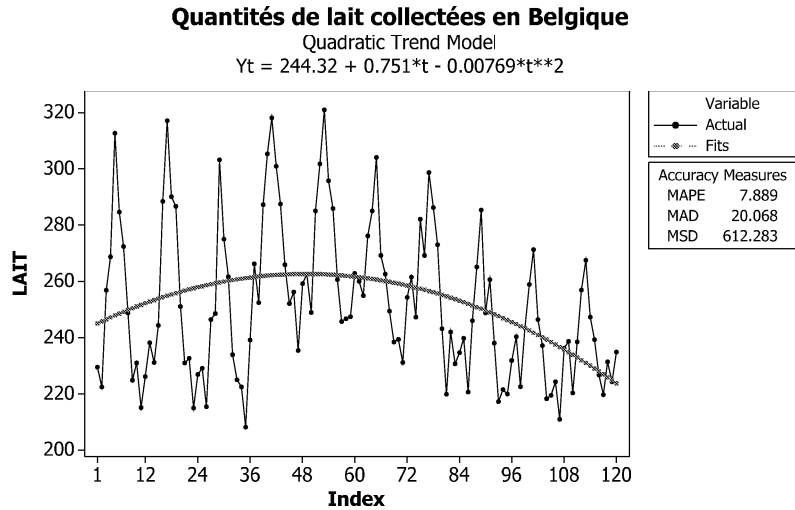


Figure 2 – Quantités de lait collectées tendance quadratique

avec $p = (k - 1)/2$ et t variant de $p + 1$ à $n - p$.

Par contre, si k est pair, on fait appel à la moyenne mobile symétrique d'ordre $k + 1$:

$$\hat{m}_t = \frac{1}{k} \left(\frac{1}{2} y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + \frac{1}{2} y_{t+p} \right),$$

avec $p = k/2$ et t variant également de $p + 1$ à $n - p$.

Cette dernière moyenne pondérée peut aussi être obtenue par le calcul des moyennes non pondérées d'ordre k , suivi du calcul des moyennes non pondérées d'ordre 2. Cette moyenne mobile itérée porte également le nom de moyenne mobile centrée d'ordre k .

Il faut remarquer que la série des moyennes mobiles symétriques est plus courte que la série initiale : elle comporte $2p$ valeurs en moins que la série de départ. De plus, le lissage par moyenne mobile symétrique ne permet pas la prévision. Si l'objectif poursuivi est la prévision, on utilisera éventuellement des moyennes mobiles non symétriques [COUTROT et DROESBEKE, 1984].

Pour illustrer le calcul des moyennes mobiles d'ordre 12, le tableau 1 reprend les premières observations pour les quantités de lait collectées, ainsi que les moyennes mobiles correspondantes.

Tableau 1 - Quantités de lait collectées et moyennes mobiles symétriques d'ordre 12 (données partielles)

Années	Mois	Observations	Moyennes mobiles
1995	1	229,501	-
1995	2	222,430	-
1995	3	256,836	-
1995	4	268,748	-
1995	5	312,672	-
1995	6	284,575	-
1995	7	272,330	249,769
1995	8	248,785	250,493
1995	9	224,791	250,334
1995	10	230,953	250,632
1995	11	215,102	251,636
1995	12	226,164	252,050
1996	1	238,189	252,877
1996	2	231,118	253,570
1996	3	244,331	253,921
1996	4	288,395	254,249
1996	5	317,127	254,311
1996	6	290,055	254,331

Pour le septième et le huitième mois de l'année 1995, les moyennes mobiles se calculent comme suit :

$$(229,501/2 + 222,430 + \dots + 226,164 + 238,189/2)/12 = 249,769,$$

$$(222,430/2 + 256,836 + \dots + 238,189 + 231,118/2)/12 = 250,493,$$

et ainsi de suite pour les autres mois.

Dans ce calcul des moyennes mobiles, on constate que les douze mois de l'année interviennent avec un même poids et que les mois qui précèdent le mois pour lequel se fait le calcul ont un poids identique à celui des mois qui suivent le mois en question.

La figure 3 reprend ces moyennes mobiles. Leur évolution est assez irrégulière dans le début de la série et elles décroissent ensuite de manière plus régulière. L'écart-type des écarts entre les observations et les moyennes mobiles est de 23,5, soit une valeur légèrement plus faible que l'écart-type des écarts par rapport au modèle quadratique, qui, rappelons-le, est égal à 24,7. On notera cependant que ces deux écarts-types ne sont pas calculés sur le même nombre d'écarts : pour le modèle quadratique, on a 120 écarts, alors que pour la moyenne mobile, on n'a que 108 écarts, les moyennes mobiles n'étant en effet pas calculées pour les six premiers et pour les six derniers mois.

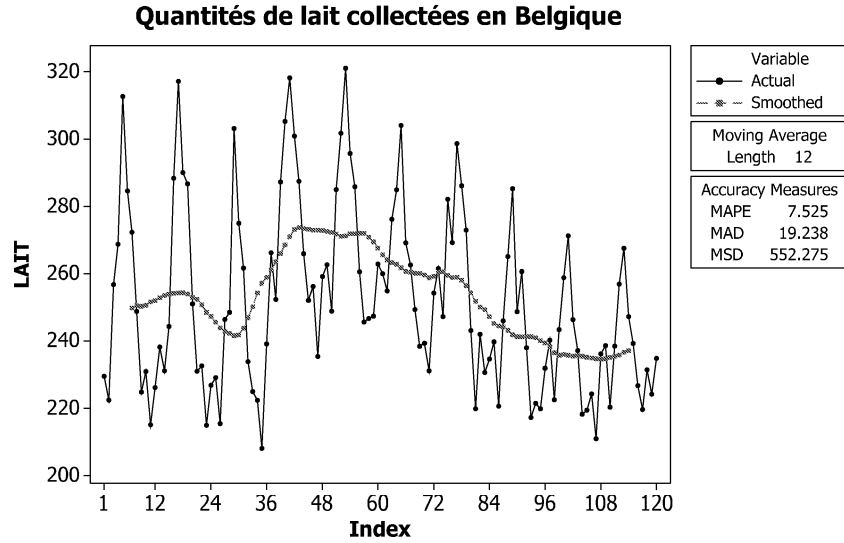


Figure 3 – Quantités de lait collectées et moyennes mobiles d'ordre 12

4. ANALYSE DE LA COMPOSANTE SAISONNIÈRE

Pour étudier la composante saisonnière, on élimine tout d'abord la tendance, en la soustrayant des données initiales si la composante saisonnière est additive :

$$y'_{ij} = y_{ij} - \hat{m}_{ij} = s_{ij} + e_{ij},$$

ou en divisant les données de départ par la tendance si la composante saisonnière est multiplicative :

$$y'_{ij} = \frac{y_{ij}}{\hat{m}_{ij}} = s_j e_{ij} \quad (\text{modèle multiplicatif})$$

ou

$$y'_{ij} = \frac{y_{ij}}{\hat{m}_{ij}} = s_j + \frac{e_{ij}}{\hat{m}_{ij}} \quad (\text{modèle mixte}).$$

On calcule alors les valeurs moyennes :

$$\bar{y}'_j = \frac{1}{l} \sum_{i=1}^l y'_{ij} \quad (j = 1, \dots, k).$$

Ces moyennes correspondent aux écarts ou aux coefficients saisonniers non corrigés. Ceux-ci sont ensuite corrigés de manière à rendre leur moyenne égale à 0 :

$$\hat{s}_j = \bar{y}'_j - \frac{1}{k} \sum_{j=1}^k \bar{y}'_j,$$

ou à 1 :

$$\hat{s}_j = \bar{y}'_j - \frac{1}{k} \sum_{j=1}^k (\bar{y}'_j - 1),$$

selon que la composante saisonnière est additive ou multiplicative.

Pour un modèle additif et une tendance donnée par une équation de régression, les écarts saisonniers peuvent également être déterminés par régression multiple à partir de variables instrumentales [CRYER, 1986 ; GOURIEROUX et MONFORT, 1990].

En effet, pour le modèle additif à tendance linéaire par exemple, on peut créer k variables artificielles $x_j (j = 1, \dots, k)$, dont les valeurs x_{jt} sont égales à 1 ou à 0 selon que l'observation y_t est ou n'est pas relative au j ème mois. Le modèle peut alors s'écrire simplement :

$$\hat{y}_t = bt + \sum_{j=1}^k c_j x_{jt}.$$

Par régression multiple, on obtient les estimations de b et des c_j . La moyenne des c_j représente le terme indépendant de la tendance :

$$b_0 = \frac{1}{k} \sum_{j=1}^k c_j,$$

et les différences entre les c_j et leur moyenne constituent les écarts saisonniers :

$$\hat{s}_j = c_j - b_0.$$

Les valeurs obtenues par la régression multiple ne sont toutefois pas identiques à celles obtenues par la méthode établie précédemment, le critère d'estimation étant différent.

La régression multiple peut également être appliquée dans le cas d'un modèle multiplicatif avec tendance décrite par une relation exponentielle, la transformation logarithmique rendant la tendance linéaire et le modèle additif.

Les procédures de calcul présentées ci-dessus peuvent aussi faire l'objet de variantes. Ainsi, dans la méthode de décomposition proposée par le logiciel Minitab, la tendance est évaluée *a priori* à partir des moyennes mobiles. Les écarts ou les rapports à cette moyenne sont alors calculés. Les écarts ou les coefficients saisonniers sont les médianes, et non les moyennes, déterminées pour

chaque mois, et corrigées de manière à ce que leur moyenne soit nulle ou égale à 1.

Pour les quantités de lait collectées et pour une tendance quadratique, le tableau 2 reprend les moyennes par mois des écarts ($y_{ij} - \hat{m}_{ij}$) et des rapports (y_{ij}/\hat{m}_{ij}). Ces moyennes correspondent aux écarts et aux coefficients non corrigés. Toutefois, dans la mesure où la moyenne des écarts saisonniers est déjà nulle et la moyenne des coefficients saisonniers est déjà égale à 1, ces deux séries de valeurs correspondent également aux écarts et aux coefficients corrigés.

Tableau 2 - Ecart saisonniers (modèle additif) et coefficients saisonniers (modèle mixte)

Mois	Ecart saisonniers	Coefficients saisonniers
Janvier	-6,6433	0,97376
Février	-19,5299	0,92260
Mars	7,5992	1,02943
Avril	21,9167	1,08656
Mai	47,1818	1,18613
Juin	21,8443	1,08584
Juillet	14,2886	1,05612
Août	-8,6038	0,96583
Septembre	-22,6710	0,91048
Octobre	-17,0106	0,93351
Novembre	-27,7220	0,89115
Décembre	-10,6499	0,95889

5. ANALYSE DE LA COMPOSANTE IRRÉGULIÈRE

La composante irrégulière ne fait pas l'objet de modélisation, puisqu'elle constitue un résidu, englobant tout ce qui n'a pas été pris en considération par les composantes précédentes.

Un examen des résidus se justifie cependant pour vérifier l'adéquation du modèle de décomposition utilisé et éventuellement pour comparer les résultats de plusieurs modèles de décomposition.

Pour un modèle additif ou mixte, les résidus sont donnés par la relation :

$$e_{ij} = y_{ij} - \hat{y}_{ij},$$

avec

$$\hat{y}_{ij} = \hat{m}_{ij} + \hat{s}_j \quad (\text{modèle additif}),$$

et

$$\hat{y}_{ij} = \hat{m}_{ij} \hat{s}_j \quad (\text{modèle mixte}).$$

Pour un modèle multiplicatif, on a :

$$e_{ij} = \frac{y_{ij}}{\hat{y}_{ij}},$$

avec

$$\hat{y}_{ij} = \hat{m}_{ij} \hat{s}_j.$$

Dans la mesure où ce dernier modèle est le plus souvent traité comme un modèle additif après transformation logarithmique, nous n'envisageons plus de manière particulière ce dernier cas.

Pour simplifier les notations, on peut, comme pour l'étude de la tendance, remplacer le double indice par l'indice simple :

$$e_{ij} = e_t.$$

Pour quantifier l'importance des résidus, on peut, notamment, calculer :

- l'écart relatif moyen absolu² :

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \frac{|e_t|}{y_t},$$

- l'écart moyen absolu³ :

$$\text{MAD} = \sum_{t=1}^n |e_t|/n,$$

- le carré moyen⁴ :

$$\text{MSD} = \sum_{t=1}^n e_t^2/n,$$

dont la racine carrée, notée RMSD ⁵, est analogue à un écart-type résiduel.

En plus de la mesure de l'importance des résidus, on peut vérifier si ceux-ci présentent un caractère aléatoire. Différentes méthodes sont disponibles et nous en retenons deux : le test du nombre de suites homogènes et le test basé sur le quotient de VON NEUMANN.

Dans le cas d'une composante irrégulière additive, le premier test est destiné à vérifier si le nombre observé p de suites de résidus de même signe est compatible avec l'hypothèse d'une répartition aléatoire des signes des résidus, compte tenu du fait que les nombres de résidus positifs et négatifs sont théoriquement identiques en moyenne. Dans le cas de séries comportant un nombre suffisant

²En anglais : *Mean absolute percentage error, MAPE.*

³En anglais : *Mean absolute deviation, MAD.*

⁴En anglais : *Mean square deviation, MSD.*

⁵En anglais : *Root mean square deviation, RMSD.*

d'observations (n supérieur à 20), la distribution de p est approximativement normale. On calcule la valeur :

$$u_{obs} = \frac{|p - \frac{n+2}{2}|}{\sqrt{\frac{n-1}{4}}},$$

et on rejette le caractère aléatoire de la succession des signes plus et moins si u_{obs} est supérieur ou égal à $u_{1-\alpha/2}$ [MORICE et CHARTIER, 1954].

Si la composante irrégulière est multiplicative, le test peut être appliqué, de la même façon, aux logarithmes des résidus. Le test de VON NEUMANN [MORICE et CHARTIER, 1954] a pour but de vérifier l'absence de corrélation entre les résidus successifs. Ce test, basé sur le caractère asymptotiquement normal du quotient :

$$q = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{2 \sum_{t=1}^n (e_t - \bar{e})^2},$$

constiste à comparer la valeur :

$$u_{obs} = \frac{|q - 1|}{\sqrt{\frac{n-2}{n^2-1}}},$$

à la valeur $u_{1-\alpha/2}$ et à rejeter l'hypothèse du caractère aléatoire des résidus si la valeur u_{obs} est supérieure ou égale à $u_{1-\alpha/2}$, une valeur de q inférieure à l'unité correspondant à une corrélation positive et une valeur supérieure à l'unité à une corrélation négative entre les résidus consécutifs.

Il faut cependant remarquer que ce test vérifie uniquement l'absence de corrélation entre e_t et e_{t-1} , c'est-à-dire la corrélation de rang 1, ce qui n'exclut pas l'existence de corrélations de rang supérieur à 1, entre e_t et e_{t-j} (j supérieur à 1). Des informations sont données à ce sujet par BOX *et al.* [1994], ainsi que dans une autre note [PALM, 1987].

Les figures 4 et 5 donnent, pour les premières observations de la série, les résultats de la décomposition, d'une part, pour le modèle additif et, d'autre part, pour le modèle mixte, la tendance étant quadratique. Pour la première observation, prise à titre d'exemple, on a les décompositions suivantes :

$$229,50 = 245,06 - 6,64 - 8,91$$

et

$$229,50 = (245,06)(0,97376) - 9,13.$$

Row	A	M	LAIT	TENDANCE	E. SAIS.	MOD_ADD	RES_ADD
1	1995	1	229.501	245.059	-6.6433	238.416	-8.9148
2	1995	2	222.430	245.787	-19.5299	226.257	-3.8274
3	1995	3	256.836	246.500	7.5992	254.099	2.7369
4	1995	4	268.748	247.197	21.9167	269.114	-0.3660
5	1995	5	312.672	247.879	47.1818	295.061	17.6109
6	1995	6	284.575	248.546	21.8443	270.390	14.1848
7	1995	7	272.330	249.197	14.2886	263.486	8.8444
8	1995	8	248.785	249.833	-8.6038	241.229	7.5560
9	1995	9	224.791	250.453	-22.6710	227.782	-2.9913
10	1995	10	230.953	251.058	-17.0106	234.048	-3.0946
11	1995	11	215.102	251.648	-27.7220	223.926	-8.8239
12	1995	12	226.164	252.222	-10.6499	241.572	-15.4083
13	1996	1	238.189	252.781	-6.6433	246.138	-7.9488
14	1996	2	231.118	253.325	-19.5299	233.795	-2.6767
15	1996	3	244.331	253.853	7.5992	261.452	-17.1209
16	1996	4	288.395	254.365	21.9167	276.282	12.1129
17	1996	5	317.127	254.863	47.1818	302.045	15.0824
18	1996	6	290.055	255.345	21.8443	277.189	12.8660

Figure 4 – Décomposition par le modèle additif (données partielles)

Row	A	M	LAIT	TENDANCE	C. SAIS.	MOD_MULT	RES_MULT
1	1995	1	229.501	245.059	0.97376	238.628	-9.1272
2	1995	2	222.430	245.787	0.92260	226.764	-4.3338
3	1995	3	256.836	246.500	1.02943	253.755	3.0807
4	1995	4	268.748	247.197	1.08656	268.596	0.1523
5	1995	5	312.672	247.879	1.18613	294.016	18.6561
6	1995	6	284.575	248.546	1.08584	269.882	14.6931
7	1995	7	272.330	249.197	1.05612	263.182	9.1478
8	1995	8	248.785	249.833	0.96583	241.296	7.4893
9	1995	9	224.791	250.453	0.91048	228.033	-3.2419
10	1995	10	230.953	251.058	0.93351	234.365	-3.4119
11	1995	11	215.102	251.648	0.89115	224.255	-9.1533
12	1995	12	226.164	252.222	0.95889	241.854	-15.6904
13	1996	1	238.189	252.781	0.97376	246.148	-7.9585
14	1996	2	231.118	253.325	0.92260	233.718	-2.5998
15	1996	3	244.331	253.853	1.02943	261.324	-16.9934
16	1996	4	288.395	254.365	1.08656	276.384	12.0107
17	1996	5	317.127	254.863	1.18613	302.299	14.8278
18	1996	6	290.055	255.345	1.08584	277.264	12.7906

Figure 5 – Décomposition par le modèle mixte (données partielles)

Le tableau 3 reprend les différentes caractéristiques calculées en vue de quantifier la qualité des deux modèles. Les premiers paramètres, basés sur la valeur des résidus, montrent une légère supériorité du modèle mixte sur le modèle additif. La situation est cependant inversée pour le test des suites homogènes et les deux modèles sont équivalents pour le test de VON NEUMANN. On doit bien constater que pour les deux modèles considérés, les résidus ne présentent pas le caractère aléatoire souhaité.

Tableau 3 - Critères d'adéquation du modèle additif et du modèle mixte

Critères	Modèle additif	Modèle mixte
MAPE	4,0	3,9
MAD	9,96	9,77
MSD	143,08	137,34
RMSD	12,0	11,7
Nombre de résidus positifs	62	63
Nombre de résidus négatifs	58	57
Nombre de suites homogènes (p)	26	24
Test des suites u_{obs}	6,4	6,7
Quotient de VON NEUMANN (q)	0,26	0,26
Test de VON NEUMANN u_{obs}	8,2	8,2

Malgré son caractère imparfait, nous retiendrons le modèle mixte, car il ne semble pas qu'il soit possible de trouver un autre modèle de décomposition simple qui soit nettement meilleur. La figure 6 donne la représentation de la décomposition par le modèle mixte.

La figure 7 donne les prédictions qu'on aurait pu réaliser, à l'issue de l'année 2004, pour les 12 mois de l'année 2005. Le tableau reprend les valeurs de la tendance, les coefficients saisonniers et les prédictions, qui sont le produit de la tendance et du coefficient saisonnier. Ces prédictions sont également confrontées aux quantités de lait qui ont effectivement été collectées et les erreurs de prédiction ont été calculées.

A l'examen de ce tableau, on remarque que les prévisions sont systématiquement inférieures aux observations. Cette sous-estimation résulte du fait que la tendance à la baisse du niveau moyen des quantités de lait collectées, observée depuis l'année 2000, ne s'est pas poursuivie en 2005.

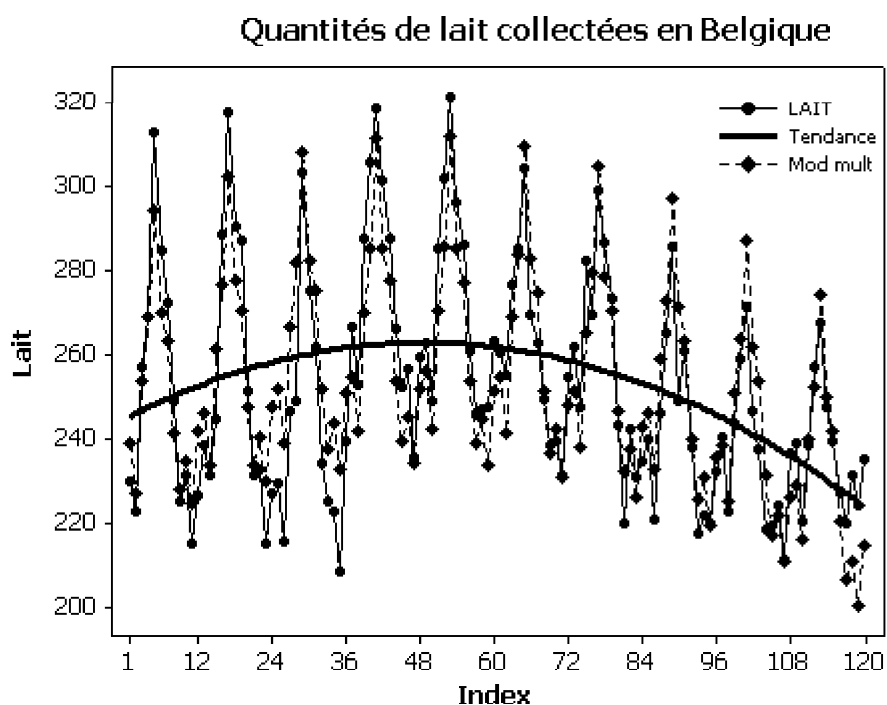


Figure 6 – Quantités de lait collectées : modèle mixte avec tendance quadratique

M	TENDANCE	C. SAIS	PREVISION	OBS.	ERREUR
1	222.586	0.97376	216.745	240.664	23.9191
2	221.468	0.92260	204.327	219.785	15.4581
3	220.335	1.02943	226.820	234.079	7.2592
4	219.186	1.08656	238.159	256.598	18.4386
5	218.022	1.18613	258.601	261.017	2.4160
6	216.842	1.08584	235.456	245.439	9.9825
7	215.647	1.05612	227.749	246.614	18.8647
8	214.437	0.96583	207.109	239.995	32.8859
9	213.211	0.91048	194.124	228.020	33.8956
10	211.970	0.93351	197.875	231.634	33.7585
11	210.713	0.89115	187.776	223.544	35.7675
12	209.441	0.95889	200.832	226.000	25.1679

Figure 7 – Prédictions pour 2005 par le modèle mixte

6. CONCLUSIONS

L'étude des séries chronologiques par les méthodes de décomposition est une approche relativement simple dans son principe : elle consiste à isoler et à estimer l'impact des diverses composantes de la série.

Des difficultés peuvent cependant se présenter lors de l'application de ce principe. Le choix de la relation mathématique permettant de décrire la tendance et le choix du mode d'association des composantes ne sont, en effet, pas toujours évidents.

De plus, les méthodes de décomposition envisagées au cours des paragraphes précédents ne permettent pas toujours de décomposer les séries chronologiques de façon tout à fait satisfaisante, comme le montre l'exemple traité dans les paragraphes précédents, les hypothèses sous-jacentes aux différents modèles étant relativement restrictives. Il en résulte que les prévisions réalisées à l'aide de ces modèles ne sont pas toujours de bonne qualité et, si l'objectif de l'étude de la série est uniquement la prévision, le recours aux méthodes signalées dans l'introduction (méthode de lissage et procédure de BOX et JENKINS notamment) peut être utile. Une description de ces méthodes est donnée dans PALM (1987 et 2007).

Les méthodes de décomposition peuvent, par contre, être très intéressantes lorsque l'objectif est uniquement d'éliminer une ou plusieurs composantes d'une série chronologique. C'est le cas, par exemple, pour l'étude d'indices économiques, lorsqu'on souhaite éliminer des séries les variations saisonnières [CALOT, 1965 ; FOURASTIER et GRAIS, 1984]. On notera aussi que, dans ce contexte, de nombreuses méthodes plus élaborées ont été proposées. Parmi celles-ci, on citera notamment la méthode X11, ainsi que ses diverses variantes. Une description détaillée en est donnée par LADIRAY et QUENNEVILLE [1999].

BIBLIOGRAPHIE

- BOX G., JENKINS G. et REINSEL G. [1994]. *Time series analysis, forecasting and control*. Englewood Cliffs, Prentice-Hall, 598 p.
- CALOT G., [1965]. *Cours de statistique descriptive*. Paris, Dunod, 519 p.
- CHATFIELD C. [2003]. *The analysis of time series : an introduction*. Boca Raton, Chapman&Hall/CRC, 333 p.
- COUTROT B. et DROESBEKE F. [1984]. *Les méthodes de prévision*. Paris, Presses Universitaires de France, 128 p.
- CRYER J. D. [1986]. *Time series analysis*. Boston, Duxbury, 286 p.

- DAGNELIE P. [2007]. *Statistique théorique et appliquée - Vol. 1 : Statistique descriptive et bases de l'inférence statistique*. Bruxelles, De Boeck et Larcier, 511 p.
- DRAPER N. R. et SMITH H. [1988]. *Applied regression analysis*. New York, Wiley, 706 p.
- FOURASTIER J. et GRAIS B. [1984]. *Les indices statistiques*. Paris, Masson, 149 p.
- GOURIEROUX C., MONTFORT A. [1990]. *Séries temporelles et modèles dynamiques*. Paris, Economica, 780 p.
- KIVELIOVITCH M. et VIALAR J. [1957]. *Les séries chronologiques et la théorie du hasard*. Paris, Publications scientifiques et Techniques du Ministère de l'Air, 129 p.
- LADIRAY D., QUENNEVILLE B. [1999]. *Comprendre la méthode X11*. Document PDF, <<http://www.census.gov/ts/papers/x11.pdf>>, 192 p.
- MORICE E. et CHARTIER F. [1954]. *Méthodes statistiques - Vol. 2*. Paris, Imprimerie Nationales, 555 p.
- PALM R. [1986]. Etude des résidus de régression : principes et application. *Notes Stat. Inform.* Gembloux 86/1, 13 p.
- PALM R. [1987]. Etude des séries chronologiques par la méthode de BOX et JENKINS. *Notes stat. Inform.* Gembloux 87/2, 40 p.
- PALM R. [2007]. Etude des séries chronologiques par les méthodes de lissage. *Notes stat. Inform.* Gembloux, 2007/1, 22 p.

La collection

NOTES DE STATISTIQUE ET D'INFORMATIQUE

réunit divers travaux (documents didactiques, notes techniques, rapports de recherche, publications, etc.) émanant de l'Unité de Statistique, Informatique et Mathématique appliquées de la Faculté universitaire des Sciences agronomiques et du Département de Biométrie, Gestion des données et Agrométéorologie du Centre wallon de Recherches agronomiques (Gembloux - Belgique).

La liste des notes disponibles peut être obtenue sur simple demande à l'adresse ci-dessous :

*Faculté universitaire des Sciences agronomiques
Unité de Statistique et Informatique
Avenue de la Faculté d'Agronomie, 8
B-5030 GEMBLoux (Belgique)
E-mail : statinfo@fsagx.ac.be*

Plusieurs notes sont directement accessibles à l'adresse Web suivante, section Publications :

<http://www.fsagx.ac.be/si/>

En relation avec certaines notes, des programmes spécifiques sont également disponibles à la même adresse, section Macros.

Quelques titres récents sont cités ci-après :

- CARLETTI I., PREVOT H. [2006]. Traitement des données par le logiciel SAS : introduction au module de base. *Notes Stat. Inform.* (Gembloux) 2006/3, 31 p.
- DUYME F., CLAUSTRIAUX J.J. [2006]. La régression logistique binaire. *Notes Stat. Inform.* (Gembloux) 2006/4, 24 p.
- PALM R. [2007]. Etude des séries chronologiques par les méthodes de lissage. *Notes Stat. Inform.* (Gembloux) 2007/1, 22 p.
- PALM R. [2007]. L'analyse des correspondances multiples : principes et application. *Notes Stat. Inform.* (Gembloux) 2007/2, 28 p.
- PALM R. [2008]. Détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée selon la norme ISO 5725-2. *Notes Stat. Inform.* (Gembloux) 2008/1, 22 p.
- CHARLES C., LECHARLIER L., RENAUD F. [2008]. Introduction à LATEX. *Notes Stat. Inform.* (Gembloux) 2008/2, 21 p.
- CHARLES C. [2008]. Introduction à OCTAVE. *Notes Stat. Inform.* (Gembloux) 2008/3, 19 p.