

MACROS MINITAB POUR L'ANALYSE DES SERIES CHRONOLOGIQUES PAR DECOMPOSITION

R. PALM¹

1. Introduction

Deux commandes Minitab principales sont à la disposition de l'utilisateur pour la décomposition des séries chronologiques : la commande TREND et la commande DECOMP.

L'utilisation conjointe de ces deux commandes s'impose pour l'analyse de séries lorsque la tendance n'est pas linéaire, ce qui alourdit inévitablement l'analyse. C'est la raison pour laquelle l'Unité de Statistique, Informatique et Mathématique appliquées de la FUSAGx² propose aux utilisateurs du logiciel Minitab une macro, basée sur les commandes TREND et DECOMP, qui réalise, de manière automatique, la décomposition de la série chronologique, en fonction d'options choisies par l'utilisateur. Cette macro ainsi que sa notice d'utilisation sont disponibles sur le site web de l'Unité :

www.fsagx.ac.be/si/

en cliquant sur le lien Macros, puis sur le thème en question.

L'objectif de cette note est de présenter cette macro et d'illustrer son utilisation par un exemple.

1. Professeur à la Faculté universitaire des Sciences agronomiques de Gembloux.
2. Faculté universitaire des Sciences agronomiques de Gembloux.

Nous donnons d'abord le principe général des méthodes de décomposition (paragraphe 2). Ensuite, nous examinons les commandes TREND et DECOMP, qui sont utilisées dans la macro proposée (paragraphe 3 et 4). Nous présentons alors la macro (paragraphe 5) et nous donnons un exemple d'application (paragraphe 6) avant de conclure (paragraphe 7).

2. Méthodes de décomposition

On considère qu'une observation donnée d'une série chronologique est le résultat de l'agrégation de plusieurs composantes, dont les plus courantes sont la tendance, la composante saisonnière et la composante irrégulière.

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

La composante saisonnière est une composante périodique dont la période est généralement dépendante de notre environnement. Ainsi, pour des données économiques récoltées mensuellement, la période sera égale à 12.

La composante irrégulière prend en compte ce qui n'est pas repris par les composantes précédentes ; elle correspond au résidu du modèle.

Pour certaines séries, l'une ou l'autre composante peut être absente, ce qui simplifie l'étude de la série. A l'inverse, certaines séries peuvent contenir plus d'une composante saisonnière. Cette situation ne sera cependant pas examinée ici.

Plusieurs modes de combinaisons des composantes peuvent être envisagés. En désignant par *OBS* l'observation, par *T* la tendance, par *S* la composante saisonnière et par *E* la composante irrégulière, on peut distinguer les trois modèles classiques suivants :

$$OBS = T + S + E \text{ (modèle additif),}$$

$$OBS = T * S * E \text{ (modèle multiplicatif),}$$

et

$$OBS = T * S + E \text{ (modèle mixte).}$$

Le choix de l'un ou l'autre mode d'association dépend de la nature de la série, mais dans de nombreux cas pratiques, le choix ne sera pas évident *a priori*. On pourra éventuellement considérer plusieurs modèles et choisir celui qui paraît le plus satisfaisant, sur la base de l'examen des résidus.

3. La commande TREND

La commande TREND permet l'ajustement d'une équation de tendance à une série chronologique. Les quatre modèles suivants sont proposés :

$$\begin{aligned}
 y_t &= \beta_0 + \beta_1 t + e_t && \text{(tendance linéaire),} \\
 y_t &= \beta_0 + \beta_1 t + \beta_2 t^2 + e_t && \text{(tendance quadratique),} \\
 y_t &= \beta_0 \beta_1^t + e_t && \text{(tendance exponentielle),} \\
 y_t &= (10^a) / (\beta_0 + \beta_1 \beta_2^t) + e_t && \text{(tendance avec courbe en S).}
 \end{aligned}$$

Dans ces relations, y_t représente la valeur observée de la série au temps t , β_0 , β_1 et β_2 sont des paramètres et a est un nombre entier considéré comme un facteur d'échelle. Les e_t sont les résidus par rapport à la tendance.

La commande TREND permet également le calcul des valeurs ajustées et des résidus ainsi que des prédictions. Ces résultats peuvent être enregistrés dans le fichier Minitab. Différents graphiques des résidus sont également proposés, de même qu'une visualisation des données initiales et du modèle ajusté.

Trois paramètres permettant de quantifier la qualité du modèle sont calculés à partir des n résidus :

- l'écart relatif moyen absolu³ :

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

- l'écart moyen absolu⁴ :

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

- le carré moyen des résidus⁵ :

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

Enfin, une option particulière permet de tenir compte d'une information *a priori* sur les paramètres du modèle retenu. Si l'utilisateur fournit des paramètres *a priori* β_{j0} et des poids α_j , les paramètres retenus pour le modèle final seront donnés par les relations :

$$\hat{\beta}_j = \alpha_j \hat{\beta}_{jc} + (1 - \alpha_j) \beta_{j0},$$

3. En anglais : *Mean absolute percentage error*.

4. En anglais : *Mean absolute deviation*.

5. En anglais : *Mean square deviation*.

$\hat{\beta}_j$ étant l'estimation finale et $\hat{\beta}_{jc}$ étant l'estimation calculée à partir des données.

4. La commande DECOMP

Cette commande permet le calcul de la composante saisonnière pour un modèle additif et pour un modèle mixte.

La détermination de cette composante se fait de la manière suivante.

- a) Minitab calcule les moyennes mobiles d'ordre k , k étant la période du phénomène saisonnier.
- b) Pour une composante saisonnière additive, Minitab calcule les écarts entre les observations et la moyenne mobile ; pour une composante saisonnière multiplicative, le logiciel divise les observations par la moyenne mobile.
- c) Pour chaque unité de temps du cycle saisonnier (par exemple, pour chaque mois de l'année dans le cas des données mensuelles), Minitab détermine la médiane des écarts par rapport à la moyenne mobile ou des rapports à la moyenne mobile. Il ajuste ensuite ces médianes de manière à ce que leur moyenne soit égale à zéro pour le modèle additif ou à l'unité pour le modèle mixte. Ces médianes ajustées constituent alors les écarts ou les coefficients saisonniers.

La commande DECOMP permet de prendre en compte ou non la présence d'une tendance. Dans ce dernier cas, la tendance considérée est linéaire. Des options permettent d'enregistrer dans le fichier Minitab différentes caractéristiques. Le tableau 1 résume la manière dont sont obtenus les résultats, en fonction du modèle retenu. La première colonne du tableau reprend la dénomination utilisée dans la version anglaise du logiciel, ainsi que le nom des colonnes correspondantes dans le fichier Minitab si l'enregistrement des résultats a été demandé. Les symboles utilisés dans les autres colonnes du tableau 1 représentent des constantes ou des variables, selon les cas. Ainsi, par exemple, *MOY* est la moyenne des observations de la série ; il s'agit donc d'une constante, qui ne dépend pas de la période de temps t considérée. A l'inverse, *OBS* représente les observations, qui, bien entendu, varient avec l'indice t .

Pour les modèles avec tendance, nous avons déjà signalé que Minitab ne considère que le cas d'une tendance linéaire. Celle-ci est calculée à partir de la série désaisonnalisée, en exprimant la variable DESES en fonction du numéro d'ordre t de l'observation. Il en résulte que la tendance sera différente pour le modèle additif et pour le modèle multiplicatif puisque, dans le premier cas, la variable à expliquer est la variable contenant les observations dont on a soustrait les écarts saisonniers alors que, dans le second cas, il s'agit des observations divisées par les coefficients saisonniers.

La commande DECOMP propose également différentes représentations graphiques ainsi que les paramètres MAPE, MAD et MSD, définis à partir des résidus par rapport à la tendance, par les formules présentées au paragraphe 3.

Caractéristiques	Modèle additif		Modèle multiplicatif	
	Sans tendance	Avec tendance	Sans tendance	Avec tendance
Trend line (TREN)	<i>MOY</i>	$T = a + bt$	<i>MOY</i>	$T = a + bt$
Detrended data (DETR)	$OBS - MOY$	$OBS - T$	OBS / MOY	OBS / MOY
Seasonals (SEAS)	<i>ES</i>	<i>ES</i>	<i>CS</i>	<i>CS</i>
Seasonally adjusted data (DESE)	$OBS - ES$	$OBS - ES$	OBS / CS	OBS / CS
Fits (FITS)	$MOY + ES$	$T + ES$	$MOY * CS$	$T * CS$
Residuals (RESI)	$OBS - (MOY + ES)$	$OBS - (T + ES)$	$OBS - (MOY * CS)$	$OBS - (T * CS)$
Forecasts (FORE)	$MOY + ES$	$T + ES$	$MOY * CS$	$T * CS$

Tableau 1. Mode de calcul des caractéristiques par la commande DECOMP de Minitab (*MOY* : moyenne de la série chronologique, *OBS* : valeur observée, *T* : valeur de la tendance, *ES* : écart saisonnier, *CS* : coefficient saisonnier).

5. La macro SCDECOMP

La macro DECOMP, présentée au paragraphe 4 a l'inconvénient de ne pas prendre en compte, de façon automatique, une éventuelle tendance non linéaire. C'est la raison pour laquelle une macro spécifique, combinant les potentialités des commandes TREND et DECOMP a été mise au point. Par diverses options, cette macro permet la prise en compte d'une tendance linéaire, quadratique, exponentielle ou sigmoïde ou encore d'une tendance donnée par les moyennes mobiles. Dans ce dernier cas cependant, les prédictions ne peuvent pas être fournies.

Les modèles peuvent être additifs ou mixtes. L'établissement de différents graphiques est en option, de même que l'enregistrement des valeurs de la tendance, de la composante saisonnière et des résidus.

La tendance est toujours calculée à partir des observations de départ et non à partir des observations corrigées des variations saisonnières, comme dans la commande DECOMP (paragraphe 4).

La macro détermine également diverses caractéristiques liées à l'analyse des résidus. Ainsi, outre les paramètres déjà décrits précédemment (MAPE, MAD et MSD), elle donne encore le paramètre RMSD, qui est la racine carrée de MSD et qui correspond à un écart-type résiduel ainsi que la part de la variance de la série $\hat{\sigma}_y^2$ qui est expliquée par le modèle :

$$R_{sq} = 100 \left(1 - \frac{\text{MSD}}{\hat{\sigma}_y^2} \right).$$

Elle fournit aussi les informations relatives au test des suites et au test de VON NEUMANN. Ces tests sont décrits dans PALM [2009].

Pour des informations plus détaillées, nous renvoyons le lecteur à la notice d'utilisation de la macro SCDECOMP.

6. Un exemple d'utilisation

Pour illustrer l'utilisation de la macro SCDECOMP, nous reprenons les données relatives aux quantités de lait de vache collectées en Belgique de janvier 1995 à décembre 2005 (en milliers de tonnes), examinées dans PALM [2009]. L'année 2005 ne sera pas prise en compte pour la détermination du modèle mais sera utilisée pour la validation des prévisions.

Un examen préliminaire de la série chronologique a montré que celle-ci présente, à première vue, une tendance d'abord croissante puis décroissante. Le modèle de tendance quadratique sera donc retenu. Un examen plus approfondi a, par ailleurs, montré que le modèle mixte est légèrement supérieur au modèle additif [PALM, 2009].

La figure 1 donne les commandes utilisées et les résultats de l'exécution pour le modèle mixte avec tendance quadratique.

La première partie de la figure résulte de l'exécution, au sein de la macro SCDECOMP, de la commande DECOMP en l'absence de tendance. La figure reprend les coefficients saisonniers et les paramètres relatifs à la série des résidus calculés de la manière suivante :

$$E = OBS - (MOY * CS).$$

Les prévisions données dans la figure 1 correspondent aux coefficients saisonniers *CS* multipliés par la moyenne *MOY*.

La seconde partie de la figure provient de la commande TREND, incluse dans la macro SCDECOMP, et appliquée à la série des observations de départ. Les paramètres de précision sont dans ce cas relatifs aux écarts :

$$E = OBS - T,$$

la tendance étant exprimée par l'équation du second degré.

La troisième partie de la figure concerne l'étude des résidus du modèle complet qui s'écrivent :

$$E = OBS - T * CE.$$

Enfin, la dernière partie de la figure donne les informations sur les erreurs de prédiction pour l'année 2005, utilisée pour la validation. Les quatre paramètres repris sont ceux qui ont été définis au paragraphe 5.

La figure 2 reprend le graphique de la série initiale, de la série ajustée, ainsi que la tendance générale.

Une discussion des résultats est donnée dans PALM [2009]. On notera cependant que de petites différences existent entre les résultats donnés par la macro SCDECOMP et les résultats présentés dans PALM [2009]. Les divergences résultent du mode de calcul des coefficients saisonniers : la commande DECOMP, utilisée dans la macro SCDECOMP, prend comme coefficient saisonnier la médiane des rapports entre les observations et la moyenne mobile, alors que dans l'exemple traité dans PALM [2009], c'est la moyenne des rapports qui a été utilisée.

COMMANDES

```
name c10 "psaison" c11 "ptend" c12 "py"
name c13 "ysaison" c14 "ytend" c15 "yest"
```

```
%SCDECOMP c1 12;
multiplicatif;
quadratique;
outy      c13 c14 c15;
outp      c10 c11 c12;
Valid c6;
nograph.
```

RESULTATS

Analyse de la composante saisonnière
(commande Decomp avec option Notrend)

Multiplicative Model

```
Data      LAIT
Length    120
NMissing  0
```

Seasonal Indices

Period	Index
1	0.97984
2	0.93764
3	1.03304
4	1.09092
5	1.17455
6	1.08872
7	1.06516
8	0.95920
9	0.90416
10	0.92220
11	0.89380
12	0.95078

Accuracy Measures

```
MAPE      5.271
MAD       13.381
MSD      240.480
```

Forecasts

Period	Forecast
121	247.259
122	236.610
123	260.685
124	275.289
125	296.395
126	274.734
127	268.790
128	242.052
129	228.163
130	232.714
131	225.548
132	239.926

Figure 1. Commandes Minitab et résultats de l'analyse de la série chronologique des quantités de lait collectées (modèle mixte avec tendance quadratique).

Analyse de la tendance (commande Trend)

Data LAIT
 Length 120
 NMissing 0

Fitted Trend Equation

$$Y_t = 244.32 + 0.751*t - 0.00769*t**2$$

Accuracy Measures

MAPE 7.889
 MAD 20.068
 MSD 612.283

Forecasts

Period	Forecast
121	222.586
122	221.468
123	220.335
124	219.186
125	218.022
126	216.842
127	215.647
128	214.437
129	213.211
130	211.970
131	210.713
132	209.441

Analyse des résidus de la décomposition (Modèle mixte)

Critères d'adéquation du modèle

MAPE 3.86845
 MAD 9.67682
 MSD 141.056
 RMSD 11.8767
 Rsq 80.5484

Signes +,signes -,nombre de suites, uobs, Prob

NP 61.0000
 NM 59.0000
 P 26.0000
 UOBSP 6.41689
 PROBP 0.000000000

Quotient de von Neumann, uobs, Prob

Q 0.276380
 UOBSQ 7.99348
 PROBQ 0.000000000

Données de validation

VMAPE 9.16947
 VMAD 21.4700
 VMSD 590.810
 VRMSD 24.3066

Figure 1. (Suite) Commandes Minitab et résultats de l'analyse de la série chronologique des quantités de lait collectées (modèle mixte avec tendance quadratique).

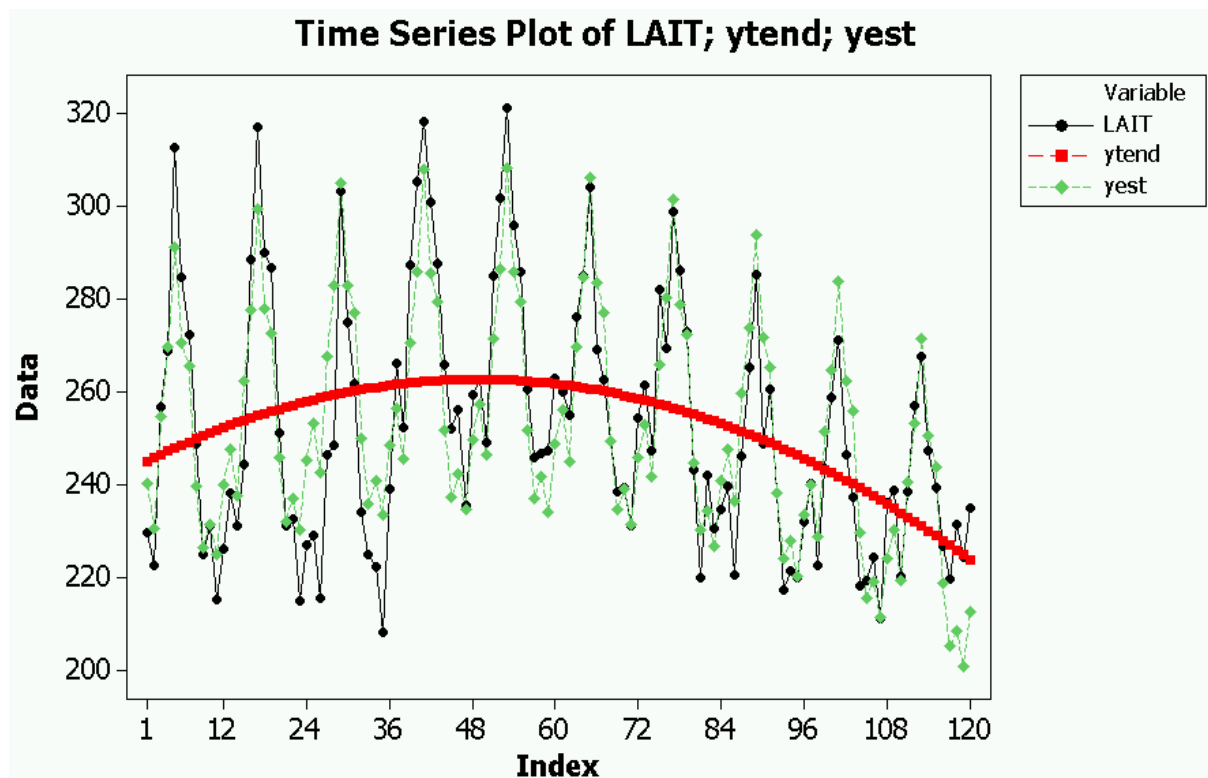


Figure 2. Etude de la série par décomposition : série initiale, série ajustée et tendance (modèle mixte avec tendance quadratique).

7. Conclusion

La macro proposée permet de simplifier la tâche de l'utilisateur souhaitant décomposer une série chronologique. Ce dernier peut dès lors assez facilement comparer les résultats provenant de différents modèles afin de retenir celui qu'il juge le plus adéquat.

Si la macro offre à l'utilisateur des choix quant aux modèles de décomposition retenus, elle impose par contre le mode d'estimation des composantes. Ainsi le mode de calcul de la composante saisonnière est imposé par la commande DECOMP (détermination de la médiane des écarts ou de rapports à la tendance plutôt que de la moyenne). De même, nous avons choisi de calculer la tendance à partir de la série initiale et non de la série différenciée.

Les conséquences de l'utilisation de ces méthodes d'estimation plutôt que d'autres solutions alternatives sont cependant, pour la plupart des séries, sans importance pratique.

8. Bibliographie

PALM R. [2009]. Etude des séries chronologiques par les méthodes de décomposition. Notes Stat. Inform. (Gembloux) 2009/ (à paraître).