

MACROS MINITAB POUR L'ANALYSE DES SERIES CHRONOLOGIQUES PAR LISSAGE

R. PALM¹

1. Introduction

Les méthodes de prévision par lissage sont des méthodes empiriques d'extrapolation qui donnent aux observations un poids décroissant avec l'ancienneté de l'observation.

Il existe différentes méthodes de lissage, chacune étant adaptée à un type particulier de série chronologique, et nous examinerons successivement le lissage exponentiel simple pour les séries sans tendance et sans composante saisonnière (paragraphe 2), le lissage de HOLT pour les séries avec tendance mais sans composante saisonnière (paragraphe 3) et, enfin, le lissage de HOLT et WINTERS pour les séries avec tendance et composante saisonnière (paragraphe 4).

Pour chacun de ces lissages, nous donnons quelques caractéristiques de la commande Minitab correspondante et nous présentons deux macros. La première analyse l'influence des paramètres de lissage sur les erreurs de prédiction à l'horizon 1 calculées pour le passé et la deuxième réalise la mise à jour du modèle de prévision au fur et à mesure de l'acquisition de données nouvelles. Ces macros sont présentées assez brièvement et leur utilisation est illustrée par un exemple. Des informations plus détaillées sur ces macros sont données dans les notices d'utilisation.

Les macros, les notices d'utilisation et les fichiers de données d'essai sont disponibles sur le site web de l'Unité de Statistique, Informatique et Mathématique appliquées de la FUSAG² :

www.fsagx.ac.be/si/

en cliquant sur le lien **MACROS**, puis sur le thème en question.

¹ Professeur à la Faculté universitaire des Sciences agronomiques de Gembloux

² Faculté universitaire des Sciences agronomiques de Gembloux

Nous supposons que le lecteur maîtrise les principes de base des différents lissages abordés. Si cela devait ne pas être le cas, nous lui suggérons la lecture de la note consacrée à ce sujet [PALM, 2007], qui est également disponible sur le site de l'Unité.

2. Lissage exponentiel simple

2.1. La commande SES de Minitab

Le lissage exponentiel simple se réalise avec la commande SES. La valeur par défaut du paramètre de lissage α est égale à 0,2, mais l'utilisateur peut imposer n'importe quelle valeur comprise entre 0 et 2. Il peut également demander l'optimisation de ce paramètre, par l'ajustement du modèle ARIMA (0, 1, 1).

Par défaut, Minitab utilise la moyenne des six premières observations pour la valeur lissée initiale $a(0)$. L'utilisateur a cependant la possibilité de fixer le nombre d'observations prises en compte pour la détermination de cette valeur initiale. Toutefois, si le paramètre de lissage est optimisé par l'ajustement du modèle ARIMA (0, 1, 1), la valeur initiale est déterminée à partir de ce modèle ARIMA, ce qui signifie que toutes les observations interviennent dans le calcul de la valeur initiale.

Enfin, la commande SES permet de réaliser des prévisions, d'enregistrer différentes caractéristiques dans le fichier et d'établir diverses représentations graphiques.

Trois paramètres permettant de quantifier l'adéquation du lissage sont déterminés :

- l'écart relatif moyen absolu³ :

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

- l'écart moyen absolu⁴ :

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

- le carré moyen des résidus⁵ :

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

3 En anglais : *Mean absolute percentage error*

4 En anglais : *Mean absolute deviation*

5 En anglais : *Mean square deviation*

Dans ces relations, n est le nombre d'observations de la série et e_t est le résidu, c'est-à-dire l'erreur de prévision à l'horizon 1 :

$$e_t = y_t - \hat{y}_{t-1}(1).$$

2.2. La macro LES

Cette macro calcule les paramètres MAPE, MAD et MSD définis ci-dessus pour différentes valeurs du paramètre de lissage α , ainsi que la moyenne quadratique des résidus RMSD, qui est la racine carrée de MSD. L'utilisateur fixe la valeur minimum, la valeur maximum et le pas de variation du paramètre α . Différentes options permettent d'imprimer et d'enregistrer les résultats.

A titre d'illustration, nous reprenons les données relatives aux rendements en blé observés en Belgique de 1980 à 2005, déjà utilisées dans PALM (2007). Cette série a été divisée en deux parties. Les années 1980 à 2000 sont utilisées pour la détermination du modèle et les années 2001 à 2005 sont utilisées pour la validation. La figure 1 donne l'évolution du carré moyen des erreurs de prévision en fonction de α . La commande qui a permis d'obtenir ce graphique est donnée en annexe. La macro signale également que la valeur de α qui minimise le carré moyen est égale à 0,51 et que, pour cette valeur, la moyenne quadratique des erreurs de prévision vaut 7,1.

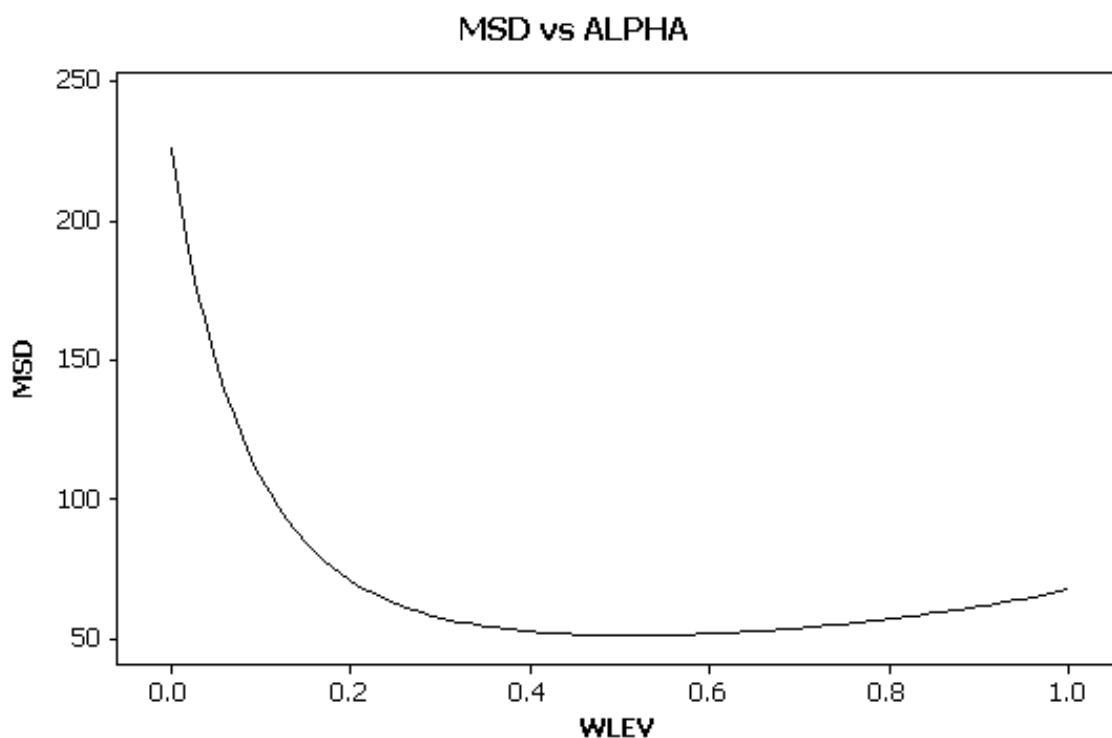


Figure 1. Rendements en blé de 1980 à 2005 : évolution du carré moyen des erreurs de prévision en fonction de α .

2.3. La macro LESMAJ

Cette macro permet la mise à jour du niveau moyen de la série, des prévisions à l'horizon 1 et des résidus, au fur et à mesure de l'acquisition de nouvelles données. Cette macro s'utilise normalement après l'exécution de la commande SES de Minitab.

Ainsi, pour les rendements de blé, supposons qu'on souhaite optimiser la constante de lissage par l'ajustement d'un modèle ARIMA (0, 1, 1) sur les données antérieures à 2001 et qu'on souhaite ensuite réaliser des prévisions à l'horizon 1, sans recalculer la constante de lissage, mais en tenant compte des données collectées au fur et à mesure que les années passent.

On a défini, dans le fichier de données, une colonne contenant la série des rendements de 1980 à 2000. On utilise, sur cette variable, la commande SES, avec l'option de calcul de α par le modèle ARIMA et on demande l'enregistrement des valeurs du niveau moyen (SMO01), des valeurs ajustées (FITS1) et des résidus (RESI1). On applique ensuite la macro LESMAJ à la série des rendements observés de 1980 à 2005 et on complète les colonnes SMO01, FITS1 et RESI1, obtenues par la commande SES, la constante de lissage étant celle donnée par la commande SES ($\alpha = 0,553515$).

La figure 2 reprend les résultats obtenus et les commandes utilisées sont données en annexe. Pour les cinq années, la moyenne quadratique des résidus est égale à 3,8. Cette valeur, non reprise dans la figure 2, est fournie par la macro.

On notera que les résultats sont différents de ceux qu'on aurait obtenus en utilisant la commande SES sur les données concernant les années allant de 1980 à 2005. Dans ce dernier cas en effet, la constante de lissage aurait été optimisée sur l'ensemble des données au lieu d'être optimisée sur les données antérieures à 2001. La valeur de α ne serait donc pas identique. De même, la valeur initiale $a(0)$ aurait également été différente, puisqu'elle aurait été déterminée à partir de 26 observations et non de 21 observations.

Par contre, lorsque le lissage se fait avec la valeur fixée par défaut ($\alpha = 0,2$) ou avec une valeur fixée par l'utilisateur, les mises à jour réalisées par la macro LESMAJ sont identiques aux valeurs lissées lorsqu'on exécute la commande SES sur la série des rendements allant de 1980 à 2005. Dans ce cas, le seul intérêt de la macro est le calcul du carré moyen des erreurs de prévision pour les années 2001 à 2005.

3. Lissage exponentiel double

3.1. La commande DES de Minitab

Le lissage double se réalise avec la commande DES. Les valeurs des paramètres de lissage α_1 et α_2 , respectivement pour le niveau moyen et la tendance linéaire, sont fixées par défaut à 0,2, mais peuvent être modifiées par l'utilisateur. Ces valeurs peuvent encore être optimisées par l'ajustement du modèle ARIMA (0, 2, 2).

Row	A_80-05	BLE_80-05	SMO01	FITS1	RESI1
1	1980	46.870	48.5234	50.5730	-3.7030
2	1981	52.138	50.5241	48.5234	3.6146
3	1982	58.704	55.0518	50.5241	8.1799
4	1983	52.953	53.8901	55.0518	-2.0988
5	1984	69.544	62.5547	53.8901	15.6539
6	1985	63.130	62.8731	62.5547	0.5753
7	1986	68.344	65.9013	62.8731	5.4709
8	1987	55.839	60.3317	65.9013	-10.0623
9	1988	60.065	60.1841	60.3317	-0.2667
10	1989	68.473	64.7721	60.1841	8.2889
11	1990	61.130	62.7562	64.7721	-3.6421
12	1991	67.567	65.4190	62.7562	4.8108
13	1992	65.412	65.4151	65.4190	-0.0070
14	1993	71.861	68.9830	65.4151	6.4459
15	1994	70.232	69.6743	68.9830	1.2490
16	1995	72.832	71.4221	69.6743	3.1577
17	1996	90.164	81.7960	71.4221	18.7419
18	1997	79.311	80.4205	81.7960	-2.4850
19	1998	80.357	80.3854	80.4205	-0.0635
20	1999	84.449	82.6346	80.3854	4.0636
21	2000	79.178	80.7213	82.6346	-3.4566
22	2001	80.500	80.5988	80.7213	-0.2213
23	2002	82.749	81.7890	80.5988	2.1502
24	2003	84.934	83.5298	81.7890	3.1450
25	2004	89.802	87.0016	83.5298	6.2722
26	2005	82.733	84.6388	87.0016	-4.2686

Figure 2. Rendements en blé : lissage exponentiel simple avec calcul de α sur les années 1980 à 2000 et mise à jour des résultats pour les années 2001 à 2005.

Pour déterminer la valeur initiale du niveau moyen $a(0)$ et de la tendance $b(0)$, Minitab ajuste une droite de régression à l'ensemble des données. L'ordonnée à l'origine correspond alors à $a(0)$ et le coefficient de régression correspond à $b(0)$. Toutefois, si les paramètres de lissage sont optimisés par l'ajustement d'un modèle ARIMA (0, 2, 2), les valeurs initiales $a(0)$ et $b(0)$ sont définies à partir de ce modèle. Dans les deux cas, les valeurs initiales sont donc déterminées en utilisant l'ensemble des observations disponibles.

Comme pour le lissage simple, différentes options permettent de réaliser des prédictions, d'enregistrer les résultats dans le fichier et d'établir des représentations graphiques. De même, les paramètres MAPE, MAD et MSD, définis au paragraphe 2.1, sont calculés.

3.2. La macro LHOLT

La macro LHOLT calcule les valeurs des paramètres MAPE, MAD, MSD et RMSD, pour un ensemble de couples α_1 et α_2 . L'utilisateur peut fixer les domaines et le pas de variation de ces deux constantes de lissage et différentes options permettent d'imprimer et d'enregistrer les résultats.

La figure 3 illustre l'utilisation de cette macro pour les données relatives aux rendements en blé observés de 1980 à 2000. La commande est donnée en annexe. Pour les couples α_1 et α_2 testés (α_1 et α_2 variant de 0 à 1, par pas de 0,1 pour l'option par défaut), la moyenne quadratique des erreurs de prévision minimum est égale à 5,4 et correspond à :

$$\alpha_1 = 0 \text{ et } \alpha_2 = 0.$$

MSD vs ALPHA1 (x-axis) and ALPHA2 (y-axis)

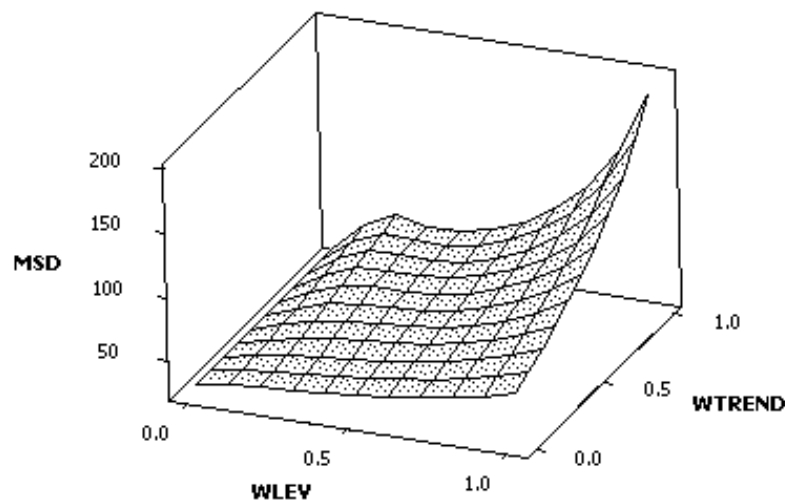


Figure 3. Rendements en blé de 1980 à 2005 : évolution du carré moyen des erreurs de prévision en fonction de α_1 et α_2 .

En réalité, les paramètres de lissage utilisés ne sont pas strictement égaux à zéro, car Minitab n'accepte pas de valeurs nulles, mais ont été fixés à 0,00001. Ces paramètres de lissage indiquent qu'il n'y a pas lieu de mettre à jour les valeurs initiales $a(0)$ et $b(0)$.

3.3. La macro LHOLTMAJ

Cette macro réalise la mise à jour du niveau moyen de la série, de la tendance, des prévisions à l'horizon 1 et des résidus, au fur et à mesure de l'acquisition de nouvelles données. Elle s'utilise normalement après l'exécution de la commande DES de Minitab.

La figure 4 donne les résultats obtenus pour la série chronologique des rendements en blé, en faisant les mises à jour pour 2001 à 2005. Les commandes utilisées sont données en annexe. Pour cette illustration, les valeurs par défaut ont été utilisées ($\alpha_1 = 0,2$ et $\alpha_2 = 0,2$).

Pour les cinq années, la moyenne quadratique des erreurs de prévision, non reprise dans la figure 4, est égale à 4,1.

Row	A_80-05	BLE_80-05	LEVEL	TREN1	FITS1	RESI1
1	1980	46.870	50.8505	1.37142	51.8457	-4.9757
2	1981	52.138	52.2052	1.36806	52.2220	-0.0840
3	1982	58.704	54.5994	1.57329	53.5732	5.1308
4	1983	52.953	55.5287	1.44451	56.1727	-3.2197
5	1984	69.544	59.4874	1.94734	56.9732	12.5708
6	1985	63.130	61.7738	2.01515	61.4347	1.6953
7	1986	68.344	64.6999	2.19735	63.7889	4.5551
8	1987	55.839	64.6856	1.75502	66.8973	-11.0583
9	1988	60.065	65.1655	1.49999	66.4407	-6.3757
10	1989	68.473	67.0270	1.57229	66.6655	1.8075
11	1990	61.130	67.1054	1.27352	68.5993	-7.4693
12	1991	67.567	68.2166	1.24104	68.3790	-0.8120
13	1992	65.412	68.6485	1.07922	69.4576	-4.0456
14	1993	71.861	70.1544	1.16455	69.7277	2.1333
15	1994	70.232	71.1015	1.12107	71.3189	-1.0869
16	1995	72.832	72.3445	1.14545	72.2226	0.6094
17	1996	90.164	76.8247	1.81241	73.4899	16.6741
18	1997	79.311	78.7719	1.83936	78.6372	0.6738
19	1998	80.357	80.5604	1.82919	80.6113	-0.2543
20	1999	84.449	82.8015	1.91157	82.3896	2.0594
21	2000	79.178	83.6061	1.69017	84.7131	-5.5351
22	2001	80.500	84.3370	1.49832	85.2962	-4.7962
23	2002	82.749	85.2180	1.37486	85.8353	-3.0863
24	2003	84.934	86.2611	1.30851	86.5929	-1.6589
25	2004	89.802	88.0161	1.39780	87.5696	2.2324
26	2005	82.733	88.0777	1.13057	89.4139	-6.6809

Figure 4. Rendements en blé : lissage double avec calcul des valeurs initiales sur les années 1980 à 2000 et mise à jour des résultats pour les années 2001 à 2005.

4. Lissage de HOLT et WINTERS

4.1. La commande TSWIN de Minitab

Pour les séries avec tendance et saisonnalité, la commande TSWIN permet d'ajuster soit un modèle multiplicatif, soit un modèle additif. Les paramètres de lissage du niveau moyen α_1 , de la tendance α_2 et de la saisonnalité α_3 sont fixés, par défaut à 0,2, mais l'utilisateur peut modifier ces valeurs. Contrairement au lissage exponentiel simple et au lissage de HOLT, aucune option ne permet d'optimiser ces paramètres de lissage.

Les valeurs initiales du niveau moyen $a(0)$ et de la tendance $b(0)$ sont obtenues par l'ajustement d'une droite de régression exprimant les observations en fonction du temps, comme pour le modèle de HOLT. Les valeurs initiales pour les coefficients saisonniers sont obtenues par régression sur des variables artificielles, en utilisant comme variable à expliquer la série après élimination de la tendance. Les valeurs initiales sont donc déterminées sur la base de l'ensemble des observations dans la série.

Comme pour les lissages précédents, des options permettent de réaliser des prévisions, d'enregistrer les résultats dans le fichier et d'établir des représentations graphiques. Les paramètres MAPE, MAD et MSD sont également calculés.

4.2. La macro LHW

La macro LHW calcule les valeurs des paramètres MAPE, MAD, MSD et RMSD pour un ensemble de triplets α_1 , α_2 et α_3 . L'utilisateur choisit le modèle additif ou le modèle multiplicatif et définit le domaine et le pas de variation des constantes de lissage. Par défaut, les trois constantes varient de 0 à 1, par pas de 0,1, ce qui correspond à 1331 triplets.

Un graphique reprend les triplets α_1 , α_2 et α_3 pour lesquels la valeur de MSD est telle que :

$$\text{MSD} < k_1 * \text{MSD}_{\min},$$

k_1 étant une constante, fixée par l'utilisateur (par défaut, $k_1 = 1,1$) et MSD_{\min} étant la valeur minimum de MSD pour l'ensemble des triplets testés.

Un autre graphique donne l'évolution de MSD en fonction des paramètres de lissage pour les triplets tels que :

$$\text{MSD} < k_2 * \text{MSD}_{\min},$$

k_2 étant une constante fixée par l'utilisateur (par défaut, $k_2 = 1,2$) et MSD_{\min} étant la valeur minimum de MSD pour l'ensemble des triplets testés.

Pour illustrer l'utilisation de cette macro, nous reprenons les quantités de lait de vache collectées en Belgique, de janvier 1995 à décembre 2005 (en milliers de tonnes), analysées précédemment (PALM, 2007). L'ajustement du modèle se fait sur les données de 1985 à 2004 et l'année 2005 est utilisée pour la validation. A partir des commandes reprises en annexe, on a trouvé que le meilleur triplet, pour le modèle additif, est le suivant :

$$\alpha_1 = 0,8, \alpha_2 = 0,1 \text{ et } \alpha_3 = 0,$$

la moyenne quadratique des erreurs de prévision étant, pour ce triplet, égale à 9,2. La figure 5 donne l'évolution de MSD en fonction des paramètres de lissage pour tous les triplets pour lesquels la valeur de MSD est inférieure à :

$$(1,2) (9,2)^2 = 101,6.$$

4.3. La macro LHWMAJ

La macro LHWMAJ réalise la mise à jour du niveau moyen de la série, de la tendance, de la saisonnalité, des prévisions à l'horizon 1 et des résidus, au fur et à mesure de l'acquisition des nouvelles données. Elle s'utilise typiquement après l'exécution de la commande TSWIN de Minitab.

La figure 6 donne les résultats (partiels) obtenus pour la série des quantités de lait collectées, en faisant la mise à jour pour l'année 2005 et en utilisant le modèle additif avec les paramètres de lissage suivants :

$$\alpha_1 = 0,8, \alpha_2 = 0,1 \text{ et } \alpha_3 = 0.$$

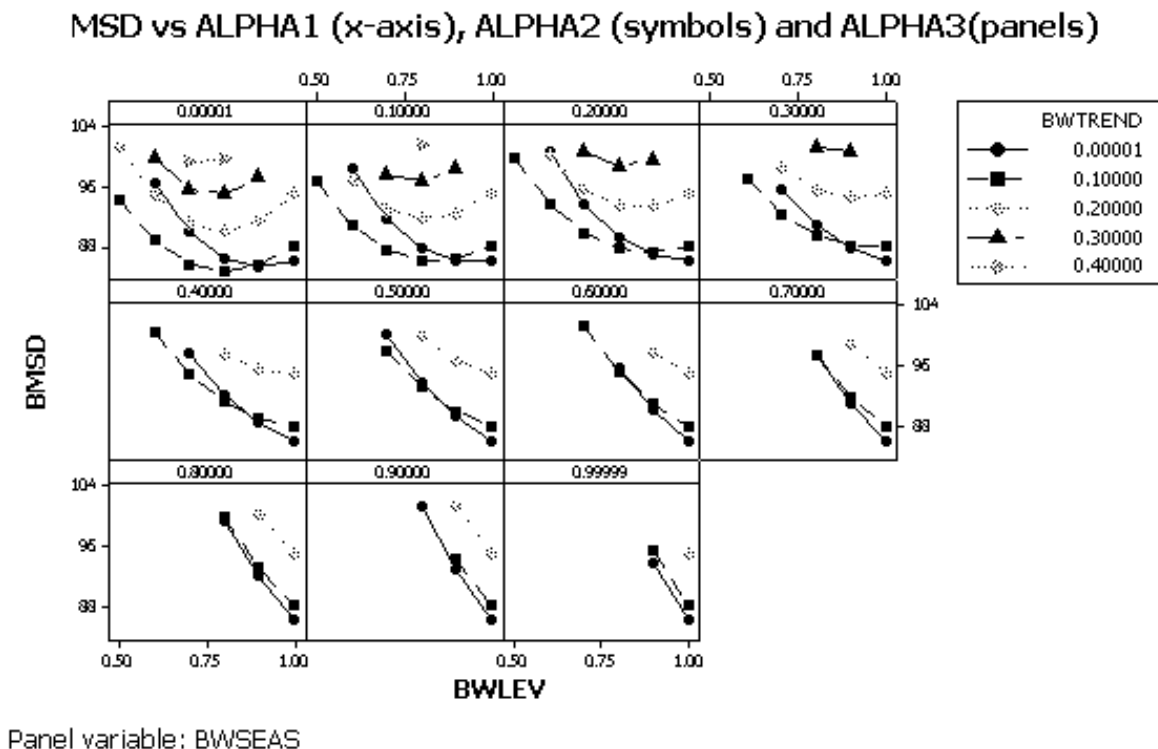


Figure 5. Quantités de lait collectées : évolution de MSD en fonction des paramètres de lissage ($MSD < 1,2 * MSD_{min}$), pour le modèle additif.

Pour les douze mois de l'année 2005, la moyenne quadratique des erreurs, donnée par la macro mais non reprise dans la figure 6, est de 11,4.

5. Conclusions

Les différentes macros proposées permettent d'étendre les fonctionnalités offertes par les commandes SES, DES et TSWIN de Minitab pour l'étude des séries chronologiques par lissage.

Elles sont d'une utilisation simple tout en étant assez souples, grâce aux différentes options proposées.

Ces macros peuvent être utilisées notamment pour la validation des méthodes de lissage sur des données qui n'ont pas été utilisées pour l'établissement des modèles. Elles ont également un rôle didactique, dans la mesure où elles permettent de bien apprécier l'incidence sur les résultats des paramètres de lissage utilisés.

Row	LAIT_95-05	M	T	LEVE1	TREN1	SEAS1	FITS1	RESI1
109	238.669	1	109	245.602	0.68962	-6.7844	239.413	-0.7445
110	220.316	2	110	241.186	0.17907	-19.5940	226.698	-6.3819
111	238.469	3	111	232.971	-0.66038	7.5966	248.962	-10.4930
112	256.927	4	112	234.435	-0.44789	21.9603	254.271	2.6561
113	267.516	5	113	223.005	-1.54611	47.2562	281.244	-13.7278
114	247.305	6	114	224.589	-1.23318	21.9341	243.393	3.9117
115	239.265	7	115	224.580	-1.11068	14.3783	237.734	1.5312
116	226.693	8	116	232.872	-0.17046	-8.5295	214.940	11.7527
117	219.628	9	117	240.345	0.59385	-22.6274	210.074	9.5539
118	231.387	10	118	246.908	1.19079	-17.0132	223.925	7.4617
119	224.209	11	119	251.216	1.50251	-27.7861	220.313	3.8965
120	234.879	12	120	247.080	0.93864	-10.7909	241.927	-7.0484
121	240.664	1	121	247.562	0.89305	-6.7844	241.234	-0.5699
122	219.785	2	122	241.194	0.16693	-19.5940	228.861	-9.0764
123	234.079	3	123	229.458	-1.02337	7.5966	248.958	-14.8788
124	256.598	4	124	233.397	-0.52714	21.9603	250.395	6.2029
125	261.017	5	125	217.583	-2.05587	47.2561	280.126	-19.1092
126	245.439	6	126	221.909	-1.41762	21.9341	237.461	7.9782
127	246.614	7	127	229.887	-0.47810	14.3784	234.870	11.7440
128	239.995	8	128	244.701	1.05116	-8.5294	220.879	19.1157
129	228.020	9	129	249.668	1.44275	-22.6274	223.125	4.8949
130	231.634	10	130	249.140	1.24563	-17.0132	234.098	-2.4640
131	223.544	11	131	251.141	1.32119	-27.7861	222.600	0.9445
132	226.000	12	132	239.925	0.06747	-10.7910	241.671	-15.6715

Figure 6. Quantités de lait collectées : modèle additif de HOLT et WINTERS ($\alpha_1 = 0,8$, $\alpha_2 = 0,1$ et $\alpha_3 = 0$), avec calcul des valeurs initiales sur les années 1995 à 2004 et mise à jour des résultats pour l'année 2005 (résultats partiels).

6. Bibliographie

PALM R. (2007). Etude des séries chronologiques par les méthodes de lissage. *Notes Stat. Inform.* (Gembloux) 2007/1, 22p.

Annexe : liste des commandes Minitab.

Figure 1

%LES c4

Figure 2

Name c5 "SMO01" c6 "FITS1" c7 "RESI1"

SES C4;

Smoothed 'SMO01';

Fits 'FITS1';

Residuals 'RESI1'.

%LESMAJ c2 c5 c6 c7;

WEIGHTS .553515;

NOPRINT.

PRINT c1 c2 c5 c6 c7

Figure 3

%LHOLT c4

Figure 4

Name c5 "SMO01" c6 "LEVE1" c7 "TREN1" c8 "FITS1" c9 "RESI1"

DES C4;

Weight 0.2 0.2;

Smoothed 'SMO01';

Level 'LEVE1';

Trend 'TREN1';

Fits 'FITS1';

Residuals 'RESI1'.

%LHOLTMAJ c2 c6 c7 c8 c9;

WEIGHTS .2 .2;

ORIGIN 21;

NOPRINT.

PRINT c1-c2 c6-c9

Annexe (suite) : liste des commandes.

Figure 5

```
%LHW c6;  
  SLENGTH 12;  
  MODEL 0;  
  NOTRIPLET.
```

Figure 6

```
Name c7 "SMO01" c8 "LEVE1" c9 "TREN1" c10 "SEAS1" c11 "FITS1" c12 "RESI1"  
TSWint 'LAIT_95-04' 12;  
  Additive;  
  Weight 0.8 0.1 0.0;  
  Smoothed 'SMO01';  
  Level 'LEVE1';  
  Trend 'TREN1';  
  Seasonal 'SEAS1';  
  Fits 'FITS1';  
  Residuals 'RESI1';  
  First 1.
```

```
%LHWMAJ c1 c8 c9 c10 c11 c12;  
  MODEL 0;  
  WEIGHTS .8 .1 .0;  
  SLENGTH 12;  
  ORIGIN 120;  
  NOPRINT.  
PRINT c1 c4 c5 c8-c12
```