

NOTES DE STATISTIQUE ET D'INFORMATIQUE

2007/1

ETUDE DES SÉRIES CHRONOLOGIQUES
PAR LES MÉTHODES DE LISSAGE

R. PALM

Faculté universitaire des Sciences agronomiques
Unité de Statistique, Informatique et Mathématique appliquées
GEMBLoux
(Belgique)

ETUDE DES SÉRIES CHRONOLOGIQUES PAR LES MÉTHODES DE LISSAGE

R. PALM*

RÉSUMÉ

Cette note décrit brièvement les méthodes de prévision par lissage lors de l'analyse des séries chronologiques.

Y sont successivement abordés, le lissage exponentiel simple, pour les séries sans tendance et sans composante saisonnière, le lissage de HOLT, pour les séries avec tendance et sans composante saisonnière et le lissage de HOLT et WINTERS, pour les séries avec tendance et composante linéaire.

Les méthodes sont chaque fois illustrées par un exemple traité avec le logiciel Minitab.

SUMMARY

This note briefly describes forecasting procedures for time series based on smoothing techniques.

The following procedures are examined: exponential smoothing for non-seasonal time series showing no trend, HOLT procedure for non-seasonal time series with trend and HOLT-WINTERS procedure for time series containing trend and seasonal variation.

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

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. Elles sont destinées à la prévision à court terme, c'est-à-dire pour quelques intervalles de temps au-delà des dernières observations.

*Professeur à la Faculté universitaire des Sciences agronomiques de Gembloux (Unité de Statistique, Informatique et Mathématique appliquées).

Les méthodes de lissage ne nécessitent qu'un volume relativement faible de calcul et sont caractérisées par une grande simplicité de la mise à jour des prévisions au fur et à mesure de la disponibilité de nouvelles observations.

Il existe différentes méthodes de lissage, chacune étant adaptée à un type particulier de série chronologique et nous étudierons successivement les lissages adaptés aux séries sans tendance et sans composante saisonnière (paragraphe 2), aux séries avec tendance mais sans composante saisonnière (paragraphe 3) et, enfin, aux séries avec tendance et composante saisonnière (paragraphe 4). Nous tirerons alors quelques conclusions (paragraphe 5).

Nous nous limitons à une présentation assez sommaire des principales méthodes de lissage. Le lecteur souhaitant approfondir le sujet trouvera des informations complémentaires dans de nombreux ouvrages consacrés à l'étude des séries chronologiques, parmi lesquels nous pouvons citer CHATFIELD [2003], GOURIEROUX et MONFORT [1990] et MELARD [1990], ainsi que dans l'article de synthèse de GARDNER [1985]. Nous ne considérons que des séries chronologiques univariées discrètes, dont les observations sont prises à des intervalles de temps égaux ou approximativement égaux, comme, par exemple, des données annuelles ou mensuelles.

2. SÉRIES SANS TENDANCE ET SANS COMPOSANTE SAISONNIÈRE

Si la série à laquelle on s'intéresse ne présente ni tendance ni saisonnalité marquées, on peut raisonnablement considérer que les prévisions faites au temps t pour les temps $t + 1, t + 2, \dots, t + h$ sont identiques. En désignant par $\hat{y}_t(h)$ la prévision réalisée au temps t pour l'horizon h , on a, par conséquent :

$$\hat{y}_t(1) = \hat{y}_t(2) = \dots = \hat{y}_t(h) = a(t).$$

Au temps t , les prévisions sont donc égales à la constante $a(t)$, quel que soit l'horizon de prévision. La valeur de cette constante est fonction de t et est remise à jour au fur et à mesure de l'acquisition de nouvelles données.

Pour déterminer cette valeur constante au temps t , on calcule une moyenne pondérée des observations passées, en accordant à celles-ci des poids de plus en plus petits au fur et à mesure que l'information vieillit. De façon plus précise, on a :

$$a(t) = \alpha y_t + \alpha(1 - \alpha) y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \alpha(1 - \alpha)^3 y_{t-3} + \dots,$$

α étant une constante, comprise entre 0 et 1.

La méthode porte le nom de *lissage exponentiel* car les poids :

$$w_j = \alpha(1 - \alpha)^j,$$

qui constituent une progression géométrique et qui sont fonction de la constante de lissage α , se situent sur une courbe exponentielle.

De façon stricte, pour que $a(t)$ soit réellement une moyenne pondérée des observations du passé, il faut que la somme des w_j soit égale à l'unité, ce qui n'est vrai que si le nombre d'observations du passé est égal à l'infini.

Aussi, l'équation est-elle généralement présentée par la formule de récurrence :

$$a(t) = \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + \alpha(1 - \alpha)^2 y_{t-3} + \dots],$$

c'est-à-dire également :

$$a(t) = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}(1).$$

Cette relation montre que, au temps t , la prévision peut être réalisée en conservant en mémoire uniquement la prévision réalisée au temps $t - 1$ pour l'horizon 1 et la valeur α .

La valeur $\hat{y}_{t-1}(1)$, qui est la valeur prédite au temps $t - 1$ pour le temps t , est aussi appelée valeur lissée au temps $t - 1$. Elle est égale à $a(t - 1)$ et est parfois représentée par le symbole y_{t-1}^* .

Pour débiter le processus de lissage par la formule de récurrence, il faut disposer d'une estimation initiale $y_0(1)$.

En pratique, on peut utiliser, par exemple, la première observation ou la moyennes des k premières observations :

$$\hat{y}_0(1) = y_1$$

ou

$$\hat{y}_0(1) = (y_1 + \dots + y_k)/k.$$

Le choix de cette estimation préliminaire n'a que relativement peu d'incidence sur les prévisions, du moins si t est suffisamment grand.

Le paramètre α permet de donner un poids plus ou moins grand aux données récentes suivant que la série dépend surtout de celles-ci ou que, au contraire, c'est son historique qui est prépondérant. Il en résulte que si α est grand, les réponses aux changements de la série seront rapides. Par contre, si α est petit, les prévisions seront caractérisées par une grande inertie.

La valeur du paramètre α peut être fixée arbitrairement ou bien être déterminée de manière à minimiser la somme des carrés des erreurs de prévision à l'horizon 1 :

$$\text{SCE}_\alpha = \sum_{t=k}^n [y_t - \hat{y}_{t-1}(1)]^2.$$

Dans cette relation, k est une constante entière positive, dont le rôle est d'exclure éventuellement de la somme des carrés, les erreurs de prévision pour le début de la série, car celles-ci sont fortement influencées par la valeur initiale $\hat{y}_0(1)$, qui est toujours déterminée de façon quelque peu arbitraire. Cette somme des carrés peut évidemment être remplacée par le carré moyen des erreurs de prévision, obtenu en divisant la somme des carrés par le nombre de termes qu'elle comporte.

La somme de carrés des écarts ci-dessus est calculée pour différentes valeurs de α comprises entre 0 et 1 et variant, par exemple, par pas de 0,1. On sélectionne ensuite la valeur de α qui minimise SCE_α .

Une autre solution pour l'optimisation de α repose sur l'équivalence qui existe entre le lissage exponentiel et le modèle ARIMA(0, 1, 1) défini dans le cadre des modèles de BOX et JENKINS, dont une présentation simplifiée est donnée par PALM [1987b]. Le modèle ARIMA(0, 1, 1) est caractérisé par un paramètre, généralement noté θ , qui est lié à la constante de lissage α par la relation :

$$\alpha = 1 - \theta.$$

L'ajustement du modèle ARIMA(0, 1, 1) à la série chronologique permet d'obtenir une estimation de θ et donc de α .

En pratique, on trouve souvent une valeur de α comprise entre 0,01 et 0,30. Si la minimisation de la somme des carrés des écarts des erreurs de prévision à l'horizon 1 conduit à une valeur de α proche de l'unité, il y a lieu de soupçonner l'existence d'une tendance ou d'une saisonnalité et, dans ce cas, il convient d'utiliser d'autres techniques de lissage (paragraphe 3 et 4).

Considérons, à titre d'illustration, la série chronologique des rendements en blé observés en Belgique de 1980 à 2005, exprimés en quintaux par hectare¹.

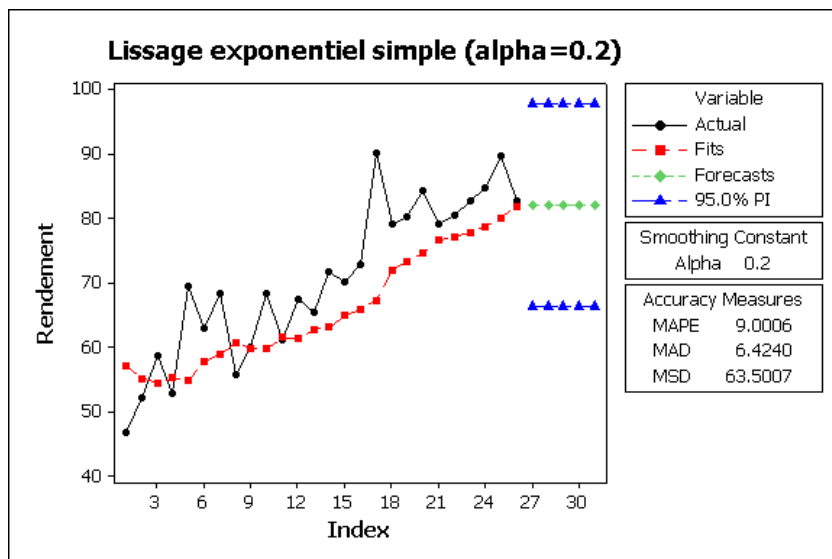


Figure 1. Rendements en blé : lissage exponentiel simple ($\alpha = 0,2$).

La figure 1 donne les résultats du lissage, ainsi que les prévisions pour les années 2006 à 2010, obtenues en fixant α à 0,2, qui est la valeur par défaut proposée par Minitab. On constate que ce lissage est inadéquat, les prévisions à l'horizon 1 étant, de manière assez systématique, inférieures aux observations.

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

Cette sous-estimation s'explique par le fait que le paramètre de lissage est assez faible alors que la série présente une tendance croissante marquée. En accordant un poids plus important au présent et un poids plus faible au passé, c'est-à-dire en augmentant α , on doit s'attendre à des écarts plus faibles entre les observations et les prévisions à l'horizon 1.

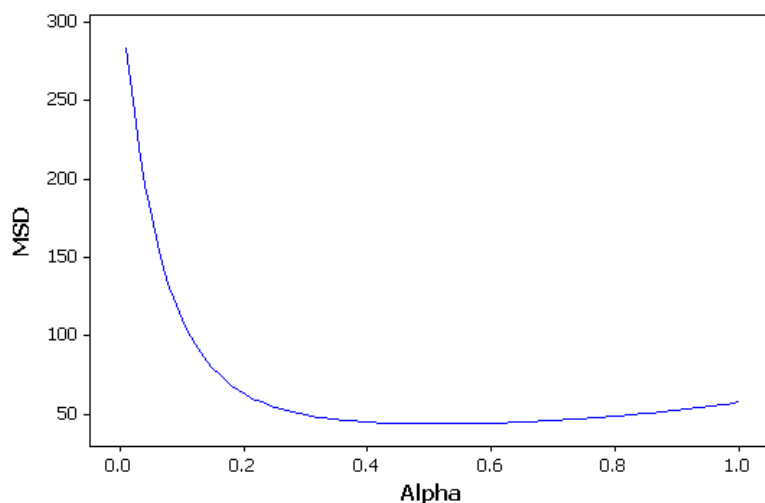


Figure 2. Rendements en blé: évolution du carré moyen des erreurs de prévision en fonction de α .

La figure 2 donne l'évolution du carré moyen des erreurs de prévision² à l'horizon 1 :

$$\text{MSD} = \frac{1}{n} \sum_{t=1}^n [y_t - \hat{y}_{t-1}(1)]^2,$$

en fonction de la valeur de α . Cette figure a été établie en faisant varier α de 0 à 1, par pas de 0,01. On peut constater que la valeur minimum se trouve entre 0,4 et 0,6 et que, au sein de cet intervalle la valeur du paramètre MSD varie très peu. De façon plus précise, le minimum se situe en $\alpha = 0,51$. Cette valeur est très proche de la valeur déduite de l'ajustement du modèle ARIMA(0, 1, 1) qui est égale à 0,55 (figure 3). Par rapport au premier modèle (figure 1), un poids plus faible est par conséquent attribué au passé lointain, ce qui permet de tenir compte, du moins partiellement, de l'augmentation des rendements avec le temps. Une meilleure prise en compte de cette tendance peut se faire, comme nous le verrons au paragraphe suivant, par l'utilisation d'un modèle plus adapté.

Pour ce dernier lissage, la figure 4 reprend, outre les données de départ, les valeurs lissées, les valeurs ajustées et les résidus. Ces résultats sont obtenus en prenant comme valeur initiale $a(0)$, une valeur résultant de l'ajustement du modèle ARIMA(0, 1, 1), qui est égale à 50,61.

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

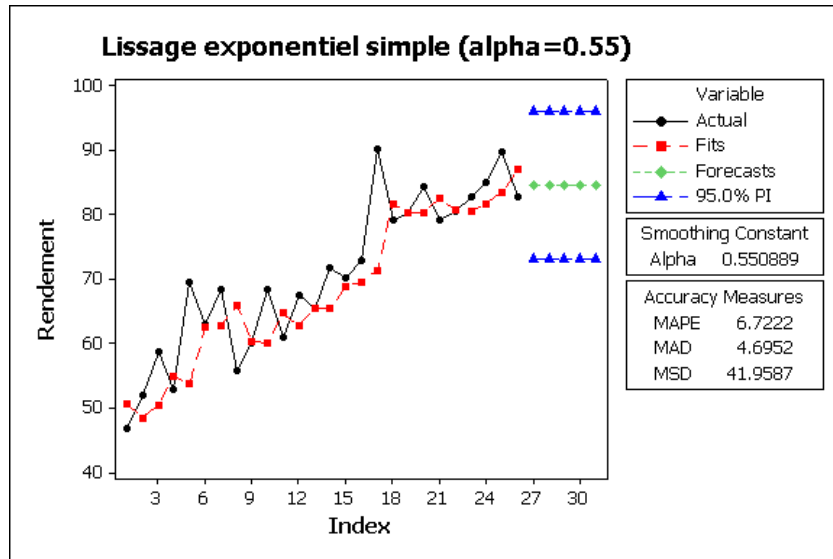


Figure 3. Rendements en blé : lissage exponentiel simple avec détermination de α par l'ajustement du modèle ARIMA(0, 1, 1).

On peut en déduire les valeurs lissées³ en $t = 1$, $t = 2$, etc. :

$$y_1^* = \alpha y_1 + (1 - \alpha)\hat{y}_0(1) = (0,55)(46,87) + (0,45)(50,61) = 48,55,$$

$$y_2^* = \alpha y_2 + (1 - \alpha)\hat{y}_1(1) = (0,55)(52,14) + (0,45)(48,55) = 50,53,$$

etc.

Il en résulte que les valeurs prédites⁴ aux temps $t = 0$, $t = 1$, ..., à l'horizon 1 sont :

$$\hat{y}_0(1) = y_0^* = 50,61,$$

$$\hat{y}_1(1) = y_1^* = 48,55,$$

$$\hat{y}_2(1) = y_2^* = 50,53,$$

etc.

Les valeurs prédites ou ajustées sont donc égales aux valeurs lissées calculées pour l'intervalle de temps précédent.

Les résidus⁵ sont donnés par la relation :

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

Ils correspondent donc aux erreurs de prévision à l'horizon 1.

3. En anglais : *Smoothed values*.

4. En anglais : *Fits*.

5. En anglais : *residuals*.

Row	ANNEE	ble_RT	SM001	FITS1	RESI1
1	1980	46.870	48.5477	50.6056	-3.7356
2	1981	52.138	50.5256	48.5477	3.5903
3	1982	58.704	55.0310	50.5256	8.1784
4	1983	52.953	53.8862	55.0310	-2.0780
5	1984	69.544	62.5119	53.8862	15.6578
6	1985	63.130	62.8524	62.5119	0.6181
7	1986	68.344	65.8777	62.8524	5.4916
8	1987	55.839	60.3475	65.8777	-10.0387
9	1988	60.065	60.1919	60.3475	-0.2825
10	1989	68.473	64.7539	60.1919	8.2811
11	1990	61.130	62.7575	64.7539	-3.6239
12	1991	67.567	65.4070	62.7575	4.8095
13	1992	65.412	65.4098	65.4070	0.0050
14	1993	71.861	68.9637	65.4098	6.4512
15	1994	70.232	69.6624	68.9637	1.2683
16	1995	72.832	71.4085	69.6624	3.1696
17	1996	90.164	81.7407	71.4085	18.7555
18	1997	79.311	80.4022	81.7407	-2.4297
19	1998	80.357	80.3773	80.4022	-0.0452
20	1999	84.449	82.6204	80.3773	4.0717
21	2000	79.178	80.7240	82.6204	-3.4424
22	2001	80.500	80.6006	80.7240	-0.2240
23	2002	82.749	81.7841	80.6006	2.1484
24	2003	84.934	83.5194	81.7841	3.1499
25	2004	89.802	86.9804	83.5194	6.2826
26	2005	82.733	84.6406	86.9804	-4.2474
Forecasts					
Period	Forecast	Lower	Upper		
27	84.6406	73.1374	96.1437		
28	84.6406	73.1374	96.1437		
29	84.6406	73.1374	96.1437		
30	84.6406	73.1374	96.1437		
31	84.6406	73.1374	96.1437		

Figure 4. Rendements en blé: valeurs lissées, valeurs prédites et résidus pour le lissage exponentiel simple ($\alpha = 0,55$).

Minitab propose trois paramètres permettant de quantifier l'adéquation du lissage :

– l'écart relatif moyen absolu⁶ :

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

– l'écart moyen absolu⁷ :

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

– le carré moyen des résidus, dont il a déjà été question auparavant :

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

La racine carrée de ce dernier paramètre est donc analogue à un écart-type résiduel. Pour le lissage avec $\alpha = 0,55$, cet écart-type est égal à 6,5 quintaux par hectare, MSD étant égal à 41,96 (figure 3).

La figure 4 montre que la valeur lissée correspondant à l'année 2005 ($t = 26$) est égale à :

$$y_{26}^* = 84,64.$$

Cette valeur correspond donc aux prévisions faites au temps $t = 26$, pour les horizons 1, 2, etc., c'est-à-dire aux prévisions faites en 2005 pour 2006, 2007,...

$$\hat{y}_{26}(1) = \hat{y}_{26}(2) = \dots = \hat{y}_{26}(5) = 84,64.$$

Les limites de confiance sont approximativement égales à :

$$\hat{y}_{26}(k) \pm u_{1-\alpha/2} \sqrt{\text{MSD}} = 84,64 \pm 1,96 \sqrt{41,96},$$

soit 71,96 et 97,32.

3. SÉRIES AVEC TENDANCE ET SANS COMPOSANTE SAISONNIÈRE

Dans les cas où la série présente une tendance et si, en outre, on peut considérer que celle-ci est localement linéaire, on peut adopter la formule de prévision suivante :

$$\hat{y}_t(h) = a(t) + b(t)h.$$

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

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

On considère qu'au temps t , le niveau moyen de la série est égal à $a(t)$ et qu'aux environs de t , la série présente une tendance linéaire dont le coefficient angulaire est égal à $b(t)$.

Les valeurs $a(t)$ et $b(t)$ sont constamment remises à jour, par les relations suivantes :

$$a(t) = \alpha_1 y_t + (1 - \alpha_1)[a(t-1) + b(t-1)]$$

et
$$b(t) = \alpha_2 [a(t) - a(t-1)] + (1 - \alpha_2)b(t-1).$$

La première relation montre que $a(t)$ peut être considéré comme le résultat d'un lissage exponentiel du niveau moyen de la série, car :

$$a(t-1) + b(t-1) = \hat{y}_{t-1}(1),$$

et la seconde relation indique que $b(t)$ est le résultat d'un lissage exponentiel du coefficient angulaire de la droite représentant la tendance. En effet, la quantité :

$$a(t) - a(t-1),$$

est la différence entre les niveaux moyens de la série aux temps t et $t-1$ tandis que $b(t-1)$ est la valeur estimée du coefficient angulaire au temps $t-1$.

Les $a(t)$ correspondent aussi aux valeurs lissées :

$$y_t^* = a(t).$$

Ce modèle de prévision est connu sous le nom de modèle de HOLT. Si les constantes de lissage α_1 et α_2 sont liées à un même paramètre α , par les relations suivantes :

$$\alpha_1 = \alpha(2 - \alpha) \quad \text{et} \quad \alpha_2 = \alpha/(2 - \alpha),$$

le modèle est dénommé modèle de BROWN ou encore lissage exponentiel double, car on peut montrer que, dans ce cas particulier, les paramètres $a(t)$ et $b(t)$ peuvent être mis à jour en utilisant deux lissages exponentiels simples successifs de la série.

En effet, si y_t^* est le résultat d'un premier lissage :

$$y_t^* = \alpha y_t + (1 - \alpha)y_{t-1}^*$$

et si y_t^{**} est le résultat du lissage de la série y_t^* :

$$y_t^{**} = \alpha y_t^* + (1 - \alpha)y_{t-1}^{**},$$

on peut montrer que :

$$a(t) = 2 y_t^* - y_t^{**} \quad \text{et} \quad b(t) = \frac{\alpha}{1 - \alpha}(y_t^* - y_t^{**}).$$

Comme pour le lissage exponentiel simple, il faut se donner des valeurs initiales $a(0)$ et $b(0)$, afin de pouvoir débiter le processus de prévision. Pour le premier paramètre, on peut utiliser, par exemple, la relation suivante :

$$a(0) = y_1 - b(0)$$

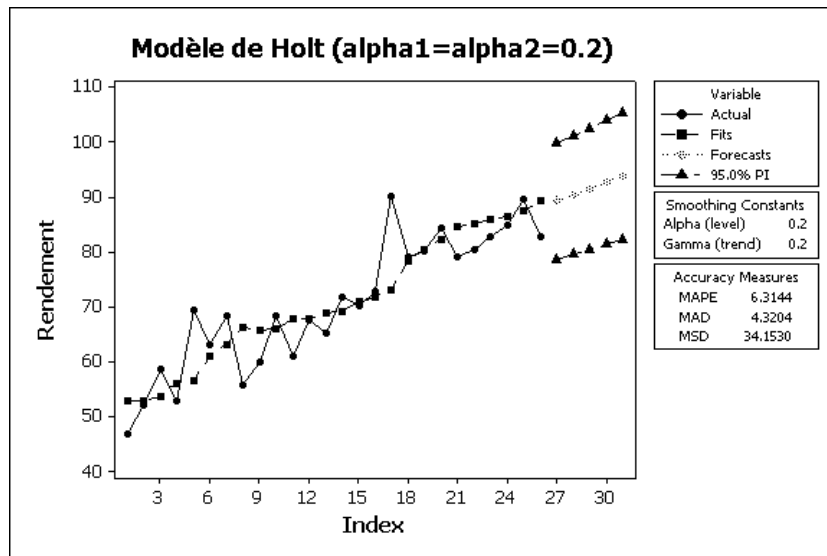


Figure 5. Rendements en blé: modèle de HOLT ($\alpha_1 = \alpha_2 = 0,2$).

et $b(0)$ peut, par exemple, être égal au coefficient de régression linéaire simple calculé sur les cinq ou six premières valeurs de la série. Une autre solution, utilisée dans le logiciel Minitab, consiste à ajuster une régression linéaire simple exprimant la caractéristique étudiée en fonction du temps et à prendre comme valeurs initiales $a(0)$ et $b(0)$, l'ordonnée à l'origine et la pente de la droite de régression.

Quant aux constantes de lissage, elles sont fixées *a priori* par l'utilisateur ou, au contraire, déterminées de manière à minimiser la somme des carrés des erreurs de prévision à l'horizon 1, comme dans le cas du lissage exponentiel simple. Elles peuvent également être estimées par l'ajustement d'un modèle ARIMA(0, 2, 2).

Pour les rendements en blé, les figures 5 et 6 donnent les résultats du lissage, en prenant pour α_1 et α_2 les valeurs proposées par défaut par Minitab :

$$\alpha_1 = \alpha_2 = 0,2.$$

Pour démarrer le processus de lissage, Minitab a calculé la régression linéaire simple du rendement y en fonction de t :

$$y = 51,431 + 1,4304 t.$$

L'ordonnée à l'origine de cette relation correspond au niveau moyen de la série au temps $t = 0$ et la pente de la droite est une estimation de la tendance au temps $t = 0$. On a donc :

$$a(0) = 51,43 \quad \text{et} \quad b(0) = 1,43.$$

Row	ANNEE	ble_RT	SM001	LEVE1	TREN1	FITS1	RESI1
1	1980	46.870	51.6632	51.6632	1.19071	52.8615	-5.9915
2	1981	52.138	52.7107	52.7107	1.16207	52.8539	-0.7159
3	1982	58.704	54.8391	54.8391	1.35532	53.8728	4.8312
4	1983	52.953	55.5461	55.5461	1.22567	56.1944	-3.2414
5	1984	69.544	59.3262	59.3262	1.73656	56.7718	12.7722
6	1985	63.130	61.4762	61.4762	1.81924	61.0628	2.0672
7	1986	68.344	64.3052	64.3052	2.02119	63.2955	5.0485
8	1987	55.839	64.2289	64.2289	1.60169	66.3264	-10.4874
9	1988	60.065	64.6775	64.6775	1.37107	65.8306	-5.7656
10	1989	68.473	66.5334	66.5334	1.46805	66.0485	2.4245
11	1990	61.130	66.6272	66.6272	1.19319	68.0015	-6.8715
12	1991	67.567	67.7697	67.7697	1.18305	67.8204	-0.2534
13	1992	65.412	68.2446	68.2446	1.04142	68.9527	-3.5407
14	1993	71.861	69.8010	69.8010	1.14442	69.2860	2.5750
15	1994	70.232	70.8028	70.8028	1.11589	70.9454	-0.7134
16	1995	72.832	72.1013	72.1013	1.15242	71.9186	0.9134
17	1996	90.164	76.6358	76.6358	1.82883	73.2537	16.9103
18	1997	79.311	78.6339	78.6339	1.86269	78.4646	0.8464
19	1998	80.357	80.4687	80.4687	1.85710	80.4966	-0.1396
20	1999	84.449	82.7504	82.7504	1.94203	82.3258	2.1232
21	2000	79.178	83.5896	83.5896	1.72145	84.6924	-5.5144
22	2001	80.500	84.3488	84.3488	1.52901	85.3110	-4.8110
23	2002	82.749	85.2521	85.2521	1.40386	85.8778	-3.1288
24	2003	84.934	86.3115	86.3115	1.33498	86.6559	-1.7219
25	2004	89.802	88.0776	88.0776	1.42120	87.6465	2.1555
26	2005	82.733	88.1457	88.1457	1.15057	89.4988	-6.7658

Forecasts

Period	Forecast	Lower	Upper
27	89.2962	78.7115	99.881
28	90.4468	79.6415	101.252
29	91.5974	80.5546	102.640
30	92.7479	81.4518	104.044
31	93.8985	82.3341	105.463

Figure 6. Rendements en blé : valeurs lissées, niveaux moyens, valeurs de la tendance, valeurs prédites et résidus pour le modèle de HOLT ($\alpha_1 = \alpha_2 = 0, 2$).

Au temps $t = 0$, la prévision à l'horizon 1 s'écrit :

$$\hat{y}_0(1) = 51,43 + 1,43 = 52,86.$$

La mise à jour du niveau moyen et de la pente conduit aux résultats suivants :

$$\begin{aligned} a(1) &= \alpha_1 y_1 + (1 - \alpha_1)[a(0) + b(0)] \\ &= (0,2)(46,87) + (0,8)[51,43 + 1,43] = 51,66, \end{aligned}$$

$$\begin{aligned} b(1) &= \alpha_2[a(1) - a(0)] + (1 - \alpha_2)b(0) \\ &= 0,2[51,66 - 51,43] + (0,8)(1,43) = 1,19, \end{aligned}$$

et la prévision à l'horizon 1 s'écrit :

$$\hat{y}_1(1) = 51,66 + 1,19 = 52,85.$$

Et ainsi de suite pour $t = 2, 3, \dots$

Au temps $t = 26$, on a :

$$a(26) = 88,15 \quad \text{et} \quad b(26) = 1,15.$$

Les prévisions à l'horizon h s'écrivent :

$$\hat{y}_{26}(h) = a(26) + h b(26),$$

soit

$$\hat{y}_{26}(1) = 88,15 + 1,15 = 89,30,$$

$$\hat{y}_{26}(2) = 88,15 + (2)(1,15) = 90,45,$$

etc.

Pour l'exemple présenté ci-dessus, nous avons utilisé les valeurs proposées par défaut par Minitab ($\alpha_1 = \alpha_2 = 0,2$). On peut, comme pour le lissage exponentiel simple, rechercher le couple de valeurs qui minimise le carré moyen des erreurs de prévision à l'horizon 1. La figure 7 donne l'évolution de ce carré moyen en fonction de α_1 et de α_2 . Elle a été établie à partir des résultats des lissages obtenus en faisant varier α_1 et α_2 de 0 à 1, par pas de 0,1. Au total 121 couples ont donc été testés. On peut constater que le minimum se situe en $\alpha_1 = \alpha_2 = 0$. Cela signifie donc que le meilleur ajustement est obtenu sans mise à jour du niveau moyen et de la pente, c'est-à-dire en considérant que les valeurs estimées à l'horizon 1 se trouvent sur une droite (figure 8). Cette droite est la droite de régression ajustée à l'ensemble des observations pour la détermination des valeurs initiales $a(0)$ et $b(0)$. On notera que pour obtenir la figure 8, les valeurs de α_1 et α_2 ont été, en réalité, fixées à 0,0001, le logiciel n'acceptant pas les valeurs nulles. En termes d'écart-type résiduel, la régression linéaire conduit à une valeur de 5,1 et le lissage de HOLT avec $\alpha_1 = \alpha_2 = 0,2$ à une valeur de 5,8. Ces valeurs sont à comparer avec la valeur 6,5, obtenue pour le lissage simple (figure 3).

Par rapport au lissage simple, la prise en compte de la tendance, qui pour cet exemple est manifeste, améliore l'ajustement, mais cette amélioration est toutefois limitée. Cela montre le pouvoir d'adaptation du lissage simple, lorsque la valeur de α est suffisamment grande.

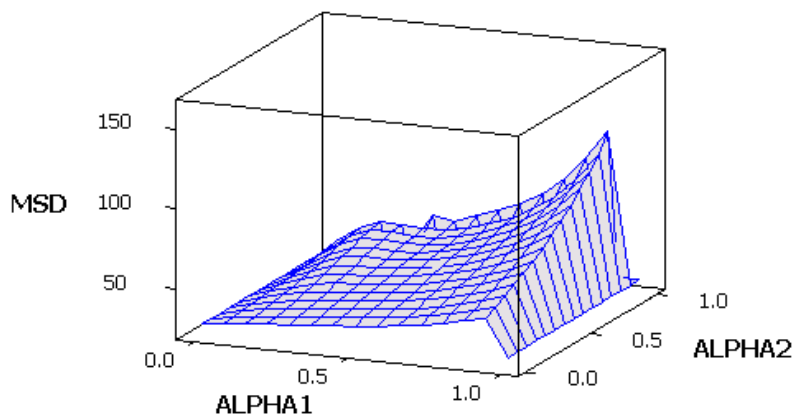


Figure 7. Rendements en blé : évolution du carré moyen des erreurs de prévision à l'horizon 1, en fonction des paramètres de lissage α_1 et α_2 .

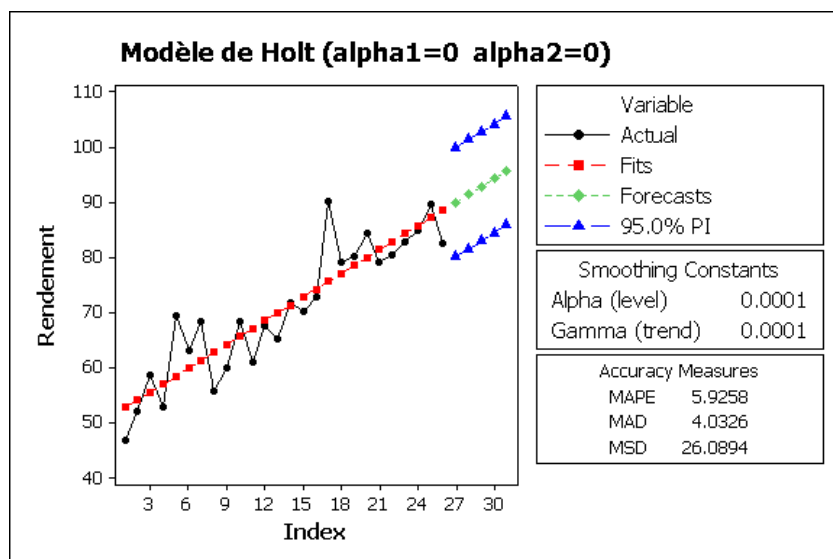


Figure 8. Rendements en blé : modèle de HOLT $\alpha_1 = \alpha_2 = 0,000$.

4. SÉRIES AVEC TENDANCE ET COMPOSANTE SAISONNIÈRE

Une composante saisonnière peut être introduite dans le modèle de HOLT, qui s'appelle alors modèle de HOLT et WINTERS. Ainsi, pour des données mensuelles et dans le cas d'une composante saisonnière additive, la formule de prévision

s'écrit :

$$\hat{y}(h) = a(t) + b(t) + s(t + h - 12).$$

Le troisième terme, $s(t+h-12)$, représente la composante saisonnière qui correspond au mois $t+h$, mais qui a été estimée 12 intervalles de temps auparavant. En fait, la formule ci-dessus n'est valable que pour des prévisions à horizon inférieur ou égal à un an. Si l'horizon est compris entre un et deux ans, la composante saisonnière sera égale à $s(t+h-24)$, et ainsi de suite pour des horizons plus éloignés encore. Notons cependant que de telles prévisions, qui ne sont plus des prévisions à court terme, seront le plus souvent entachées d'erreurs importantes, les rendant inutiles en pratique.

Les composantes du modèle sont mises à jour par les relations suivantes :

$$a(t) = \alpha_1[y_t - s(t-12)] + (1 - \alpha_1)[a(t-1) + b(t-1)]$$

$$b(t) = \alpha_2[a(t) - a(t-1)] + (1 - \alpha_2)b(t-1)$$

et

$$s(t) = \alpha_3[y_t - a(t)] + (1 - \alpha_3)s(t-12).$$

Les formules de mise à jour des coefficients $a(t)$ et $b(t)$ sont identiques aux formules de mise à jour des coefficients du modèle de HOLT, à l'exception de la quantité y_t , qui est remplacée, dans le modèle de HOLT et WINTERS, par la différence :

$$y_t - s(t-12),$$

correspondant en fait à la valeur désaisonnalisée au temps t . Quant aux écarts saisonniers $s(t)$, ils sont également mis à jour par un lissage exponentiel simple, puisque la quantité :

$$y_t - a(t)$$

représente l'écart saisonnier au temps t .

Si la composante saisonnière est multiplicative, le modèle de HOLT et WINTERS s'écrit :

$$\hat{y}_t(h) = [a(t) + b(t)h]s(t + h - 12)$$

et les formules de mise à jour sont les suivantes :

$$a(t) = \alpha_1 \frac{y_t}{s(t-12)} + (1 - \alpha_1)[a(t-1) + b(t-1)]$$

$$b(t) = \alpha_2 [a(t) - a(t-1)] + (1 - \alpha_2)b(t-1)$$

et

$$s(t) = \alpha_3 \frac{y_t}{a(t)} + (1 - \alpha_3)s(t-12).$$

Ces formules sont identiques aux formules de mise à jour du modèle additif, sauf en ce qui concerne la valeur désaisonnalisée au temps t et la composante saisonnière au temps t qui valent respectivement :

$$\frac{y_t}{s(t-12)} \quad \text{et} \quad \frac{y_t}{a(t)}.$$

En pratique, le choix entre le modèle additif et le modèle multiplicatif se fait selon les mêmes principes que dans le cas des méthodes de décomposition [PALM,

1987a] et les formules ci-dessus doivent évidemment être adaptées, si la période du phénomène saisonnier ne comporte pas 12 données.

Comme pour les autres procédures de lissage, les modèles de HOLT et WINTERS nécessitent des valeurs initiales $a(0)$ et $b(0)$ ainsi que les douze coefficients saisonniers $s(0), s(-1), \dots, s(-11)$ et, ici aussi, diverses procédures peuvent être envisagées. Des informations à ce sujet sont données, notamment, par CHATFIELD et YAR [1988]. Dans le logiciel Minitab, les valeurs initiales pour le niveau moyen et la tendance sont obtenues par l'ajustement d'une régression linéaire exprimant la caractéristique étudiée en fonction du temps. Les valeurs initiales pour les coefficients saisonniers sont calculées par régression multiple sur des variables instrumentales, la variable expliquée étant la variable étudiée dont on a éliminé l'effet de la tendance.

Quant aux constantes de lissage, elles peuvent être fixées *a priori*, ou, au contraire être calculées de manière à minimiser la somme des carrés des erreurs de prévision à l'horizon 1.

L'utilisateur peut en effet tester plusieurs triplets α_1, α_2 et α_3 , afin de retenir le triplet qui donne lieu à la plus faible somme des carrés des erreurs de prévision. Mais il faut bien admettre que cette procédure est rapidement fastidieuse, le nombre de triplets à tester devenant vite trop élevé, puisque, pour un même nombre de n valeurs différentes de chacun de ces trois paramètres, le nombre de triplets à tester est égal à n^3 . Si on ne souhaite pas tester plusieurs triplets, et en l'absence de toute information relative à la série, on pourra s'inspirer de la recommandation de CHATFIELD et YAR [1988] qui proposent d'utiliser, dans ces cas, les valeurs suivantes :

$$\alpha_1 = 0,4, \quad \alpha_2 = 0,1 \quad \text{et} \quad \alpha_3 = 0,4,$$

ou encore les valeurs proposées par défaut par Minitab :

$$\alpha_1 = \alpha_2 = \alpha_3 = 0,2.$$

A titre d'illustration, considérons la série chronologique des quantités de lait de vache collectées en Belgique de janvier 1995 à décembre 2005 (en milliers de tonnes)⁸.

La figure 9 donne une représentation de la série et la figure 10 reprend les résultats partiels du lissage, en considérant un modèle additif et en utilisant comme constantes de lissage les valeurs proposées par CHATFIELD et YAR ($\alpha_1 = 0,4, \alpha_2 = 0,1$ et $\alpha_3 = 0,4$). Les données de 2005 sont exclues des calculs et seront utilisées, par la suite, pour la validation.

Illustrons d'abord les formules de mise à jour, en prenant comme point de départ les résultats relatifs au mois de décembre 2003, soit en $t = 108$.

Pour $t = 108$, on a (figure 10) :

$$a(108) = 239,200, \quad b(108) = 0,3340.$$

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

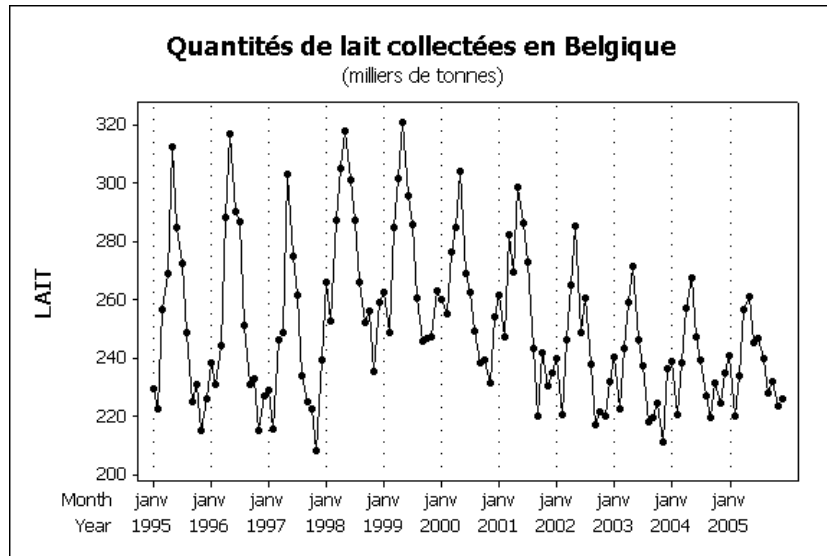


Figure 9. Quantités de lait collectées en Belgique de janvier 1995 à décembre 2005.

Les mises à jour, pour le mois de janvier 2004, c'est-à-dire pour $t = 109$, s'écrivent :

$$\begin{aligned} a(109) &= \alpha_1[y_{109} - s(109 - 12)] + (1 - \alpha_1)[a(108) + b(108)] \\ &= 0,4[238,669 - (-5,9066)] + (1 - 0,4)(239,200 + 0,3340) = 241,551 \end{aligned}$$

$$\begin{aligned} b(109) &= \alpha_2[a(109) - a(108)] + (1 - \alpha_2)b(108) \\ &= 0,1(241,551 - 239,200) + (1 - 0,1)(0,3340) = 0,536 \end{aligned}$$

$$\begin{aligned} s(109) &= \alpha_3[y_{109} - a(109)] + (1 - \alpha_3)[s(109 - 12)] \\ &= 0,4(238,669 - 241,551) + (1 - 0,4)(-5,9066) = -4,697. \end{aligned}$$

De même, pour février 2004, on a :

$$\begin{aligned} a(110) &= \alpha_1[y_{110} - s(110 - 12)] + (1 - \alpha_1)[a(109) + b(109)] \\ &= 0,4[220,316 - (-20,7137)] + (1 - 0,4)(241,551 + 0,53565) = 241,664 \end{aligned}$$

$$\begin{aligned} b(110) &= \alpha_2[a(110) - a(109)] + (1 - \alpha_2)b(109) \\ &= 0,1(241,664 - 241,551) + (1 - 0,1)(0,53565) = 0,493 \end{aligned}$$

$$\begin{aligned} s(110) &= \alpha_3[y_{110} - a(110)] + (1 - \alpha_3)[s(110 - 12)] \\ &= 0,4(220,316 - 241,664) + (1 - 0,4)(-20,7137) = -20,967 \end{aligned}$$

winters' Method for LAIT - Additive Method

Accuracy Measures

MAPE 3.398
MAD 8.601
MSD 111.863

LAIT	M	T	SMOOTH	LEVEL	TREND	SEAS1	FITS1	RES1
240.286	1	97	234.808	244.130	-0.02113	-5.9066	234.558	5.7285
222.503	2	98	223.698	243.639	-0.06810	-20.7117	221.677	-1.1743
213.411	3	99	252.795	239.845	-0.44076	6.9201	252.727	-9.3163
258.874	4	100	260.302	239.009	0.48025	20.7203	259.861	-0.9873
271.741	5	101	286.035	232.811	-1.05199	43.5757	285.535	-14.2935
246.376	6	102	255.547	228.512	-1.37675	20.7870	254.495	-8.1190
237.139	7	103	248.753	221.040	-1.78624	17.7841	247.376	-10.2372
218.224	8	104	218.867	221.711	-1.74051	-1.8988	217.081	1.1431
219.389	9	105	199.486	228.628	-0.87479	-17.0305	197.746	21.6430
274.710	10	106	213.100	232.555	0.39460	-17.6468	212.225	12.0047
210.980	11	107	207.661	211.645	-0.24634	-24.0013	207.268	3.7116
236.180	12	108	221.927	239.200	0.33401	-8.2418	221.676	14.5037
238.669	1	109	233.294	241.551	0.53565	-1.6967	233.628	5.0411
220.326	2	110	220.837	241.664	0.49338	-20.9673	221.373	-1.0569
258.469	3	111	248.584	237.914	0.06905	4.3741	249.077	-10.6083
256.927	4	112	258.134	237.472	0.01800	19.9140	258.203	-1.2762
267.526	5	113	281.048	232.071	-0.52399	40.3233	281.066	-33.5496
247.305	6	114	257.858	229.535	-0.72511	19.5801	252.334	-5.0285
239.265	7	115	247.319	225.878	-1.01829	16.0251	246.594	7.3292
226.693	8	116	221.980	227.153	-0.78903	-2.5232	220.961	5.7317
219.678	9	117	210.122	230.482	0.37724	-14.5597	209.111	10.2947
231.387	10	118	217.835	235.676	0.17994	9.3037	217.458	13.9294
224.209	11	119	211.675	240.798	0.67410	-21.0363	211.855	17.3542
234.879	12	120	232.556	242.131	0.74006	-7.8460	233.230	1.6489

FORECAST

Period	FORECAST	Lower	Upper
121	238.275	217.103	259.246
122	222.644	206.212	245.077
123	248.726	224.680	272.772
124	265.006	239.141	290.870
125	286.153	258.307	314.004
126	266.152	236.187	296.117
127	263.337	231.150	295.524
128	245.529	211.033	280.024
129	234.232	197.359	271.106
130	240.228	200.920	279.537
131	229.736	187.445	271.026
132	243.166	198.855	287.478

Figure 10. Quantités de lait collectées : valeurs lissées, niveaux moyens, valeurs de la tendance et de la composante saisonnière, valeurs prédites et résidus pour le modèle additif de HOLT-WINTERS ($\alpha_1 = 0, 4$, $\alpha_2 = 0, 1$ et $\alpha_3 = 0, 4$).

et ainsi de suite pour les mois successifs.

Si, en décembre 2004, on souhaite réaliser des prévisions pour les mois suivants, on a :

$$\begin{aligned}\widehat{y}_{120}(1) &= a(120) + b(120) + s(120 + 1 - 12) \\ &= 242,131 + 0,7401 - 4,6967 = 238,175\end{aligned}$$

$$\begin{aligned}\widehat{y}_{120}(2) &= a(120) + b(120)(2) + s(120 + 2 - 12) \\ &= 242,131 + (2)(0,7401) - 20,9673 = 222,644\end{aligned}$$

etc.

On retrouve bien, aux arrondis près, les valeurs données dans la figure 10.

Pour le modèle multiplicatif avec les mêmes constantes de lissage que celles utilisées ci-dessus, on obtient les résultats repris à la figure 11.

Les mises à jour, pour $t = 109$, s'écrivent :

$$\begin{aligned}a(109) &= \alpha_1 \frac{y_{109}}{s(109 - 12)} + (1 - \alpha_1)[a(108) + b(108)] \\ &= 0,4 \frac{238,669}{0,97643} + (1 - 0,4)[238,403 + 0,19859] = 240,933\end{aligned}$$

$$\begin{aligned}b(109) &= \alpha_2[a(109) - a(108)] + (1 - \alpha_2)b(108) \\ &= 0,1(240,933 - 238,403) + (1 - 0,1)(0,19859) = 0,4317\end{aligned}$$

$$\begin{aligned}s(109) &= \alpha_3 \frac{y_{109}}{a(109)} + (1 - \alpha_3)[s(109 - 12)] \\ &= 0,4 \frac{238,669}{240,933} + (1 - 0,4)(0,97643) = 0,9821,\end{aligned}$$

et les prévisions réalisées en décembres 2004, pour les mois suivants, sont égales à :

$$\begin{aligned}\widehat{y}_{120}(1) &= [a(120) + b(120)]s(120 + 1 - 12) \\ &= (241,911 + 0,64141)(0,98210) = 238,210\end{aligned}$$

$$\begin{aligned}\widehat{y}_{120}(2) &= [a(120) + b(120)(2)]s(120 + 2 - 12) \\ &= [241,911 + (0,64141)(2)](0,91623) = 222,821\end{aligned}$$

etc.

Le carré moyen des erreurs de prévision à l'horizon 1 est de 111,9 pour le modèle additif et de 117,0 pour le modèle multiplicatif. Pour les constantes de lissage retenues, le modèle additif est donc légèrement meilleur. Pour rechercher le triplet α_1 , α_2 et α_3 , qui minimise le carré moyen, on a fait varier les trois

Winters' Method for LAIT - Multiplicative Method

Accuracy Measures

MAPE 3.379
 MAD 8.624
 MSD 117.048

LAIT	M	T	SNOO2	LEVE2	TREN2	SEAS2	FITS2	RESI2
240.286	1	97	234.016	243.648	-0.08218	0.97643	233.671	6.6147
222.503	2	98	223.795	243.036	-0.13514	0.91732	223.719	-1.2161
243.411	3	99	251.184	239.946	-0.43058	1.02589	251.045	-7.6337
258.874	4	100	258.841	239.700	-0.41214	1.07924	258.377	0.4972
271.241	5	101	283.817	235.204	-0.82049	1.17172	283.329	-12.0877
246.376	6	102	255.793	231.248	-1.13404	1.07869	254.901	-8.5248
237.139	7	103	249.875	225.853	-1.56014	1.06832	248.650	-11.5105
218.224	8	104	222.208	223.297	-1.65972	0.98123	220.673	-2.4495
219.389	9	105	203.703	229.180	-0.90554	0.95026	202.189	17.2001
224.230	10	106	215.400	232.394	-0.49351	0.94987	214.549	9.6814
210.980	11	107	209.746	232.645	-0.41907	0.90428	209.300	1.6796
236.180	12	108	221.854	238.403	0.19859	0.96844	221.455	14.7254
238.669	1	109	232.783	240.933	0.43175	0.98210	232.977	5.6916
220.316	2	110	221.012	240.889	0.38414	0.91623	221.408	-1.0918
238.469	3	111	247.126	237.744	0.03124	1.01676	247.520	-9.0510
256.927	4	112	256.584	237.890	0.04272	1.07956	256.617	0.3096
267.516	5	113	278.739	234.084	-0.34212	1.16016	278.789	-11.2731
247.305	6	114	252.504	231.951	-0.52121	1.07369	252.135	-4.8296
239.265	7	115	247.797	228.444	-0.81982	1.05994	247.740	-7.9751
226.693	8	116	224.156	228.986	-0.68359	0.98473	223.351	3.3419
219.628	9	117	213.017	231.419	-0.37199	0.93778	212.381	7.2467
231.387	10	118	219.818	236.067	0.13008	0.96199	219.465	11.9223
274.209	11	119	213.470	240.896	0.59991	0.91486	213.587	10.6215
234.879	12	120	233.293	241.911	0.64141	0.96944	233.874	1.0048

Forecasts

Period	Forecast	Lower	Upper
121	238.210	217.082	259.338
122	222.821	200.329	245.313
123	247.921	223.811	272.031
124	263.926	237.993	289.860
125	284.375	256.452	312.298
126	263.869	233.825	293.914
127	261.169	228.896	293.442
128	243.270	208.682	277.858
129	232.272	195.300	269.244
130	238.887	199.473	278.300
131	227.769	185.867	269.671
132	241.979	197.549	286.409

Figure 11. Quantités de lait collectées : valeurs lissées, niveaux moyens, valeurs de la tendance et de la composante saisonnière, valeurs prédites et résidus pour le modèle multiplicatif de HOLT-WINTERS ($\alpha_1 = 0,4$, $\alpha_2 = 0,1$ et $\alpha_3 = 0,4$).

constantes de lissage de 0 à 1 par pas de 0,1. Pour ces 1331 triplets, le carré moyen minimum correspond au triplet :

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

tant pour le modèle additif que pour le modèle multiplicatif, les carrés moyens étant respectivement égaux à 84,7 et 85,9. Il faut cependant noter que beaucoup d'autres triplets, parfois très différents, conduisent à un carré moyen du même ordre de grandeur. La figure 12 schématise, pour le modèle additif, tous les triplets pour lesquels le carré moyen ne dépasse pas de 10 pourcents le carré moyen minimum, soit 93,17. Tous ces points sont tels que :

$$0,6 \leq \alpha_1 \leq 1 \quad \text{et} \quad 0 \leq \alpha_2 \leq 0,2.$$

Par contre, α_3 peut varier de 0 à 1.

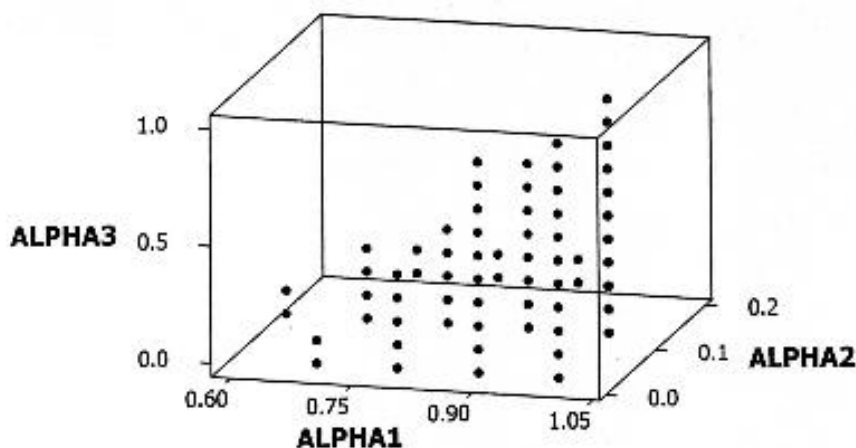


Figure 12. Quantités de lait collectées : ensemble des triplets α_1, α_2 et α_3 conduisant à un carré moyen de l'erreur de prévision inférieur à 93,17 pour le modèle additif de HOLT-WINTERS

Par ailleurs, il faut noter également que l'utilisation du triplet optimal ne garantit pas nécessairement que les prévisions faites à l'avenir auront un carré moyen des erreurs de prévision qui sera minimum. Le modèle pour lequel les erreurs de prévision à l'horizon 1 pour les données du passé sont les plus faibles ne conduit en effet pas automatiquement à des prévisions pour l'avenir qui sont meilleures que celles obtenues pour un autre choix des constantes de lissage.

Ainsi, si on utilise les deux modèles additifs examinés ci-dessus pour réaliser, à partir de décembre 2004, des prévisions à l'horizon 1, 2, ..., 12, c'est-à-dire jusqu'en décembre 2005, et qu'on calcule le carré moyen des erreurs de prédiction, on obtient 175,9 pour le premier modèle ($\alpha_1 = 0,4, \alpha_2 = 0,1$ et $\alpha_3 = 0,4$) et 356,8 pour le second modèle ($\alpha_1 = 0,8, \alpha_2 = 0,1$ et $\alpha_3 = 0$). Sur cette base, on peut donc conclure que le second modèle est moins bon. On peut noter aussi que,

globalement, les erreurs de prédiction pour les horizons supérieurs à 1 sont bien plus importantes que les erreurs de prédiction à l'horizon 1. En effet, pour les 12 mois de 2005, les prévisions à l'horizon 1 conduiraient à des carrés moyens de 77,4 pour le premier modèle et de 129,3 pour le deuxième modèle.

5. CONCLUSIONS

Comme nous l'avons signalé dans l'introduction, les prévisions par lissage sont assez simples à mettre en œuvre, ce qui explique qu'elles sont largement utilisées.

Elles nécessitent cependant de la part de l'utilisateur un certain nombre de choix : choix du modèle, choix des constantes de lissage et choix des valeurs initiales pour débiter le processus.

Le choix du modèle pourra se faire sur la base de l'examen du graphique de la série, qui permettra de vérifier la présence ou non d'une tendance et d'un effet saisonnier et, si ce dernier est présent, de vérifier s'il est additif ou multiplicatif.

Les constantes de lissage peuvent soit être fixées *a priori* soit être déterminées de manière à optimiser un critère. Dans le premier cas, l'utilisateur se base sur des valeurs proposées dans les logiciels ou recommandées dans la littérature. Dans le second cas, le critère d'optimisation généralement retenu est la minimisation du carré moyen de l'erreur de prévision à l'horizon 1. Pour le lissage exponentiel simple et le lissage double, les constantes de lissage optimales peuvent être déterminées par l'ajustement d'un modèle ARIMA bien particulier. Par contre, pour les modèles avec saisonnalité, l'optimisation nécessite l'exploration directe de l'espace à trois dimensions des constantes de lissage. On gardera cependant à l'esprit que les valeurs optimisées sur le passé ne sont pas nécessairement celles qui minimisent les erreurs des prévisions lors de l'utilisation pratique des méthodes.

Quant aux valeurs initiales permettant de débiter le processus de lissage, leur choix n'est sans doute pas crucial. De plus, l'utilisateur d'un logiciel n'a pas nécessairement la possibilité d'intervenir dans ce choix et est, dans ce cas, contraint d'adopter les valeurs calculées par le logiciel.

BIBLIOGRAPHIE

- BOX G.E.P., JENKINS G.M. [1976]. *Times series analysis : forecasting and control*. San Francisco, Holden-Day, 575 p.
- CHATFIELD C. [2003]. *The analysis of time series : an introduction*. London, CRC Press, 352 p.
- CHATFIELD C., YAR M. [1988]. HOLT-WINTERS forecasting: some practical issues. *The Statistician* 37, 129-140.
- GARDNER E.S. [1985]. Exponential smoothing: the state of the art. *J. Forecast.* 4, 1-28.

- GOURIEROUX C., MONFORT A. [1990]. *Séries temporelles et modèles dynamiques*. Paris, Economica, 780 p.
- MELARD G. [1990]. *Méthodes de prévision à court terme*. Bruxelles, Editions de l'Université, 468 p.
- PALM R. [1987a]. Etude des séries chronologiques par les méthodes de décomposition. *Notes Stat. Inform.* Gembloux, 87/1, 25 p.
- PALM R. [1987b]. Etude des séries chronologiques par la méthode de BOX et JENKINS. *Notes Stat. Inform.* Gembloux, 87/2, 40 p.

La collection

NOTES DE STATISTIQUE ET D'INFORMATIQUE

réunit divers travaux (documents didactiques, notes techniques, rapports de recherche, publications, etc.) émanant des services de statistique et d'informatique de la Faculté universitaire des Sciences agronomiques et du Centre de Recherches agronomiques de 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 :

- PALM R., IEMMA A.F. [2002]. Conditions d'application et transformations de variables en régression linéaire. *Notes Stat. Inform.* (Gembloux) 2002/1, 34 p.
- BROSTAUx Y. [2002]. Introduction à l'environnement de programmation statistique R. *Notes Stat. Inform.* (Gembloux) 2002/2, 22 p.
- PALM R. [2003]. Le positionnement multidimensionnel : principes et application. *Notes Stat. Inform.* (Gembloux) 2003/1, 31 p.
- CLAUSTRIAUX J.J. [2006]. Un regard sur les activités de l'Unité de Statistique et Informatique de la Faculté universitaire des Sciences agronomiques de Gembloux (Belgique). *Notes Stat. Inform.* (Gembloux) 2006/1, 9 p.
- CARLETTI I., CLISSEN V., CLAUSTRIAUX J.J. [2006]. Introduction au logiciel Minitab sous WINDOWS. *Notes Stat. Inform.* (Gembloux) 2006/2, 23 p.
- 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.