

Précision des valeurs estimées à l'aide de tarifs de cubage d'arbres

R. PALM

*Faculté des Sciences agronomiques de l'Etat, Chaire de Statistique,
Gembloux, Belgique*

Résumé

La précision du volume estimé d'un ensemble d'arbres à partir d'une équation de cubage dépend essentiellement de la variation résiduelle (c'est-à-dire de la variation du volume des arbres qui ont des dimensions identiques), de la variation liée à l'équation de cubage, du nombre et de la répartition en catégories de grosseurs des arbres à cuber.

L'importance relative de ces différents facteurs a été étudiée, d'une part, en fonction des caractéristiques de l'échantillon utilisé pour la construction du tarif et, d'autre part, en fonction des caractéristiques de l'ensemble d'arbres dont on estime le volume.

1. Introduction

Parmi les avantages que l'on accorde à la construction des tarifs de cubage par voie mathématique, on cite généralement la possibilité de calculer la précision des estimations de volume faites à partir de ces tarifs.

Il faut cependant bien admettre que les indications accompagnant les tarifs sont rarement suffisantes pour permettre à l'utilisateur de chiffrer la précision à laquelle il peut s'attendre lors de l'estimation du volume d'un arbre, d'une coupe ou d'un peuplement.

L'objectif de cette étude est de détailler les différents facteurs qui influencent cette précision.

Le problème sera envisagé dans le cas des tarifs de cubage établis par la méthode des moindres carrés pondérés. En effet, à cause de l'inégalité des variances résiduelles (c'est-à-dire des variances des volumes d'arbres de mêmes dimensions), cette technique conduit, en général, à de meilleurs résultats que la méthode des moindres carrés non pondérés (PALM, 1981 a, 1981 b).

Après un rapide exposé des principes généraux (paragraphe 2), nous analyserons les deux composantes de la variance d'une estimation individuelle : la variation résiduelle (paragraphe 3) et la variation liée à l'équation (paragraphe 4). Ensuite, nous étudierons le problème, plus pratique, de la précision du volume estimé d'un ensemble d'arbres (paragraphe 5). Quelques conclusions termineront ce texte (paragraphe 6).

2. Principes généraux

Supposons que, à partir des observations relatives à un échantillon de n arbres, on désire calculer l'équation de régression suivante :

$$v = b_1 x_1 + b_2 x_2 + \dots + b_p x_p,$$

v étant le volume et x_1, x_2, \dots, x_p les différentes variables explicatives, telles que, par exemple, la circonférence à 1,30 m de hauteur, la hauteur totale ou des fonctions de ces deux caractéristiques.

L'ajustement de cette relation par la méthode des moindres carrés pondérés nécessite, au préalable, une estimation, du moins à une constante près, des variances résiduelles, $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, relatives aux n arbres de l'échantillon.

Rappelons que, pour un arbre donné de l'échantillon, la variance résiduelle, σ_i^2 , mesure la variabilité des volumes de tous les arbres de la population à laquelle on s'intéresse, dont les dimensions sont identiques à celles de l'arbre en question. Si $x_{i1}, x_{i2}, \dots, x_{ip}$ sont les valeurs des variables explicatives associées à l'arbre i , cette variance s'écrit donc :

$$\sigma_i^2 = \text{var}(v \mid (x_{i1}, \dots, x_{ip})).$$

Pour estimer ces variances, une solution consiste à admettre que les résidus de régression sont normaux, de moyenne nulle et que leur variance peut être exprimée en fonction du volume, \hat{v}_i , par la relation :

$$\hat{\sigma}_i^2 = k \hat{v}_i^\lambda.$$

En prenant comme estimations préliminaires les valeurs \hat{v}_i obtenues par régression non pondérée, les paramètres k et λ peuvent être estimés par la méthode du maximum de vraisemblance (PALM, 1981 a).

Disposant d'une estimation des variances résiduelles, la méthode des moindres carrés généralisés permet d'obtenir une estimation du vecteur des paramètres (THEIL, 1971) :

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}^{-1}\mathbf{v}.$$

Dans cette relation, \mathbf{X} est la matrice, de dimension $(n \times p)$, des valeurs des p variables explicatives, \mathbf{W} est la matrice diagonale, de dimension $(n \times n)$, dont les éléments diagonaux sont les variances résiduelles estimées relatives aux différentes observations et \mathbf{v} est le vecteur-colonne des n volumes observés.

Lorsqu'on utilise l'équation ainsi établie pour estimer le volume d'un arbre donné, la variance de cette estimation s'écrit (THEIL, 1971) :

$$\text{var}(\hat{v}_1) = \mathbf{x}'_1 (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \mathbf{x}_1 + \hat{\sigma}_1^2,$$

\hat{v}_1 étant le volume estimé, \mathbf{x}_1 étant le vecteur-colonne des valeurs des variables explicatives relatives à l'estimation \hat{v}_1 et $\hat{\sigma}_1^2$ étant la variance résiduelle correspondant au vecteur \mathbf{x}_1 .

La variance d'une estimation individuelle peut donc être divisée en deux composantes. La première composante mesure la variabilité liée à l'équation de cubage et la seconde composante mesure la variabilité du volume des arbres qui ont des variables explicatives identiques. L'étude de ces deux composantes sera détaillée aux paragraphes 3 et 4.

Pour le calcul de la variance du volume estimé d'une série d'arbres, il est nécessaire de tenir compte de la corrélation qui existe entre les différents volumes estimés, puisque ceux-ci sont obtenus à partir de la même équation de cubage.

La variance de la somme des volumes de k arbres peut s'écrire (BOUCHON, 1974) :

$$\text{var} \left(\sum_{i=1}^k \hat{v}_i \right) = \left(\sum_{i=1}^k \mathbf{x}_i \right) (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \left(\sum_{i=1}^k \mathbf{x}_i \right) + \sum_{i=1}^k \hat{\sigma}_i^2$$

Il en résulte que la variance du volume moyen estimé est égale à :

$$\text{var} \left(\frac{\sum_{i=1}^k \hat{v}_i}{k} \right) = \left(\frac{\sum_{i=1}^k \mathbf{x}'_i}{k} \right) (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \left(\frac{\sum_{i=1}^k \mathbf{x}_i}{k} \right) + \frac{1}{k^2} \sum_{i=1}^k \hat{\sigma}_i^2$$

ou encore :

$$\text{var} (\hat{v}) = \bar{\mathbf{x}}' (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \bar{\mathbf{x}} + \frac{1}{k^2} \sum_{i=1}^k \hat{\sigma}_i^2$$

$\bar{\mathbf{x}}$ étant le vecteur des moyennes des variables explicatives pour l'ensemble des arbres à cuber.

De cette expression, il ressort que, pour une espèce forestière et un type de tarif donnés, la variance du volume moyen estimé dépend du nombre et de la répartition en catégories de grosseurs des arbres dont on désire estimer le volume (paragraphe 5).

3. La variation résiduelle

La variance résiduelle $\hat{\sigma}_i^2$ mesure donc la variabilité du volume des arbres dont les dimensions correspondent aux valeurs du vecteur \mathbf{x}_i . Cette variance n'est pas constante, mais augmente avec les dimensions des arbres, ce qui justifie d'ailleurs l'utilisation de la régression pondérée.

Pour un vecteur \mathbf{x}_i donné, c'est-à-dire pour un arbre de dimensions données, la variance résiduelle dépend d'abord du type de tarif construit. Ainsi, elle sera plus importante dans le cas d'un tarif exprimant le volume uniquement en fonction de la circonférence à 1,30 m que dans le cas d'un tarif exprimant le volume en fonction de la circonférence et de la hauteur. Cette variabilité peut éventuellement être réduite encore si on introduit dans l'équation une autre caractéristique telle qu'un paramètre de forme ou la circonférence à 6 ou 7 m. de hauteur.

A titre d'illustration, nous avons repris les données relatives à 891 chênes, récoltés antérieurement par A. THILL (THILL & PALM, 1979). Les deux équations de cubage suivantes ont été ajustées par la méthode des moindres carrés pondérés décrite au paragraphe 2 :

$$v = b_0 + b_1c + b_2c^2 + b_3c^3, \quad (1)$$

$$v = b_0 + b_1c + b_2c^2 + b_3c^3 + b_4h + b_5c^2h. \quad (2)$$

Dans ces relations, v représente le volume du bois fort, c , la circonférence à 1,30 m. et h la hauteur totale.

Quelques ordres de grandeur de la variation résiduelle sont donnés, pour ces deux tarifs, dans le tableau 1. Dans le cas du tarif à deux entrées, la hauteur prise en considération pour le calcul de la variation résiduelle est, pour les différentes circonférences du tableau, la hauteur moyenne déduite d'une courbe exprimant la hauteur en fonction de la circonférence.

TABLEAU 1

Ecart-type résiduel ($\hat{\sigma}$) et coefficient de variation (CVr) estimés en fonction de la circonférence à 1,30 m., dans le cas des tarifs de cubage à une et à deux entrées relatifs au chêne.

Estimated residual standard deviation ($\hat{\sigma}$) and residual coefficient of variation (CVr) for oak, in relation with the girth at 1.30 m. height, for one entry and two entries volumes tables.

Circonférence à 1,30 m. (cm)	Tarif à 1 entrée		Tarif à 2 entrées	
	$\hat{\sigma}$ (dm ³)	CVr %	$\hat{\sigma}$ (dm ³)	CVr (%)
40	14	18,4	9	12,1
80	79	17,0	44	9,7
120	191	16,3	102	8,6
160	346	15,8	177	8,0
200	541	15,5	265	7,6
240	772	15,2	361	7,2
280	1 037	15,0	469	7,0
320	1 334	14,8	598	6,7
360	1 660	14,7	768	6,5

Pour un type de tarif donné, la variation résiduelle dépend également de l'essence considérée et de la zone de validité du tarif. Pour une région donnée, elle est, par exemple, plus faible dans le cas de l'épicéa que dans le cas du hêtre ou du chêne. De même pour ces deux essences, on peut s'attendre à ce que la variation résiduelle soit plus faible dans le cas d'un tarif valable uniquement en futaie que dans le cas d'un tarif valable à la fois en futaie et en taillis-sous-futaie.

Par contre, la variation résiduelle ne dépend pas de l'effectif et de la structure de l'échantillon utilisé pour la construction du tarif, sauf en ce qui concerne la précision de son estimation ; elle ne peut donc être réduite que par l'augmentation du nombre d'entrées ou par la réduction de la zone de validité du tarif.

4. La variation liée à l'équation

Pour un arbre de dimensions données, la variance liée à l'équation dépend de la matrice des variances et covariances des paramètres de l'équation de régression :

$$(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}.$$

Cette matrice peut encore s'écrire :

$$(\mathbf{Q}'\mathbf{Q})^{-1},$$

\mathbf{Q} étant la matrice de dimension $(n \times p)$, obtenue en divisant les valeurs des variables explicatives relatives à chacun des n arbres de l'échantillon par l'écart-type résiduel estimé correspondant (CUNIA, 1964 ; DRAPER & SMITH, 1966). Dans ces conditions, la matrice $(\mathbf{Q}'\mathbf{Q})$ est donc la matrice des sommes des carrés et des produits des variables explicatives transformées. Par la mise en évidence de la constante n , on peut remplacer les sommes des carrés et des produits par leur moyenne et on obtient :

$$(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} = (\mathbf{Q}'\mathbf{Q})^{-1} = \frac{1}{n}(\mathbf{P}'\mathbf{P})^{-1},$$

$(\mathbf{P}'\mathbf{P})$ étant la matrice des moyennes des carrés et des produits des variables explicatives transformées.

Il apparaît donc clairement, d'une part, que la variance liée à l'équation est inversement proportionnelle à l'effectif de l'échantillon et, d'autre part, qu'elle dépend de la structure de l'échantillon.

A titre d'exemple, la variabilité liée à l'équation a été calculée pour le tarif de cubage à une entrée mentionné au paragraphe 3 (équation (1)). La figure 1 donne l'évolution de cette variabilité, exprimée en pour cent du volume moyen estimé correspondant, en fonction de la grosseur des arbres.

L'examen de cette figure montre que ce coefficient de variation (CVe) est relativement constant dans un intervalle de circonférences donné. En effet, pour des circonférences allant de 60 à 220 cm, il est compris entre 0,6 et 1 p. 100. On constate également que lorsqu'on s'écarte de cet intervalle, le coefficient de variation augmente rapidement.

Une étude basée sur la simulation d'une série d'échantillons d'arbres correspondant à des distributions nettement différentes a montré qu'il existe une relation linéaire très étroite, d'une part, entre la longueur de l'intervalle en question et l'écart-type des circonférences des arbres de l'échantillon et, d'autre part, entre le point central de cet intervalle et la circonférence moyenne des arbres de l'échantillon (PALM, 1981 a).

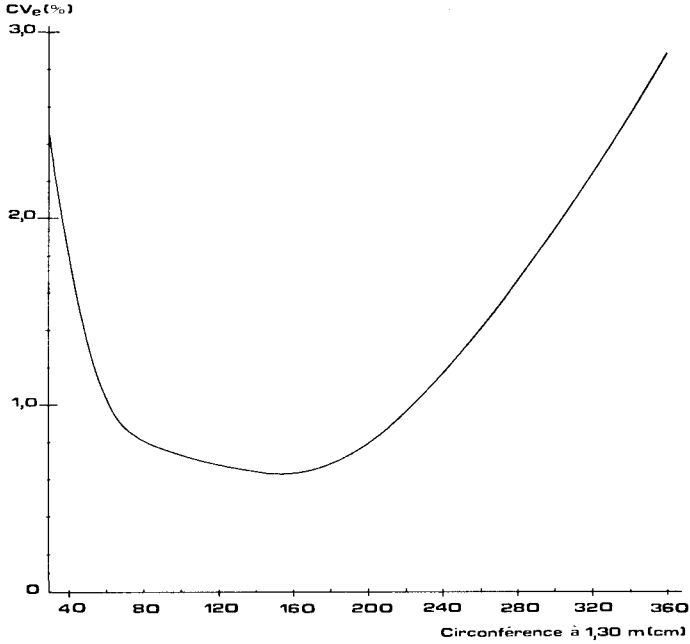


FIG. 1

Coefficient de variation lié à l'équation de régression pour le tarif de cubage à une entrée relatif au chêne.

Coefficient of variation related to the regression equation for one entry volume table for oak.

Dans l'ensemble, on remarque donc une nette similitude entre la précision de l'équation au niveau d'une catégorie de grosseurs donnée et la fréquence relative des observations dans cette catégorie de grosseurs au niveau de l'échantillon. Les extrapolations donneront par conséquent lieu à de très fortes imprécisions.

5. Précision du volume estimé d'un ensemble d'arbres

La relation exacte présentée au paragraphe 2 pour le calcul de la variance du volume moyen d'un ensemble d'arbres est difficilement utilisable en pratique, notamment lorsque se pose le problème d'obtenir, a priori, l'ordre de grandeur de la précision à laquelle on peut s'attendre lors d'un inventaire.

Une estimation rapide, mais plus grossière, de la variance du volume moyen d'un ensemble d'arbres peut être obtenue si on admet que le vecteur des moyennes, \bar{x} , est approximativement égal au vecteur x_* relatif à l'arbre de surface terrière moyenne, et que la moyenne des variances résiduelles est proche de la variance résiduelle de l'arbre de surface terrière moyenne :

$$\frac{1}{k} \sum_{i=1}^k \hat{\sigma}_i^2 \approx \hat{\sigma}_*^2$$

Dans ces conditions, on a, en effet :

$$\text{var}(\hat{v}) \simeq \mathbf{x}'_*(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1} \mathbf{x}_* + \frac{1}{k} \hat{\sigma}_*^2.$$

Dans le but de les comparer, la méthode exacte et la méthode approchée ont été appliquées au calcul de la précision du volume d'une série de groupes de 10, 100 et 1 000 arbres correspondant à diverses distributions. Les calculs ont été réalisés en utilisant le tarif de cubage à une entrée présenté au paragraphe 3 (équation (1)).

L'examen des résultats a montré que, pour des groupes de 10 et de 100 unités, la méthode approchée conduit à une sous-estimation systématique du coefficient de variation du volume moyen estimé. En moyenne, cette sous-estimation est de 7 à 8 p. 100 mais, dans les cas les plus défavorables, elle atteint 35 p. 100 du coefficient de variation calculé par la méthode exacte. De plus, cette sous-estimation est essentiellement fonction du coefficient de variation de la circonférence des arbres dont on estime le volume : plus les circonférences sont dispersées, plus la sous-estimation est importante.

Pour les échantillons de 1 000 observations, par contre, la méthode rapide peut donner lieu à des surestimations. Et, contrairement aux cas précédents, l'erreur n'est pas liée au coefficient de variation des circonférences, mais dépend surtout de la circonférence moyenne : aux moyennes faibles correspondent des surestimations et aux moyennes élevées correspondent des sous-estimations.

On a également vérifié que le remplacement de l'arbre de surface terrière moyenne par l'arbre de circonférence moyenne ou, comme le propose SVALOV (1978), par l'arbre de volume moyen, conduit aux mêmes résultats.

Enfin, on a aussi constaté que les erreurs liées à l'utilisation de la méthode approchée sont du même ordre de grandeur dans les cas d'autres essences forestières : ces erreurs dépendent donc essentiellement de la répartition des arbres dans les différentes catégories de grosseurs.

Bien qu'elle conduise à une erreur pouvant atteindre, dans des situations extrêmes, près de 40 p. 100 du coefficient de variation réel, l'estimation rapide sera cependant, en pratique, largement satisfaisante, du moins dans la plupart des cas. Il ne faut, en effet, pas perdre de vue que, même si elle est importante par rapport au coefficient de variation réel, l'erreur commise reste faible, et est souvent négligeable par rapport au volume moyen.

L'ordre de grandeur de la précision du volume moyen estimé d'un ensemble d'arbres peut donc être obtenu en ne tenant compte que du nombre d'arbres à cuber et de la dimension de l'arbre moyen.

A titre d'exemple, nous avons calculé, dans le cas du tarif à une entrée relatif au chêne, à quelle précision on aurait pu s'attendre lors du cubage d'un ensemble de k arbres, en supposant que le tarif ait été établi sur la base d'un échantillon de 200, 500, 1 000 ou 2 000 arbres. Pour cela, nous avons simplement tenu compte du fait que la variance liée à l'équation est inversement proportionnelle à l'effectif n de l'échantillon. Nous avons donc supposé que les différents échantillons considérés sont identiques, du point de vue de la répartition relative des observations par catégories de grosseurs, à

l'échantillon effectivement récolté sur le terrain. Les résultats, exprimés en pour cent du volume moyen, sont donnés dans le tableau 2, pour trois valeurs différentes de la circonférence moyenne des arbres à cuber.

TABLEAU 2

*Coefficient de variation du volume estimé de k arbres cubés
à l'aide d'un tarif établi à partir de n arbres,
en fonction de la circonférence moyenne des arbres à cuber (c).*

*Coefficient of variation of the estimated volume of k oaks,
in relation with their average girth (c)
and with the number of trees used to fit the regression equation (n).*

c	n	k					
		1	5	10	50	100	1 000
40	200	18,8	9,1	7,0	4,7	4,3	3,9
	500	18,6	8,6	6,3	3,6	3,1	2,5
	1 000	18,5	8,4	6,1	3,1	2,5	1,8
	2 000	18,5	8,3	6,0	2,9	2,2	1,4
160	200	15,9	7,2	5,2	2,6	2,1	1,4
	500	15,8	7,1	5,1	2,4	1,8	1,0
	1 000	15,8	7,1	5,0	2,3	1,7	0,8
	2 000	15,8	7,1	5,0	2,3	1,6	0,7
280	200	15,4	7,6	5,9	4,1	3,8	3,6
	500	15,2	7,1	5,2	3,1	2,7	2,3
	1 000	15,1	6,9	5,0	2,6	2,2	1,7
	2 000	15,0	6,8	4,9	2,4	1,9	1,2

De ce tableau, il ressort que la précision du volume estimé d'un nombre réduit d'arbres (k inférieur à 5 ou 10, par exemple) dépend peu du nombre d'arbres (n) utilisés pour la construction du tarif. Au contraire, lorsque le nombre d'arbres cubés est très grand (k supérieur à 100), la précision du volume moyen dépend de l'effectif n. La variance résiduelle est en effet prépondérante dans le premier cas, tandis que la variance de l'équation détermine essentiellement la précision dans le second cas.

C'est donc la variance de l'équation qui détermine la précision maximum que l'on peut atteindre lorsqu'on cube un grand nombre d'arbres. Le fait de négliger cette variance, comme cela se fait couramment (BOUCHON, 1974), conduit donc à une surestimation de la précision.

6. Conclusions

Les conclusions que l'on peut dégager de cette étude ont trait, d'une part, à la récolte des données en vue de la construction d'un tarif de cubage et, d'autre part, à l'utilisation de ce tarif.

En ce qui concerne la récolte des données, l'importance du nombre d'arbres mesurés, et surtout de la répartition de ceux-ci par catégories de grosseurs, a été mise en évidence. En effet, bien que, pour une distribution de grosseurs donnée, la précision des équations double lorsque l'effectif quadruple, il est impossible, en pratique, d'obtenir un tarif dont la précision soit suffisante pour les catégories de grosseurs mal représentées au niveau de l'échantillon. Il est, au contraire plus avantageux d'élargir la zone de validité du tarif en augmentant le nombre d'arbres dans les catégories de grosseurs extrêmes, plutôt que d'accroître, de façon considérable, l'effectif total.

Lorsqu'on utilise un tarif de cubage donné pour estimer le volume d'un ensemble d'arbres, la précision du volume moyen de ces arbres dépend, d'une part, de leur nombre et, d'autre part, de leur répartition en catégories de grosseurs. Une valeur approchée de cette précision peut cependant être obtenue en ne tenant compte que des dimensions de l'arbre moyen et du nombre d'arbres à cuber.

Pour un nombre réduit d'arbres, la précision du volume moyen estimé dépend essentiellement de la variance résiduelle des volumes. Au contraire, pour un grand nombre d'arbres, elle dépend avant tout de la variance de l'équation de cubage utilisée, c'est-à-dire des caractéristiques de l'échantillon sur la base duquel le tarif a été construit. Dans ce cas, il faut notamment, ne pas perdre de vue que, pour les arbres dont les dimensions se situent à la limite ou en dehors de la zone couverte par l'échantillon récolté pour l'établissement du tarif, les estimations de volumes sont automatiquement très peu précises, quel que soit l'effectif de cet échantillon.

Remerciements

Nous remercions le Pr P. DAGNELIE pour tous les conseils qu'il a bien voulu nous donner, ainsi que M. A. THILL qui nous a permis de disposer de ses données.

Summary

Precision of estimated values by tree volume tables

The precision of the estimated volume of a group of trees by the use of a volume equation principally depends on the residual variation (i.e. the volume variation of the trees which have the same dimensions), on the variation due to the volume equation and on the number and repartition in size classes of the trees to be estimated.

The importance of these different factors has been studied, on the one hand, in relation with the characteristics of the sample used to determine the volume table, and, on the other hand, in relation with the characteristics of the group of trees to be estimated.

Reçu pour publication le 11 novembre 1982.

Références bibliographiques

- BOUCHON J., 1974. *Les tarifs de cubage*. Nancy, Inst. Nat. Rech. For. et Ec. Nat. Génie Rural, Eaux et Forêts, 135.
- CUNIA T., 1964. Weighted least squares method and construction of volume tables. *For. Sci.*, **10**, 180-191.
- DRAPER N.R., SMITH H., 1966. *Applied regression analysis*. New York, Wiley, 407.
- PALM R., 1981 a. *Contribution méthodologique au cubage des arbres et à la construction de tables de cubage et d'assortiments*. Gembloux, Thèse, Fac. Sci. Agron., 295.
- PALM R., 1981 b. Calcul et choix des équations de cubage d'arbres. *Bull. Rech. agron. Gembloux*, **16**, 379-398.
- SVALOV S.N., 1978. Algorithms for calculating the standard errors of methods for evaluating standing timber in cruising. *Sov. agric. Sci.*, **12**, 50-54.
- THEIL H., 1971. *Principles of econometrics*. Amsterdam, North Holland Publishing Co., 736.
- THILL A., PALM R., 1979. Etude dendrométrique des chênes indigènes (*Quercus robur* L. et *Quercus petraea* (Mattuschka) Lieblein). *Bull. Rech. agron. Gembloux*, **14**, 267-284.