

# **L'ajustement de modèles mathématiques aux données longitudinales de croissance**

par

R. HAUSPIE

Nationaal Fonds voor Wetenschappelijk Onderzoek

## **Introduction**

L'étude longitudinale de la croissance implique le recueil de plusieurs données concernant les mêmes sujets et obtenues à des temps différents de la croissance. Dans la plupart des enquêtes longitudinales, le nombre de ces occasions par sujet, ainsi que la grandeur des intervalles entre les occasions successives, varient d'un individu à l'autre. C'est ainsi que se pose le problème méthodologique d'établir de façon objective la courbe de croissance d'un individu à partir d'un certain nombre de mensurations qui ne sont en quelque sorte que des témoins discrets d'un processus qui est continu dans le temps. Plusieurs chercheurs ont tâché de résoudre ce problème en utilisant un modèle mathématique qui décrit une partie ou l'ensemble de la courbe de croissance. Cette méthode a principalement pour but de résumer toutes les observations d'un enfant en un nombre limité de paramètres d'une fonction. Ayant la même signification pour chaque courbe, ces paramètres permettent une comparaison facile et directe entre individus. De plus, il est aisé à partir de ces courbes de déterminer des paramètres biologiques, tels que, par exemple, la taille et la vitesse atteintes au début de la poussée de croissance pubertaire ou à l'âge de la vitesse maximale, etc.

Plusieurs modèles ont été proposés pour décrire la courbe de croissance staturale, telle que la courbe de Jentsch pour la période de l'enfance (DEMING et WASHBURN, 1963), les modèles logistique et Gompertz dans la période de l'adolescence (MARUBINI, RESELE et BARGHINI 1971 ; MARUBINI, RESELE, TANNER et WHITEHOUSE 1972, DEMING 1957), et finalement la fonction double logistique (BOCK, WAINER, PETERSEN, THISSEN, MURRAY et ROCHE 1973, RARICK, WAINER, THISSEN et SEEFELDT 1975, THISSEN, BOCK, WAINER et ROCHE 1976), la fonction triple logistique (BOCK et THISSEN 1976) et les fonctions proposées par PREECE et BAINES (1978) qui couvrent la période de croissance à partir de l'enfance jusqu'à l'âge adulte.

Le développement de nouveaux modèles de croissance, ainsi que l'apparition de programmes d'ordinateur efficaces, qui simplifient la pratique de l'ajustement d'équations non linéaires, ont contribué à une utilisation plus générale de ces techniques dans l'analyse de données longitudinales. Le but de cette étude est de comparer pour quelques-uns de ces modèles leur aptitude à décrire le processus de croissance staturale. Nous avons analysé les fonctions logistique et Gompertz dans la période de l'adolescence et les fonctions double logistique et Preece Baines I et III à partir de l'enfance jusqu'à l'âge adulte. Nous les avons appliquées dans une analyse préliminaire de données longitudinales de croissance de garçons bengalis.

### Matériel et méthodes

#### Matériel

L'échantillon actuel est composé de 68 garçons issus d'une enquête longitudinale menée dans deux villages bengalis, près de Calcutta aux Indes (Sarsuna et Barisha) dans la période 1952-1966. Les 68 sujets ont été sélectionnés dans un échantillon de 303 sujets, parce qu'ils présentaient des données de croissance staturale sur des périodes suffisamment longues, afin de pouvoir ajuster les modèles mathématiques envisagés. Les cinq fonctions étudiées exigent notamment que les données couvrent au moins la période de l'adolescence et montrent une tendance de fin de croissance. L'échantillon a été décrit précédemment par HAUSPIE, DAS, PREECE et TANNER (1980). En vue de la présente analyse, il est toutefois important de rappeler que les enfants ont été mesurés à intervalles d'une année (dans la majorité des cas à 15 jours près de leurs dates d'anniversaire) et à intervalles de six mois entre 10 et 14 ans. Toutes les mesures ont été prises entre 07.00 et 09.00 heures du matin par la même personne, le Professeur S. R. Das, qui a mené l'enquête pendant 14 ans. La taille a été mesurée par la technique décrite par MARTIN (1928), utilisant un anthropomètre standardisé.

#### Modèles

Les fonctions logistique et Gompertz ont été ajustées aux cycles de l'adolescence. Une discussion détaillée des caractéristiques de ces modèles a été donnée par MARUBINI *et coll.* (1972) pour la fonction logistique et par DEMING (1957) et MARUBINI *et coll.* (1971) pour la fonction Gompertz. Nous allons nous limiter à la présentation des équations :

$$\begin{aligned} \text{LOGISTIQUE} & : y = P + K/[1 + \exp(a-bt)] \\ \text{GOMPERTZ} & : y = P + K.\{\exp[-\exp(a-bt)]\} \end{aligned}$$

avec  $y$  = taille (cm),  $t$  = âge (années),  $P$ ,  $K$ ,  $a$  et  $b$  les quatre paramètres de la fonction qui sont à estimer pour chaque sujet. Ces deux fonctions ne peuvent être ajustées qu'aux données correspondant à l'adolescence et exigent donc une estima-

tion de l'âge du début de la poussée de croissance pubertaire à partir duquel les données peuvent être ajustées.

L'addition de deux facteurs logistiques a fourni une équation appelée double logistique qui permet de décrire la courbe de croissance staturale depuis l'enfance jusqu'à la maturité (BOCK *et coll.*, 1973). L'équation a la forme suivante :

$$\text{DOUBLE LOGISTIQUE : } y = \frac{a_1}{1 + \exp[-b_1(t-c_1)]} + \frac{f-a_1}{1 + \exp[-b_2(t-c_2)]}$$

avec  $y$  = taille (cm),  $t$  = âge (années),  $a_1$ ,  $b_1$ ,  $c_1$ ,  $b_2$  et  $c_2$  les cinq paramètres. La taille adulte ( $f$ ) doit être fournie à la fonction. Si la valeur n'est pas connue, elle peut être estimée comme sixième paramètre.

PREECE et BAINES (1978) ont développé une série de trois modèles non linéaires, décrivant le processus entier de la croissance humaine. Le modèle II étant relativement instable, nous n'avons analysé que les modèles I et III, respectivement à 5 et 6 paramètres ; la forme des équations est la suivante :

$$\text{PREECE BAINES I : } y = h_1 - \frac{2(h_1-h_\theta)}{\exp[s_0(t-\theta)] + \exp[s_1(t-\theta)]}$$

$$\text{PREECE BAINES III : } y = \frac{4(h_1-h_\theta)}{\{\exp[p_0(t-\theta)] + \exp[p_1(t-\theta)]\} \{1 + \exp[q_1(t-\theta)]\}}$$

avec  $y$  = taille (cm),  $t$  = âge (années),  $h_1$ ,  $h_\theta$ ,  $s_0$ ,  $s_1$  et  $p_0$ ,  $p_1$ ,  $q_1$  les paramètres des fonctions. Un avantage important de ces modèles est que la taille adulte ne doit pas être connue, mais est estimée par la fonction ( $h_1$ ).

### Programmation

Toutes les fonctions ont été ajustées par une technique des moindres carrés non linéaires, développée par POWELL (1969) pour le «Harwell Subroutine Library». Cette routine a été incluse dans un programme FORTRAN qui fournit pour chaque enfant les estimations des constantes, les erreurs d'estimation, un certain nombre de paramètres biologiques et des statistiques, telles que la variance résiduelle et le test «runs» (SIEGEL 1956). Le programme se termine en fournissant des statistiques sur l'ensemble de l'échantillon, telles que les moyennes et écarts-type des paramètres, les variances résiduelles totales calculées pour tous les individus, et le test de Bartlett, pour vérifier l'homogénéité des variances résiduelles. Les calculs ont été effectués au Centre de Calcul de l'Université Libre de Bruxelles et au London University Computer Centre.

## Résultats

Pour chacun des 68 garçons, nous avons pu ajuster les modèles logistique et Gompertz. L'ajustement des trois autres modèles a échoué dans quelques cas,

donnant des courbes biologiquement non valables. Ceci était dû à la trop courte durée de la période sur laquelle les données s'étendaient pour ces individus. C'est ainsi que seulement dans 62 cas la fonction double logistique, dans 63 cas le modèle Preece Baines I et dans 57 cas le modèle Preece Baines III ont donné des courbes de croissance valables.

La qualité de l'ajustement des fonctions a été estimée par la grandeur de la variance résiduelle. Le rapprochement de la courbe estimée aux points observés est meilleur lorsque la variance résiduelle est plus petite. Le tableau 1 montre, pour les différents modèles, la valeur de la variance résiduelle totale et les degrés de liberté. Les variances résiduelles obtenues pour les différentes fonctions ont été comparées par le test de Wilcoxon pour des échantillons appariés, car le test de Bartlett avait révélé une hétérogénéité significative des variances résiduelles pour certains modèles. Le tableau 2 montre une valeur de t qui est significative ( $P < .05$ ) pour la comparaison des modèles logistique/Gompertz indiquant que la variance résiduelle après ajustement de la fonction logistique est généralement inférieure à celle obtenue avec la fonction Gompertz.

TABLEAU 1

Variances résiduelles totales, test «runs» pour cinq modèles mathématiques de croissance

Modèles	Nombre de paramètres	Nombre de sujets ajustés	Variance résiduelle totale	Degrés de liberté	% de test de «runs» significatif ( $P < .05$ )
Logistique	4	68	.353	395	0
Gompertz	4	68	.450	395	0
Double-logistique	5	62	.725	592	6.6
Preece-Baines I	5	63	.459	580	0
Preece-Baines III	6	57	.558	517	1.8

TABLEAU 2

Comparaison de la qualité des ajustements de cinq modèles par le test de Wilcoxon pour des échantillons appariés

Modèles	t	Probabilité	
Logistique-Gompertz	- 2.316	.021	*
Double logistique - Preece Baines I	- 2.404	.016	*
Double logistique - Preece Baines III	- .310	.757	NS
Preece Baines I - Preece Baines III	- 5.009	.000	***

Signification : \* =  $P < .05$  ; \*\*\* =  $P < .001$  ; NS = non significatif.

La grandeur de la variance résiduelle n'est qu'un critère de la qualité de l'ajustement d'une équation, le test «runs» décrit par SIEGEL (1956) en est un autre. Ce test est indépendant de la grandeur des résidus mais mesure pour chaque courbe si les signes des résidus consécutifs s'alternent d'une façon aléatoire. Un changement trop fréquent des signes des résidus consécutifs pourrait être une indication d'une variation cyclique dans les données (variation saisonnière, par exemple), qui ne peut être décrite par le modèle ; l'inverse indiquerait des erreurs systématiques de l'ajustement. Ce test n'était significatif pour aucune des courbes logistique ou Gompertz.

Les trois autres modèles qui couvrent des périodes d'âge plus étendues s'ajustent relativement bien à nos données. Néanmoins, nous constatons que la variance résiduelle totale est plus grande pour la fonction double logistique que pour les modèles Preece Baines (tableau 1). D'après le test de Wilcoxon la variance résiduelle est significativement plus grande pour la fonction double logistique que pour le modèle Preece Baines I, mais la différence n'est pas significative quand on compare la fonction double logistique au modèle Preece Baines III (tableau 2).

Parmi les 62 courbes double logistique, quatre (6.6%) présentaient des tests «runs» significatifs, ce qui est légèrement supérieur à ce que l'on pourrait attendre en admettant le seuil de 5%. Dans 3 cas, le test indiquait des erreurs systématiques dans l'ajustement. La variance résiduelle totale pour le modèle Preece Baines I est inférieure à celle du modèle III. Le test de Wilcoxon montre que le modèle I donne des ajustements significativement meilleurs que le modèle III. Aucune des courbes du modèle I n'a présenté des tests «runs» significatifs tandis que, pour le modèle III, une courbe sur 57 (1.8%) montrait une alternance des signes des résidus consécutifs significativement trop élevée.

Les figures 1 et 2 montrent un exemple de chaque type de courbes. Les cas représentés correspondent aux cas ayant la variance résiduelle la plus proche de la variance résiduelle totale pour le modèle en question. Ces exemples sont une illustration du rapprochement très net des courbes aux points observés.

### Discussion

Cette étude a confirmé que les fonctions logistique et Gompertz sont très robustes et aptes à être ajustées aux données longitudinales de la croissance staturale dans la période de l'adolescence. Comme les trois autres fonctions analysées, elles exigent que les données indiquent l'achèvement de la croissance et couvrent au moins la période de l'adolescence. Nos résultats confirment l'observation faite par MARUBINI *et coll.* (1971) et par HAUSPIE *et coll.* (1980), notamment que la fonction logistique est légèrement meilleure que la fonction Gompertz pour ajuster le cycle de l'adolescence.

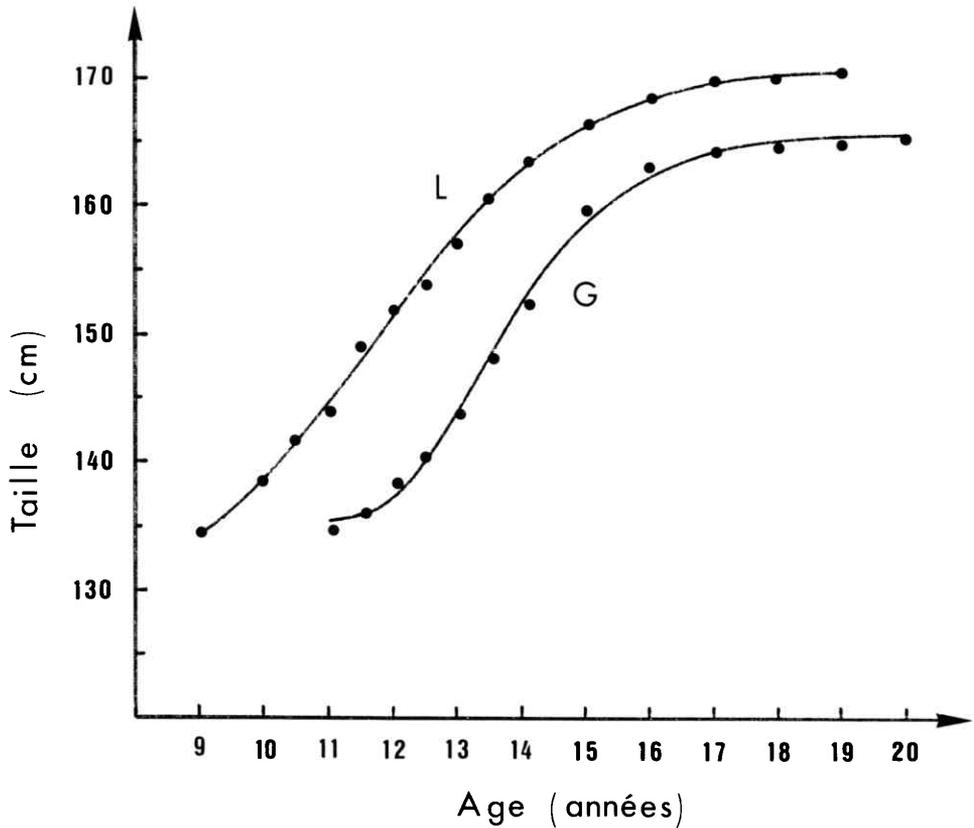


Fig. 1. — Courbes individuelles de croissance staturale : L : modèle logistique ajusté aux données du sujet n° 147 (variance résiduelle = .345 cm<sup>2</sup>) ; G : modèle Gompertz ajusté aux données du sujet n° 149 (variance résiduelle = .443 cm<sup>2</sup>).

Si les données longitudinales s'étendent au-delà du cycle de l'adolescence, on a intérêt à utiliser le modèle Preece Baines I qui nécessite l'estimation de 5 paramètres mais n'exige pas la définition graphique du début de la poussée de croissance pubertaire comme dans le cas de la fonction logistique et Gompertz. Le modèle Preece Baines I est, lui aussi, meilleur que la fonction double logistique, car cette dernière montre une variance résiduelle significativement plus grande et a tendance à donner des erreurs systématiques dans l'ajustement. Les mêmes observations ont été faites par HAUSPIE *et coll.* (1980) dans un échantillon longitudinal de 35 filles belges. D'ailleurs, pour la fonction double logistique il faut connaître avec précision la valeur de la taille adulte, sinon elle doit être estimée par l'équation comme sixième paramètre. Le modèle Preece Baines III a une constante de plus que le modèle I, mais dans notre échantillon il ne donne en général pas des

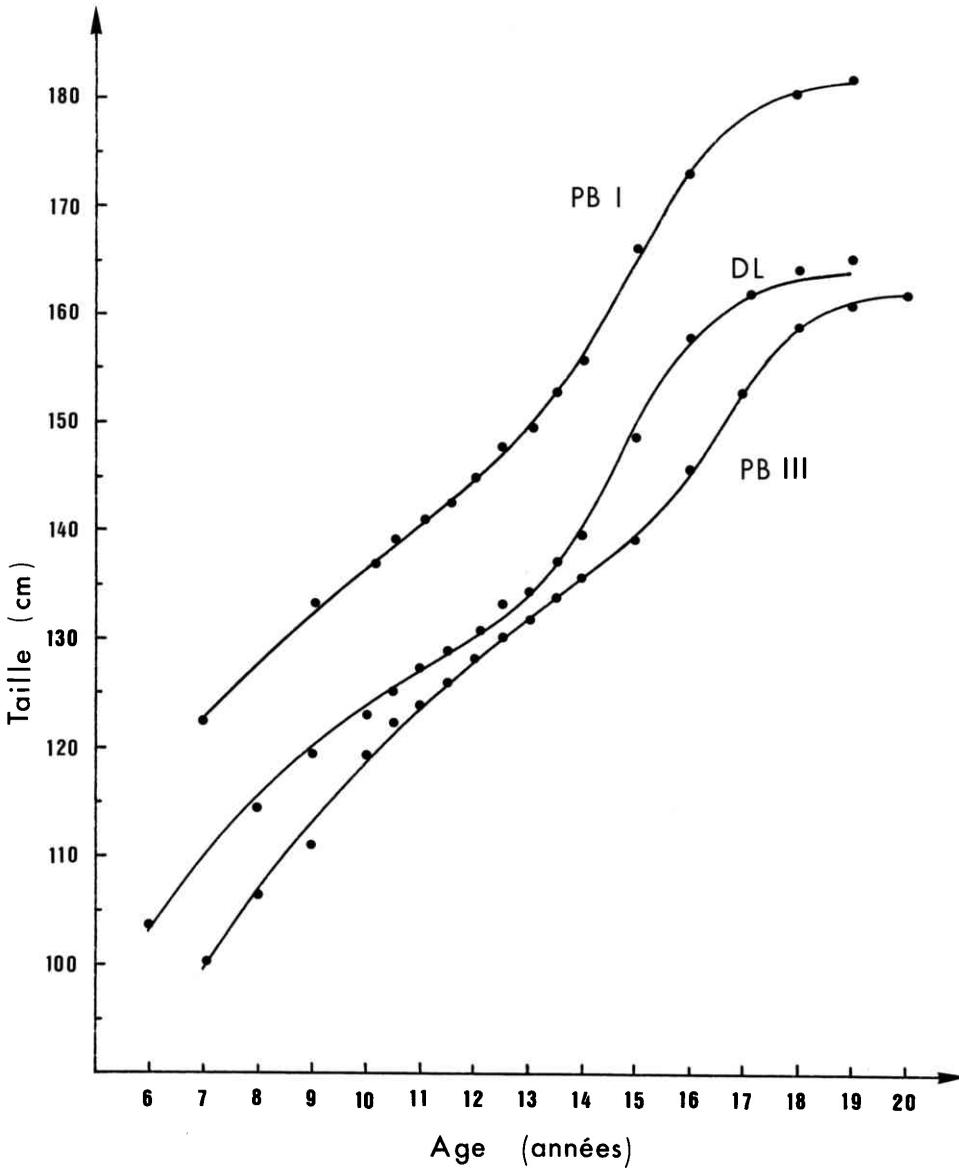


FIG. 2. — Courbes individuelles de croissance staturale : DL : modèle double logistique ajusté aux données du sujet n° 206 (variance résiduelle =  $.812 \text{ cm}^2$ ) ; PB I : modèle Preece Baines I ajusté aux données du sujet n° 40 (variance résiduelle =  $.458 \text{ cm}^2$ ) ; PB III : modèle Preece Baines III ajusté aux données du sujet n° 121 (variance résiduelle =  $.579 \text{ cm}^2$ ).

ajustements meilleurs que le modèle I quoique la différence entre les variances résiduelles ne soit pas statistiquement significative.

Ceci est en contradiction avec les observations de PREECE et BAINES (1978) qui trouvaient dans tous les cas des ajustements plus satisfaisants avec le modèle III qu'avec le modèle I. Dans notre étude cependant, nous avons préféré le modèle I, d'autant plus que le modèle III présentait également une convergence beaucoup plus lente lors du processus de minimalisation.

### Remerciements

L'auteur tient à remercier le Professeur J. M. Tanner et le Professeur C. Susanne pour leurs commentaires constructifs. Ce travail a pu être réalisé grâce à un «NATO Research Fellowship» et grâce à la collaboration du Professeur S. R. Das.

### RÉFÉRENCES

- BOCK, R. D. et THISSEN, D. M.  
1976 Fitting multicomponent models for growth in stature.  
*Proceedings of Ninth International Biometrics Conference*, Boston, p. 431.
- BOCK, R. D., WAINER, H., PETERSEN, A., THISSEN, D., MURRAY, J. et ROCHE, A.  
1973 A parameterization for individual human growth curves.  
*Human Biology*, **45** : 63-80.
- DEMING, J.  
1957 Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth.  
*Human Biology*, **29** : 83-122.
- DEMING, J. et WASHBURN, A. H.  
1963 Application of the Jenss curve to the observed pattern of growth during the first eight years of life in forty boys and girls.  
*Human Biology*, **35** : 484-506.
- HAUSPIE, R. C., DAS, S. R., PREECE, M. A. et TANNER, J. M.  
1980 A longitudinal study of the growth in height of boys and girls of West Bengal (India) aged six months to 20 years.  
*Ann. hum. Biol.* **7** : 347-358.
- HAUSPIE, R. C., WACHHOLDER, A., BARON, G., CANTRAINED, F., SUSANNE, C. et GRAFFAR, M.  
1980 A comparative study of the fit of four different functions to longitudinal data of growth in height of Belgian girls.  
*Ann. hum. Biol.* **7** : 347-358.
- MARUBINI, E., RESELE, L. F. et BARGHINI, G.  
1971 A comparative fitting of the Gompertz and logistic functions to longitudinal height data during adolescence in girls.  
*Human Biology*, **43** : 237-252.

- MARUBINI, E., RESELE, L. F., TANNER, J. M. et WHITEHOUSE, R. H.  
1972 The fit of the Gompertz and logistic curves to longitudinal data during adolescence on height, sitting height, and biacromial diameter in boys and girls of the Harpenden Growth Study.  
*Human Biology*, **44** : 511-524.
- POWELL, M. J. D.  
1969 *Harwell Subroutine Library Routine VAO5A*. A.E.R.E.
- PREECE, M. A. et BAINES, M. J.  
1978 A new family of mathematical models describing the human growth curve.  
*Ann. hum. Biol.*, **5** : 1-24.
- RARICK, G. L., WAINER, H., THISSEN, D. et SEEFELDT, V.  
1975 A double logistic comparison of growth patterns of normal children and children with Down's syndrome.  
*Ann. hum. Biol.*, **2** : 339-346.
- SIEGEL, S.  
1956 *Non-parametric statistics for behavioural sciences*.  
London, McGraw Hill, 311 p.
- THISSEN, D., BOCK, R. D., WAINER, H. et ROCHE, A.  
1976 Individual growth in stature : a comparison of four studies in the U.S.A.  
*Ann. hum. Biol.*, **3** : 529-542.

*Adresse de l'auteur* : R. HAUSPIE  
Laboratorium Antropogenetica  
Vrije Universiteit Brussel  
Pleinlaan 2  
1050 Brussel.