

Choice of a model for height–growth curves in maritime pine

F Danjon ¹, JC Hervé ²

¹ INRA, Laboratoire Croissance et Production, Pierroton, 33610 Cestas;
² Université Claude-Bernard-Lyon I, Laboratoire de Biométrie, Génétique
et Biologie des Populations (CNRS–URA 243), 69600 Villeurbanne, France

(Received 28 April 1993; accepted 31 March 1994)

Summary — A modelling procedure is presented for height–growth curves in maritime pine (*Pinus pinaster* Ait). We chose to fit 4 parameter nonlinear functions. Some of the parameters were fixed or estimated globally (1 value for all curves in a data set). The models were reparametrized to ensure good identifiability and better characterization of the data. The structural properties of parametrizations were investigated using sensitivity functions and the models were compared using a test file. We show that the estimation of 4 parameters for each curve is not possible in practice and that even the estimation of only 3 parameters should be avoided, in particular with the Lundqvist–Matern model or with short growth curves. With 2 local parameters, the Lundqvist–Matern model appears slightly more suitable than the Chapman–Richards model.

height–growth curves / nonlinear regression / *Pinus pinaster* / parametrization

Résumé — **Choix d'un modèle pour l'étude des courbes de croissance en hauteur du pin maritime.** Une procédure de modélisation est présentée pour l'étude des courbes de croissance en hauteur de pins maritimes (*Pinus pinaster* Ait). Nous avons choisi l'ajustement à des fonctions non linéaires à 4 paramètres. Certains paramètres ont été fixés ou estimés globalement (une valeur commune à toutes les courbes). Les modèles ont été reparamétrés, de façon à améliorer l'identifiabilité ainsi que la caractérisation des données. Les propriétés des modèles et des paramétrisations ont été examinées à l'aide des fonctions de sensibilité. Les modèles ont été comparés sur un fichier test. Nous montrons que l'estimation de 4 paramètres pour chaque courbe est pratiquement impossible, et que même l'estimation de seulement 3 paramètres doit être évitée, en particulier avec le modèle de Lundqvist-Matern ou avec des courbes courtes. En revanche, avec 2 paramètres locaux, le modèle de Lundqvist-Matern semble un peu mieux adapté que le modèle de Chapman-Richards, ce dernier sous-estimant les hauteurs aux âges avancés.

courbe de croissance en hauteur / régression non linéaire / *Pinus pinaster* / paramétrisation

INTRODUCTION

Nonlinear growth functions have been used to assess the genetic variability of height–growth curves of forest trees (Namkoong *et al.*, 1972; Buford and Bukhart, 1987; Sprinz *et al.*, 1987; Magnussen, 1993). A well-known advantage of these models is that they can provide an efficient summary of the data *via* a small number of meaningful parameters, the significance of which does not change with the trials.

Our aim is to select a model to be used on several data sets of individual height–age curves of maritime pines (*Pinus pinaster* Ait) aged between 20 and 80 years. Most of the work was carried out on 22-year-old progeny tests, especially to investigate their genetic variability. From an examination of nearly 4 000 curves we observed that they generally have a regular sigmoidal shape, with an inflexion point at about 10 years and an asymptote between 20 and 50 m (Danjon, 1992). It therefore seems possible to describe all the curves by a sigmoidal growth function.

However, fitting the model by nonlinear regression may pose a number of practical difficulties, especially if the curves are short. Ill-conditioning is a commonly encountered problem (see, *eg*, Seber and Wild, 1989, chapter 3), resulting in highly correlated and unsound estimates, which can greatly affect the use of the method (Rozenberg, 1993). The problem may partly come from the data, but also from the model itself, and/or from the parametrization used; this last point is often neglected in applications.

In order to detect and avoid these potential shortcomings, a preliminary investigation was carried out and is presented in this paper. Different models and different parametrizations of the same model are compared on a test file of long growth series. The objectives were to check the model's ability to fit the full growth profile and to char-

acterize the general behaviour of the models, noting the properties that are inherent in the models themselves and those that depend on the parametrization.

MODELLING PROCEDURE

Model functions

Debouche (1979) recommend the use of Lundqvist–Matern (Matern, 1959) and Chapman–Richards (Richards, 1959) variable-shape functions. Both curves have 4 parameters, which have the following meanings: A = asymptote; r = related to relative growth rate; m = shape parameter; and a position parameter (location of the curve on the time axis).

With height at time 0 (h_0) as position parameter, the Lundqvist–Matern model (LM1) is (h = height; t = time):

$$h(t) = A \exp - \left[\frac{rt}{m} + \left(\log \frac{A}{h_0} \right)^{\frac{1}{m}} \right]^{-m}$$

and the Chapman–Richards model (CR1) is:

$$h(t) = A [1 - q \exp(-rt)]^{1/1-m}$$

with $q = 1 - (h_0/A)^{1-m}$

Number of parameters

As the curves are sometimes rather short, estimating all 4 parameters for each curve may be wasteful (Day, 1966): the precision of each estimation will be low, with high correlations between the estimates for each curve (which we will call 'e-correlations'), and a poor convergence of the numerical procedures in many cases. Hence, to produce reliable estimations,

some parameters must be fixed at a given value or estimated globally for the population (one value for the whole set of curves) with minimum total sum of squares as a criterion.

Because the age of the trees are known and because we use height at age zero (h_0) as position parameter, the latter can be fixed to zero. As suggested by Day (1966), scale parameters (asymptote and growth rate) are considered specific to each individual whereas the shape parameter (m) may be estimated globally for the population.

Parametrization

The original equations were reparametrized to gain 'stable parameters' (Ross, 1970). Such parameters vary little in the whole region of best fittings. They are simple expressions of physical characters of a curve, and only have a major influence on a limited portion of the curve.

For the LM model, the maximum growth rate is given by:

$$R_M = rA \exp \left[\left(1 + \frac{1}{m}\right) \left(\log\left(1 + \frac{1}{m}\right) - 1\right) \right]$$

Three parameters are related to this essential characteristic of the curve, which is likely to induce e-correlations between parameters and instability. To avoid these problems, R_M will be used as a parameter, instead of r .

The shape parameter m locates the inflexion point on the h -axis at a proportion $p = \exp^{-(1+1/m)}$ of the final size. This expres-

sion can be inverted to yield m as a function of p . It is hence possible to use p directly as shape parameter instead of m in order to make the interpretation of the estimated value easier¹. This leads to the following new form of the LM model (LM2) where R_M is called r_{LM2} and p is called m_{LM2} for homogeneity of notation:

$$h(t) = A \exp - \left[\left(\log \frac{A}{h_0} \right) (1 + \log m_{LM2}) \right. \\ \left. - \frac{r_{LM2} t (1 + \log m_{LM2})}{A \exp[(\log m_{LM2})(\log(-\log m_{LM2})) - 1]} \right]^{\frac{1}{1 + \log m_{LM2}}}$$

In the same way, for the CR2 model, r_{CR1} is changed to r_{CR2} , the maximum growth rate:

$$r_{CR2} = r_{CR1} A^{\frac{m}{1-m}}$$

But in this case, the relative height of the inflexion point is $p = m^{m/1-m}$, and there is no closed form solution for m in terms of p . This precludes the use of p for the CR model. Keeping m , the new form of the CR model (CR2) is as follows:

$$h(t) = A \left[1 - q \exp(-r_{CR2} A^{\frac{-m}{1-m}} t)^{\frac{1}{1-m}} \right]$$

After reparametrization of both models, all parameters have a direct physical meaning, except m in CR2.

Sensitivity functions

Seber and Wild (1989, p 118) state that "one advantage of finding stable parameters lies

¹ This transformation is made for this practical reasons but, being univariate, it has essentially no effect on the precision and on e-correlations with other parameters. Notably, the sensitivity functions of m and p (see below) are identical, apart from a multiplicative constant, and the first-order estimates of e-correlations will be strictly equal under either parametrization. Nevertheless, the transformation may have second-order effects on the precision by reducing the parametric nonlinearity, but we did not investigate this point.

in forcing us to think about those aspects of the model for which the data provide good information and those aspects for which there is little information". Sensitivity functions are a convenient means of studying the repartition of information along the time scale.

For a model $f(t, \theta)$, depending on the parameter vector θ , the sensitivity function of a parameter θ_i is the partial derivative of the model function with respect to θ_i (Beck and Arnold, 1977):

$$\Phi_i(t) = \partial f(t, \theta) / \partial \theta_i$$

and indicates how the growth curve is modified at time t by a small change $\Delta\theta_i$ in the parameter value θ_i :

$$\Delta f(t) \approx \Phi_i(t) \cdot \Delta \theta_i$$

Formally, the importance of the sensitivity function may be appreciated by considering that the asymptotic variance-covariance matrix of the estimates is proportional to $(X^t X)^{-1}$, where X is a rectangular matrix whose columns are the sensitivity functions of each estimated parameter, evaluated at each observed time.

If the sensitivity functions of 2 parameters are proportional on a given sampling interval, the 2 parameters have essentially the same effect on the corresponding part of the curve and their e-correlation will be high. Additionally, the precision of estimation of a given parameter is better when its sensitivity function is higher (in absolute value) in the observed time range.

Chapman–Richards model

It can be seen on figure 1a that, for CR1, the sensitivity functions of A , r and m are nearly proportional on the $[0, 25]$ time interval. Figure 1b shows that this feature disappears in the second parametrization, which concentrates the effects of m in the

early ages, and those of A in the latter part of the growth curve. This is likely to reduce e-correlations between A and r , and r and m .

It should be noted that fitting trees under 20 years old will result in imprecise estimates for both parametrizations: for CR1, precision will be low for all parameters because of e-correlations between all of them, while for CR2, imprecision will essentially concern A , because its sensitivity function is very small and negative in this time range.

Lundqvist–Matern model

The features of the different parametrizations are essentially the same as for the Chapman–Richards model. The major differences are that, for the LM2 model, the maximum of Φ_m is after 50 years and the rise of Φ_A is slower than for CR2 (fig 1c,d).

The former happens because, in the LM model, m controls both the beginning of the curve and its convergence rate to the asymptote. This is a special property of the LM model, and is not shared by the CR model. It is potentially misleading since a single parameter controls 2 distinct features of the curve, between which no evident biological link exists. It is also likely to increase e-correlation between A and m , compared to the CR model.

The latter illustrates that although the convergence rate to the asymptote depends on m (the curve converges to its asymptote in $t^{-m_{LM}}$ when $t \rightarrow +\infty$), it is always under-exponential, while it is exponential for the CR model. Both features are intrinsic properties of the LM model, which do not depend on the parametrization.

MATERIAL AND METHODS

The models were tested with a data set containing 44 trees belonging to 13 good growing stands,

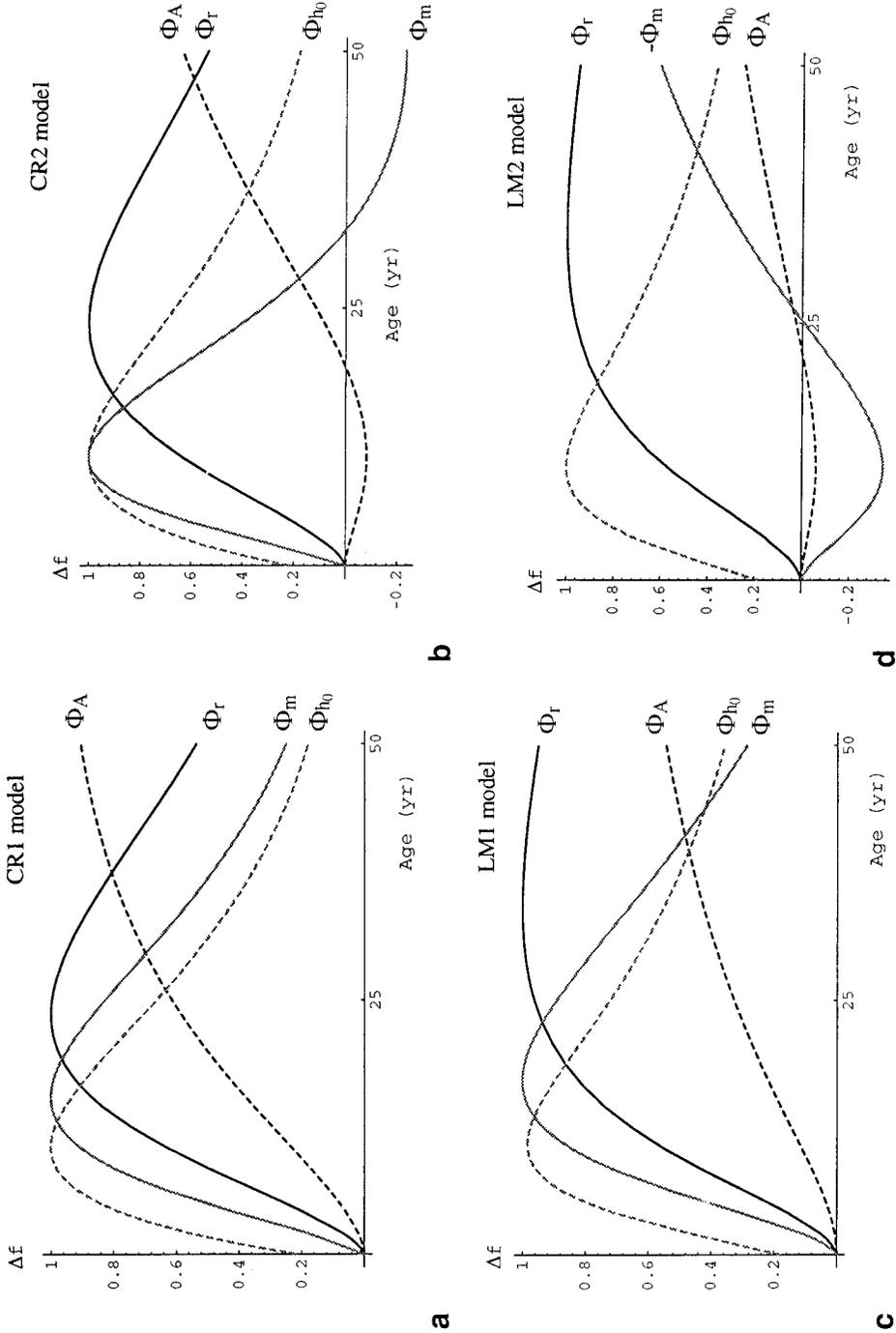


Fig 1. Sensitivity functions for CR1 (a), CR2 (b), LM1 (c) and LM2 (d) models. The functions have been normalized so that their maximum value over $[0, +\infty[$ is +1. The graphics have been drawn with the following values, typical of a good growing tree: for the Chapman-Richards model: $A = 30$ m, $r_{CR1} = 0.06 \text{ yr}^{-1}$, $r_{CR2} = 0.9 \text{ m}\cdot\text{yr}^{-1}$, $m = 0.5$, $h_0 = 0.1$ m; for the Lundqvist-Matern model: $A = 55$ m, $r_{LM1} = 0.03 \text{ yr}^{-1}$, $r_{LM2} = 0.9 \text{ m}\cdot\text{yr}^{-1}$, $m_{LM1} = 1$, $m_{LM2} = 0.1352$, $h_0 = 0.16$ m.

sampled in the Landes de Gascogne area and aged more than 35 years to get the main part of the curve. This selection was made because further studied tests are all good growing stands and because we suspect that potential drawbacks of the different models, although always present, may not be fully appreciable on short curves. Half of the trees were measured by stem analysis (stems sectioned at 2-m intervals, see Carmean, 1972), and for the remaining trees annual height increments were assessed using branch whorls as morphological markers (Kremer, 1981). Measures started at about age 5 years, the zero point was included in the analysis. Two trees had non-sigmoidal curves.

Nonlinear regression was made with a special software which use ordinary least-squares estimation and the Gauss–Marquardt algorithm following the implementation recommended by Moré (1977).

The quality of fit was appreciated by graphical displays including plots of the observed points together with the regression curve, plots of residuals *versus* time and plots of bivariate distribution of parameter estimates with ellipses representing first-order asymptotic approximations of confidence regions (as in Corman *et al*, 1986). The ellipse area was related to the precision of estimation. An inclination and a lengthening of the ellipse indicates a high *e*-correlation. These graphical representations provide a synthetic overview of estimation quality which cannot be so easily assessed by marginal standard errors and *e*-correlations. Note that residuals and residual sum of squares do not vary with the

parametrization, but depend only on the model functions (LM or CR).

RESULTS AND DISCUSSION

Number of local parameters

All estimations with 4 local parameters yield very high *e*-correlations, indicating over-parametrization. With 3 local parameters (*A*, *r* and *m*), convergence for 5 trees with LM1 and for 1 tree with other models could not be obtained and *e*-correlations were all higher than 0.8 (table I).

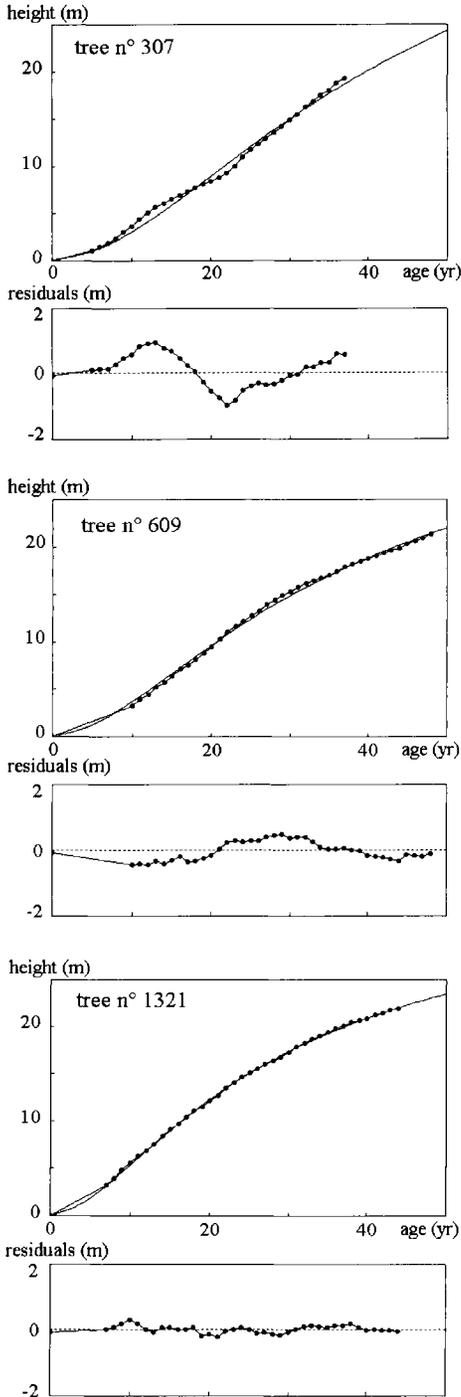
The origin of the strong correlation between *A* and *m* (0.98 for LM1 and LM2) in the Lundqvist–Matern model has been previously investigated with the sensitivity functions and, consequently, the use of 3 local parameters with this model should be considered with care and restricted to long growth series. Only fitting with 2 local parameters (*A* and *r*) is carried out in the sequel.

Typical examples of fit are shown in figure 2. No evidence of systematic behaviour of residuals exist (fig 3), and so the basic hypothesis concerning the sigmoidal shape of curves prove to be reasonable. Further-

Table I. Mean of absolute values of *e*-correlations as a function of model, parametrization and number of local parameters.

<i>Number of local parameters</i>	<i>e</i> -correlation	<i>Chapman–Richards model</i>		<i>Lundqvist–Matern model</i>	
		CR1	CR2	LM1	LM2
3	(<i>A</i> , <i>r</i>)	(–) 0.95	(–) 0.83	(–) 0.996	(–) 0.84
3	(<i>A</i> , <i>m</i>)	(–) 0.85	(–) 0.85	(–) 0.980	(–) 0.98
3	(<i>r</i> , <i>m</i>)	0.96	0.84	0.990	0.81
2	(<i>A</i> , <i>r</i>)	(–) 0.94	(–) 0.56	(–) 0.95	(–) 0.55

(–) Mean correlation is negative.



more, the constant shape imposed by the global estimation of m seems acceptable.

Effect of reparametrization

For both models, the mean e-correlation between A and r is close to 1 with the first parametrization (table 1). Following reparametrization the correlation decreases to approximately 0.5.

On the CR1 plot of the bivariate distribution of A and r (fig 4), a nonlinear trend between A and r is visible, and the confidence ellipses are large compared with the distance between curves and oriented along the trend. With CR2, ellipses are smaller, with no general trend being observed. Similar observations have been made concerning LM1 and LM2 (not shown). These considerations show that the second parametrizations are certainly more appropriate to appreciate true differences between curves.

Comparison of the LM2 and CR2 models

The position parameter (h_0) was first fixed at zero for both models, which resulted in good fit with CR but gave rise to positive residuals around 3 years for all trees with LM model: the Lundqvist–Matern model starts slowly, the lag phase at the beginning of the curve seems too long for maritime pine, and best fitting is generally obtained with a very low non-zero value of h_0 (a few cm or less). Indeed, with the test file, a global estimation of the position parameter (h_0) yielded

Fig 2. Typical fits with the LM model (residuals versus age at the bottom): tree 307 (bad fit), tree 609 (medium fit) and tree 1321 (good fit); m and h_0 globally estimated ($m_{LM2} = 0.16$ and $h_0 = 9.2$ cm).

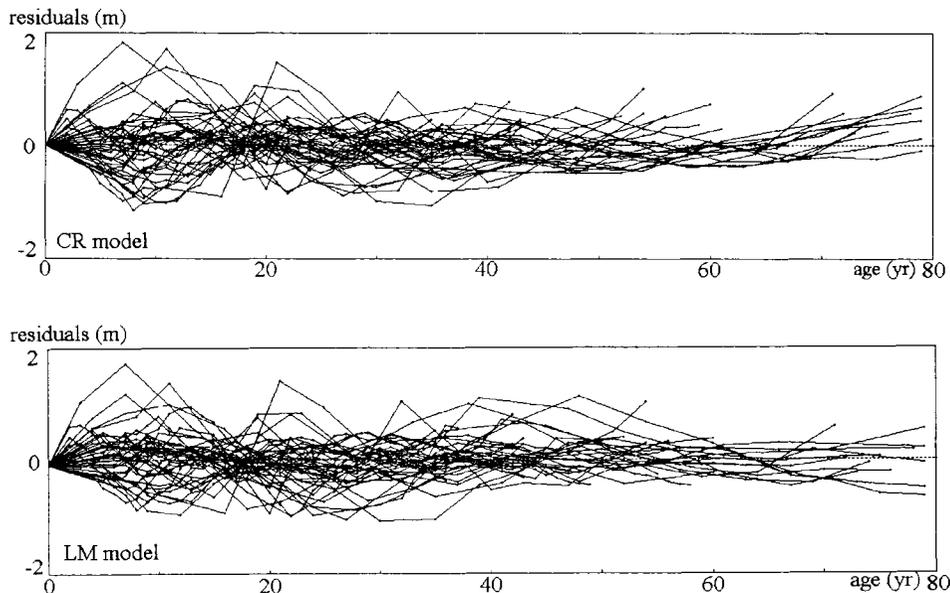


Fig 3. Residuals versus age for all trees with CR and LM models; m and h_0 globally estimated (for CR2, $m_{CR2} = 0.43$ and $h_0 = 0$ cm; for LM2, see fig 2).

about 0 cm for CR but 10 cm for LM, so h_0 was fixed to 10 cm for the LM model.

Mean, standard deviation and mean standard errors are quite similar for r , but not for A (table II). There is a general tendency for A to be about 30% greater for LM2 than for CR2. This is a consequence of the faster

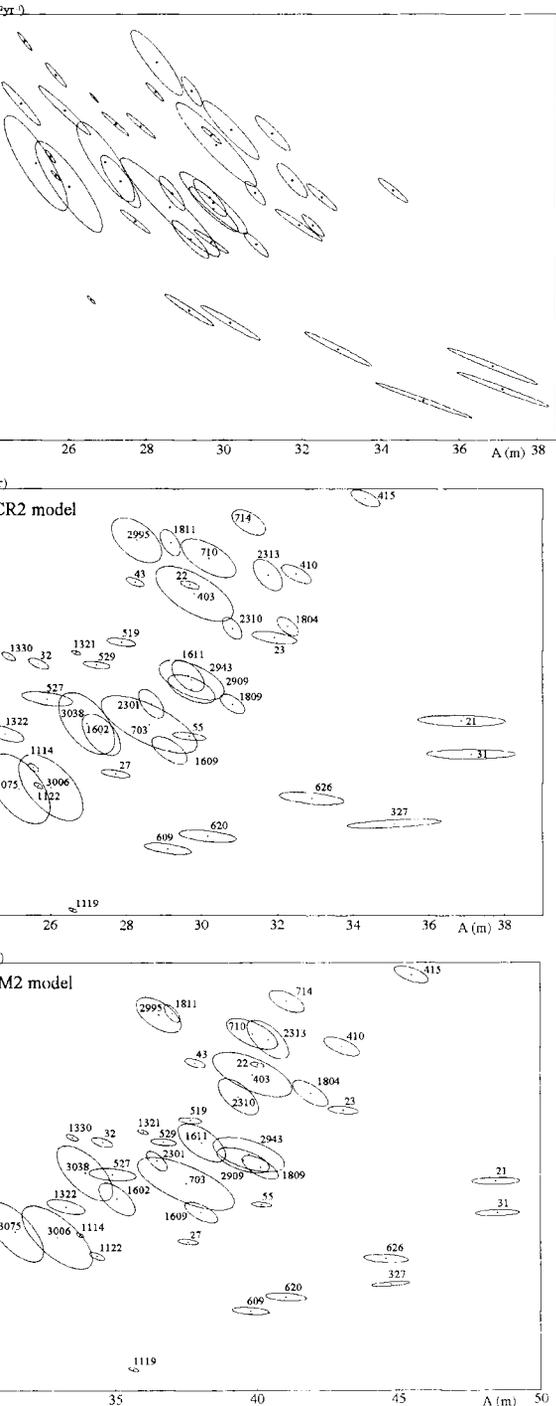
convergence of the Chapman–Richards model to its asymptote (exponential) compared with that of the Lundqvist–Matern model (under-exponential).

Examination of the residuals (fig 3) reveals another consequence of this intrinsic difference between the 2 models: the

Table II. Comparison of Lundqvist–Matern (LM2) and Chapman–Richards (CR2) model.

Model	Global sum of squares (m^2)	A (in m)					r (in cm/year)				
		mean	s-e	s-d	min	max	mean	s-e	s-d	min	max
LM2	118	39.4	0.61	5.3	32	46	72.7	0.66	7.1	56	88
CR2	134	30.3	0.59	5.5	25	54	69.2	0.66	6.6	54	84

s-e = mean estimation standard error; s-d = standard deviation; min = minimum; max = maximum.



pattern of the residuals is rather similar under the 2 models, nevertheless, there is a visible tendency for the last CR2 residuals to be positive. Indeed, the mean of the last observed residual of each curve is significantly positive (22 cm, $p = 0.9995$) for CR2, which is not the case for LM2 (5 cm, $p = 0.85$). Therefore, it seems that the CR model joins its asymptote too quickly, underestimating height for old ages.

The maxima of the asymptote estimates are rather high, but not completely unrealistic. Furthermore, they are obtained for the non-sigmoid curves (by removing them, the maxima decrease to 37 and 48 m). However, the estimated asymptotes should not (and need not) be considered as estimations of ultimate heights of trees, because such an interpretation involves extrapolations of the models far beyond the last observed points. In any case, we have no real interest in the prediction of growth after 80 years; we use this parameter to characterize the later part of the curves.

Comparing residual sum of squares, LM2 is a little better than CR2, and the precision of estimations and e -correlations are close for the 2 models (table I and II). The relative positions of each curve on the A - r plane (fig 4) are very similar: correlations between the estimations obtained with the 2 models are high (0.95 for A and 0.996 for r). As long as one is not concerned with extrapolation towards old ages, the 2 models (with only 2 local parameters) are likely to yield similar results.

Fig 4. Bivariate distribution of asymptote (A) and growth rate parameter (r) with 50% confidence ellipses for CR1, CR2 and LM2 models, with trees 307 and 320 removed. For CR1, m was estimated globally ($m = 0.43$) and h_0 was fixed to zero. For CR2 and LM2, m and h_0 were globally estimated (see fig 3). Each tree is identified by a number.

CONCLUSION

The analysis was made with rather long series. However, the classical parametrizations (CR1 and LM1) always yield high e-correlations and even after reparametrization e-correlations remain high with 3 local parameters. This is especially true with the Lundqvist–Matern model. We have emphasized the dual influence of the shape parameter in this case, which partially explains the high e-correlation. For this model, a variable shape parameter between curves will also lead to interpretative difficulties (asymptotes are not comparable when the convergence rate varies). Examination of the sensitivity functions indicates that, handling shorter growth series, it will be even more essential to use the reparametrized functions and to keep only 2 local parameters.

With 2 local parameters, the Lundqvist–Matern function appears slightly better than the Chapman–Richards one, yielding a lower sum of squares, as a result of a closer fit to the last part of the curves. With 8 other data sets (Danjon, 1992), the advantage of the LM model is conserved. This seems to indicate that the exponential slowing down of growth that characterized the Chapman–Richards function is too fast and does not well describe maritime pine final growth. Nevertheless, it is a small effect and, in contrast, the Lundqvist–Matern does not fit the very beginning of growth while the CR model does. On a practical ground, when 2 local parameters are sufficient, and for descriptive purposes, the 2 models will lead to similar conclusions. However, they will probably differ in extrapolation, and this requires further study.

ACKNOWLEDGMENTS

The authors wish to thank B Lemoine and A Kremer for providing data, and 2 anonymous referees

for helpful remarks, which greatly improved the presentation of the paper.

REFERENCES

- Beck JV, Arnold KJ (1977) *Parameter Estimation in Engineering and Science*. J Wiley & Sons, New York, USA
- Buford MA, Burkhart HE (1987) Genetic improvement effects on growth and yield of loblolly pine plantations. *For Sci* 33, 707-724
- Carmean WH (1972) Site index curves for upland oaks in the central states. *For Sci* 18, 109-120
- Corman A, Carret G, Pavé A, Flandrois JP, Couix C (1986) Bacterial growth measurement using an automated system: mathematical modelling and analysis of growth kinetics. *Ann Inst Pasteur Microbiol* 137B, 133-143
- Danjon F (1992) Variabilité génétique des courbes de croissance en hauteur du pin maritime (*Pinus pinaster* Ait). PhD Thesis, Université de Lyon I, France
- Day NE (1966) Fitting curves to longitudinal data. *Biometrics* 22, 276-291
- Debouche C (1979) Présentation coordonnée de différents modèles de croissance. *Rev Stat Appl* 27, 5-22
- Kremer A (1981) Déterminisme génétique de la croissance en hauteur du pin maritime (*Pinus pinaster* Ait). I. Rôle du polycyclisme. *Ann Sci For* 38, 199-222
- Magnussen S (1993) Growth differentiation in white spruce crop tree progenies. *Silvae Genet* 42, 258-266
- Matern B (1959) Some remarks on the extrapolation of height growth. Forest Rest Inst Sweden Statistical Report n° 2, Vallentuna
- More JJ (1977) The Levenberg–Marquardt algorithm: implementation and theory. In: *Numerical Analysis, Lecture Notes in Mathematics 630* (GA Watson ed). Springer, Berlin, 105-116
- Namkoong G, Usanis RA, Silen RR (1972) Age-related variation in genetic control of height growth in douglas-fir. *Theor Appl Genet* 42, 151-159
- Richards FJ (1959) A flexible growth function for empirical use. *J Exp Bot* 10, 290-300
- Ross GJS (1970) The efficient use of function minimization in nonlinear maximum-likelihood estimation. *Appl Stat* 19, 205-221
- Rozenberg P (1993) Comparaison de la croissance en hauteur entre 1 et 25 ans de 12 provenances de douglas (*Pseudotsuga menziesii* (Mirb) Franco). *Ann Sci For* 50, 363-381
- Seber GAF, Wild CJ (1989) *Nonlinear Regression*. J Wiley & Sons, New York
- Sprinz PT, Talbert CB, Strub MR (1987) Height-age trends from an Arkansas seed source study. *For Sci* 35, 677-691