

Prediction of Growth: A Hierarchical Bayesian Approach

ELJA ARIAS

Rolf Nevanlinna Institute
University of Helsinki
Finland

LIPING LIU

Department of Probability and Statistics
Peking University
P.R. China

NIKO MAGLAPERIDZE

Georgian Statistical Association
Tbilisi
Republic of Georgia

Summary

A nonparametric hierarchical growth curve model is proposed. Different levels in the model hierarchy are intended to correspond to different sources of variation in an individual's growth. The nonparametric character of the model offers considerable flexibility in fitting the growth curves to empirical data. Here the emphasis is on prediction, and for this purpose the adopted Bayesian inferential approach seems particularly natural and efficient. A Markov chain Carlo method is used to perform the numerical computations. As an illustration of the techniques, we consider the growth of children, during their first two years.

Key words: Growth curve; Hierarchical modelling; Predictive distribution; Bayesian credible interval.

1. Introduction

The prediction of an individual's growth, based on its earlier size measurements and on data describing the growth of other "similar" individuals, forms an interesting challenge to a statistician. Parametric models considered in the literature, which moreover are often assumed to be of a polynomial form, tend to be specific to some particular species or populations. It also seems to be the general experience from such models that relatively simple models apply only over a short range

of time, and that even then it is difficult to attach a biological meaning to their coefficients. Therefore, in order to allow for more flexibility and a more widely applicable approach, it seems desirable to consider nonparametric or semi-parametric models. Furthermore, if the emphasis is on the model's ability to predict the future growth of some particular (set of) individuals, it is attractive to look for a solution where the uncertainties relating to such growth are expressed probabilistically. This gives a practical justification for adopting an essentially nonparametric Bayesian approach in the inferential problem, and for formulating such probabilities explicitly in terms of predictive distributions.

The literature concerning statistical modelling of growth is vast. A good overview is given in TANNER (1989). The literature has been dominated by generalized multivariate analysis-of-variance models originally introduced by POTTHOFF and ROY (1964), and subsequently studied e.g. by RAO (1965, 1977, 1984, 1987), KHATRI (1966), GRIZZLE and ALLEN (1969), GEISSER (1970, 1980, 1981), LEE and GEISSER (1972, 1975), FEARN (1975), LEE (1982, 1988, 1990), LISKI and NUMMI (1990), among others. These models are basically following the parametric linear models tradition, assuming that an individual's growth is a polynomial function of time (typically, age) but that the coefficients can be specific to a group of such individuals. Both non-Bayesian and Bayesian inferential methods have been used in the estimation of such models; for the latter, see e.g. GEISSER (1970) and LEE (1982). To give an example of nonlinearly parametrized growth curve models, BERKEY (1982) developed an empirical Bayes approach for fitting the Jenss growth curve model. In comparison, the nonparametric approach to growth curve modeling has been much less used. GASSER et al. (1984) discussed certain advantages of nonparametric models over the parametric ones, mostly in the context of human growth, while KNEIP and GASSER (1992) modelled the structural pattern common to all members in a sample of curves, proposed some estimators and discussed their asymptotic behaviour. RAO (1987) discussed direct and inverse regression methods under nonparametric growth curve models, using Bayesian and empirical Bayesian methods. TIAN et al. (1994) proposed the use of restricted cubic splines, and compared this with the methods of RAO (1987). BOULARAN et al. (1994): developed a two-stage nonparametric method, splitting the growth curve into two additive parts, one representing a common structure of all individuals, and the other being specific to each individual in the population. BARRY (1995) assumed that the growth curves are Gaussian, specifying a prior distribution such that the curves vary slowly over time, and developed a test of a homogeneity hypothesis. To our knowledge, the present paper is the first to combine a nonparametric modelling approach of growth curves with full Bayesian estimation and prediction.

We now outline the questions considered in this paper. Suppose we observe the growth of n individuals from the same population, making measurements of some particular characteristic, such as height or weight, over a period of time. Let $Y_{k,i}$ denote a measurement made on individual k at time $t_{k,i}$. Denoting the observation interval by $[0, T_{\max}]$, we have measurements of the first individual at times

$0 < t_{1,1}, \dots, t_{1,p_1} \leq T_{\max}$, of the second at times $0 < t_{2,1}, \dots, t_{2,p_2} \leq T_{\max}$, etc. Note that, unlike in the more classical multivariate growth curve models, we do not require that the measurement times for different individuals coincide, or even that there is the same number of measurements on all. The observations can now be collected into n vectors

$$\begin{aligned} Y_1 &\equiv (Y_{1,1}, \dots, Y_{1,p_1}); & Y_2 &\equiv (Y_{2,1}, \dots, Y_{2,p_2}); & \dots; \\ Y_n &\equiv (Y_{n,1}, \dots, Y_{n,p_n}). \end{aligned} \tag{1.1}$$

Let us first write the true size of individual k at time t in the form $f_k(t)$, where f_k is an unknown function of t (growth curve). The observations are then assumed to have the structure

$$Y_{k,t} = f_k(t) + \eta_{k,t}, \quad k = 1, \dots, n, \tag{1.2}$$

where the $\eta_{k,t}$'s are Gaussian errors, independent of each other and independent of the growth curves.

The main problem in this paper is the prediction of the future growth of a particular individual, say the n -th in the previous sample. Supposing we have already made observations $Y_n \equiv (Y_{n,1}, \dots, Y_{n,p_n})$ on this individual, at respective times $t_{n,1}, \dots, t_{n,p_n}$, where $p_n \geq 0$, our task would typically be to predict the size $f_n(\tilde{t})$ at a time $\tilde{t} \in (t_{n,p_n}, T_{\max}]$. (We take $p_n = 0$ to mean that there are no previous observations on this particular individual in the data.) Alternatively, we may be interested in predicting the time it will take until this individual has grown to some target size.

In the next section we construct the growth curve model, which has a hierarchical structure, and assign prior distributions to variables which are not given or observed. Section 3 outlines the Metropolis-Hastings algorithm, which is used in the numerical calculations. We apply the model to a real data set in section 4, and end the paper with a brief discussion in section 5.

2. The model

It is a natural idea in the context described that characteristics assigned to individuals need to be separated from population characteristics. The former can be viewed as representing variability among individuals belonging to the same population. The latter, on the other hand, will be interpreted as unknown structural parameters or functions describing the population from which the individuals are sampled (see KNEIP and GASSER (1992)). This leads us to consider the general framework of hierarchical Bayesian models.

More precisely, we assume that the growth curve $f_k(t)$ of individual k can be represented in the form of the composite function $g(A_k(t))$. Here we assume that

(i) g is an unknown non-negative, continuous and strictly increasing function satisfying $g(0) = 0$.

(ii) The random function $W_k(t)$ driving an individual's growth is taken to be an Ornstein-Uhlenbeck process $W_k(t) = e^{-\gamma t} \int_0^t e^{\gamma s} dB_k(s)$, where γ is a random parameter with prior distribution $\exp(\gamma_0)$ and $B_k(s)$ is a Brownian motion having covariance function $E(B_k(s_1) B_k(s_2)) = \sigma_B^2 (s_1 \wedge s_2)$, with σ_B^2 having an inverse gamma distribution $IG(b, c)$. Here γ_0 , b and c are given hyperparameters.

(iii) $A_k(t) = \int_0^t h(W_k(s)) ds$, $k = 1, \dots, n$ is a time transformation, where h is a given (deterministic) positive increasing link function.

(iv) g , the B_k 's ($k = 1, 2, \dots, n$), γ and σ_η^2 are assumed to be independent. We also assume that B_1, B_2, \dots, B_n are mutually independent given σ_B^2 .

The intuitive idea behind this admittedly rather elaborate model is the following: The increasing process $A_k(t)$ driven by a Brownian motion is a cumulative operational time (or "physiological age") for individual k corresponding to age t . The integrand $h(W_k(t))$ is always positive and intended to describe the relative speed of an "operational time clock" with respect to t . Depending on the considered application, different choices of the link function h (e.g., the exponential function) can turn out to provide good fit to the data. Since h is increasing, a bigger value of W_k corresponds to a faster speed of the clock. Moreover, the increments of the \hat{A} process $A_k(t) = \int_0^t h(W_k(s)) ds$ are serially correlated, i.e., slow (fast) speed at a time point is associated with slow (fast) speed at neighbouring time points. In order to facilitate the calibration of $A_k(t)$ with calendar time t , we have assumed W_k to be an Ornstein-Uhlenbeck process with unknown drift towards the origin. The drift term turns out to be particularly important in prediction problems which involve extrapolation beyond the last observation time. In the special case where $\gamma = 0$, W_k collapses to the Brownian motion B_k . Since A_k is determined uniquely by W_k , and again W_k is determined uniquely by γ and B_k , we use γ and the B_k 's as our model parameters.

Theoretically, the processes $B_k(\cdot)$ driving the individual growth curves are infinite dimensional parameters of the model, but in practice we need discrete approximations. This is done by simply dividing $[0, T_{\max}]$ into N subintervals, for some (fixed) large enough N . Denoting $B_k(\cdot) \equiv \left(B_k\left(\frac{T_{\max}}{N}\right), \dots, B_k(T_{\max}) \right)$, now if the $B_k(\cdot)$'s, and therefore $W_k(\cdot)$'s are available, then by rounding the observation times $\{t_{k,l}, l = 1, \dots, p_k, k = 1, \dots, n\}$ to the closest grid points (multiples of T_{\max}/N), we can approximate the $A_k(t_{k,l})$'s easily by the corresponding Riemann sums.

Consider then the nonparametrically defined function g which represents the common growth curve for the considered population. In practice as an approximation, we assume g to be an unknown piecewise linear function, with positive slopes. The points at which the slope changes could be chosen to form a sequence of ordered random variables (see ARJAS and GASBARRA (1994)). For simplicity of computation, however, we let these points to be deterministic and equally-spaced

on the considered interval. Since the operational age $A_k(t)$ can increase faster than the time t , $A_k(T_{\max})$ might be larger than T_{\max} . To avoid complications we define g on $[0, CT_{\max}]$, where $C > 1$ is a large enough constant. Supposing that the interval $[0, CT_{\max}]$ has been divided into N_0 subintervals, say $[0, t_1], [t_1, t_2], \dots, [t_{N_0-1}, t_{N_0}]$, where $t_l = lCT_{\max}/N_0$ ($l = 1, 2, \dots, N_0$), we postulate the prior distribution of g in terms of its increments as follows: the increment $\Delta g(l) = g(t_l) - g(t_{l-1})$, is an independent random variable from the exponential distribution $\text{Exp}((t_l - t_{l-1})^{-1} \mu)$, where μ is a given hyperparameter. It is easily seen that the prior expectation of the increment is proportional to the length of the segment, and therefore $E_{\text{prior}}(g(t))$ is a linear function of t . No special shape of the common growth curve is assumed.

As for the Gaussian measurement errors in (1.2), we assume a priori that $\eta_k \equiv (\eta_{k,t_{k,1}}, \dots, \eta_{k,t_{k,p_k}})$, $k = 1, 2, \dots, n$, are i.i.d. Normal random variables with mean 0 and variance σ_η^2 . They are also assumed to be independent of all other random parameters. The prior uncertainty about the measurement error variance σ_η^2 is described by an inverse gamma distribution $IG(a_\eta, b_\eta)$ (see GELFAND and SMITH (1990)), with hyperparameters a_η and b_η being given.

As a summary, we can say that the random parameters g , γ , σ_B^2 and σ_η^2 are designed to describe the properties of the common growth curve, while the B_k 's reflect heterogeneity between individuals. The choice $\sigma_B^2 = 0$ would correspond to complete homogeneity. The fact that we are leaving the parameters σ_B^2 and γ unspecified, controlling them only through the corresponding hyperpriors $IG(b, c)$ and $\exp(\gamma_0)$, means that the estimated level of heterogeneity will effectively be "data driven".

A convenient way to summarize this hierarchical model is to use the chain independence graph displayed in Figure 1 (see LANGE (1992)). As usual, unknown

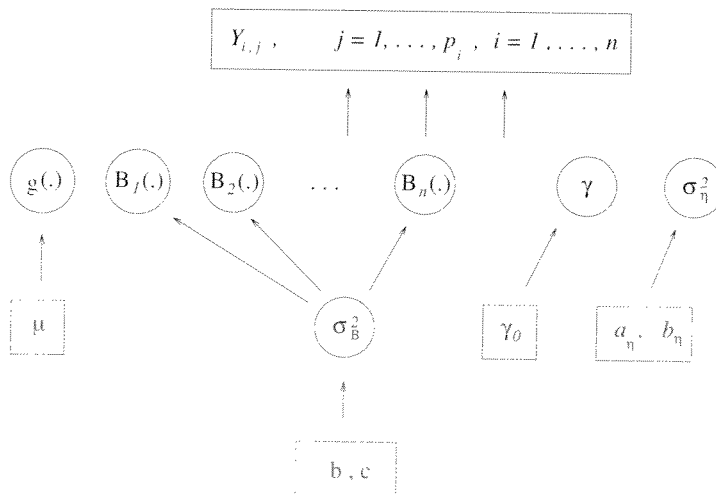


Fig. 1. Hierarchical structure of the model

parameters in the graph are surrounded by circles, and data and given constants by squares. The collection of all variables can be partitioned into four “levels”. In Figure 1 the levels are ordered vertically, and connections between them are indicated by arrows. The variables on each level are conditionally independent given all the variables from which there are arrows pointing at them.

3. A Markov Chain Monte Carlo method

Let $\pi(x | y)$, where $x = (x_1, \dots, x_p) \in \mathcal{X}$, denote the posterior distribution of parameter x , given (fixed) data y . Often $\pi(x | y)$ is known up to a multiplicative constant. We shall be mainly interested in evaluating expectations $E_\pi(f(x))$ of function f under π . In particular, such expectations include probabilities of events under π , in which case f is an indicator function.

It is well known that numerical Markov chain Monte Carlo (MCMC) methods can often be used successfully in practice in connection with large hierarchical Bayesian models, for approximating integrals with respect to the desired distribution (see ROBERTS and SMITH (1994), also SMITH and ROBERTS (1993)). If $x^{(1)}, x^{(2)}, \dots$ denotes a realization of the chain, under mild conditions, ergodicity implies that

$$\frac{1}{m} \sum_{i=1}^m f(x^{(i)}) \xrightarrow[\text{a.s.}]{} E_\pi(f(x)) \quad \text{as } m \rightarrow \infty. \quad (3.1)$$

In our model we have the following vector of parameters $x = (g, B_k, \gamma, \sigma_\eta^2, \sigma_B^2, k = 1, \dots, n)$, and data (Y_1, \dots, Y_n) . To start the algorithm we first generate $x^{(0)}$ from the prior. Then we move from $x^{(0)}$ to $x^{(1)}$, updating each coordinate from the posterior distribution when the other coordinates are fixed at their current values. In the general step of the Metropolis-Hastings algorithm, $x^{(i)}$ is replaced by $x^{(i-1)}$, where

$$x^{(i)} = (g^{(i)}, B_k^{(i)}, \gamma^{(i)}, (\sigma_\eta^2)^{(i)}, (\sigma_B^2)^{(i)}, k = 1, \dots, n).$$

The parameters can be updated one at a time, in a way which corresponds to the hierarchy displayed in Figure 1: For example, first parameters from level 1, then from level 2, etc. Any other order is equally possible. Under mild conditions, the Markov chain $x^{(1)}, x^{(2)}, \dots$, obtained in this way is ergodic with the posterior distribution of x as the limit distribution (see ROBERTS and SMITH (1994)).

In the practical execution of the algorithm, after discretization, the increments $\Delta g(l) = g(t_l) - g(t_{l-1})$ are used as model parameters, instead of the value $g(t_l)$. One advantage of this is that the $\Delta g(l)$'s are independent of each other under the prior distribution. Although we can obtain the $\Delta g(l)$'s and the $g(t_l)$'s from each other by simple telescoping, from a computational point of view the two approaches are quite different. In the latter, we just update the segment values $g(t_l)$

one by one, while in the former, after a new simulated value of $g(t)$ has been obtained, all other values beyond this one will be shifted up or down by the same amount. In this sense, the former approach is thought to have better mixing properties. The same reasoning applies to the B_k 's.

It is not easy to draw a new value for $\Delta g(l)$ directly from its conditional distribution. For this reason we use the Metropolis-Hastings algorithm, drawing a proposal for a new $\Delta g(l)$ from its prior distribution. The corresponding Hastings ratio will then coincide with the likelihood ratio corresponding to the new and the old value of $\Delta g(l)$. A similar procedure is used to update the increments $\Delta_k(l) = B_k \left(\frac{l}{N} T_{\max} \right) - B_k \left(\frac{l-1}{N} T_{\max} \right)$ of the B_k 's (see BESAG et al. (1995)).

For updating σ_{η}^2 and σ_B^2 , the conjugacy property of the normal density and the IG prior is available. Actually, it is easy to show that, given all other parameters, the posterior density of σ_{η}^2 is

$$IG \left(a_{\eta} + \frac{1}{2} \sum_{k=1}^n p_k, b_{\eta} + \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^{p_k} (Y_{k,j} - g^{(i)}(A_k^{(i)}(t_{k,j})))^2 \right),$$

where we have denoted $A_k^{(i)}(t) = \int_0^t h(W_k^{(i)}(s)) ds$. Similarly, the conditional density of σ_B^2 is

$$IG \left(b + \frac{nNT_{\max}}{2}, c + \frac{N}{2} \sum_{k=1}^n \sum_{j=1}^{NT_{\max}} (\Delta_k^{(i+1)}(j))^2 \right)$$

(see GELFAND and SMITZ (1990)).

Finally, to update γ , a proposal is first generated from its prior distribution, and then the corresponding Hastings ratio is calculated from the corresponding likelihood ratio.

As described in the introduction, we are mainly interested in estimating quantities of the form $f_n(\tilde{t})$, the true size of individual n at time \tilde{t} . Notice that given $x^{(i)}$, $f_n^{(i)}(\tilde{t}) = g^{(i)}(A_n^{(i)}(\tilde{t}))$ is uniquely determined. Taking f to be

$$f(x) = 1_B(g(A_n(\tilde{t})))$$

in (3.1), where B can be any Borel set contained in R^+ , we get the following Monte Carlo approximation

$$P(f_n(\tilde{t}) \in B \mid Y_1, \dots, Y_n) \approx \frac{1}{m} \sum_{i=1}^m 1_B(f_n^{(i)}(\tilde{t})). \tag{3.2}$$

Consequently, a simple method for obtaining the approximate predictive distribution of $f_n(\tilde{t})$ is to calculate the values $f_n^{(i)}(\tilde{t})$ at the end of each iteration of the program, and take the sets B simply to be intervals.

In a similar fashion, we can study the time it takes for an individual to grow to a certain target level Y_n^* . Denote such a time by t_n^* ; again it is a function of x . By

calculating the time $t_n^{*(i)}$ from the corresponding simulated $x^{(i)}$, we obtain its predictive distribution approximately by

$$P(t_n^* \in B \mid Y_1, \dots, Y_n) \approx \frac{1}{m} \sum_{i=1}^m 1_B(t_n^{*(i)}), \quad (3.3)$$

where B is again a Borel set in R^+ .

In case one wants to predict the value of a possible future measurement $Y_{n,\bar{i}}$, the measurement error should be taken into account. Since $Y_{n,\bar{i}}$, given $x^{(i)}$, is a Gaussian random variable with known mean and variance, the predictive distribution can be expressed as

$$\begin{aligned} P(Y_{n,\bar{i}} \in B \mid Y_1, \dots, Y_n) &= \int_x P(Y_{n,\bar{i}} \in B \mid x) d\pi(x \mid Y_1, \dots, Y_n) \\ &\approx \frac{1}{m} \sum_{i=1}^m P(Y_{n,\bar{i}} \in B \mid x^{(i)}). \end{aligned} \quad (3.4)$$

Usually there would be little direct interest in such probabilities if compared to (3.2); however, (3.4) can be useful in model assessment (see section 4).

We can also try to predict the growth curve(s) of one or more ‘‘hypothetical’’ individuals, about which we have no data. This can be done in practice by simply adjusting the number of individuals to include the hypothetical one(s), and by setting the corresponding p_n to be 0.

4. An example: the growth of children

Finally, we consider the empirical problem of predicting the future growth of small children from earlier observations. We do not claim that the numerical results derived here contribute anything new to the vast substantive knowledge concerning human growth; rather, we want to use this as an illustration of the potential of the proposed techniques in the context of irregularly behaving and irregularly measured growth curves.

The data set was a random subsample from a larger cohort, describing the growth of children born in Northern Finland during 1985 and 1986. The sample consists of 86 children, all boys. The data are given in the appendix, and in a graphical form in Figure 2. The maximum period of the follow-up was 753 days after birth. The height of each child was measured at the time of birth, and then 2–13 times afterwards. The time intervals between measurements varied, but typically more frequent measurements were made when the child was younger.

To fit our model to the data, we set the date of conception as the time origin, so that each child under investigation would begin to grow from height 0. As an approximation, the duration of pregnancy was assumed to be same, 280 days for

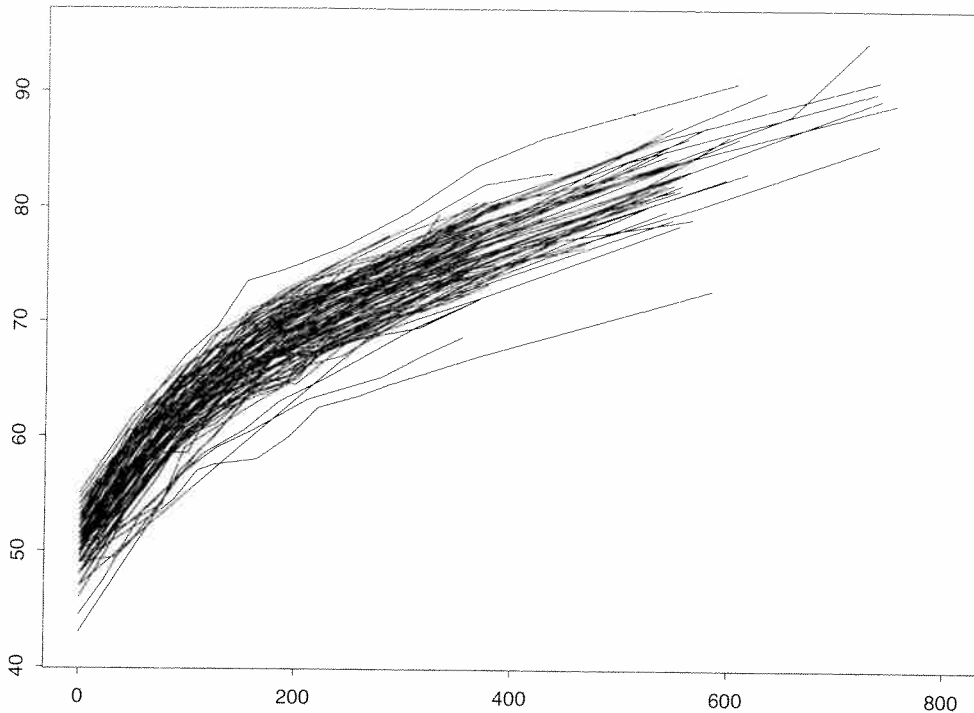


Fig. 2. Growth measurements of 86 boys

all children. In fact, the numerical estimates are rather insensitive to the exact choice of such a time origin. We used 30 days as a time unit, and T_{\max} was taken to be 35. The interval $[0, T_{\max}]$ was divided into $N = 105$ subintervals for discretisation of B_k . To approximate g , we chose $C = 2$, so that $[0, 2T_{\max}]$ was divided into $N_0 = 210$ subintervals.

Since the true durations of the pregnancies were unknown, they could in principle be treated as missing data, and then be estimated together with other unobservables by using Bayesian data augmentation methods. This was not done here, however.

The hyperparameters were given the values: $\mu = 0.4$, $b = 3$, $c = 0.2$, $a_\eta = 3$, $b_\eta = 2$ and $\gamma_0 = 0.5$. As mentioned before, these particular choices had little influence on the empirical results which were obtained.

There are many possible candidates for the link function h . Here we chose it to be twice the logistic link function, $h(x) = 2 \frac{\exp(x)}{1 + \exp(x)}$, so the "speed" of growth ranges in $(0, 2)$, and the "initial speed" is always $h(0) = 1$ (since $W_k(0) = 0$ by assumption).

The program was coded in C and run in a Sun Ultra 1 workstation. A few hundred iterations seemed enough for the estimates to reach reasonable stability. Actually, in each trial we made 4500 iterations in total, ignoring the first 1500 to

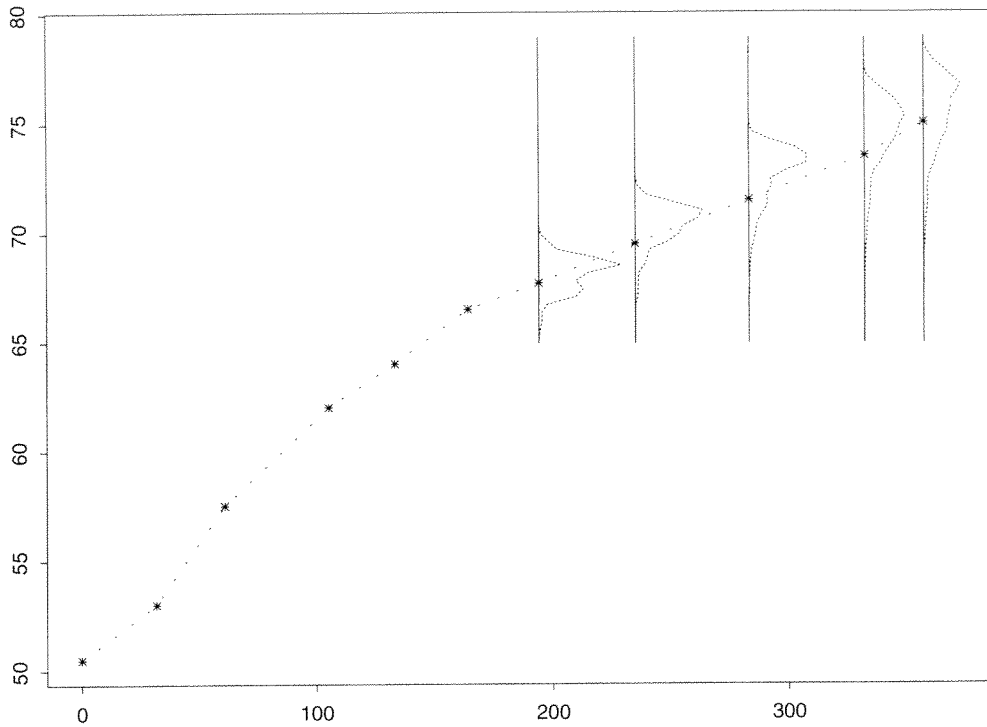


Fig. 3. Predictive distributions of heights at times 194, 235, 283, 332 and 357. Asterisks (*) are real measurements of height. Only the first six measurements were pretended to be known and were used in the prediction

account for burn-in and using the remaining 3000 in the estimation. For the purpose of making predictions, growth curves were sampled at each iteration of the algorithm, as described in section 3.

Our first task was provide predictive distributions of future values $f_n(\tilde{t})$, as shown in (3.2). This was done to the boy indexed by 86, about whom there were eleven height measurements in our data set. Pretending that we only know the first six measurements made at $t = 0, 32, 61, 105, 133$ and 164 days, we used these, as well as the complete data on the remaining 85 boys, to get the simulated $x^{(i)}$. The length $f_{86}(\tilde{t})$ was then predicted for $\tilde{t} = 194, 235, 283, 332$ and 357 days respectively, which were the ages at which the five deleted length measurements were made. In Figure 3 we plotted their predictive distribution densities, with * indicating the real measurements. The pretended last observation is $Y_{86}(164) = 66.5$. For example, it can be seen that, according to our model, at $\tilde{t} = 194$ there is a small (approximately 0.01) probability that the true length $f_{86}(194)$ is in fact less than the measurement 66.5 a month earlier. The explanation is that $Y_{86}(164) = 66.5$ contains a random measurement error, and there is a small probability that this error is positive and actually exceeds the true growth during the next month.

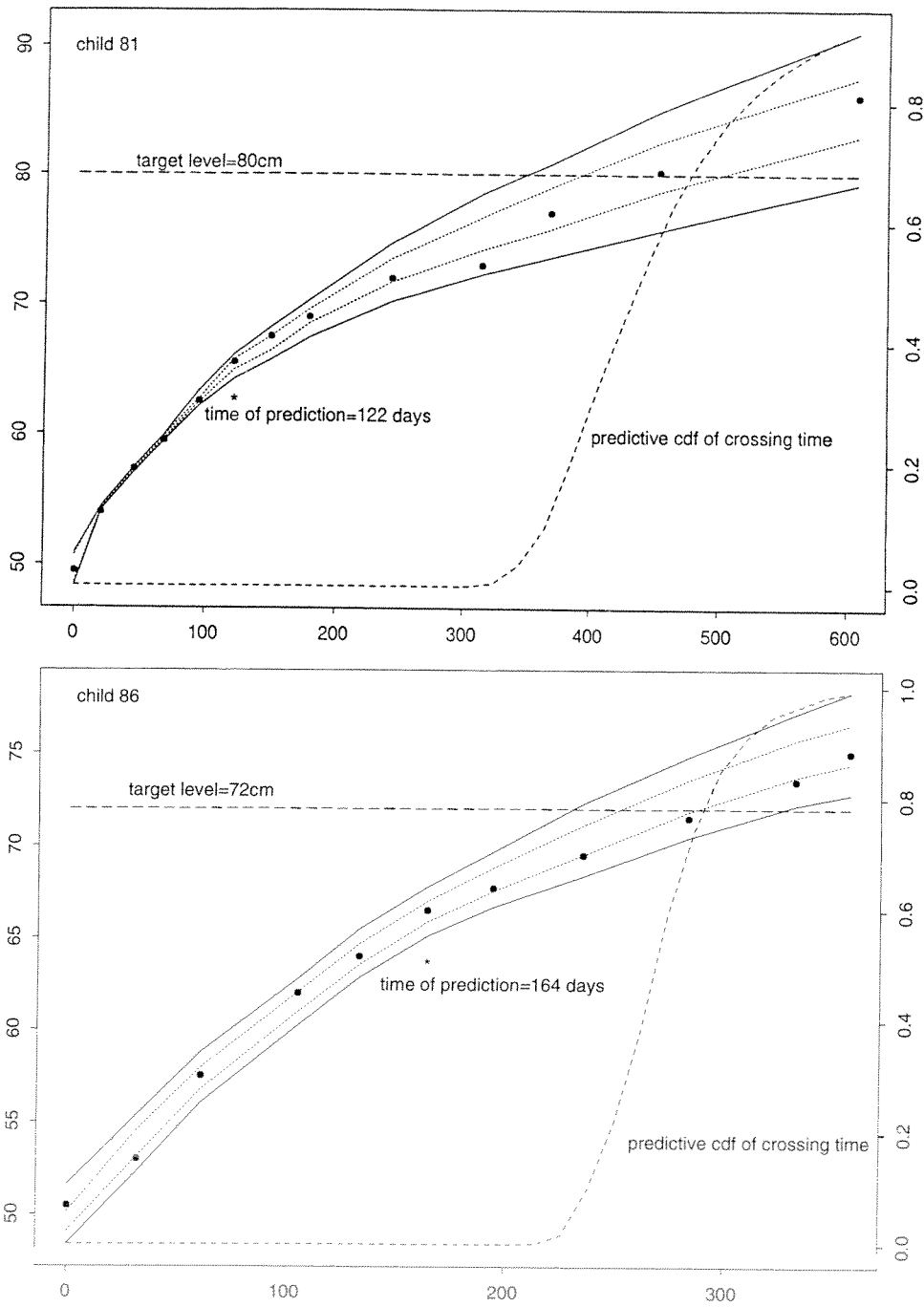


Fig. 4. Prediction of growth and of crossing time of target level. The lines connect the pointwise 90 percent (solid lines) and 50 percent (dotted lines) credible intervals determined at the times of measurement

We also determined the symmetric 50 percent and 90 percent Bayesian point-wise credible intervals for the boys indexed by 81 and 86. These are displayed in Figure 4. The credible intervals were calculated at times where real measurements were made, and their end points were connected with splines. (Note that they do not form credible bands over time for the entire growth curve). The individual measurement data were used up to the times marked by *; the data beyond that are only used for comparison and were not involved in the prediction. These intervals become naturally wider as the prediction time becomes more distant from time *.

We then considered the predictive distribution of the time at which a child reaches a certain target height. Looking again at the child indexed by 86, we wanted to predict, at the age of 164 days, the age at which he would reach the height of 72 cm. A similar prediction was made for child 81, now at the age of 122 days, and choosing the target height to be 80 cm. The results, based on (3.3) and given in the form of predictive cumulative distribution functions, are shown in Figure 4. Note that the former prediction is much more accurate than the latter for an obvious reason: child 86 was older, and the target was much closer to the present height.

In order to make a somewhat more systematic assessment of the predictive performance of the model, a cross-validation technique was used. We compared some real measurements in the data with their corresponding predictive distributions, obtained by pretending that they were not available to us. Since we have only 86 children in the data set, in order to make better use of it, we run the program many times, always making the predictions for a different subset of children. In each trial we considered 10 children (once 6 children), forming for each the predictive distributions for one to seven measurements which had been deleted from the data, using the earlier measurements on these same children and the complete data on the others, and comparing these predictions with real observations. Altogether 406 predictive distributions were formed in this way. A simple check showed that 56.6 (91.9) percent of the actual measurements fell inside the corresponding 50 (90) percent credible intervals, which indicates reasonable agreement with the nominal values. Unfortunately, the fact that the prediction errors for any one child tend to be correlated over time prevents us from carrying out a more formal significance test.

5. Discussion

The growth curve model described here is both nonparametric and highly non-linear. As a consequence, it is difficult to summarize the empirical results from a modelling experiment and data analysis, let alone do it in the usual form of point estimates and confidence intervals. Rather, what we hope to have accomplished here is an ability to produce probability predictions, in a numerical form, for future observables. We believe that similar methods could also be used with other

growth problems, subject to minor modifications of the model. For example, we have studied the possibility of using a somewhat simpler variant of this general model as a description of crack growth in fatigue problems.

There are also obvious ways in which this model could be extended. For example, had our data set contained observations from both boys and girls, we could have assumed that σ_b^2 and σ_g^2 are common to all, but replaced g by two random functions, one for boys and the other for girls.

Finally, it should be stressed that the present model is really only suitable for describing individuals whose growth at the end of the observation interval is not close to saturation. This is because differences between individuals are only accounted for by differences between the corresponding operational times $A_k(t)$, and asymptotically they all tend to $+\infty$ as $t \rightarrow +\infty$. According to the model, therefore, if $T_{\max} = +\infty$, all individuals would grow towards the common asymptote at $g(+\infty)$. If we wanted to consider growth to "adulthood", in which case the growth of an individual typically first slows down and then stops completely but where fully grown individuals end up being of different sizes, the model would have to be extended further. One way of doing this would be by modulating the function g by individual multiplying factors.

Acknowledgements

We are grateful to Prof. Paula Rantakallio for permission to use the data, and to Mr. Markku Koironen and Mr. Tommi Härkönen for their help. This research was supported by grants from the Academy of Finland.

Appendix. Growth data of Boys

Appendix A. Times of measurement

0 39 53 88 121 145 179 204 250 303 361
 0 61 97 129 164 193 370 570
 0 34 47 78 110 125 157 201 227 270 317 362 598
 0 26 51 74 94 123 153 186 226 277 324 367 513 753
 0 61 99 127 158 213 257 299 365 545
 0 58 100 129 157 177 216 294 331
 0 32 136 165 193 235 275 368 562
 0 47 70 97 125 161 199 230 277 327 363 546
 0 59 72 87 118 133 154 184 212 272 364
 0 28 60 97 128 160 188 222 299 560
 0 44 74 162 191 231 305 378 540 738
 0 30 57 89 120 155 190 225 257 306 358 547

0 57 360
0 44 76 89 124 154 160 243 305 377 431 615
0 35 62 92 126 156 183 244 308 372 435
0 53 96 124 160 195 249 307 362
0 46 73 95 136 164 201 231 313 368
0 55 97 131 166 195 257 386
0 76 97 182 209
0 94 244 353
0 34 66 93 127 160 210 279 315 353
0 59 88 120 155 184 246 359
0 51 94 122 330
0 73 87 91 125 156 175 238 273 324 365
0 51 96 129 160 192 227 276 339 376
0 124 150 206 272 369
0 94 168 303 366 486
0 47 94 125 156 184 219 271
0 42 95 219 368
0 61 91 138 171 199 250 297 356 540
0 19 39 64 86 122 146 176 213 242 373 540
0 47 70 104 131 166 197 306 377 540
0 63 92 124 168 203 237 268 379
0 58 93 126 157 183 211 240 274 352 373 525
0 48 91 125 157 192 198 223 274 318 366 554
0 62 93 124 152 180 242 275 367
0 63 91 157 192 221 340 559
0 68 86 125 152 183 210 240 384
0 33 63 90 120 152 188 215 273 335 377 655 727
0 89 120 150 183 244 362
0 363 547
0 92 126 151 190 220 298 367 555
0 91 122 153 185 216 313
0 52 172 200 255
0 69 95 143 174 208 248 304 394
0 38 67 103 152 184 214 258 297 363 550
0 95 232 268 577
0 46 84 123 159 195 242 299
0 46 86 117 151 183 214 284
0 45 77 100 126 165 178 206 280 317 366 568 737
0 39 61 95 128 163 211 289 544 739
0 35 170 245 367 595
0 31 43 94 162 254 388
0 64 92 122 166 242 302 373
0 60 89 131 165 202 258 295 348 540

0 75 88 109 124 164 194 220 257 285 368 583
 0 36 63 93 126 154 187 252 303 365 559
 0 23 59 93 115 152 184 226 309 359
 0 43 72 92 107 135 163 212 251 308 343 365 520
 0 33 61 95 131 144 186 299 382 632
 0 46 91 186 222
 0 29 61 97 134 183 215 256 279 315 356
 0 30 53 63 140 176 216 295 378 504 735
 0 89 123 153 184 212 276 313 369
 0 55 98 127 179 208 272 365
 0 38 45 56 71 93 126 176 220 553
 0 47 61 79 93 131 162 197 250 313 398 565
 0 59 88 118 148 189 245 311 382
 0 17 66 104 145 178 209 245 292 360
 0 66 96 129 156 187 205 233 265 318 425
 0 21 41 77 96 127 162 194 232 274 335 432
 0 53 81 120 151 180 210 263 319 369 531
 0 60 91 127 158 186 244 275 362
 0 55 91 127 190 239 295 364
 0 18 95 125 188 234 266 312 370
 0 16 25 51 67 94 127 179 210
 0 43 91 121 153 184 223 279 329 391
 0 59 93 125 155 185 216 246 302 373
 0 43 69 91 122 153 181 211 262
 0 33 48 97 126 154 188 242 301 363 426 606
 0 20 45 68 95 122 150 179 243 314 368 454 607
 0 39 97 123 151 181 221 270 305 370 438
 0 55 92 132 160 196 223 237 275 309 339 370 409 464
 0 24 49 77 106 162 198 234 276 310 366 555
 0 66 96 126 158 188 214 250 279 367
 0 32 61 105 133 164 194 235 283 332 357

Appendix B. Measured Heights

52.0 58.5 60.0 65.5 68.0 70.0 71.0 75.2 76.5 79.0
 50.5 56.5 61.3 64.0 66.7 69.0 78.0 84.5
 49.0 49.5 50.7 54.5 58.7 60.0 63.7 65.0 68.3 70.0 71.0 73.0 86.2
 54.0 57.5 60.0 61.5 64.5 67.5 68.5 71.0 73.0 75.0 77.0 80.0 83.0 89.0
 52.5 60.0 64.0 66.0 69.0 71.0 74.5 76.0 77.5 87.0
 53.0 61.0 65.0 68.0 68.0 69.3 72.0 74.2 79.4
 50.0 55.0 64.4 66.4 69.0 70.5 73.6 77.0 83.2

48.0 54.0 57.0 60.0 61.0 64.0 67.0 67.5 69.5 72.0 73.0 79.0
 52.5 62.0 63.0 64.5 66.0 67.0 69.6 70.0 73.0 74.2 79.0
 52.0 57.2 61.0 65.0 69.0 70.0 71.5 72.5 77.0 86.0
 51.0 58.0 62.5 70.0 72.0 74.5 78.0 80.5 86.0 91.0
 52.0 55.5 58.5 63.5 64.0 66.5 69.0 69.7 71.5 74.0 75.5 82.0
 50.0 59.0 76.5
 51.5 57.0 60.0 62.5 64.0 65.5 66.0 70.5 73.5 75.0 77.3 83.0
 52.0 58.0 62.5 64.5 67.0 71.0 72.0 75.5 78.5 82.0 83.0
 51.0 57.5 60.7 64.5 67.0 70.5 73.5 75.0 78.5
 49.0 53.0 56.5 60.0 62.5 64.5 65.5 68.0 69.5 72.0
 53.5 61.0 63.5 67.0 70.0 71.0 74.0 80.5
 50.0 54.0 59.0 64.0 66.5
 49.0 61.5 71.5 77.0
 46.0 51.5 54.0 56.6 59.0 60.5 63.2 65.2 67.0 68.7
 49.0 57.0 60.0 64.0 66.0 67.0 70.0 72.5
 55.0 62.0 66.0 67.5 75.5
 51.0 62.0 61.5 61.8 65.6 67.1 68.5 73.0 73.5 74.5 76.5
 50.0 54.4 60.6 62.5 64.2 65.2 68.0 70.0 71.7 73.3
 50.5 62.2 63.5 67.0 69.7 75.5
 51.0 62.5 67.0 73.7 76.5 80.0
 49.0 54.5 59.9 62.5 66.0 67.0 67.0 70.0
 48.0 53.5 59.5 67.5 75.0
 50.5 58.0 60.0 63.0 65.0 64.5 68.5 70.5 73.0 79.7
 51.5 52.0 56.0 59.5 60.5 64.0 65.0 66.0 68.0 69.5 74.3 82.0
 51.0 60.0 62.5 65.0 67.5 69.5 71.0 76.0 78.8 84.5
 53.5 62.0 65.0 65.4 70.0 72.5 73.5 75.0 79.0
 50.0 56.7 60.8 63.5 64.5 66.3 67.5 69.0 70.5 74.0 76.0 81.5
 50.5 56.0 61.0 62.5 63.5 64.7 66.0 67.5 69.0 70.5 72.0 78.5
 43.0 52.0 57.0 60.0 63.0 67.5 68.0 69.0 74.0
 51.5 59.0 62.7 64.5 68.5 70.0 74.5 83.0
 53.0 61.0 63.8 65.7 67.5 69.0 69.8 70.5 75.5
 51.0 56.0 60.5 62.5 65.0 68.5 71.0 72.0 73.5 77.0 78.0 88.0 94.4
 51.5 63.7 66.0 67.5 70.5 73.0 78.0
 47.0 76.5 81.5
 48.0 59.5 62.0 63.0 64.4 67.0 71.0 75.0 82.0
 51.5 62.0 65.0 67.0 69.0 70.5 75.0
 49.5 58.0 68.0 68.5 71.0
 54.0 61.5 65.0 69.3 71.0 72.0 74.0 76.0 79.5
 52.0 56.5 59.0 63.0 67.5 69.5 71.5 73.8 76.0 78.0 84.0
 50.5 62.5 69.5 72.5 87.0
 52.0 59.0 62.0 64.0 66.5 68.5 70.7 73.8
 51.0 60.0 62.5 67.0 70.0 72.5 74.0 77.5
 51.0 56.0 58.5 58.5 63.1 65.3 66.7 68.5 71.0 72.5 74.2 80.0 85.5

47.0 53.0 55.5 60.5 63.0 66.5 69.0 74.0 82.5 89.4
 49.0 54.0 67.0 69.0 75.0 82.5
 48.0 51.0 53.5 60.0 65.0 70.0 76.2
 48.0 56.5 59.0 61.5 65.0 67.0 70.5 74.2
 50.5 60.0 63.0 66.0 69.0 71.0 75.0 76.0 85.0
 49.0 53.5 54.5 57.0 57.5 58.0 60.0 62.5 63.5 64.5 67.0 72.8
 50.5 56.0 59.0 62.5 65.0 68.0 70.0 72.0 74.0 77.0 83.0
 44.5 47.5 53.5 56.5 58.5 60.5 63.5 65.0 69.5 71.6
 52.0 56.5 60.0 61.0 64.0 65.0 65.5 67.0 69.0 71.0 73.0 73.5 80.0
 52.0 57.0 61.0 65.0 67.0 69.9 71.5 76.0 80.0 90.0
 48.0 55.0 61.0 68.0 71.5
 51.0 54.0 58.0 61.4 65.6 69.1 70.0 71.6 71.6 73.5 75.0 77.5
 52.0 58.0 59.0 61.5 68.0 70.5 72.0 74.5 78.0 84.0 90.0
 50.0 61.5 63.5 64.8 67.2 68.3 71.2 72.8 74.3
 50.5 61.0 66.0 68.0 72.0 73.5 75.0 80.4
 53.0 59.0 60.5 60.5 60.5 63.0 65.0 68.0 71.0 81.5
 47.0 54.0 56.0 58.0 59.5 62.3 66.0 67.7 71.0 73.8 76.5 79.0
 50.0 57.5 60.5 64.0 66.0 67.0 70.5 74.0 75.0
 49.5 53.5 58.0 62.0 66.5 67.0 68.8 70.5 71.7 76.0
 49.0 57.0 61.5 63.5 67.0 69.0 69.5 69.5 70.5 72.0 74.0 79.0
 49.5 54.0 57.0 60.0 62.0 65.5 67.5 69.0 71.7 73.0 75.0 77.5
 48.0 55.0 58.5 62.0 66.5 66.9 70.0 73.0 76.5 79.7 86.0
 53.0 60.7 65.0 66.5 68.0 71.1 73.1 75.0 78.5
 51.5 59.2 63.0 65.5 70.0 72.4 73.7 76.5
 52.0 52.0 63.8 65.5 69.0 71.5 73.0 75.0 77.5
 52.5 56.0 56.5 60.0 61.0 63.5 66.5 69.0 71.0
 52.0 59.5 65.0 68.6 70.5 71.9 72.5 75.0 77.8 80.5
 53.0 60.2 64.5 66.8 70.0 72.0 73.5 74.8 76.3 80.5
 51.0 59.5 61.0 61.9 65.5 68.0 70.0 71.0 74.0
 54.5 59.0 61.0 67.0 69.5 73.5 74.5 76.5 79.5 83.5 86.0 90.8
 49.5 54.0 57.3 59.6 62.5 65.5 67.5 69.0 72.0 73.0 77.0 80.2 86.0
 50.5 56.0 62.0 63.5 65.0 67.0 67.5 70.0 73.0 74.0 77.0
 49.0 57.1 61.0 64.5 66.0 67.5 69.5 70.0 71.6 72.5 74.0 74.5 75.5 76.5
 51.5 54.7 57.0 62.5 63.0 68.5 71.0 72.0 74.5 76.5 78.5 84.0
 50.0 59.5 62.7 65.0 70.0 72.0 73.5 73.8 78.0
 50.5 53.0 57.5 62.0 64.0 66.5 67.7 69.5 69.5 71.5 73.5 75.0

References

- ARJAS, E. and GASBARRA, D., 1994: Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4**, 505–524.
 BARRY, D., 1995: A Bayesian model for growth curve analysis. *Biometrics* **51**, 639–655.
 BERKEY, C. S., 1982: Bayesian approach for a nonlinear growth model. *Biometrics* **38**, 953–961.

- BESAG, J., GREEN, P., HIGDON, D., and MENGERSON, K., 1995: Bayesian computation and stochastic systems. *Statistical Science* **10**, 3–66.
- BOULARAN, J., FERRE, L., and VIEU, F., 1994: Growth curves: a two-stage nonparametric approach. *Journal of Statistical Planning and Inference* **38**, 327–350.
- FEARN, T., 1975: A Bayesian approach to growth curves. *Biometrika* **62**, 89–100.
- GASSER, T., MULLER, H.-G., KOHLER, W., MOLINARI, L., and PRADER, A., 1984: Nonparametric regression analysis of growth curves. *The Annals of Statistics* **12**, 210–229.
- GEISSER, S., 1970: Bayesian analysis of growth curves. *Sankhyā A* **32**, 53–64.
- GEISSER, S., 1980: Growth curve analysis. In: P. R. Krishnaiah (ed.): *Handbook of Statistics 1: Analysis of Variance*. North-Holland, Amsterdam, 89–115.
- GEISSER, S., 1981: Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. *Biometrika* **68**, 243–250.
- GELFAND, A. E. and SMITH, A. F. M., 1990: Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.
- GRIZZLE, J. E. and ALLEN, D. M., 1969: Analysis of growth and dose response curves. *Biometrics* **23**, 357–381.
- KNEIP, A. and GASSER, T., 1992: Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics* **20**, 1266–1305.
- KHATRI, C. G., 1966: A note on MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics* **18**, 75–86.
- LANGE, N., 1992: Graphs and stochastic relaxation for hierarchical Bayes modelling. *Statistics in Medicine* **11**, 2001–2016.
- LEE, J. C. and GEISSER, S., 1972: Growth curve prediction. *Sankhyā A* **34**, 393–412.
- LEE, J. C. and GEISSER, S., 1975: Applications of growth curve prediction. *Sankhyā A* **37**, 239–256.
- LEE, J. C., 1982: Classification of growth curves. In: P. R. Krishnaiah and L. N. Kanal (eds.): *Handbook of Statistics 2: Classification Pattern Recognition and Reduction of Dimensionality*. North-Holland, Amsterdam, 121–137.
- LEE, J. C., 1988: Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association* **83**, 432–440.
- LINDLEY, D. V. and SMITH, A. F. M., 1972: Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–42.
- LISKI, E. and NUMMI, T., 1990: Prediction in growth curve models using the EM algorithm. *Computational Statistics & Data Analysis* **10**, 99–108.
- POTTHOFF, R. F. and ROY, S. N., 1964: A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.
- RAO, C. R., 1965: The theory of least squares when the parameters are stochastic and its application to the analysis of the growth curves. *Biometrika* **52**, 447–458.
- RAO, C. R., 1977: Prediction of future observations with special reference to linear models. In: P. R. Krishnaiah, (ed.): *Multivariate Analysis II* Academic Press, New York, 87–103.
- RAO, C. R., 1984: Prediction of future observations in polynomial growth curve models. In: *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions*. Indian Statistical Institute, Calcutta, 512–520.
- RAO, C. R., 1987: Prediction of future observations in growth curve models. *Statistical Science* **2**, 434–471.
- ROBERTS, G. O. and SMITH, A. F. M., 1994: Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* **49**, 207–216.
- SHOHOJI, T., KANEFUJI, K., SUMIYA, T., and QIN, T., 1991: A prediction of individual growth of height according to an empirical Bayesian approach. *Annals of the Institute of Statistical Mathematics* **43**, 607–619.
- SMITH, A. F. M. and ROBERTS, G. O., 1993: Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **55**, 3–23.

of time, and that even then it is difficult to attach a biological meaning to their coefficients. Therefore, in order to allow for more flexibility and a more widely applicable approach, it seems desirable to consider nonparametric or semi-parametric models. Furthermore, if the emphasis is on the model's ability to predict the future growth of some particular (set of) individuals, it is attractive to look for a solution where the uncertainties relating to such growth are expressed probabilistically. This gives a practical justification for adopting an essentially nonparametric Bayesian approach in the inferential problem, and for formulating such probabilities explicitly in terms of predictive distributions.

The literature concerning statistical modelling of growth is vast. A good overview is given in TANNER (1989). The literature has been dominated by generalized multivariate analysis-of-variance models originally introduced by POTTHOFF and ROY (1964), and subsequently studied e.g. by RAO (1965, 1977, 1984, 1987), KHATRI (1966), GRIZZLE and ALLEN (1969), GEISSER (1970, 1980, 1981), LEE and GEISSER (1972, 1975), FEARN (1975), LEE (1982, 1988, 1990), LISKI and NUMMI (1990), among others. These models are basically following the parametric linear models tradition, assuming that an individual's growth is a polynomial function of time (typically, age) but that the coefficients can be specific to a group of such individuals. Both non-Bayesian and Bayesian inferential methods have been used in the estimation of such models: for the latter, see e.g. GEISSER (1970) and LEE (1982). To give an example of nonlinearly parametrized growth curve models, BERKEY (1982) developed an empirical Bayes approach for fitting the Jenss growth curve model. In comparison, the nonparametric approach to growth curve modeling has been much less used. GASSER et al. (1984) discussed certain advantages of nonparametric models over the parametric ones, mostly in the context of human growth, while KNEIP and GASSER (1992) modelled the structural pattern common to all members in a sample of curves, proposed some estimators and discussed their asymptotic behaviour. RAO (1987) discussed direct and inverse regression methods under nonparametric growth curve models, using Bayesian and empirical Bayesian methods. TIAN et al. (1994) proposed the use of restricted cubic splines, and compared this with the methods of RAO (1987). BOULARAN et al. (1994): developed a two-stage nonparametric method, splitting the growth curve into two additive parts, one representing a common structure of all individuals, and the other being specific to each individual in the population. BARRY (1995) assumed that the growth curves are Gaussian, specifying a prior distribution such that the curves vary slowly over time, and developed a test of a homogeneity hypothesis. To our knowledge, the present paper is the first to combine a nonparametric modelling approach of growth curves with full Bayesian estimation and prediction.

We now outline the questions considered in this paper. Suppose we observe the growth of n individuals from the same population, making measurements of some particular characteristic, such as height or weight, over a period of time. Let $Y_{k,i}$ denote a measurement made on individual k at time $t_{k,i}$. Denoting the observation interval by $[0, T_{\max}]$, we have measurements of the first individual at times

Prediction of Growth: A Hierarchical Bayesian Approach

ELJA ARJAS

Rolf Nevanlinna Institute
University of Helsinki
Finland

LIPING LIU

Department of Probability and Statistics
Peking University
P.R. China

NIKO MAGLAPERIDZE

Georgian Statistical Association
Tbilisi
Republic of Georgia

Summary

A nonparametric hierarchical growth curve model is proposed. Different levels in the model hierarchy are intended to correspond to different sources of variation in an individual's growth. The nonparametric character of the model offers considerable flexibility in fitting the growth curves to empirical data. Here the emphasis is on prediction, and for this purpose the adopted Bayesian inferential approach seems particularly natural and efficient. A Markov chain Carlo method is used to perform the numerical computations. As an illustration of the techniques, we consider the growth of children, during their first two years.

Key words: Growth curve; Hierarchical modelling; Predictive distribution; Bayesian credible interval.

1. Introduction

The prediction of an individual's growth, based on its earlier size measurements and on data describing the growth of other "similar" individuals, forms an interesting challenge to a statistician. Parametric models considered in the literature, which moreover are often assumed to be of a polynomial form, tend to be specific to some particular species or populations. It also seems to be the general experience from such models that relatively simple models apply only over a short range