PERGAMON Mathematical and Computer Modelling 31 (2000) 217-225

MATHEMATICAL
AND
COMPUTER
MODELLING

www.elsevier.nl/locate/mcm

Computer Modeling of Dynamically Changing Distributions of Random Variables

E. O. VOIT
Department of Biometry and Epidemiology
Medical University of South Carolina, P.O. Box 250551, Charleston, SC 29425, U.S.A.

A. SORRIBAS
Departament de Ciències Mèdiques Bàsiques
Universitat de Lleida, 25198 Lleida, Spain

Abstract—Physiological characteristics of populations are often described with two types of data and with analyses that coexist in mutual independence. One type targets global measures like mean and standard deviation, prevalence, incidence, or distributions of weights, sizes, or some physiological markers, while the other type describes physiological or metabolic processes in individuals, such as growth or the accumulation of a metabolite. The paper describes a methodological framework for bridging the gap between the two aspects. It shows how a dynamic model that characterizes the accumulation process in individuals and is represented by a set of ordinary differential equations can be reformulated as a transformation function that changes the statistical distribution of the physiological marker over time. The method and its limitations are illustrated with data describing growth and size distributions of loblolly pine (*Pinus taeda*). © 2000 Elsevier Science Ltd. All rights reserved.

Keywords—Dynamical model, Statistical distribution, Growth, Trend, Computational analysis.

INTRODUCTION

The design of a mathematical model usually starts with two considerations: should the model be discrete or continuous, and should it be deterministic or stochastic. The answers to these questions lay the ground rules for the types of functions and for the bag of computational tools that can be used for analysis. In particular, the decision of whether the model should be deterministic or stochastic has immediate and far-reaching consequences, since a deterministic model usually leads to differential equations, whereas a stochastic model usually falls in the domain of statistics.

Of course, there is overlap. Stochastic systems can be rather predictable—one only has to think of ideal gases in which molecules follow complicated, stochastic paths, while the overall behavior of the gas follows the quite simple, deterministic rules that are known as "gas laws." On the other hand, some completely deterministic systems can yield unpredictable, "chaotic" behaviors that are best analyzed with statistical means.

Typical approaches for dealing with true combinations of stochastic and deterministic effects are encountered in the area of time series and stochastic processes (e.g., [1,2]). The stochastic

---

The authors are grateful to W. Harms and the USDA Forest Service, Southeastern Forest Experiment Station, Charleston, SC for making the tree growth data available for this and other analyses. Initial phases of this work were supported by a grant from the USDA Forest Service.

0895-7177/00/$ - see front matter Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX
PII: S0895-7177(00)00041-8

component is often modeled as white or colored noise, while the deterministic time trend component may be an algebraic function, a difference equation, or a differential equation. In cases of simple algebraic trends, the preferred strategy of analysis is separating the deterministic component from the superimposed random effects and analyzing the random effects with specially developed statistical methods. In the case of differential equations, randomness can be incorporated in parameters, the initial or boundary values, or in random functions, which may model fluctuating inputs. The result is a stochastic differential equation, whose analysis can be quite difficult (for further discussion, see, e.g., [3]).

The present paper also falls in the intersection between deterministic and statistical modeling, but it takes an approach that is different from standard time series analysis. Of interest here is a situation where a random variable is subject to a potentially complicated trend which is given as a system of nonlinear ordinary differential equations. The question asked is how the distribution of the random variable, which is known at some initial time point, changes over time with respect to mean, variance, and shape. It is demonstrated that, by reformulating the problem, stochastic differential equations can be avoided, and the distributional time trends can be evaluated with an analysis of ordinary differential equations instead.

The method will be illustrated below with the prediction of size distributions of trees in an even-aged stand. Given the size distribution of young trees and the growth function for the species, the method yields projections of size distributions of older trees. As a rather different example that could be analyzed with the proposed method, consider a group of workers subjected to an incident of radiation exposure. Over the years, the amount of radiation decreases, and under simplifying assumptions, this decrease can be expected to be some monotonic function of time. Studying a cohort of these workers at some point in time, one will most likely find a distribution of radiation, with some workers showing lower and other workers higher values. The questions one can ask in the present context are: "what will the distribution of radiation be several years in the future?" or "what was the distribution of radiation shortly after the incidence?"

The two examples have in common that a deterministic process drives the development of a probability or frequency distribution. If the process is mathematically simple, the problem is well known in statistics, even though it is usually presented with the probability function as the focus. The statistical textbook approach (e.g., [4]) studies the distribution of a random variable which itself is a function of another random variable with a known distribution. The illustrative example given by Mood *et al.* [4, p. 200] is a logarithmic transformation of a beta distribution with parameters a and b ($b = 1$), which results in an exponential distribution with parameter a. In general, if the original random variable is denoted as X and if the new random variable Y is a differentiable, monotonic function $Y = \varphi(X)$ with inverse $\varphi^{-1}(Y)$, then the density of Y is

$$f_Y(y) = \left| \frac{d}{dy}\varphi^{-1}(y) \right| f_X\left(\varphi^{-1}(y)\right). \tag{1}$$

Comparing this result to the question posed in our context, the relationship between distributions does not explicitly show dynamic features. To introduce these, consider how the dynamical process, during a given, fixed time period τ, shifts values of the random variable X to new values of Y. For instance, if the application is the growth of trees according to a known, deterministic growth process, select a time period, such as one year, and study how much each tree of the cohort grows within this year. Evidently, the growth function maps the size distribution of year 1 into the size distribution of year 2; each member of the original distribution f_X has become a unique member of the shifted distribution f_Y. The shift constitutes some function $\varphi(X)$, which is not a function of time *per se*, even though time, implicitly through the magnitude of the shift, determines the numerical characteristics of the function (see Figure 1). Under appropriate adaptations, the distribution of Y can therefore be computed from the distribution of X according to transformation (1), presented above.

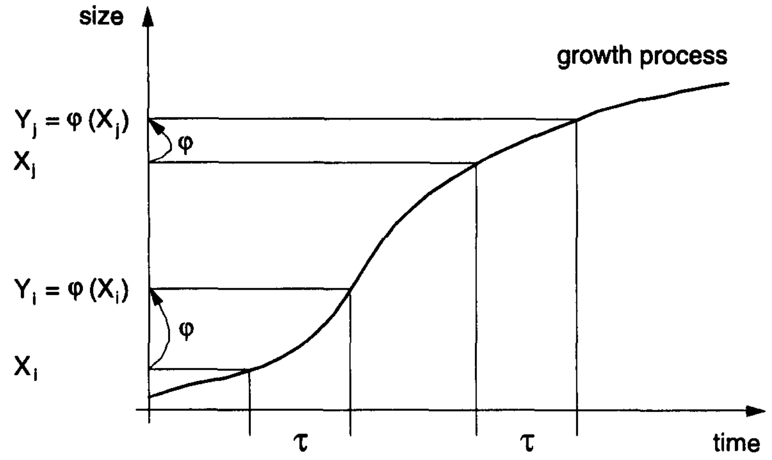

Figure 1. The function φ, which implicitly depends on the magnitude of the time shift τ, maps all admissible values of the random variable X = "size at time t_1" onto corresponding values of the new random variable Y = "size at time t_2."

If the dynamic process is given as a set of differential equations, the problem is the same in concept, but it is not obvious what the function $\varphi(X)$ is. To explore this question, let us begin with the simple example of a dynamic process that is described by a single differential equation

$$\dot{X} = \Psi(X), \qquad X_0 = X(t_0). \tag{2}$$

After a time period of τ time units, the value of the state variable X at time t has changed from $X(t)$ to $X(t+\tau)$, and this is true for any admissible value $X(t)$ and any admissible time shift τ. In particular, if $X(t)$ is the random variable of interest at time t, then the corresponding random variable τ time units later is $Y = X(t+\tau)$ (see Figure 1). The new variable Y satisfies exactly the same differential equations, with the only exception that its time scale is shifted by τ time units. For instance, we can write

$$\dot{Y} = \Psi(Y), \qquad Y_0 = Y(t_0) = X(t_0+\tau). \tag{3}$$

When we solve equations (2) and (3) simultaneously, we obtain two copies of the growth process, shifted by τ time units.

The transformation of the distribution of X into the distribution of Y, according to equation (1), requires the term $\frac{dX}{dY}$, if one recalls $X = \varphi^{-1}(Y)$. This term is readily obtained from equations (2) and (3) as

$$\frac{dX}{dY} = \frac{dX}{dt}\frac{dt}{dY} = \frac{\Psi(X)}{\Psi(Y)}. \tag{4}$$

Equation (4) constitutes a new differential equation in which Y is the independent variable and t is eliminated. It describes how X changes as a function of Y. Multiplication of the term $\Psi(X)/\Psi(Y)$ with the probability density $f_X(x)$ thus yields $f_Y(y)$. Equation (4) is initialized with $Y(t_0)$ as the initial value of the independent variable and $X(t_0)$ as the value of the state variable.

It is convenient in this context to formulate the density as the solution of a differential equation. This is possible in numerous ways, and only two are mentioned here. As a first alternative, densities can be recast exactly as systems of differential equations in which the first variable is the density or cumulative of interest and the other variables are auxiliary variables that are generated in the recasting procedure (e.g., [5,6]). The second alternative is an approximate representation in the form of a so-called *S-distribution*. In this distribution, the dependent variable is the cumulative and the independent variable is the random variable. The S-distribution has the form

$$f = \frac{dF}{dX} = \alpha\left(F^g - F^h\right), \qquad F(X_0) = F_0, \tag{5}$$

where α is a positive parameter that characterizes the spread, g and h ($g < h$) are real-valued shape parameters, and the initial value F_0 determines the location. Several articles about this distribution have appeared in the recent literature [7–10], and it suffices to state here that the S-distribution closely approximates most of the traditional distributions.

If the original random variable is S-distributed, and if it changes dynamically according to the process in equation (2), the shifted random variable has the distribution

$$\frac{dF}{dY} = \alpha\left(F^g - F^h\right)\frac{\Psi(X)}{\Psi(Y)}, \qquad F(Y_0) = F_0 \tag{6}$$

(cf. equation (1)), which is solved simultaneously with the equations describing the shift, namely,

$$\begin{aligned} \frac{dX}{dY} &= \frac{\Psi(X)}{\Psi(Y)}, & X_0 &= X(t_0), \\ \frac{dY}{dY} &= 1, & Y_0 &= Y(t_0) = X(t_0 + \tau). \end{aligned} \tag{7}$$

The same procedure holds if the distribution and the dynamic process are modeled by systems of differential equations. If the variable of interest is X_1, then the corresponding variable of the shifted system is Y_1. Simultaneous solution of both systems produces in X_1 and Y_1 two copies of the growth process, shifted by τ. Division of the entire system by the equation for Y_1 yields the required term $\frac{dX_1}{dY_1}$. The technical details of this general procedure are given elsewhere, along with some illustrative examples [11].

ANALYSIS OF TREE GROWTH

Between 1962 and 1989, the USDA Southeastern Forest Experiment Station in Charleston, SC measured growth characteristics of trees in even-aged stands of different planting densities. In 1989, Hurricane Hugo destroyed these plantations, thereby precluding further measurements. A small subset of the data is used as an illustration of the shift method described above. This subset contains the measurements of stem diameters at breast height (DBH) of trees with ages six through 26 in a plantation of 64 trees, which corresponds to a density of 1,000 trees per acre. Only the 25 trees surviving throughout the observation period are considered in this illustrative analysis. Their growth curves are shown in Figure 2. It is noted that the diameters of some trees at age six are zero or close to zero, which is due to the fact that these trees had not, or only barely, reached breast height at age six.

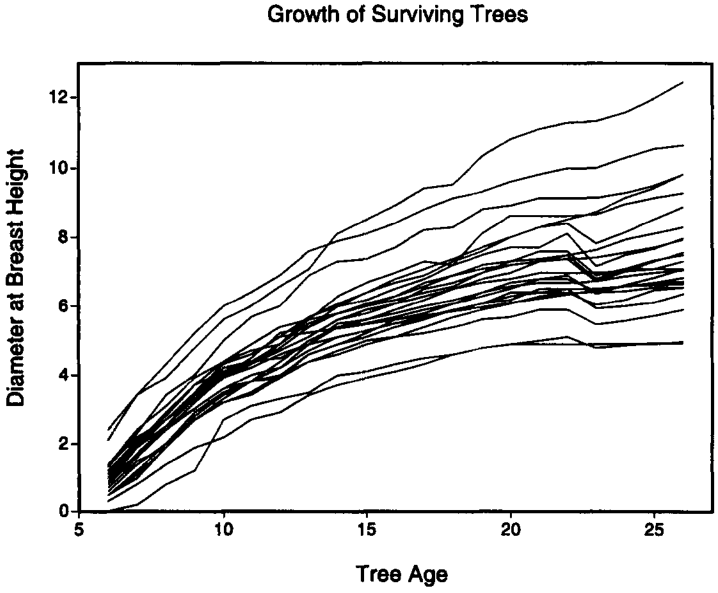

Figure 2. Growth curves of all 25 trees surviving throughout the observation period.

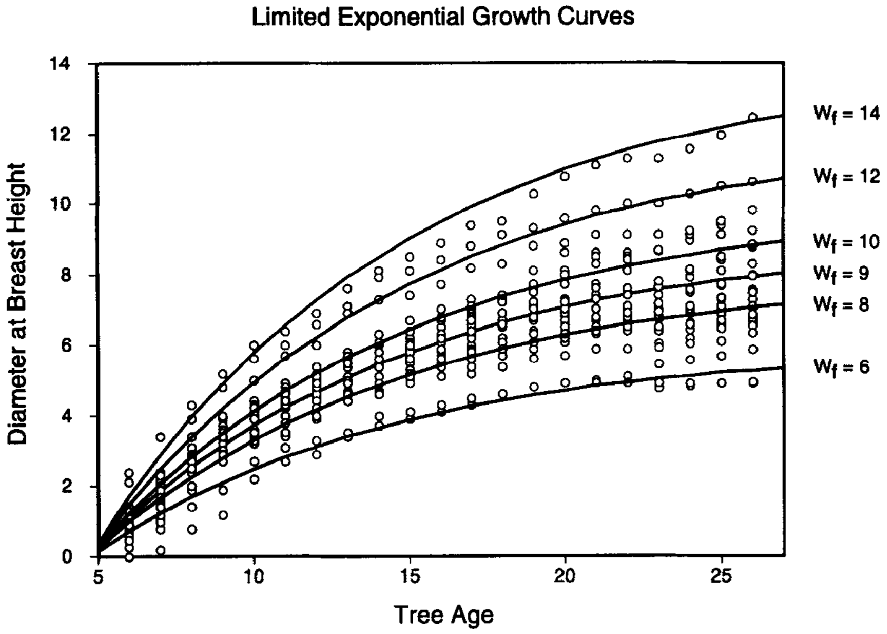

Figure 3. DBH measurements of all 25 trees surviving throughout the observation period, superimposed with limited exponential growth curves (see text for details).

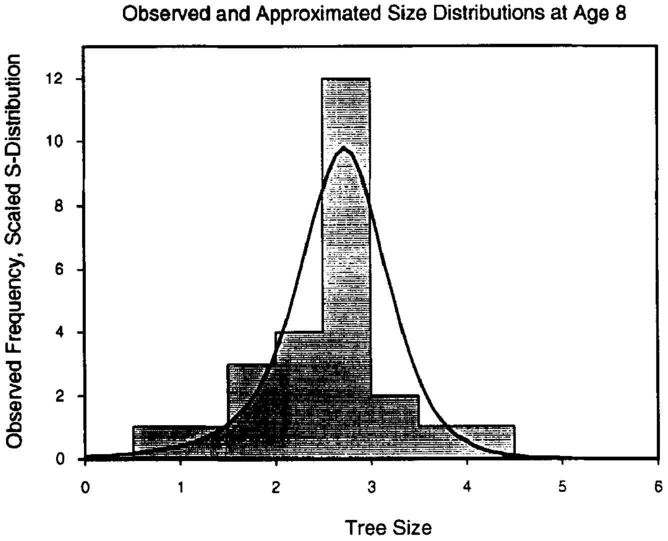

Figure 4. DBH distribution at age six, superimposed with an S-distribution model. For comparison, the S-distribution is scaled.

After investigating several alternatives, it was decided to formulate the underlying growth process with the so-called *limited exponential* growth function, which in differential form (e.g., [12]) reads

$$\dot{W} = k\,(W_f - W)\,, \qquad W(t_0) = W_0. \tag{8}$$

This function captures the decreasing growth rate and has only three parameters. A coarse parameter estimation with the program CurveFit 7.0 [13] suggested an "average" growth function with the parameters $W_f = 9$, $k = 0.1$, and the initial value $W_6 = W(t = 6) = 1.097$ at age six. The function is shown for several values of W_f in Figure 3, superimposed on the observed data.

The size distribution was estimated for age eight, when all trees have a positive DBH, as an S-distribution with the parameters $\alpha = 17$, $g = 1.5$, $h = 1.7$, and $F(0) = 0.0075$ (Figure 4). The differential equations describing the desired shifts are thus given as

$$\begin{aligned} \frac{dF}{dY} &= 17\left(F^{1.5} - F^{1.7}\right)\frac{k\left(W_f - W\right)}{k\left(W_f - Y\right)}, && F(Y_0) = F_0, \\ \frac{dW}{dY} &= \frac{k\left(W_f - W\right)}{k\left(W_f - Y\right)}, && W_0 = W(t_0), \\ \frac{dY}{dY} &= 1, && Y_0 = Y(t_0) = W(t_0 + \tau) \end{aligned} \tag{9}$$

(cf. equations (6) and (7) with the notation $W = X$).

To execute the shift operations, the initial value for the random variable W was taken as the value of the growth function for age eight, and the value of the shifted random variable Y for ages nine, ten, 11, etc., and also for ages seven, six, and five. In order to obtain the entire distribution of the shifted variable, the system was solved twice, once in positive direction and once with the independent variable running towards zero. The results exhibit shifts in the locations of distributional peaks that are in excellent agreement with the observations (Figure 5; data not shown).

However, it is obvious that the predicted distributions have systematically smaller variances than the observed distributions. In fact, the predicted variances decrease over time, while the observed variances increase. This is quite interesting since the type of the observed growth functions necessarily leads to decreases in variance. In general, one can show that monotonic functions with increasing slopes (such as an exponential growth process) yield increasing variances, whereas functions with decreasing slopes, as is the case here (cf. Figure 3), lead to distributions with decreasing variances (cf. [11]). Since both the growth process and the distributions are directly obtained from the same data set, and since the shift method is a direct mathematical extension of an established statistical method, this discrepancy must result from some of the assumptions or simplifications that underlie the proposed method or the dynamical growth model.

The most likely culprit is the considerable variability among the growth functions. Studying the individual growth patterns in Figure 2, it is evident that some trees are growing much faster than others. In particular, the size increments, given a particular size, do not exclusively depend on the size but are themselves distributed, thereby violating the assumption that the growth process is appropriately described by the ordinary differential equation above. There is no easy way to remedy this problem. Since the trees in the study are even-aged, age itself cannot be the source of size variation. In fact, from a biological standpoint, all trees in the plantation are as similar as can be achieved in a field study of the given type. Competition between neighboring trees, individual genetic make-up, and intraspecies variation are some of the likely causes, but these are difficult to capture in a simple model.

One alternative to constructing a more comprehensive model would be to abandon the proposed method entirely and to construct a stochastic model in which a tree of a given size would grow by a certain increment with a certain probability which could presumably be estimated from data as in Figure 1. Another alternative is to apply the shift method to the initial size distribution many times, each time using a slightly altered growth function. For the example under investigation, variation in the final size parameter W_f results in growth curves that include the curve used before, but also slower growing and faster growing trees (see Figure 3). By weighting, according to observed frequencies of occurrence, and then adding the distributions that result from shifting the original distribution by means of these different growth functions, one obtains an impression of the impact of variability in growth parameters on the trend in distributions. In some sense, this procedure corresponds to separately shifting subsamples from the original distribution that are characterized by the same growth parameter W_f. As an example, Figure 6 compares the observed tree sizes with the distribution that results from the shift method with superposition. It is noted

Figure 5. The DBH distribution at age eight is shifted toward other ages, using a limited exponential growth function with an average final value $W_f = 9$.

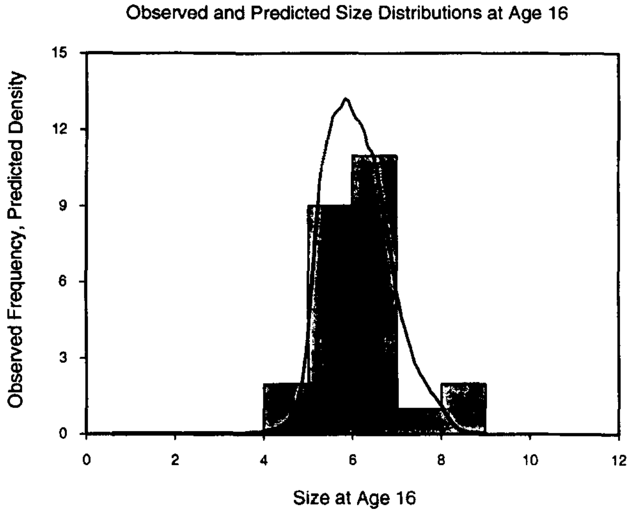

Figure 6. The DBH distribution at age eight is shifted toward age 16, using limited exponential growth functions that differ in their final values ($W_f = 6,8,10,12,14$). The resulting densities are weighted according to approximate occurrence, added and multiplied with 25, in order to facilitate comparison with the observed distribution of 25 sizes (shaded area).

that the resulting density is not quite smooth, because only five cohorts were superimposed. A finer division would result in a smoother appearance.

DISCUSSION

If a random variable accumulates or deteriorates in a more or less deterministic and monotonic fashion, the method described here can be used to predict with relative ease how the distribution of a random variable changes location, spread, and shape over time. The method is a numerical generalization of the well-established statistical technique of transforming distributions of random variables and merely requires an algorithm for solving ordinary differential equations. If the accumulation process does not show too much variation in itself, the method appears to be reliable. In particular, the method can be used for theoretical studies, including best case and worst case analyses and the simulation of what-if scenarios. While the method may at first seem to be cumbersome, it has been implemented in its entirely in a relatively simple Mathematica program which requires specification of the accumulation process, the original distribution, and the shift interval, and subsequently computes entire trends in a few minutes.

The method has been applied here to a data set in which both the growth phenomenon and the distributions over several years were available. This data set provided a good opportunity for illustrating the method as well as its limitations. Another example that has been analyzed elsewhere is the accumulation of mercury in king mackerel (*Scomberomorus cavalla*); in this case, distributions were available but the accumulation function itself was not [11]. The method has also been used to demonstrate that dynamic shifts can reverse the skewness of distributions and that it can transform unimodal distributions into bimodal distributions, a fact that has been observed in diverse contexts (e.g., [14–17]; see also [11]).

As any method, the shift method has limitations. Most significant may be the situation in which the accumulation process in itself shows variability that cannot be ignored. In such a situation, the method still predicts the trend in modes, but tends to underestimate the spread in the shifted distributions. One approach of dealing with this situation is executed here. It yields results similar to those observed and allows us to separate variability in growth parameters from the overall trend caused by the growth process itself.

If an application is suited for the proposed method, the method has certain advantages over standard techniques of time series analysis. It is very flexible with respect to the distribution of the random variable and with respect to the time trend. In fact, except for monotonicity, there is hardly a requirement on the system of differential equations that describes the time trend or the initial distribution. Yet, the entire analysis can be executed with widely available algorithms for solving ordinary differential equations. For simple trends with superimposed white noise, this method is probably overly complicated, but in situations that do not fall in any of the typical classes of trends and stochasticity, the proposed method appears to be a good alternative.

REFERENCES

1. E. Parzen, *Stochastic Processes*, Holden-Day, Oakland, CA, (1962).
2. P.J. Diggle, *Time Series Analysis: A Biostatistical Introduction*, Oxford Statistical Science Series No. 5, Oxford, (1991).
3. L. Arnold, *Stochastic Differential Equations. Theory and Applications*, Wiley and Sons, New York, (1974).
4. A.M. Mood, F.A. Graybill and D.C. Boes, *Introduction to the Theory of Statistics*, 3rd edition, McGraw-Hill, Tokyo, (1983).
5. M.A. Savageau, A suprasystem of probability distributions, *Biometrical J.* **24**, 323–330 (1982).
6. E.O. Voit and P.F. Rust, *S*-system analysis of continuous univariate probability distributions, *J. Stat. Comp. Simul.* **42**, 187–249 (1992).
7. E.O. Voit, The *S*-distribution. A tool for approximation and classification of univariate, unimodal probability distributions, *Biometrical J.* **34** (7), 855–878 (1992).
8. E.O. Voit, W.L. Balthis and R.A. Holser, Hierarchical Monte Carlo modeling with *S*-distributions: Concepts and illustrative analysis of mercury contamination in king mackarel, *Enviroment International* **21**, 627–635 (1995).
9. E.O. Voit and S. Yu, The *S*-distribution. Approximation of discrete distributions, *Biometrical J.* **36**, 205–219 (1994).
10. S. Yu and E.O. Voit, A simple, flexible failure model, *Biometrical J.* **37**, 595–609 (1995).
11. E.O. Voit, Dynamic trends in distributions, *Biometrical J.* **38**, 587–603 (1996).
12. M.A. Savageau, Growth equations: A general equation and a survey of special cases, *Math. Biosci.* **48**, 267–278 (1980).
13. K. Raner, CurveFit Version 7.0; for information, email to `Kevin@carbon.chem.csiro.au` (1992).

14. E.D. Ford, Competition and stand structure in some even-aged plant monocultures, *J. Ecology* **63**, 311–333 (1975).
15. C.L. Mohler, P.L. Marks and D.G. Sprugel, Stand structure and allometry of trees during self-thinning of pure stands, *J. Ecology* **66**, 599–614 (1978).
16. M. Kimura, Changes in population structure, productivity and dry matter allocation with the progress of wave regeneration of *Abies* stands in Japanese subalpine regions, In *Carbon Uptake andAllocation in Subalpine Ecosystems as a Key to Management*, (Edited by R.H. Waring), Forest Research Laboratory, Oregon State University, (1982).
17. D.J. Gates, R. McMurtrie and C.J. Borough, Skewness reversal of distribution of stem diameter in plantations of *Pinus radiata*, *Australian For. Res.* **13**, 267–270 (1983).