Modelling stand development with stochastic differential equations *

Oscar García†

1 Introduction

This is a progress report on the development of a general methodology for producing stand models. The methodology must not be understood as a package of computer programs which are fed with data to automatically produce a growth model. Instead, it tries to be a coherent set of ideas and techniques intended to help in the design and implementation of sound models. An intelligent use of these techniques still requires a considerable dose of skill and common sense.

The methodology consists essentially of a general approach to modelling, a class of stand models, and procedures for the estimation of parameters. General applicability is considered to decrease in this same order. The approach to modelling is believed to be essential for any kind of growth models. The models proposed, while being fairly flexible, are by no means the solution to all modelling problems. The estimation procedures are specific to the class of models already mentioned, and even then, are only one among several alternatives.

The first part of the paper develops these ideas, starting with an introduction to some concepts from System Theory, within the context of stand modelling. The second part illustrates some of the ideas with partial results from tests with a small data set. In order to preserve continuity in the presentation, some technical details, extensions, comments and other

*Pp. 315–333 *in* Elliott, D A. (*Comp.*) Mensuration for Management Planning of Exotic Forest Plantations. *New Zealand Forest Service, FRI Symposium No. 20.* 1979.

†Scientist, Forest Research Institute, N. Z. Forest Service, Rotorua

non-essential information has been collected in the *Notes* at the end of the paper.

2 Theory

2.1 The State Space Approach

The purpose of building a stand model is to predict future values of certain *outputs*, such as volumes of timber, resulting from given *inputs*, silvicultural treatments. Both inputs and outputs are functions of time. The outputs at a future time t depend not only on the inputs applied between an initial time t_0 and t, but they depend also on the *state* of the system at time t_0. "Roughly, a state of a system at any given time is the information needed to determine the behaviour of the system from that time on" (Zadeh, 1969). From a slightly different point of view, the outputs depend on the complete past history of inputs to the system. The state may then be "regarded as a kind of information storage or memory or an accumulation of past causes" (Kalman *et al.*, 1969), "some compact representation of the past activity of the system complete enough to allow us to predict, on the basis of inputs, exactly what the outputs will be, and also to update the state itself" (Padulo and Arbib, 1974).

To be more specific, we will assume that the values of the inputs and of the outputs at a given time t are sets of numbers collected in finite-dimensional vectors $\boldsymbol{u}(t)$ and $\boldsymbol{v}(t)$, respectively. In addition, the state will also be a finite-dimensional vector. The components of the state vector are called state variables. These restrictions are sufficient for most applications.

The behaviour of the system is described by a *transition function*

$$\boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t_0), \boldsymbol{u}, t - t_0) \ , \tag{2.1.1}$$

and an *output function*

$$\boldsymbol{v}(t) = \boldsymbol{g}(\boldsymbol{x}(t)) \ , \tag{2.1.2}$$

where $\boldsymbol{x}(t_0)$ is the state at time t_0, $\boldsymbol{x}(t)$ is the state at time $t \geq t_0$, $\boldsymbol{u}$ is the input (a vector function of time, with only the values for times between t_0 and t affecting the value of (2.1.1)), and $\boldsymbol{v}(t)$ is the value of the output vector at time t. In words, a future state is completely determined by an initial state, the elapsed time, and the values of the input during this time

interval. The output is a function of the current state. For the moment we are considering only deterministic systems; the extensions to probabilistic systems are discussed later.

For our stand model we might take as state variables, for example, the basal area, stocking (number of stems per hectare), and mean top height. In matrix notation, we will write the state as a column vector $\boldsymbol{x} = (B, N, H)'$, where B is basal area, N is stocking, and H is mean top height. Our inputs will be silvicultural regimes consisting of thinnings and possibly other treatments. Usually these treatments occur at discrete points in time, and their effect can be regarded as an instantaneous change in the state vector. We can then simplify the discussion by considering the development of the stand only between treatments. Thus, it is sufficient to model the stand as a *free* system, that is, a system with no inputs or, equivalently, a system with just one constant input.

Let us see how good the state vector $\boldsymbol{x} = (B, N, H)'$ might be for a stand model. (It is clear that any one-to-one transformation of $\boldsymbol{x}$ must also be regarded as an equivalent state vector). Basically, the state must describe the stand adequately for the purposes for which the model will be used. Adequately in the sense that two stands with equal states (on a given site) can always be regarded as practically equivalent in terms of their present condition and of their future behaviour. In other words, the state must determine, to a satisfactory degree of accuracy, both the future states, according to (2.1.1), and the outputs in which we are ultimately interested, through (2.1.2). The state vector $\boldsymbol{x} = (B, N, H)'$ seems adequate for estimating the outputs usually required for management purposes. Volumes for different products and standards of utilization may be estimated from $\boldsymbol{x}$ with stand volume equations, or with procedures such as the "stand volume generator" of Goulding and Shirley (1978). As a determinant of future behaviour, $(B, N, H)'$ may be considered as satisfactory if only moderate thinning intensities and pruning heights are used. After a very heavy thinning and/or pruning the amount of canopy remaining may not be enough to make full use of the site potential, at least temporarily. The growth would not be equal then to the growth of another stand which has reached the same B, N and H values following a different path. For a good prediction of the effects of heavy thinning and/or pruning it would seem necessary to include an additional state variable, such as the mean green crown level. It is easy to think of additional variables which might give a more complete description of the stand, such as various characteristics of the d.b.h. distribution, for

example. In general, the selection of a state vector must be a compromise between, among other things, the prediction requirements, on one hand, and the availability of data for fitting the model and the information necessary to use it, on the other.

Having discussed the selection of a state vector, we turn now to the transition function. Notice that the transition function (2.1.1) must satisfy some natural conditions:

(a) (consistency)

$\boldsymbol{F}(\boldsymbol{x}(t), \boldsymbol{u}, 0) = \boldsymbol{x}(t)$, for all times t, states $\boldsymbol{x}(t)$ and admissible input functions $\boldsymbol{u}$.

(b) (Composition or semigroup property)

$\boldsymbol{F}(\boldsymbol{F}(\boldsymbol{x}(t_0), \boldsymbol{u}, t_1 - t_0), \boldsymbol{u}, t_2 - t_1) = \boldsymbol{F}(\boldsymbol{x}(t_0), \boldsymbol{u}, t_2 - t_0)$, for any $t_0 < t_1 < t_2$.

That is, the result of projecting the state first from t_0 to t_1, and then from t_1, to t_2 must be the same as that of the "one go" projection from t_0 to t_2.

(c) (Causality)

$\boldsymbol{F}(\boldsymbol{x}(t_0), \boldsymbol{u}_1, t_1 - t_0) = \boldsymbol{F}(\boldsymbol{x}(t_0), \boldsymbol{u}_2, t_1 - t_0)$ if $\boldsymbol{u}_1(t) = \boldsymbol{u}_2(t)$ for $t_0 \leq t \leq t_1$.

The most practical way of specifying a transition function with these properties is through a *local transition function*.

We first make an important distinction between discrete-time systems and continuous-time systems. A *discrete-time* system is one in which the time parameter t takes only integer values. In a *continuous-time* system t can take any real values. It may seem natural to model a forest stand as a discrete-time system, defined only when t is an integral number of years. In a stand model we do not attempt to represent the seasonal pattern of growth, so that predictions are strictly valid only at a fixed date within a year, usually during the vegetative season. We will see, however, that there are certain advantages in developing the model as a continuous-time system, even if later in the applications we restrict it to integer time values.

In the discrete-time case the local transition function is completely straightforward. We simply specify how the state and input at some time

t determine the state at the next time $t+1$ on the discrete-time scale. We then observe how the "global" transition function (2.1.1) for any (integer) time interval can be obtained by repeated application of this local "one-step" description. Specifically, from (2.1.1), the state at time $t+1$ is given by

$$\boldsymbol{x}(t+1) = \boldsymbol{F}(\boldsymbol{x}(t), \boldsymbol{u}, 1) \ .$$

The effect of the input is only through its value at time t (cf. the causality property of $\boldsymbol{F}$), so that we can write the local transition function in the simpler form

$$\boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)) \ . \tag{2.1.3}$$

It is clear that this equation can be applied repeatedly to obtain the state at any future time, given an initial state and the input. In some instances it is also possible to obtain a closed analytical expression for the global transition function (2.1.1) from (2.1.3) (e.g., Goldberg, 1958; Miller, 1968). In any case, the behaviour of the system is completely determined by (2.1.3) and (2.1.2).

For continuous-time systems some mathematical technicalities are involved, but it is shown that, under some mild "smoothness" conditions on the system, a local transition function takes the form of a differential equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)) \tag{2.1.4}$$

(We use the notation $\dot{\boldsymbol{x}}$ for the time derivative $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$). The changes of state over any finite time interval, as given by (2.1.1) , can then be obtained by integration of (2.1.4). As in the discrete-time case, the system's behaviour is completely specified by (2.1.4) and (2.1.2).

Some words now about stochastic (or probabilistic) systems. In the deterministic systems discussed above, knowledge of the state $\boldsymbol{x}(t_0)$ and of the values $\boldsymbol{u}(t)$ of the input for $t_0 \leq t \leq t_1$ determine exactly the state $\boldsymbol{x}(t_1)$. In a stochastic system only the probability distribution of $\boldsymbol{x}(t_1)$ is determined. This leads us to the theory of Markov processes. Alternatively, we might fit stochastic processes into the standard deterministic System Theory by taking as the state the probability distribution of $\boldsymbol{x}(t)$, instead of $\boldsymbol{x}(t)$ (remember that the state can be any mathematical object, not necessarily a finite-dimensional vector).

Instead of using these approaches we will find more convenient to keep all the concepts about deterministic systems already mentioned, and to model

stochastic systems by allowing random inputs. That is, we select a state vector and specify the model by equations of the type (2.1.4) and (2.1.2) (or (2.1.3) and (2.1.2)), where $\boldsymbol{u}$ is a random process of given characteristics. We can also include random variables in (2.1.2), representing measurement errors.

It is worth mentioning that a restriction to deterministic models is not as strong as it might seem. It is clear that the evolution of the state of a forest stand would be represented more realistically by a stochastic system. However, the expected value or the most likely value or any other point estimate of the state vector would behave as the state of a deterministic system. It is theoretically possible to use the full probability distributions of the predictions for making decisions. But the state of our knowledge about forest stands and about the rest of the relevant aspects of a forest operation suggest that point estimates will continue to be used for some time. We will use a stochastic structure mainly for suggesting reasonable estimation procedures, but it is anticipated that the models obtained will be used mostly as deterministic models.

2.2 A Model

For concreteness, we select the state vector $\boldsymbol{x} = (x1, x2, x3)' = (B, N, H)'$, and consider a model for predicting the effects of thinning regimes. The extensions to other state vectors and other inputs is straightforward. A continuous-time model will allow us to deal with the effect of site index and of measurements taken at different dates during the year in simpler ways. Also the theory of differential equations is more developed than that of difference equations, or at least it is better known. A deterministic model is considered first.

We model then the development of a stand between thinnings as a free differential system (cf. 2.1.4):

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \; . \tag{2.2.1}$$

(from now on, as is usual, we simplify the notation by dropping the distinction between a function $\boldsymbol{x}$ and its values $\boldsymbol{x}(t)$ when there is no risk of confusion). The stand volume equations or stand volume generator (2.1.2) will not be considered as part of the stand model. The development of equations for estimating the instantaneous change of state produced by a

thinning, essentially the change in basal area due to a given change in stocking, is relatively simple and will not be discussed here.

It would be nice to use for (2.2.1) equations with a sound ecophysiological basis. Extrapolations outside the range of data are usually less dangerous this way. However, there are no satisfactory relationships of this type yet, specially so for describing natural mortality. Once we decide on using an empirical model, it is natural to use one which is mathematically convenient, as well as flexible enough for describing the observed development patterns. In particular, it is desirable that (2.2.1) could be integrated analytically.

Linear differential equations can be easily integrated, besides having some other properties which might be useful when using the model for simulation or optimisation. Flexibility can be increased by not using the state $\boldsymbol{x}$ directly in the differential equation, but some one-to-one transformation of $\boldsymbol{x}$ instead. A useful class of transformations takes the form

$$\begin{aligned} y_1 &= k_1 B^{c_{11}} N^{c_{12}} H^{c_{13}} \\ y_2 &= k_2 B^{c_{21}} N^{c_{22}} H^{c_{23}} \\ y_3 &= k_3 B^{c_{31}} N^{c_{32}} H^{c_{33}} \end{aligned} \tag{2.2.2}$$

Notice that these transformations also take out large part of the arbitrariness in selecting as states variables basal area and stocking, and not mean d.b.h. and spacing, for example. Many of the variables used in forestry are power functions of B, N and H of the type (2.2.2) (e.g., mean d.b.h., average spacing, relative spacing, Reineke's density index, volume when estimated by a logarithmic stand volume equation, etc.).

We propose then the model

$$\dot{\boldsymbol{y}} = A\boldsymbol{y} + \boldsymbol{b}\,, \tag{2.2.3}$$

where

$$\boldsymbol{y} = (y_1, y_2, y_3)'\,,$$
$$y_i = x_1^{c_{i1}} x_2^{c_{i2}} x_3^{c_{i3}}\,, \quad i = 1, 2, 3\,,$$

and A and $C = (c_{ij})$ are 3×3 matrices of coefficients and $\boldsymbol{b}$ is a three-dimensional vector of coefficients (the factors k_i in (2.2.2) are absorbed into A and $\boldsymbol{b}$). Any of the coefficients may be functions of site index.

The transformation may also be denoted more neatly as

$$\boldsymbol{y} = \boldsymbol{x}^C\,, \tag{2.2.4}$$

defining

$$\boldsymbol{x}^C = \exp(C \ln \boldsymbol{x}) \tag{2.2.5}$$

(the extension of scalar functions to vectors is understood in the standard way, as the vector of functions of the components: $\ln \boldsymbol{x} = (\ln x_1, \ln x_2, \ln x_3)'$). The inverse transformation can then be written in terms of the inverse of C as

$$\boldsymbol{x} = \boldsymbol{y}^{C^{-1}} \ .$$

We will impose the constraints $a_{31} = a_{32} = c_{31} = c_{32} = 0$. Then the third equation in (2.2.3) takes the form

$$\frac{\mathrm{d}H^{c_{33}}}{\mathrm{d}t} = a_{33} H^{c_{33}} + b_3 \ . \tag{2.2.6}$$

This is just one way of writing von Bertalanffy's growth model (Bertalanffy, 1949, 1957; Richards, 1959). This model has been very popular in recent years for describing height growth in the development of site index equations (e.g., Elliott and Goulding, 1976; Burkhart and Tennent, 1978). In a way, (2.2.3) might be regarded as a multivariate generalisation of Bertalanffy's model. This becomes clearer if we write it as

$$\frac{\mathrm{d}\boldsymbol{x}^C}{\mathrm{d}t} = A\boldsymbol{x}^C + \boldsymbol{b} \ . \tag{2.2.7}$$

Now we derive a stochastic model by introducing a random input, as suggested at the end of 2.1. The simplest way of doing this is by changing (2.2.7) to

$$\frac{\mathrm{d}\boldsymbol{x}^C}{\mathrm{d}t} = A\boldsymbol{x}^C + \boldsymbol{b} + B\dot{\boldsymbol{w}} \ , \tag{2.2.8}$$

where B is a 3×3 matrix and $\boldsymbol{w}$ is the standardised three-dimensional Wiener process (or Brownian motion, process). The relevant properties of this process are that the increments $w_i(t_2) - w_i(t_1) \quad (i = 1, 2, 3)$ are independent random variables, normally distributed with mean 0 and variance $|t_2 - t_1|$, and that increments over non-overlapping time intervals are independent (see for example Karlin, 1966).

The important assumptions here are that the random input (which may be thought as representing the effects of environmental variation and of the approximate nature of the model) enters additively in (2.2.8), and that the increments are independent. The matrix B accounts for correlation

between the components of the random process, and the proportionality of the covariance matrix to $|t_2 - t_1|$ is a consequence of the independence of the increments. It can be shown (Gihman and Skorohod, 1972, Theorem 1) that if we assume that $\boldsymbol{w}$ is continuous, the normality of the distribution follows.

It must be admitted that the assumptions of additivity and of independent increments are not very realistic. For one thing, there are climatic periodicities (Tomlinson, 1976). However, the stochastic structure of the model is intended only for the development of estimation procedures. General statistical experience suggests that in most cases the performance of estimators is not too affected by moderate deviations from the distributional assumptions.

2.3 Estimation

We are confronted now with the problem of estimating the parameters of the model using permanent sample plot data. The parameters are the coefficients of A, $\boldsymbol{b}$, C and B (or rather BB', see below). Some of the coefficients may actually be functions of the site index, containing other parameters to be estimated; more on this later.

The method of Maximum Likelihood (M.L.) provides a convenient methodology for this purpose. The M.L. estimators usually perform well, and some optimum asymptotic properties can be proven (assuming that the model used is the "true" one, whatever this means). More important, the method has two very useful characteristics. First, given a model and data, no matter how complicated the model, the method of M.L. specifies a well-defined procedure for estimating the parameters: we "simply" set up the likelihood function (the probability density for the observed data, considered as a function of the parameters) and select the values of the parameters for which the function has a maximum. Second, the following invariance property is true: if $\hat{\boldsymbol{\theta}}$ is a M.L.E. of $\boldsymbol{\theta}$ then $\boldsymbol{g}(\hat{\boldsymbol{\theta}})$ is a M.L.E. of $\boldsymbol{g}(\boldsymbol{\theta})$, for any function $\boldsymbol{g}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ (Zacks, 1971). This means that parameter transformations have no effect, and that any quantities computed from the model with the M.L. estimates substituted for the parameters will be M.L. estimates for those quantities.

To find the likelihood function we first make another assumption, that observations from different plots are statistically independent. This is an

approximation, since observations for different plots made in a same year will tend to be correlated. If the number of measurement periods is large enough we might expect the assumption to be acceptable.

We use pairs of successive measurements with no thinnings between them. Let $\boldsymbol{x}_1$ be the first measurement and $\boldsymbol{x}_2$ the second measurement for one of these pairs. Given the conditional probability density $f(\boldsymbol{x}_2|\boldsymbol{x}_1)$ of $\boldsymbol{x}_2$ given $\boldsymbol{x}_1$, from the assumptions of independent increments and of independence between plots it follows that the likelihood function is just the product of the $f(\boldsymbol{x}_2|\boldsymbol{x}_1)$ over all the pairs. It is more convenient to work with the logarithm of the likelihood function,

$$\ln L = \sum \ln f(\boldsymbol{x}_2|\boldsymbol{x}_1) \;, \tag{2.3.1}$$

where the sum is over all the pairs of successive measurements without thinnings between them.

To find $f(\boldsymbol{x}_2|\boldsymbol{x}_1)$, consider a measurement pair with $\boldsymbol{x}_1$ being the state at time t_1 and $\boldsymbol{x}_2$ the state at time t_2. According to our model, given $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ is obtained by integration of

$$\frac{\mathrm{d}\boldsymbol{x}^C}{\mathrm{d}t} = A(\boldsymbol{x}^C - \boldsymbol{a}) + B\dot{\boldsymbol{w}} \tag{2.3.2}$$

between t_1, and t_2 with initial condition $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$. Here $\boldsymbol{a} = -A^{-1}\boldsymbol{b}$, an asymptote or equilibrium point for $\boldsymbol{x}^C$. There is no loss of generality in assuming that A has distinct eigenvalues, so that it can be factorised as

$$A = P^{-1}\Lambda P \;, \tag{2.3.3}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \;,$$

the λ_i are the eigenvalues, and the rows of the 3×3 matrix P are the (left) eigenvectors of A. The eigenvalues are assumed to be real, since complex eigenvalues imply an oscillatory behaviour which is unacceptable for our system.

Writing

$$\boldsymbol{z} = P(\boldsymbol{x}^C - \boldsymbol{a}) \tag{2.3.4}$$

and $\Delta t = t_2 - t_l$, the integration of (2.3.2) is equivalent to integrating

$$\dot{\boldsymbol{z}} = \Lambda \boldsymbol{z} + PB\dot{\boldsymbol{w}} \tag{2.3.5}$$

between 0 and Δt, with the initial condition $\boldsymbol{z}(0) = P(\boldsymbol{x}_1^C - \boldsymbol{a})$. The solution takes the form

$$\boldsymbol{z}(\Delta t) = e^{\Lambda\,\Delta t}\boldsymbol{z}(0) + \boldsymbol{\epsilon}\ , \tag{2.3.6}$$

where, conditional on $\boldsymbol{x}_1$, $\boldsymbol{\epsilon}$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix

$$V = \int_0^{\Delta t} e^{\Lambda s} PBB'P' e^{\Lambda s}\,\mathrm{d}s \tag{2.3.7}$$

(Erickson, 1971).
Note:

$$e^{\Lambda s} = \begin{bmatrix} e^{\lambda_1 s} & 0 & 0 \\ 0 & e^{\lambda_2 s} & 0 \\ 0 & 0 & e^{\lambda_3 s} \end{bmatrix}\ .$$

Performing the integration in (2.3.7) we find

$$V = e^{\Lambda\,\Delta t} S e^{\Lambda\,\Delta t} - S\ , \tag{2.3.8}$$

where the elements of S are $s_{ij} = \sigma_{ij}/(\lambda_i + \lambda_j)$, and σ_{ij} are the elements of the matrix $\Sigma = PBB'P'$.

Since

$$\boldsymbol{\epsilon} = P(\boldsymbol{x}_2^C - \boldsymbol{a}) - e^{\Lambda\,\Delta t} P(\boldsymbol{x}_1^C - \boldsymbol{a})\ , \tag{2.3.9}$$

the probability density $f(\boldsymbol{x}_2|\boldsymbol{x}_1)$ can be obtained by multiplying the probability density of $\boldsymbol{\epsilon}$ by the absolute value of the Jacobian of the transformation (2.3.9). The Jacobian is the determinant

$$\begin{aligned} \left|\frac{\partial \boldsymbol{\epsilon}}{\partial \boldsymbol{x}_2}\right| &= |P| \left|\frac{\partial \boldsymbol{x}_2^C}{\partial \boldsymbol{x}_2}\right| = |P|\,|\operatorname{diag}(\boldsymbol{x}_2^C) C \operatorname{diag}(\boldsymbol{x}_2^{-1})| \\ &= |P|\,|C|\,|\operatorname{diag}(\boldsymbol{x}_2^{C-I})|\ . \end{aligned}$$

We finally get for the log-likelihood (2.3.1):

$$\begin{aligned} \ln L &= -\frac{1}{2}(np \ln 2\pi + \sum \ln |V| + \sum \boldsymbol{\epsilon}' V^{-1} \boldsymbol{\epsilon}) \\ &\quad + n[\ln \operatorname{abs}(|P| \cdot |C|)] + \mathbf{1}'(C - I) \sum \ln \boldsymbol{x}_2\ , \end{aligned} \tag{2.3.10}$$

where $\mathbf{1} = (1, 1, 1)'$, the sums are over all observation pairs, n is the number of observation pairs, $p = 3$ is the length of the state vector, V is given by (2.3.8), and $\boldsymbol{\epsilon}$ is given by (2.3.9).

The parameters for the third equation of the model, (2.2.6), are best estimated separately. Note that H alone is a perfectly adequate state vector when the outputs of interest are functions of H. The mean top height development may then be regarded as a self-contained subsystem which can be studied in isolation.

Initially, the site indices for the different plots are unknown. Some or all of the coefficients in (2.2.6) may then differ between plots, being functions of the unknown site quality. Also, assuming as usual that thinnings do not affect mean top height, all the measurements and also the ages of measurement carry information on height growth, not only the periods without thinnings as in the case of the other state variables. All this makes the problem of estimating the parameters of the height growth submodel different and more complicated than the estimation of parameters for the rest of the model. A method has been developed for modelling the height growth and site index relationships, and it will be published elsewhere.

Once estimates for the site indices and for the parameters in (2.2.6) are available (some or all of them functions of site index), they are substituted into (2.3.10). Note that $a_{33} = \lambda_3$, and $b_3 = -\lambda_3 a_3$. Since the eigenvectors are defined only up to a actor of proportionality, we normalise them arbitrarily by setting $P_{11} = P_{22} = P_{33} = 1$. The constraints $c_{31} = c_{32} = 0$ and $a_{31} = a_{32} = 0$ mentioned in section 2.2 are enforced (these last ones result in $P_{31} = P_{32} = 0$). Finally, we must ensure that V is always a symmetric positive-definite matrix. The easiest way of doing this is by substituting $V = TT'$, where T is a lower-triangular matrix. This also simplifies the computation of $|V|$ and of the quadratic form in (2.3.10) (see Martin *et al.*, 1971). Some of the remaining parameters may actually be functions of site index, containing other parameters to be estimated. Now M.L. estimates for the parameters can be obtained by finding a maximum of (2.3.10), using any of the procedures available for unconstrained nonlinear optimisation (Chambers, 1973; Murray, 1972; Jacoby *et al.*, 1972).

The selection of a good starting point for the optimisation is important. A possible approach is to first select C, by guessing or from previous experience with other sets of data, and then to fit (2.2.3) by ordinary linear regression, using central finite differences to approximate the derivatives. This is assuming that after introducing the site index we still have a model linear on the parameters. As a second stage, a restricted version of (2.3.10) may be maximised. Specifically, if Σ in (2.3.8) is taken as a diagonal matrix, then its M.L.E. can be obtained explicitly by equating to zero the partial

derivatives of (2.3.10), resulting in

$$\hat{\sigma}_{ii} = \frac{1}{n} \sum \epsilon_i^2 / R_i(\Delta t) \,, \quad (i = 1, 2, 3) \,, \tag{2.3.11}$$

where

$$R_i(\Delta t) = \frac{e^{2\lambda_i \, \Delta t} - 1}{2\lambda_i} \,.$$

V can then be eliminated from (2.3.10), so that the terms

$$\sum \ln |V| + \sum \boldsymbol{\epsilon}' V^{-1} \boldsymbol{\epsilon}$$

reduce to

$$\sum_{i=1}^{p} [n \ln \hat{\sigma}_{ii} + \sum \ln R_i(\Delta t)] + np \,. \tag{2.3.12}$$

3 Results

3.1 Data

The permanent sample plots for radiata pine in Southland Conservancy were used. The main reasons for choosing this particular set of data were that it was readily available (a preliminary analysis had been carried out by B. Manley), and that the amount of data seemed about right for developing the estimation procedures at a reasonable cost in computing resources.

The data was screened by examining scattergrams displaying the mean top height, basal area, stocking and mean d.b.h. measurements, and also annual increments for the same variables computed from successive measurements. All measurements showing inconsistent or highly suspect values were eliminated. In the case of increments where it was not clear which of the two measurements was in error both measurements were eliminated. Two complete plot series with numerous inconsistencies were left out of the study. Since the time was limited and the main purpose was to develop methods, no attempt was made at tracing inconsistencies to the original records or to the individual tree measurements. A considerable amount of data was discarded.

It seemed useful to introduce some kind of correction for reducing part of the variation caused by differences in dates of measurement. The only

data found on seasonal growth patterns for radiata pine was in Jackson *et al.* (1976) and in unpublished reports by J. Beekhuis (pers. comm.), both for the central North Island. The accumulated growth as a fraction of a year's growth was plotted against the months of measurement, with the origin in July. Corrections to be applied to the age were obtained for each month by reading from the graph back onto the growth axis. The basal area growth follows a pattern different from that of the height growth, but a reasonable compromise was found for the months of February to October. For November to January the corrections become large and different for height and basal area, so that the few measurements taken in these months were discarded. The corrections used, to be added to the age, are:

February	-0.2	July	0.0
March	-0.2	August	0.1
April	-0.1	September	0.2
May	-0.1	October	0.3
June	0.0		

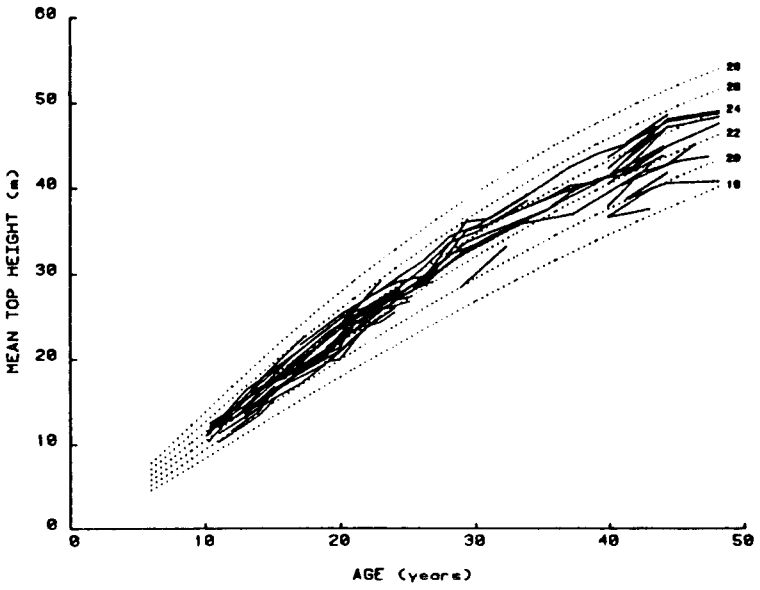

Figure 1:

The data used for the height growth model included 58 plots, with a total of 247 measurements. Figure 1 shows the data, together with the site index curves.

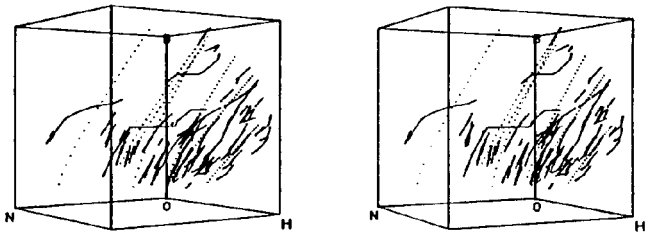

Figure 2:

For estimating the rest of the parameters, 171 pairs of measurements from 56 plots are available. These are shown in the stereogram of Fig. 2. The three-dimensional distribution and trends of the measurements of basal area (B), stocking (N) and mean top height (H) may be appreciated using a pocket-size lens stereoscope. The range of the axis is from 0 to 130 m^2/ha for B, from 0 to 3500 stems/ha for N, and from 0 to 50 m for H. Pairs of successive measurements with no thinnings between them are joined by straight line segments. The dotted curves correspond to predictions by a provisory model, and are discussed in 3.3.

It can be seen from the stereogram that the observations do not cover adequately the region of interest in the B-N-H state space. Also, most of the data seems concentrated near a surface. In addition to suggesting the likely range of validity of a stand model based on a particular set of data, this kind of graphical representation seems potentially useful for the planning of future measurements.

3.2 Height Growth

The model used may be written in the form

$$\frac{\mathrm{d}H^c}{\mathrm{d}t} = b(a^c - H^c) \;, \tag{3.2.1}$$

with an initial condition $H(t_0) = 0$. In addition to the Wiener process representing environmental variation, as in 2.2, a random variable representing measurement error was also included.

Three ways of allowing for the differences in site quality were studied in detail (with $t_0 = 0$):

(a) Parameters a and c were forced to be the same for every plot, and b free to take a different value for each plot. This produces site index curves with a common upper asymptote (the asymptote is a).

(b) Parameters b and c the same for every plot, and a free. This produces anamorphic site index curves.

(c) Parameters a and b related by $a = \alpha + \beta S$, with α, β and c common for all plots. S is the predicted height at age 20 years for each plot. and involves the free parameter b. This covers (a) and (b) as particular cases ($\beta = 0$ and $\alpha = 0$, respectively), at the cost of one extra parameter.

The "measurement" variance was allowed to take different values for each of the plots, and the "environmental" variance was taken as proportional to b, with the same proportionality constant for all plots. No appreciable improvements were observed when t_0 was allowed to take values different from zero.

The following guides may be used to compare models in terms of the likelihood function. Edwards (1972) considers that a difference of about two units in the log-likelihood might be taken as "significant". When comparing models with different numbers of parameters, Akaike (1975) and Stone (1977) suggest subtracting one unit for each additional parameter. Only differences in the log-likelihood matter; the exact values are largely irrelevant.

The following results were obtained for each of the models:

(a) Log-likelihood = 374.6
a = 74.8
c = 0.783

(b) Log-likelihood = 377.3
b = 0.0267
c = 0.775

(c) Log-likelihood = 377.9
α = -46.3
β = 5.08
c = 0.774

There are no large differences in goodness of fit between the three models, although the data gives slightly better support to hypothesis (b). Graphical

comparison of the height-age curves shows little practical differences up to the age of 30 years. Over 30 years the three models tend to diverge for low and for high site indices. This suggests that the data do not define well the trends in those regions (see Fig. 1), and any of the models must be regarded as unreliable for old stands in extreme sites.

Model (a) was selected, since it greatly simplifies the development of the rest of the stand model. The resultant site index curves are shown in Figure 1 (heights and site indices are in metres). These curves agree closely with those obtained by Burkhart and Tennent (1978) using a different approach.

3.3 Full Model

Basal area, stocking and mean top height were adopted as the state variables. Three state variables is certainly a minimum for a satisfactory stand model (Garcia, 1968, 1974). As mentioned in 2.1, it might be desirable to add the green crown level and perhaps other variables. The green crown level might improve the performance of the model when heavy thinning or pruning are involved, and also improve the estimation of outputs by explaining part of the variation in log quality and in stem form. However, only a small part of the permanent sample plots available include measurements of crown level. It seems better at this stage to use the simpler three-dimensional state vector and to assess the need for adjustments later (see 3.4).

A problem in implementing the model is to decide in what form the site index should enter in it. A convenient assumption is that the trajectories followed by stands in the state space (Fig. 2) do not depend on site index, only the speed of movement along a trajectory depending on it. In other words, the effect of the site index is a change in the time scale. Similar-assumptions have often been used in forestry, for example in Beekhuis (1966). The model of Elliott and Goulding (1976) does not make this assumption, but even so, the predicted trajectories differ remarkably little among site indices. It seems useful to adopt this as a working hypothesis and to check it later at the validation stage.

The height growth model with common asymptotes is compatible with the time-scaling effect hypothesis, and this is the main reason why it was selected. In terms of the stand model (2.3.2), the hypothesis implies that only the eigenvalues of A are functions of the site index; all the other parameters are constants. Also, the eigenvalues depend on site index through

a proportionality factor which is the same for all the eigenvalues. An easy way of handling the effect of the site index when fitting the model is then to multiply the age by this factor, which can be taken as λ_3. λ_3 is the same as $-b$ in 3.2, and is a known function of the site index for each plot.

Two algorithms were tried for maximising the log-likelihood: a quasi-Newton procedure using finite difference approximations for the derivatives, coded in Algol (Lill, 1970), and a Fortran implementation of Nelder and Mead's (1965) Simplex method (O'Neill, 1971). Both algorithms were chosen mainly because they do not require derivatives and were readily available. Quasi-Newton procedures (also known as Davidon-type or variable-metric methods) are usually very efficient, with good rates of convergence, specially near the optimum. The Simplex procedure is more robust, often succeeding with difficult functions when other methods fail, but it is generally regarded as slower, specially when the number of variables is large.

Attempts using Lill's program failed repeatedly, with the procedure exiting after a few iterations without reaching an optimum. It was found later that it also fails in the some way with a simple test function, for some starting points. No errors have been detected in the transcription from the publication, but some misprint is suspected.

With the Simplex method a rapid initial improvement in the function value was generally obtained, but convergence was extremely slow afterwards. Several attempts with different starting points using the full function (2.3.10) (20 variables) were discontinued because of slow convergence.

Somewhat better results were obtained with a stage-wise approach, as suggested at the end of 2.3. First, some values were selected for C, and initial estimates for the other parameters were obtained by linear regression with finite-difference approximations. Then the optimisation procedure was applied using (2.3.12). In addition, in this second stage the matrix A was also restricted to be diagonal (which implies $P = I$) . This reduces the number of variables (parameters) to 10. Much better values of the log-likelihood than before were obtained, but still convergence was not complete after several thousand function evaluations. In a third stage, the constraint $P = I$ was lifted, increasing the number of variables to 14, and the optimisation was restarted. The computation was discontinued after about 6000 function evaluations, with the log-likelihood still decreasing at a very slow but roughly linear rate.

This seems to confirm the slow ultimate convergence of the Simplex

method for functions of many variables. It is hoped that the use of other quasi-Newton procedures will produce much better results. The computing time on the ICL 2980 computer, with the full 20-variables function (2.3.10), is approximately 1.3 minutes per thousand function evaluations, for our sample of 171 observation pairs. The total optimisation time may be expected to be roughly proportional to the number of observations.

Although their discrepancy from the M.L. estimates is unknown, the best parameter values obtained until now way be used to illustrate some characteristics of the model and possible methods of validation.

According to the concepts explained in 2.1 and 2.2, the model defines a field of velocities and a set of trajectories or flow-lines that fill completely the B-N-H state space. Through each point in this space, passes one and only one trajectory. No matter how a stand arrives to a point, it will follow the corresponding trajectory until the next thinning. The effect of a thinning is a jump from one trajectory onto another. Observing the paths followed by stands in Figure 2, and considering the natural variability in this kind of data, these ideas do not seem entirely unrealistic.

A few trajectories predicted by the provisional model are included in Figure 2 as dotted curves. Each dot corresponds to one year growth for site index 23. They do not seem too unreasonable, except for the higher densities where one would expect heavier mortality. It might be observed that there are very few plots in this region, and their trends are somewhat erratic.

Figure 3 represents two "cuts" through Figure 2, showing projections of the predicted velocity field at the planes $H = 15$ m and $H = 30$ m. The line segments represent the directions in which stands at different points in the state space would move. The length of the line segments corresponds to two-years' growth on site 23. 0f course, reliable prediction is only possible within the area covered by the data. In particular, the increases in stocking shown for the region of high stocking and low basal area are clearly impossible. The possibility that predictions would be improved by constraining the model to feasible behaviour over all or part of the state space is still an open question.

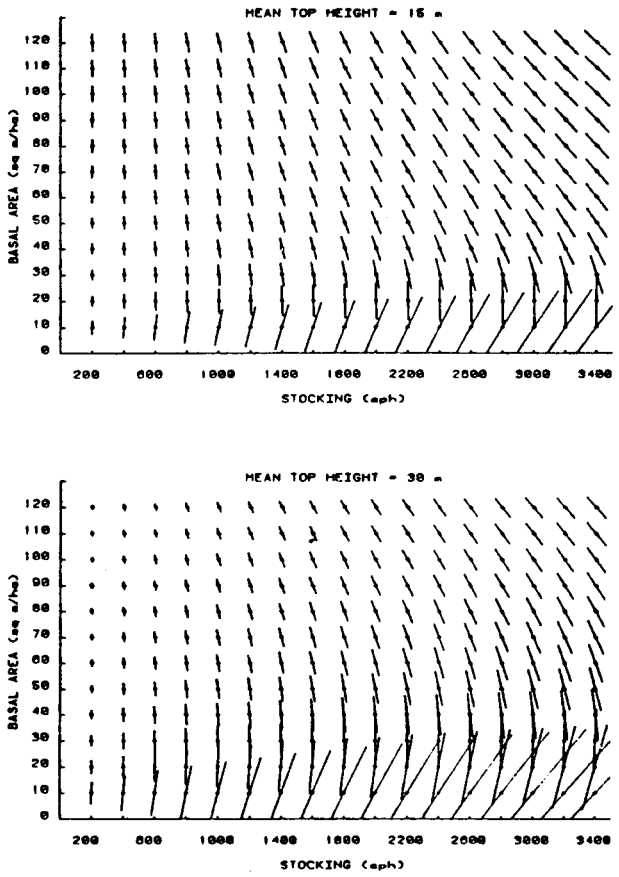

Figure 3:

3.4 Validation

The discussion is limited to demonstrating a particular technique which seems useful for assessing results obtained with models of the type described in 2.2. For other validation methods which could also be used see Goulding (1978).

From (2.3.6), note that the M.L.E. for the logarithms of the absolute values of the z_i are linear functions of time. We might then plot the $\ln|z_i|$ computed from the data versus t, and look for systematic deviations from linearity.

It seems better, instead, to plot $\ln|z_i|$ *vs.* $\ln|z_j|$ for all the combinations

of i and j. These relationships still tend to straight lines, and it might be expected that large part of the variation due to year-to-year differences in growth rates would be eliminated. In addition, with the particular assumption on the effect of site index made in 3.3, the slopes λ_i/λ_j must be the same for all plots.

The data have been plotted in this way in Figure 4. Successive measurements with no thinnings between have been joined with straight line segments. The same data is also shown three-dimensionally in the stereogram in Figure 5. In addition, in Figure 4 the plots have been grouped in three site index classes and distinguished with different line types: site indices less than 22.5 with dotted lines, 22.5 to 23.5 with dashed lines, and over 23.5 with continuous lines. The small triangles mark measurements immediately following a thinning.

In general, the trajectories seem to follow reasonably well the theoretical pattern of parallel straight lines. Some notable exceptions correspond to the plots with high density and erratic mortality already mentioned. There is no evidence of systematic differences in slope or in curvature between site index classes; the site index hypothesis seems satisfactory. The graphs also give some information on the effect of thinnings. In most cases the trajectories immediately after thinnings (triangles) are not appreciably different from those for other measurements, but there are several instances in which clearly this is not so. Identification of these measurements and examination of the plot records might show if there is anything special about them, if they correspond to exceptionally heavy thinnings or perhaps high prunings, and indicate if adjustments to the model become necessary.

Notes

Section 2.1

System Theory has been developing as an independent discipline during the 1970s and late '60s, embracing and extending what was the "state space approach" in the Control Theory of the '60s. In turn, Control Theory derived mainly from part of what was called Cybernetics in the '50s. For further details the following sequence of readings can be recommended:

Garcia (1974), Chapter 1 of Bellman and Kalaba (1965), Chapter 1 of

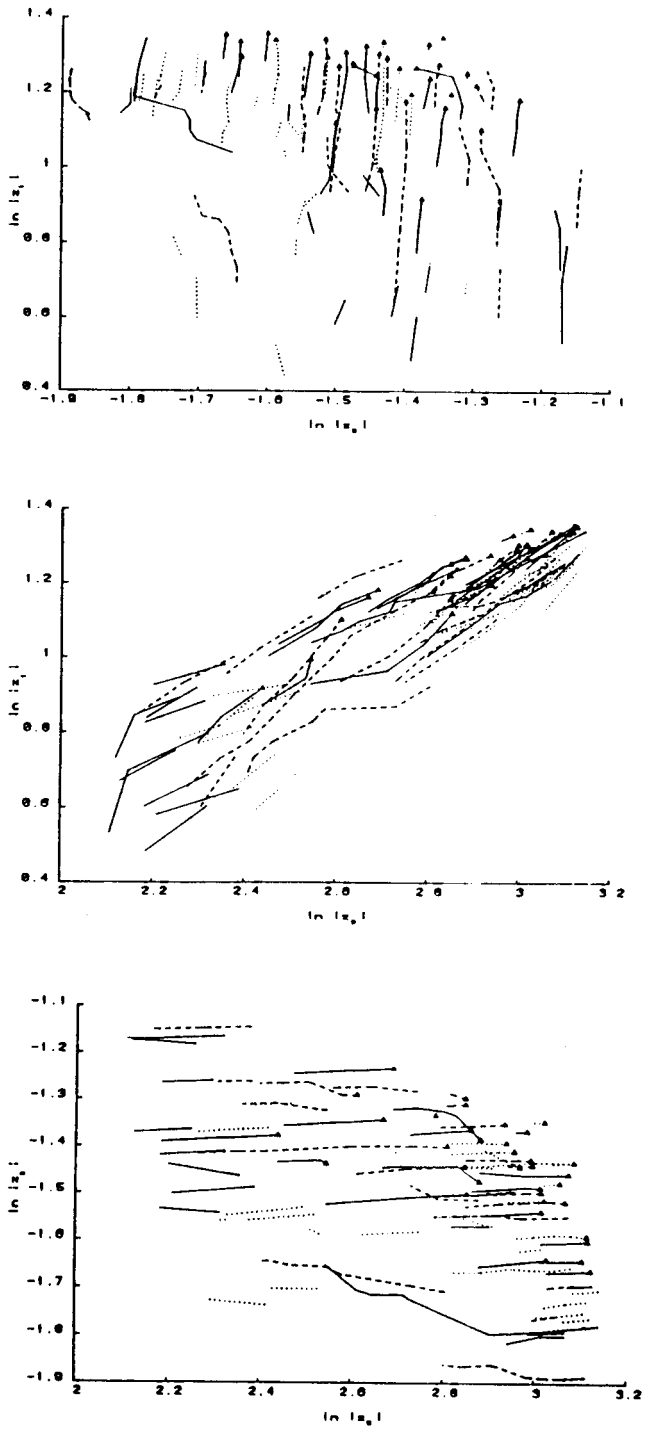

Figure 4:

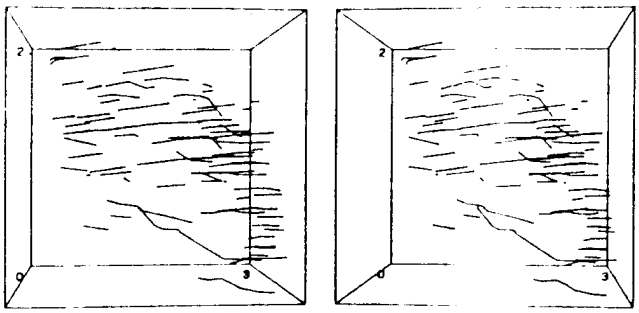

Figure 5:

Wiberg (1971), and Padulo and Arbib (1974). Patten (1971) and Bellman and Kalaba (1960) are also useful.

Many efforts in modelling stands have suffered from failures in selecting appropriate states and transition functions. The model of Beekhuis (1966) is remarkable in being perhaps the first model for managed stands with a sound structure. It had an adequate state vector, and its transition function "almost" satisfied the semigroup property. The same can be said of many later models (e.g., Elliott and Goulding, 1976), but some ones with serious deficiencies continue to be produced. It is felt that a better appreciation of system-theoretic ideas might help to clarify the issues involved (Garcia, 1968, 1974).

The discussion of systems in section 2.1 has actually been restricted to time-invariant systems. This is not considered important, since time-varying systems, that is those whose behaviour depends explicitly on time, can be formulated as time-invariant ones by including time as an additional state variable.

Section 2.2

What we have called the Bertalanffy model is frequently referred to as the Chapman-Richards model in the forestry literature. It seems to have been first proposed by Bertalanffy (1949, 1957), and later its properties extensively analysed by Richards (1959). The usual reference for Chapman is Chapman (1961), where the model is used but nothing new about it is added.

A useful reference for linear differential systems is Wiberg (1971).

The notation for the stochastic differential equation (2.2.8) follows Erickson (1971). Strictly speaking it might be considered as incorrect, because the Wiener process is not differentiable. More correct and more usual (for stochastic differential equations) would be

$$\mathrm{d}\boldsymbol{x}^C = (A\boldsymbol{x}^C + \boldsymbol{b})\,\mathrm{d}t + B\,\mathrm{d}\boldsymbol{w}$$

(Gihman and Skorohod, 1972). However, (2.2.8) has been used because it looks more familiar for most people.

Section 2.3

Instead of permanent sample plot data, stem analysis could be used if some way of estimating mortality is available. A combination of both types of data is also possible.

Many authors define a likelihood function only up to a factor of proportionality (which can depend on the data).

A maximum of the likelihood function may do not exist. In practice this usually indicates an inadequate formulation of the model. More troublesome is the possibility of local maxima (see Chambers, 1973).

All the results of sections 2.2 and 2.3 are valid for state vectors with any number of components.

It may be sometimes desirable to impose constraints on some of the parameters when maximising the log-likelihood. Many constraints can easily be implemented by transformations of parameters. For example, parameters constrained to be non-negative may be replaced by the square of a new parameter. In fact, any constraint can be enforced through transformations, possibly at the cost of adding new variables. Alternatively, a Nonlinear Programming algorithm may be used.

Section 3.2

The guides suggested for comparing models on the basis of likelihoods must be regarded only as a rough aid to intuition, and used in conjunction with

other knowledge or sources of evidence, if possible. In general, the whole problem of Statistical Inference tends to be an ill-defined one. See Barnett (1973) for a good review of some of the issues involved.

An unexpected result was that in all cases the estimate for the "environmental" variance was zero. However, the estimated correlations between the estimates for the two variances were fairly high. This suggests that the data does not allow a reliable partition of the variation into "environmental" and "measurement" components.

Section 3.3

A practical way of simulating the development of a stand with the model is as follows. First, compute the one-year transition matrix corresponding to the appropriate site index as

$$T = P^{-1} e^{\Lambda} P \ ,$$

and the vector $\boldsymbol{h} = \boldsymbol{a} - T\boldsymbol{a}$. Compute $\boldsymbol{y}(t_0) = \boldsymbol{x}^C$ for the initial state vector $\boldsymbol{x} = \boldsymbol{x}(t_0)$. Then project forward by repeatedly applying the recursion

$$\boldsymbol{y}(t+1) = T\boldsymbol{y}(t) + \boldsymbol{h} \ .$$

The states are recovered through $\boldsymbol{x} = \boldsymbol{y}^{C^{-1}}$. Note that the zeros in the matrices allow for some simplification in the computations.

It seems evident that to develop a reliable stand model a fair amount of data, well distributed over the state space, is required. A possible strategy for developing local models for regions where the data are inadequate might be the use of "adapted" models. This means that part of the parameters, for example the matrix C and perhaps also P, are taken from other region where a good model is already available. Then the optimization, possibly using the simplified log-likelihood given by (2.3.12), is carried out over the remaining parameters using the local data.

References

Akaike, H. 1973: Information Theory and an extension of the maximum likelihood principle. In: Petrov, B.N. and Czaki, G. (eds), 2nd International Symposium on Information Theory. Akademia Kiado, Budapest.

Barnett, V. 1973: Comparative Statistical Inference. Wiley, London.

Beekhuis, J. 1966: Prediction of yield and increment in *Pinus radiata* stands in New Zealand. Technical Paper No. 49, Forest Research Institute, N. Z. Forest Service.

Bellman, R. and Kalaba, R. 1960: Dynamic Programming and Adaptive Processes: Mathematical foundation. *IRE Transactions on Automatic Control, AC–5*, 5–10. (Reprinted in: Bellman, R. and Kalaba, R. (eds) Selected papers on mathematical trends in Control Theory. Dover, New York, 1964).

Bellman, R. and Kalaba, R. 1965: Dynamic Programming and modern Control Theory. Academic Press, New York.

Bertalanffy, L. von 1949: Problems of organic growth. *Nature, 163*: 156–158.

Bertalanffy, L. von 1957: Quantitative laws in metabolism and growth. *The Quart. Rev. Biol., 32*: 217–231.

Burkhart, H.B. and Tennent, R.B. 1978: Site index equations for Radiata pine in New Zealand. *N .Z. J. For. Sci. 7*: 408–416.

Chambers, J. M. 1973: Fitting nonlinear models: Numerical techniques. *Biometrika, 60*: 1–13.

Chapman, D. G. 1961: Statistical problems in population dynamics. Proc. Fourth Berkeley Symp. Math. Stat. and Prob. University of California Press, Berkeley.

Edwards, A. W. F. 1972: Likelihood. Cambridge Univ. Press.

Elliott, D. A. and Goulding, C .J. 1976: The Kaingaroa growth model for Radiata pine and its implications for maximum volume production (abstract). *N. Z. J. For. Sci., 6*: 187.

Erickson, R. V. 1971: Constant coefficient linear differential equations driven by white noise. *Ann. Math. Statist., 42*: 820–823.

Garcia V., O. 1968: Problemas y modelos en el manejo de las plantaciones forestales. U. of Chile, Esc. Ing. Forestal. (thesis).

Garcia V., O. 1974: Sobre modelos matemáticos de rodal. Informe Tecnico 48, Instituto Forestal, Chile (English translation available at the F.R.I. library).

Gihman, I. J. and Skorohod, A. V. 1972: Stochastic Differential Equations. Springer-Verlag, New York.

Goldberg, S. 1958: Introduction to Difference Equations. Wiley, New York.

Goulding, C. J. and Shirley, J .W. 1978: A method to predict the yield of log assortments for long term planning. (Paper presented at this Symposium).

Goulding, C. J. 1978: The validation of growth models used in forest mensuration. (Submitted for publication to the N. Z. J. of Forestry).

Jackson, D. S., Gifford, H. H. and Chittenden, J. 1976: Environmental variables influencing the growth of Radiata pine: (2) Effects of Seasonal draught on height and diameter increment. *N. Z. J. For. Sci. 5*: 265–286.

Jacoby, S. L. S., Kowalik, J. S. and Pizzo, J .T. 1972: Iterative methods for nonlinear optimization problems. Prentice-Hall, Englewood Cliffs, N.J.

Kalman, R. E., Falb, P. L. and Arbib, M. A. 1969: Topics in Mathematical System Theory. McGraw-Hill, New York.

Karlin, S. 1966: A first course in Stochastic Processes. Academic Press, New York.

Lill, S. A. 1970: Algorithm 46: A modified Davidon method for finding the minimum of a function using difference approximation for derivatives. *Computer J., 13*: 111–113 (and corrections in *14*, 106).

Martin, R. S., Peters, G. and Wilkinson, J. H. 1971: Symmetric decomposition of a positive definite matrix. *In*: Wilkinson, J. H. and Reinsch, C. (eds). Linear Algebra – Handbook for Automatic Computation, Volume II Springer - Verlag, Berlin.

Miller, K. S. 1968: Linear Difference Equations. Benjamin, New York.

Murray, W. 1972 (ed). Numerical methods for unconstrained optimization. Academic Press, London.

Nelder, J. A. and Mead, R. 1965: A simplex method for function minimization. *Computer J., 7*: 308–313.

O'Neill, R. 1971: Algorithm AS47 – Function minimization using a simplex procedure. *Appl. Statist., 20*: 338–345 (corrections and changes in *23*, 250–252 and *25*, 97).

Padulo, L. and Arbib, M. A. 1974: System Theory. Hemisphere Pub. Co., Washington, D.C.

Patten, B. C. 1971: A primer for ecological modelling and simulation with analog and digital computers. *In*: Patten, B. C. (ed) System Analysis and simulation in Ecology. Academic Press, New York.

Richards, F. J. 1959: A flexible growth function for empirical use. *J. Expt. Botany, 10*: 290–300.

Stone, M. 1977: An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Statist. Soc., B, 39*: 44–47.

Tomlinson, A. I. 1976: Short period fluctuations of New Zealand rainfall. *N .Z. J. of Science 19*: 149–161.

Wiberg, D. M. 1971: State space and linear systems. Shaum's Outline Series, McGraw-Hill, New York.

Zacks, S. 1971: The theory of Statistical Inference. Wiley, New York.

Zadeh, L. A. 1969: The concepts of system, aggregate, and state in system theory. *In*: Zadeh, L. A. and Polak, E. (eds). System Theory. McGraw-Hill, New York.