U.S.D.A. Forest Service
Research Paper NC-45
1971

System Identification Principles in Studies of Forest Dynamics

Rolfe A. Leary

North Central Forest Experiment Station
Forest Service
U.S. Department of Agriculture

The author is a Mensurationist for the Station at the Headquarters Laboratory in St. Paul, Minnesota, which is maintained by the USDA Forest Service in cooperation with the University of Minnesota.

NORTH CENTRAL FOREST EXPERIMENT STATION
D. B. King, Director
FOREST SERVICE — U.S. DEPARTMENT OF AGRICULTURE
Folwell Avenue
St. Paul, Minnesota 55101

CONTENTS

SYSTEM IDENTIFICATION PRINCIPLES IN STUDIES OF FOREST DYNAMICS

ROLFE A. LEARY

This paper is directed at the following problem: Given observational data on a dynamic system, what are the equations governing its behavior? This is known as a system identification or inverse problem, and in its broadest sense involves both unknown algebraic form and numerical constants in the governing equations. Clearly, knowledge of the exact, or even approximate, equations governing a system makes effective control a realistic possibility. The ability to control a dynamic forest system is of concern in the many instances where man has developed preferences for certain system states and is prepared to take action to ensure their existence.

Methods of dealing with both types of unknowns are discussed. The first part concerns concepts that are helpful in rationalizing the algebraic form of the governing equation right-hand sides. The second part deals with parameter estimation. The only type of governing equations considered are linear and nonlinear first-order ordinary differential equations.

The methods of regression analysis, so widely used in modeling forest systems, are not used in this study. In a traditional regression approach one is attempting to solve two problems simultaneously. First is that of modeling the causal mechanisms of observed system behavior. Second is the problem of making inferences about the population from which the observed systems were selected. Regression methods are particularly inadequate for the first problem. Thus, the above problems are treated separately, attention being given to the first.

The approach used here is to transform the problem from one of having several observations through time on a forest system in some inconvenient tabulated form to one of having a single point located in state space corresponding to the system state at each observation. This is essentially a problem of trajectory identification. It can, in many cases, be solved by considering it as a multipoint boundary value problem of differential equations and applying the computational procedure called quasilinearization (Bellman and Kalaba 1965, Lee 1968). The approach used here in state space is entirely analogous to that used in physical space for such problems as orbit determination (Bellman 1962).

It is assumed that the reader is interested in developing mathematical models of forest-system dynamics. The level of mathematics is variable. As a rule, the lowest level of mathematics that conveys correct meaning is used. Most of the concepts and techniques discussed are not new; they have been gleaned from numerical analysis, numerical methods, differential equations, mathematical analysis, control theory, and general systems theory. In anticipation that some readers may not be familiar with these subjects and how they relate to the problem of modeling dynamic forest systems, each topic is discussed verbally and related to example models. No attempt is made at an exhaustive treatment. Instead, the emphasis is on synthesizing the concepts into a comprehensive, workable scheme.

In the following discussion the assumed goal is to develop governing equations for the standing crop in northern hardwood forests. Plot data from the Argonne Experimental Forest in northeastern Wisconsin are used.

GENERAL CONCEPT OF A SYSTEM

The past decade or two has witnessed the evolution of a holistic or systems approach to the study of complex entities. During this time the forester's ability to talk and write about systems has in some cases outdistanced his ability to quantify or model them. The concepts and techniques that follow may help to fill the gap in modeling dynamic forest systems.

The success of the approach hinges on successfully combining the following: (1) the researcher's knowledge of the biological processes involved in the system he is studying, (2) his ability to quantify these processes, and (3) the ability of the digital computer to do the calculations necessary for successful solution of the inverse problem.

Initially, a proper definition of the entities involved is required. The following definition suits the purposes well (Hall and Fagen 1956). "A system is a set of objects together with relationships between the objects and the attributes." Objects are the parts or components of a system and attributes are properties of objects.

This study is concerned with modeling the primary production of forest systems. Although only the vegetative portion of the ecosystem is modeled, the theories and methods discussed are applicable to ecosystem production in general. Due to the limited historical data available, the examples given deal only with the component capable of producing merchantable timber.

Thus, what is being modeled is a subsystem of the total vegetation system. Where does one draw the line or boundary for his system? It is helpful to place system boundaries at natural breaks in the hierarchy of systems. Thus, the system may be restricted to all vegetation of species capable of producing merchantable timber in the northern hardwood forest. But, this would be an unnatural boundary. Hence, our system includes all woody and herbaceous vegetation in the northern hardwood type even though our examples, for reasons stated previously, consider only timber species.

CHARACTERIZING SYSTEMS

A meaningful discussion of alternative methods of characterizing systems presupposes a well-defined end objective. Usually the interest is in gaining insight into the system's inner workings and developing a predictive capability, via a mathematical model. A logical beginning would be to characterize a system by its components. In the examples from later sections, where the vegetative portion of the ecosystem is treated as a system, components may be defined on the basis of taxonomic criteria, physical size or function (in the distribution and accumulation of energy), or by some combination. For example, an individual plant, species, diameter or height class, or species x diameter class may be considered as a system component. Clearly, the number of possible ways of defining system components is enormous. For the purposes of this study, where our interest is in system productivity, a combination of physical size and function was deemed most desirable.

Once the components for the system have been defined, a question arises as to what attribute should be used to characterize each component. A most elemental approach would be to use presence or absence of qualifying individuals as the component attribute. A next logical refinement would be the frequency of occurrence of qualifying individuals. For ecological succession studies this level of refinement might be suitable. In productivity studies, it is still too elemental for reasons that will be evident in the section on dynamic models. Having gone from the binary (1-present: 0-absent) level to the discrete (positive integers), any further refinement necessitates passing to the continuous. This method of describing system components is common; basal area of qualifying individuals is an often-used attribute. In later examples, sum of diameters of trees in each component is used.

One could, of course, go further and compute volume from diameter (d.b.h.) and height and use this as the attribute. The advisability of the last step is questionable, however. As a general rule, component attributes should be the most elemental measures of individuals in the component consistent with the type of model used.

To summarize briefly, then, a system is comprised of objects or components and each component may be characterized by various attributes.

At this point it is convenient to use a slightly different approach, and consider the system components as coordinate axes of an abstract space — called state space. The component attributes specify the system location with respect to each coordinate axis. All system components plus their attributes specify the system location in state space, and are called state variables.

In later examples, species group (based on shade tolerance) and size (diameter class) were used as the basis for delineating the system components. The size attribute selected was sum of diameters. Thus, sums of tree diameters at breast height in each species group x diameter class are the state variables. Of course, the units in which each is expressed need not be the same. That is, the first m state variables may represent standing crop of m browse species and be expressed in pounds per acre, the $m + 1. . . n$ variables may represent standing crop in timber species and be expressed in inches of diameter for all qualifying individuals. The variety of expressions are, obviously, limited only by the user's needs. It is not at all difficult to conceive of system components numbering into the hundreds. Theoretically, the more state variables employed, the more refined the characterization of the system will be. However, in practical applications, one must strike a compromise between increasing realism and decreasing tractability as the number of variables increases.

The state variable approach to system characterization is well established in engineering and control theory (De Russo *et al.* 1965). There is much to be gained in terms of conciseness of expression and ease of mathematical analysis of forest systems by its adoption in our work. The combination of a state variable formulation of forest systems with the versatility of systems of nonlinear differential equations makes a powerful predictive scheme. The remainder of this paper, then, uses a state variable approach to forest system modeling.

DYNAMIC SYSTEMS

An underlying objective of this paper is to develop a capability of predicting system development through time; hence, the concern is with dynamic systems. But first it is instructive to consider a static system.

Take for example a system with state variables $Y1$ and $Y2$. The system state is given, then, by the ordered pair $(Y1, Y2)$. The only type of systems of interest to us are interactive ones, that is, where the variable $Y1$ affects and is affected by $Y2$. Thus, a static interactive linear model is given by the system of simultaneous equations

$$\begin{aligned} a_{11}Y1 + a_{12}Y2 &= b_1 \\ a_{21}Y1 + a_{22}Y2 &= b_2. \end{aligned} \qquad (1)$$

The system state is given by the values of $Y1$ and $Y2$ that satisfy these equalities.

The primary concern here is the differential equation formulation of system development through time, and the discussion that follows deals with it exclusively. As a means of introducing this formulation, consider a system characterized by a single state variable; i.e., the scalar case. The equation that specifies the instantaneous rate of change, with respect to time, in system state takes the form

$$dY/dt = f(Y,t) \tag{2}$$

where f is a suitably chosen function,
Y is the dependent variable, and
t is time, the independent variable.

The actual state, as opposed to rate of change of state, is obtained by solving equation (2) for Y.

What are the possible forms the function f may take? In general, they may be as follows:

(a) $$dY/dt = 0. \tag{3}$$

This seemingly uninteresting form, in the context of a system of equations, is very useful, and is used extensively in the later sections on parameter estimation.

(b) $$dY/dt = f(t). \tag{4}$$

This form, which states that rate of change in Y is dependent only on time, is used extensively in engineering and other physical sciences, but in the context of open systems of living organisms it is of limited usefulness.

(c) $$dY/dt = g(Y). \tag{5}$$

This equation states that the rate of change of Y is a function solely of the amount present of Y, and is the form of most biological growth functions.

(d) $$dY/dt = h(Y,t). \tag{6}$$

In the case of equation (6), rate of change in Y is related to both amount present of Y and time. No further reference is made to this form.

In this study functions f and g are points of beginning and are not, in the case of f, obtained by differentiating some other function F with respect to time. This approach allows use of differential equations of the form (4) and (5) regardless of whether their solutions may be expressed in closed form. The approach is one of selecting a function of the form f or g that has the desired qualitative properties (see section on "Some Qualitative Properties of System Equations"), and describing the system on that basis, without regard to the closed form expression for F or G if such form does, in fact, exist. G is understood to be the closed form solution of equation (5).

Some most used forms of g and the names commonly associated with them are:

$g = a Y^2 + b Y$	(7)	logistics, autocatalytic
$g = a Y \ln(\frac{A}{Y})$	(8)	Gompertz
$g = a Y^c + bY$	(9)	von Bertalanffy, Chapman-Richards.

Equation (9) is a special case of what is called the Bernoulli equation; i.e., a nonlinear equation of the form

$$dY/dt + Q(t)\ Y^n = R(t)\ Y. \qquad (10)$$

The logistics equation has $Q(t) = -a$, $R(t) = b$, and $n = 2$.

The Bernoulli equation is one of the few nonlinear equations that may be solved in closed form. This is accomplished for the von Bertalanffy function by first substituting

$$Z = Y^{1-c}$$

in (9), which gives

$$\begin{aligned} dZ/dt &= (1-c)\ Y^{-c}\ (a\ Y^c - b\ Y) \\ &= (1-c)\ (a - b\ Y^{1-c}) \\ &= (1-c)\ (a - b\ Z). \end{aligned} \qquad (11)$$

Equation (11) is linear in the variable Z and may easily be solved (see page 14). Once the solution in Z is obtained, it is expressed in terms of Y; i.e.,

$$Y = Z^{1/(1-c)}.$$

To this point, the discussion has been of a system described by a single equation. This approach is, of course, not new. The following observations and contrasts of examples from the literature may be made. In many cases the investigators dealt exclusively with the function G — i.e., the solution of the differential equation — and only incidentally mention that it was the solution of a particular differential equation. Naturally, this meant they were dealing only with equations having solutions expressible in closed form. In a more recent approach[1] the differential equation was fit to observational data, and the solution was obtained analytically.

Let us consider the case where Y is a vector (or an ordered n-tuple of numbers). Only through this treatment can the full benefit of the state variable approach be reaped.

Our model may have the following form:

$$\begin{aligned} dY1/dt &= f_1(Y1,Y2,Y3, \ldots ,Yn,t,a,b,c, \ldots.) \\ dY2/dt &= f_2(Y1,Y2,Y3, \ldots ,Yn,t,a,b,c, \ldots.) \\ &\ \ \vdots \\ dYn/dt &= f_n(Y1,Y2,Y3, \ldots ,Yn,t,a,b,c, \ldots.) \end{aligned} \qquad (12)$$

where Yi, i = 1,2, ... ,n are the dependent variables,
t is time, and
a,b,c, ... are parameters of state such as soil and climate (Lotka 1928).

[1] *Moser, J. W., Jr. Growth and yield models for uneven-aged forest stands. (Unpublished Ph.D. thesis on file at Purdue Univ., Lafayette.)*

Because *a,b,c* . . . are usually time invariant, they may be dropped, leaving

$$dYi/dt = f_i(Y1,Y2,Y3, \ldots ,Yn,t),\ i = 1,2, \ldots ,n. \qquad (13)$$

If equation (13) is modified further by eliminating t from the right-hand side,

$$dYi/dt = f_i(Y1,Y2,Y3, \ldots ,Yn),\ i = 1,2, \ldots ,n. \qquad (14)$$

This is called an autonomous system of equations. All the examples that follow are in terms of autonomous systems of equations.

Equation (14) states that the rate of change of state variable $Y1$ may be a function of the amount present — in our case standing crop — in $Y1,Y2,Y3, \ldots , Yn$, and likewise for the other state variables $Y2,Y3,Y4, \ldots Yn$. It follows from this formulation that system development, as reflected by changing state variables, is influenced by system position in state space. In the context of the later examples, a model such as (14) indicates the diameter growth of individuals in a specific diameter x species group class may be influenced by the amount present in each (or none) of the other classes, depending upon the form of the f_i, $i = 1,2,3, \ldots n$.

Returning to the question concerning the origin of the functions f_i, the system of equations (14) is a point of beginning. This being the case, it is clear that the researcher must rationalize the form of the right-hand sides (r.h.s.) of (14) on the basis of his knowledge of the processes involved. Two general principles in this regard are: (1) the r.h.s. of (14) should possess qualitative properties, individually and as a set, that do not violate biological principles, and (2) the r.h.s. of (14) should reflect the causal pathways between interacting parts of the forest system.

The process of rationalizing the form of the r.h.s. of (14) may conveniently be broken down into two phases: (1) specifying the state variables that should be included in the r.h.s. of each equation in (14), and (2) specifying the algebraic form of the relationships between the state variables selected in phase 1.

An example of phase 1 from Darlington Woods in Indiana follows:[2] The problem is rationalizing the r.h.s. of an interactive dynamic model that has timber size classes as system components.

The biological basis for the r.h.s. is that of competition for the ecosystem resources of space and light (solar radiation). It is based on the following two premises: (1) the amount of light (relative to full sunlight) received by a forest system (stand) component affects its growth and development, and (2) the amount of solar radiation received by a forest stand component is primarily a function of its position in the stand.

If a stand is uneven-aged with resultant diversity in tree size, the trees in the lower strata receive radiation of a different intensity and quality than trees in the overstory. If a stand is even-aged, a larger percentage of the trees receive nearly the same intensity and quality radiation because they tend to be more nearly the same height.

---

[2] *Leary, R. A. A multidimensional model of even-aged forest growth. (Unpublished Ph.D. thesis on file at Purdue Univ., Lafayette.)*

It is constructive to consider this problem in the manner of Hutchinson (1957) in terms of the relationships between ecological niches of potential competitors. An ecological niche is here defined as all of the ecosystem resource levels that a species is currently capable of using. Hutchinson (1957) distinguishes two types of niches. The fundamental niche of a particular forest component is all of the ecosystem resource levels that a component is capable of using at the present time in the absence of competition. The fundamental niche of a particular forest component may thus be shown as a region or set of values in space, the co-ordinate axes of which correspond to the various resources. If only two ecosystem resources are considered, the fundamental niche may be shown as a region in the x - y plane.

The concept of a realized niche arises when competition for ecosystem resources exists. The realized niche of a competitor may be defined as the set of ecosystem resource levels it is capable of using and has available for use free from competition. It may also be defined as the fundamental niche minus the intersection subset.

If N_1 is the fundamental niche of forest component $Y1$, and N_2 is the fundamental niche of forest component $Y2$, the realized niche of component $Y1$ is

$N_1 - \{N_1 \cap N_2\} = RN_1$. ($\{N_1 \cap N_2\}$ is the set of all ecosystem resource levels (a,b) that are simultaneously in both sets N_1 and N_2.)

Likewise, the realized niche of component $Y2$ is

$$N_2 - \{N_1 \cap N_2\} = RN_2.$$

Of special interest is $\{N_1 \cap N_2\} = N_2$. In such cases, $RN_2 = \square$, where $\square$ is the null (empty) set. A similar condition could, of course, exist for N_1; i.e., $RN_1 = \square$.

The two ecosystem resources that were used as a basis for the model may be shown graphically as a region in two-dimensional space (Miller 1967) (fig. 1). Let us assume the diagonally hatched area represents the set of light and space resource levels that stand component $Y1$ is presently capable of using in the absence of competition. This area therefore represents the fundamental niche of $Y1$.

If the horizontally hatched area is the set of space and light resource levels that stand component $Y2$ is capable of using in the absence of competition, it is the fundamental niche of component $Y2$.

It is clear from figure 1 that the fundamental niche of the lower stratum component ($Y2$) is a proper subset of the fundamental niche of the larger component ($Y1$). Miller (1967) describes two possible outcomes of competition between components $Y1$ and $Y2$ where $N_2 \subset N_1$. The first is where competition proceeds in favor of $Y1$ and given adequate time, only $Y1$ survives. The second is where $Y2$ survives in all regions of the x - y plane corresponding to $N_1 \subset N_2$, and both components survive.

The hypothesis underlying the proposed model is that the growth and development of a stand component is influenced by the amount of the contained component present, and the amount of the containing component(s) present in the stand.

Thus, for stands comprised of n components, if Y_i indicates the amount of the i^{th} stand component and if N_i indicates the fundamental niche of the i^{th} component, and if

$N_1 \subset N_2$ and $N_1 \subset N_3$ and $N_1 \subset N_4$ and ... $N_1 \subset N_n$, then

$dY1/dt = f_1(Y1,Y2,Y3, \ldots ,Yn)$.

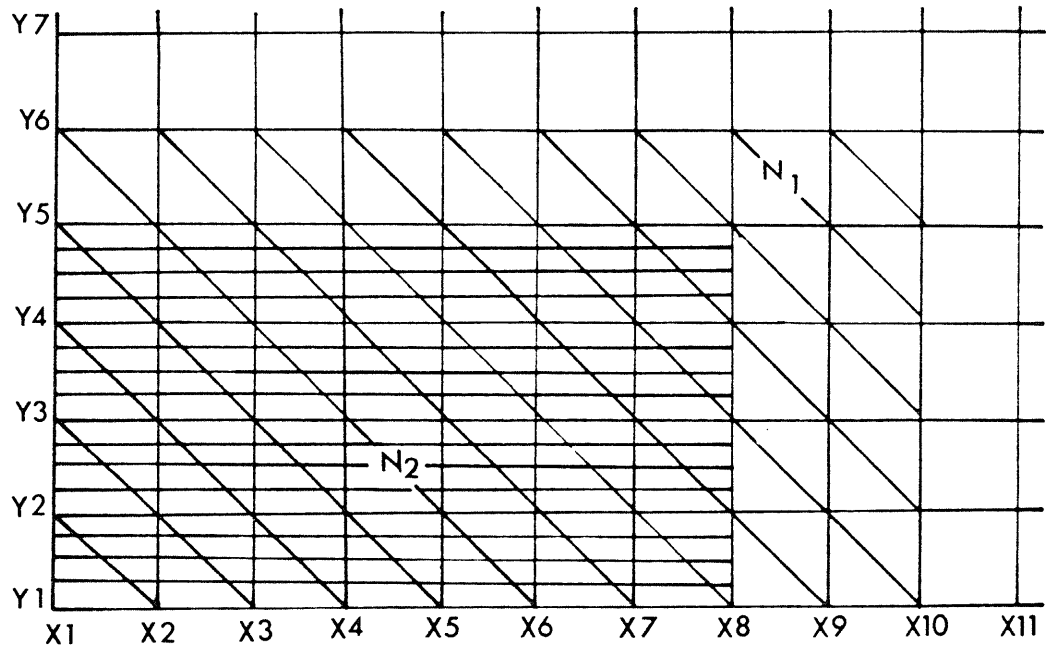


Figure 1. — Fundamental niches N_1 and N_2 of stand components $Y1$ and $Y2$ relative to ecosystem resources of space (Y) and light (X).

Likewise, if

$$N_2 \subset N_3 \text{ and } N_2 \subset N_4 \text{ and } N_2 \subset N_5 \text{ and } \ldots N_2 \subset N_n, \text{ then}$$
$$dY2/dt = f_2(Y2, Y3, \ldots, Yn).$$

If we continue in this manner and consider a model for a pure stand divided into five height classes (components),

$$\begin{aligned} dY1/dt &= f_1(Y1,Y2,Y3,Y4,Y5) \\ dY2/dt &= f_2(\ \ Y2,Y3,Y4,Y5) \\ dY3/dt &= f_3(\ \ \ \ Y3,Y4,Y5) \\ dY4/dt &= f_4(\ \ \ \ \ \ Y4,Y5) \\ dY5/dt &= f_5(\ \ \ \ \ \ \ \ Y5) \end{aligned} \qquad (15)$$

where Yi is the standing crop in the i^{th} height class,
Yl is the shortest component, and
t is time, in years.

Implicit in the above argument is that the understory does not affect overstory development. This is, of course, not universally true.

As an example of phase 2 applied to a reduced system, consider a simple linear additive relationship between components, such as

$$\begin{aligned} dY1/dt &= a_{11}Y1 + a_{12}Y2 + a_{13}Y3 \\ dY2/dt &= \ \ \ \ \ \ a_{22}Y2 + a_{23}Y3 \\ dY3/dt &= \ \ \ \ \ \ \ \ \ \ a_{33}Y3 \end{aligned} \qquad (16)$$

or nonlinear multiplicative relationships such as

$$
\begin{aligned}
dY1/dt &= (a_{11}Y1^{a_{12}} + a_{13}Y1)\, e^{a_{14}Y2 + a_{15}Y3} \\
dY2/dt &= (a_{21}Y2^{a_{22}} + a_{23}Y2)\, e^{a_{24}Y3} \\
dY3/dt &= (a_{31}Y3^{a_{32}} + a_{33}Y3).
\end{aligned} \qquad (17)
$$

One must often compromise between a desire for the realism that would favor equation (17), and the often limited availability of historical data that would favor equation (16). Involved is the question concerning the evolution of r.h.s. from first approximations based on limited data to more refined r.h.s. as more information is gathered concerning the processes involved. It has been suggested that there is a need for a dynamical yield (standing crop) function revised periodically to incorporate results from new practices (Moser and Hall 1969). This may involve updating parameters as well as substituting more realistic r.h.s.

SOME QUALITATIVE PROPERTIES OF SYSTEM EQUATIONS

Earlier is was stated that " . . . the researcher must rationalize the form of the r.h.s. . . . on the basis of his knowledge of the processes involved." It is not sufficient to simply hypothesize a governing equation form and then test it with observational data. A valuable and potentially enlightening first step is to examine the behavior of governing equation solutions to eliminate from further consideration all equations that violate irrefutable biological "laws." Because the goal in rationalizing the governing equations is to identify the causal pathways of system component interaction, nonviolation of these biological "laws" is a necessary, but not sufficient, condition for attaining this goal.

It is instructive to ask what, in terms of a system model, is measured. Certainly parameters are not measured directly nor is instantaneous change. Rather, the standing crop (system state) or some attribute thereof is measured. Because the solution may not be expressable in closed form, the following discussion is concerned with inferring qualitative properties of governing equation solutions from the r.h.s. of the governing equation. One-dimensional models are discussed first.

The properties the solution should possess are often apparent. For solutions describing undisturbed standing-crop development, such as standing crop in survivor trees, these properties might be required: (1) nonnegativity for nonnegative values of the independent variable, (2) upper boundedness for nonnegative values of the independent variable, and (3) monotonic nondecreasing for nonnegative values of the independent variable. Clearly, properties (2) and (3) imply an upper asymptote. If property (1) was extended to include all values of the independent variable, then (1), (2), and (3) imply an upper and lower asymptote.

If the discussion is extended to include standing crop development in general, not just survivor standing crop, there is no universal agreement on the suitability of an upper asymptote. MacKinney *et al.* (1937) state, " . . . the yield (standing crop) curve is limited between zero . . . at inception and a finite maximum . . . at that advanced age before the stand commences to break up." Implicit is that the amount of standing crop does not remain at the maximum level indefinitely.

If it does not remain at the maximum (asymptotic) level indefinitely it may decrease monotonically or through oscillations. A decreasing standing crop is not considered in this study.

Let us discuss solutions of one-dimensional models, types given by equations (4) and (5), in terms of properties (1), (2), and (3). As an example of the form in equation (5) consider equation (9), and for form in equation (4) consider a polynomial approximation to a current annual increment curve such as equation (18),

$$dY/dt = at^2 + bt + c \qquad (18)$$
$$a<0,\ b>0,\ c>0.$$

Consider the requirement that the solution be nonnegative. In the case of equation (9); i.e.,

$$dY/dt = a\,Y^c + b\,Y$$
$$a<0,\ c>0,\ b>0,$$

it is clear that given a positive initial condition, standing crop cannot be negative. If Y ever becomes zero, it cannot deviate from zero. Equation (18) contains no such restriction, however, and standing crop may be negative.

Consider next the property of upper boundedness. How is this property of the solution inferred from equations (9) and (18)? Clearly, boundedness is a property of the solution of (9) if at any point

$$a\,Y^c + b\,Y = 0 \qquad (19)$$

The upper bound corresponds to the value of Y at which equation (19) is true. In equation (18), standing crop at time t_1 corresponds to the area under the polynomial curve between $t = 0$ and $t = t_1$. Upper boundedness is a property of the solution if $a < 0$. This will generally be the case; hence, equations of the form (18) possess upper boundedness.

In many cases, especially when the state variables represent physical dimensions of surviving trees, the solution should possess a monotonic increasing character, hence, property (3). How is this property reflected in governing equations such as (9) and (18)? Monotonicity is a property of the solution if the r.h.s. of the governing equation is greater than zero for all $t > 0$. In the case of equation (9), it cannot be otherwise. For equation (18), monotonicity is a property of the solution only if $a > 0$. If the coefficients a, b, and c have been determined from actual data, this will usually not be the case; hence, monotonicity is not a property of governing equations like equation (18).

These properties, nonnegativity, upper boundedness, and positive monotonicity, are but a few of many qualitative properties of solutions that may be inferred by considering the governing equations. Note that the solutions of governing equations (9) and (18) are expressible in closed form, hence their qualitative properties need not have been investigated via the governing equations. These examples were used since it is easy to verify the conclusions.

As an example of a governing equation, without a solution expressible in closed form, consider

$$dY/dt = a\,Y\,e^{bY} \qquad (20)$$
$$a>0,\ b<0.$$

Given a positive initial condition, $Y(0)$, the solution is clearly monotonic increasing. The upper bound occurs at that value of Y for which

$$a\,Y\,e^{bY} = 0.$$

There exists no value of Y for which this equality is true, hence there is no upper bound. If $Y(0) = 0$, it cannot deviate from zero, hence $Y(0) \geq 0$ insures that the solution of governing equation (20) is monotonic nondecreasing. Further examination would reveal other properties.

A governing equation solution may be characterized as an element of one or more of the following sets (fig. 2). It is apparent that the solution of equation (9) is an element of the set formed by the intersection of sets 1, 2, and 3 (region A in figure 2). Likewise, the solution of equation (20) is a member of sets 1 and 3 but not 2; i.e., region B.

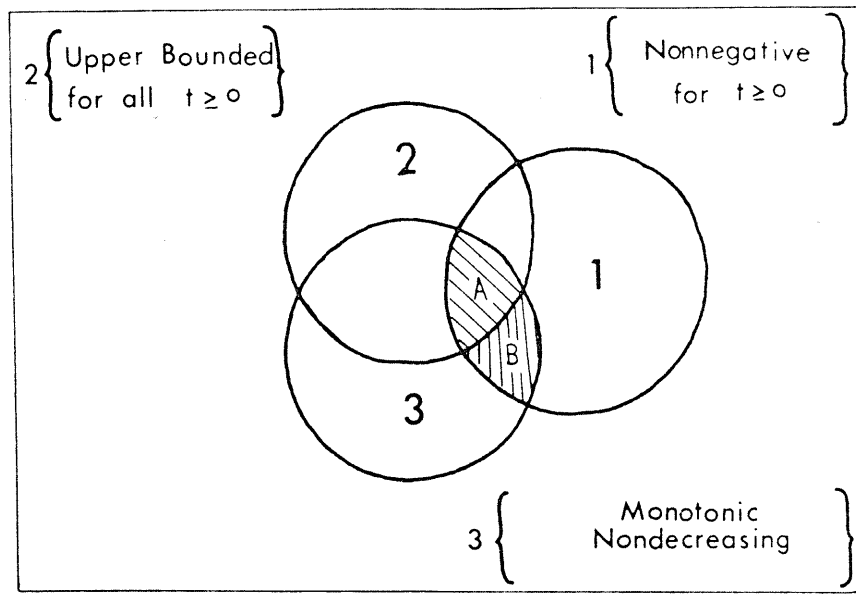


Figure 2. — Relationship of properties 1, 2, and 3 for solution of governing equation.

What conclusions may be drawn from the previous discussion? One is that governing equations, the solutions of which are contained in region A, are suitable only for describing forest system components such as standing crop in survivor trees. They are not, therefore, suitable for describing total standing crop because this may involve an initial increase but subsequent decline (MacKinney *et al.* 1937).

It may be said that figure 2 characterizes a function (solution) space, a space in which the elements (points) are functions (solutions of differential equations). This is in contrast with the more familiar value space where the elements correspond to values of functions. Also, in contrast with the value space, which is finite-dimensional, many function spaces of interest are infinite-dimensional (Rosen 1967). It is obvious there are infinitely many functions in region A that are solutions of the differential equation

$$dY/dt = -aY^c + bY. \qquad (21)$$

One need only let the parameters a, b, and c vary over the positive real numbers.

The quasilinearization method used later allows choosing a starting point in function (solution) space, and constructing a sequence of functions (solutions) that in many cases converges to the function desired.

Consider briefly the qualitative properties of systems of equations. Again the purpose is to examine the r.h.s. of the differential equations and on that basis infer properties of the solution.

Of all the "irrefutable" biological laws, perhaps the one most commonly violated in modeling forest systems is that of nonnegativity. This, of course, need not be the case as shown by the following nonlinear multiplicative system:

$$\begin{aligned} dY1/dt &= (a_{11}Y1^{a_{12}} + a_{13}Y1)\, e^{a_{14}Y2 + a_{15}Y3} \\ dY2/dt &= (a_{21}Y2^{a_{22}} + a_{23}Y2)\, e^{a_{24}Y3} \\ dY3/dt &= (a_{31}Y3^{a_{32}} + a_{33}Y3). \end{aligned} \tag{22}$$

It is immediately apparent, assuming the coefficients are based on actual field data, that $Y1$, $Y2$, and $Y3$ cannot become negative. This is so because whenever $Y1$, $Y2$, or $Y3$ becomes zero, the corresponding derivative becomes zero, and furthermore cannot deviate from zero again.

The same cannot be said for the following linear additive system:

$$\begin{aligned} dY1/dt &= a_{11}Y1 + a_{12}Y2 + a_{13}Y3 \\ dY2/dt &= a_{21}Y1 + a_{22}Y2 + a_{23}Y3 \\ dY3/dt &= a_{32}Y2 + a_{33}Y3. \end{aligned} \tag{23}$$

Clearly, even though $Y1$ may be zero, $dY1/dt$ is still influenced by $Y2$ and $Y3$. In later applications $Y2$ and $Y3$ have inhibiting influences on $Y1$, so the coefficients a_{12} and a_{13} are negative, thus giving a negative derivative and eventually a negative standing crop, $Y1$. Leary[2] compensated for this problem by eliminating from the system (23) any component that became negative, a procedure suggested by Lotka (1928).

The property of boundedness for all $t \geq 0$ is a possible point of disagreement. In fact it is a very restrictive condition to impose on a solution. Although the statement may be made that everything worldly is finite, hence functions that describe standing crop should be bounded above, a question arises as to how one treats something like standing crop in surviving trees. By definition, surviving trees are living trees, and living trees do not remain static in size. Hence, it appears that by definition, functions that describe standing crop of surviving trees should be monotonic nondecreasing and unbounded above. An example of a nonlinear multiplicative model with possibly unbounded solutions is

$$\begin{aligned} dY1/dt &= a_{11}Y1\, e^{a_{12}(Y1 + Y2 + Y3)} \\ dY2/dt &= a_{21}Y2\, e^{a_{22}(Y2 + Y3)} \\ dY3/dt &= a_{31}Y3\, e^{a_{32}(Y3)} \end{aligned} \tag{24}$$

Notice the above system is nonnegative and may be unbounded above, although well behaved for all t. Due to its relative simplicity, nonnegativity, and conformity to the requirements for functions representing survivor standing crop, this is the system used in later sections to demonstrate the quasilinearization method of solving the inverse problem.

EXAMINING SYSTEM DEVELOPMENT

Let us briefly discuss methods of examining system development: i.e., of solving the governing equations. In what follows, the unknown model parameters are assumed known. The purpose of discussing these topics now is that many things common to methods of solution are used in the methods of parameter estimation discussed later.

Solution methods may be categorized as analytic, graphic, and numeric. The concepts and approaches contained in these methods are as follows:

(1) Analytic solutions: (a) the role of initial and boundary conditions in isolating a unique solution to a single equation or system of equations, (b) the solution of linear equations, and (c) the principle of superposition, which forms an integral part of the quasilinearization method.

(2) Graphic (geometric) solutions: the construction of direction fields is a useful approach for (a) exhibiting the approximate system development for a wide range of initial conditions in an easily understood manner, (b) understanding the boundary condition formulation of the inverse problem, and (c) exhibiting convergence of the quasilinearization method to the correct system parameters.

(3) Numeric solutions: the algorithms used in obtaining numerical solutions are an efficient means of examining the large systems of nonlinear equations encountered in modeling forest systems.

Several of the above areas are the subjects of books in themselves. The purpose is, therefore, to introduce the ideas involved through examples, give a brief discussion of applications, and refer the reader to authoritative sources for a more thorough discussion.

Analytic Solutions

It is perhaps well to distinguish between an analytic formulation of the problem and its analytic solution. An analytic formulation is used in all cases regardless of the solution method (analytic, graphic, or numeric). Conceivably, problems could be formulated verbally, which might be a good beginning for someone with little experience in this area, but they must be converted to analytic form for solution. As a simple example consider: "The growth rate of an organism is directly proportional to the amount of it that is present." Expressing this in analytic form gives

$$dY(t)/dt = a\,Y(t), \tag{25}$$

where a is a constant of proportionality.

An analytic solution, in the context of the above example, is a function $Y(t)$ which, when differentiated with respect to time, results in equation (25). Such a function is easily obtained by direct integration of both sides of

$$dY(t)/aY(t) = dt; \text{ i.e.,}$$

$$\int \frac{dY(t)}{aY(t)} = \int dt, \text{ or}$$

$$\ln Y(t) = at + C_4, \text{ where } C_4 = a(C_2 - C_1), \text{ or}$$

$$Y(t) = e^{at}\, e^{C_4}, \text{ or}$$

$$Y(t) = C_5\, e^{at}, \tag{26}$$

where $C_5 = e^{C_4}$ and is the constant of integration.

The role of C_5 in the complete solution (26) may be seen from graphs of $Y(t) = C_5 e^{at}$ for various values of C_5 (fig. 3). Clearly, there is a family of solutions. How does one

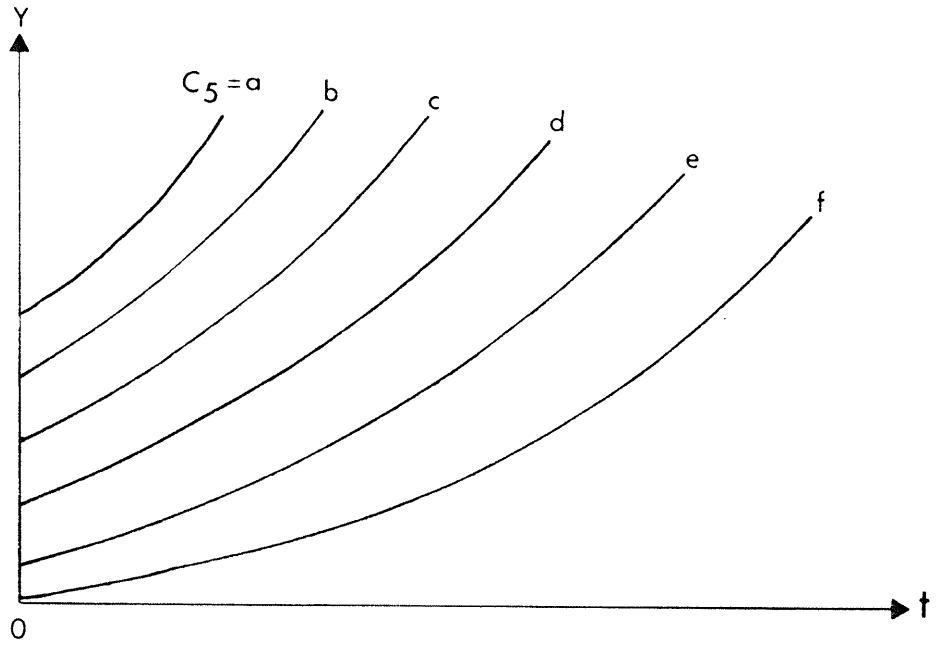


Figure 3. — Family of solutions of equation 25.

specify a particular element of the family shown in figure 3 as the unique solution? This is accomplished by specifying a constraint on the solution $Y(t)$ at a value of t. The common case is where one specifies $Y(t) = b$ when $t = 0$. This is an initial condition requirement that $Y(t) = C_5 e^{at}$ must meet; i.e.,

$$Y(0) = b = C_5 e^{a0}, \text{ or } b = C_5. \tag{27}$$

Thus, the unique solution of

$$dY(t)/dt = a\,Y(t), \text{ with}$$
$$Y(0) = b, \text{ is}$$
$$Y(t) = b\,e^{at}.$$

For a more involved example, consider the generalization of von Bertalanffy's equation (9), for which a solution for Z in terms of t is desired. Starting with equation (11) and multiplying through by $-b\,dt/(a - bZ)$ gives

$$-b\,dZ/(a - bZ) = -b(1-c)dt. \tag{28}$$

Integrating both sides gives

$$\ln(a - bZ) + C_1 = -b(1-c)t + C_2. \tag{29}$$

Solving for Z, we have

$$Z = \frac{a}{b} - \frac{C_3}{b} e^{-b(1-c)t},$$
$$C_3 = e^{C_2 - C_1}.$$

Thus

$$Y = \left[\frac{a}{b} - \frac{C_3}{b} e^{-b(1-c)t}\right]^{\frac{1}{1-c}} \tag{30}$$

If $Y_t = Y_0$ at $t = 0$,

$$Y_t = Y_0 = \left[\frac{a}{b} - \frac{C_3}{b} e^{0}\right]^{\frac{1}{1-c}}, \text{ or } Y_0{}^{1-c} = \left[\frac{a}{b} - \frac{C_3}{b}\right], \text{ and}$$

$$Y = \left[\frac{a}{b} - \left(\frac{a}{b} - Y_0{}^{1-c}\right) e^{-b(1-c)t}\right]^{\frac{1}{1-c}} \tag{31}$$

The above examples were chosen so that they would be simple, yet exhibit what is meant by analytic solutions and how they are obtained; i.e., using basic rules of integration, derived in elementary calculus. There are many special techniques for finding analytic solutions of differential equations, for instance, use of integrating factors, method of undetermined coefficients, method of variation of parameters, and others. They are mentioned here because these are the techniques that are taught in the standard courses in differential equations. But for our purposes they are of limited value.

The reader should clearly understand the role that initial conditions play in obtaining unique solutions, because later two-point and multipoint boundary value problems are discussed that presuppose an understanding of these ideas.

Solutions to homogeneous and nonhomogeneous linear differential equations are required to implement the quasilinearization method of solving multipoint boundary value problems. Hence, it is appropriate to say something about the form of their solution.

Let $L(D)$ be a linear differential operator (Rainville 1964, Kaplan 1964),

$$L(D) = a_n(t)D^n + a_{n-1}(t)D^{n-1} + \ldots + a_1(t)D + a_0,$$

where $D = d/dt$, $D^k = d^k/dt^k$,

and consider the differential equations

$$\begin{aligned} L(D)Y &= R(t) \qquad (L) \\ L(D)Y &= 0 \qquad (LH). \end{aligned} \tag{32}$$

LH is called the homogeneous or complementary equation associated with L.

It is possible to write down the entire set of solutions of L in the following way: (1) Determine any one solution of L, called a particular solution, $Y = Y_p(t)$. (2) Determine the set of all solutions to LH. Let it be $Y = Y(t,c_1,c_2,c_3,\ldots,c_n)$ where the c_i are arbitrary constants. (3) Then $Y = Y_p(t) + Y(t,c_1,c_2,\ldots,c_n)$ is the complete set of solutions to L. The function $Y(t,c_1,c_2,\ldots,c_n)$, called the complementary function, or Y_c, has a standard form also. It is $Y_c = c_1y_{h1}(t) + c_2Y_{h2}(t) + \ldots + c_ny_{hn}(t)$, where $y_{h1}(t) \ldots y_{hn}(t)$ are n linearly independent solutions of LH.

As an example consider:

$$\begin{aligned} Y'' + Y &= t^2 \qquad (L) \\ Y'' + Y &= 0 \qquad (LH) \end{aligned} \tag{33}$$

1. $Y = t^2-2$ is a solution of L

2. $y_{h1} = \cos t$, $y_{h2} = \sin t$ are linearly independent solutions of LH, therefore

$$Y = t^2 - 2 + c_1 \cos t + c_2 \sin t \tag{34}$$

is the complete solution of L. The arbitrary constants c_1 and c_2 may be specified by initial or boundary conditions.

Consider a system of linear first order equations similar to those of equation (16):

$$\begin{aligned} dY1/dt &= a_{11}Y1 + a_{12}Y2 + a_{13}Y3 \\ dY2/dt &= a_{21}Y1 + a_{22}Y2 + a_{23}Y3 \\ dY3/dt &= a_{31}Y1 + a_{32}Y2 + a_{33}Y3. \end{aligned} \tag{35}$$

It is known that the solution of linear constant coefficient differential equations such as (35) contain terms of the form e^{rt}. Let them be as follows:

$$\begin{aligned} Y1 &= A\, e^{rt} \\ Y2 &= B\, e^{rt} \\ Y3 &= C\, e^{rt}. \end{aligned}$$

Using these as trial solutions of (35) gives, upon substitution,

$$\begin{aligned} rA\, e^{rt} &= a_{11}A\, e^{rt} + a_{12}B\, e^{rt} + a_{13}C\, e^{rt} \\ rB\, e^{rt} &= a_{21}A\, e^{rt} + a_{22}B\, e^{rt} + a_{23}C\, e^{rt} \\ rC\, e^{rt} &= a_{31}A\, e^{rt} + a_{32}B\, e^{rt} + a_{33}C\, e^{rt}. \end{aligned} \tag{36}$$

Dividing through each equation by e^{rt} and transposing gives

$$\begin{aligned} (a_{11}-r)A + a_{12}B + a_{13}C &= 0 \\ a_{21}A + (a_{22}-r)B + a_{23}C &= 0 \\ a_{31}A + a_{32}B + (a_{33}-r)C &= 0. \end{aligned} \tag{37}$$

A homogeneous system of algebraic equations has nontrivial solutions if and only if the determinant of the coefficient matrix equals zero; i.e.,

$$\begin{vmatrix} a_{11}-r & a_{12} & a_{13} \\ a_{21} & a_{22}-r & a_{23} \\ a_{31} & a_{32} & a_{33}-r \end{vmatrix} = 0. \tag{38}$$

Expanding the determinant gives a polynomial in r. The three roots of this polynomial, called the characteristic polynomial, are known as the characteristic numbers or eigenvalues of the matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{39}$$

For simplicity in the following discussion, it is assumed that the r_i, $i = 1, 2, 3$ are real and distinct.

The eigenvalues of a matrix A are those scalar numbers r for which

$$Ax = rx, \tag{40}$$

where x is a column vector,
A is a matrix, and
r is a scalar.

This may be put into a more familiar form as

$$(A - rI)x = 0, \tag{41}$$

where I is the identity matrix and 0 is the zero column vector. This is a homogeneous matrix equation that has nontrivial solutions if and only if

$$\det(A - rI) = 0.$$

Clearly, this is the same form as equation (38). Associated with each eigenvalue r is a vector x for which equation (40) holds. This vector is called the associated eigenvector.

Returning to equation (35) a general solution may be written down directly. It is

$$\begin{matrix} Y1 \\ Y2 \\ Y3 \end{matrix} = k_1 \begin{bmatrix} G_{11} \\ G_{12} \\ G_{13} \end{bmatrix} e^{r_1 t} + k_2 \begin{bmatrix} G_{21} \\ G_{22} \\ G_{23} \end{bmatrix} e^{r_2 t} + k_3 \begin{bmatrix} G_{31} \\ G_{32} \\ G_{33} \end{bmatrix} e^{r_3 t} \tag{42}$$

where r_1, r_2, and r_3 are the eigenvalues of matrix (39), and
G_{1i}, G_{2i}, and G_{3i} are the associated eigenvectors.

The constants k_i, $i = 1, 2, 3$ are determined by initial conditions on $Y1$, $Y2$, and $Y3$.

The relevance of this discussion may be seen as follows. It is clear from equation (40) that, for a specified value of r, say, $r = 2.5$, multiplying the vector x by the matrix A has the same effect as multiplying x by the scalar 2.5. Multiplying a vector by a positive scalar increases its length, but does not change its direction. Hence, eigenvalues serve as expanders or contractors of the eigenvectors in equation (42). The eigenvectors G_{1i}, G_{2i}, and G_{3i} are linearly independent (a fundamental result from linear algebra for distinct eigenvalues), and span the 3-space. Hence, they form a basis of this space (Perlis 1952). This means that any point in this space may be located by a linear combination of the basis vectors, G_{1i}, G_{2i}, G_{3i}. (The most familiar basis for 3-space is the orthogonal set

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}).$$

Consider again equation (42). At a particular time, say t_1, the vectors G_{ij} have a specified length and direction. Using basic rules for vector addition (i.e., the parallelogram law), the sum of three vectors is again a vector that locates a point in space, in this case, our state space. After an elapse of time, say of length Δt, this point will have moved to a different position in state space, the distance from that at time t being determined by

$$e^{r_1\Delta t},\ e^{r_2\Delta t},\ \text{and}\ e^{r_3\Delta t}.$$

Clearly, the eigenvalues r_1, r_2, and r_3 determine how much the system state changes in a specified time period Δt.

At this point the following question seems relevant: Does it not stand to reason that, everything else held constant, forest stands on good sites should progress faster through state space than stands on poor sites? If this is so, does it not follow that a model such as (35) fit to historical data from stands on good sites would have larger eigenvalues associated with its coefficient matrix than the same model fit to historical data from stands on poor sites?

Computing the eigenvalues for models comprised of many equations can be bothersome, because the equations will usually be integrated numerically. To secure the essential information desired concerning the size of the eigenvalues and to circumvent the problem of computing them, use is made of the following result from linear algebra (Faddeev and Faddeeva 1963). The trace of a matrix is identical to the sum of its eigenvalues. In terms of the above example,

$$\text{trace (A)} = a_{11} + a_{22} + a_{33} = r_1 + r_2 + r_3.$$

Thus, the trace of a coefficient matrix for a model such as (35) may well be a measure of the productivity of the site on which the stand is growing.

Graphic or Geometric Solutions

Perhaps in the strictest sense of the word "solution," graphic methods are not sufficiently precise. Nonetheless, they are useful in giving a picture of the solution's behavior over a wide range of initial conditions. Before discussing the basic method, recall that the solution of

$$dY(t)/dt = aY(t) \qquad (43)$$

is a locus of points in 2-space corresponding to different values of Y and t. This was seen in figure 3.

The method of tangents or direction fields gives a representation of this locus of points by the construction, at a suitable grid of points in Y - t space, of a short line segment with the appropriate slope, $dY/dt|_{Y,t}$. If the grid points are sufficiently close together, the locus of points corresponding to the true solution would possess a set of lines tangent to it. The logic

in using the method of tangents is that one constructs the "tangent" field, and from it infers the locus of points that would be the solution for any given starting value (initial condition). Figure 4 shows a schematic direction field for equation (43). Obviously, the slope increases as Y increases, and time is not a factor in its determination.

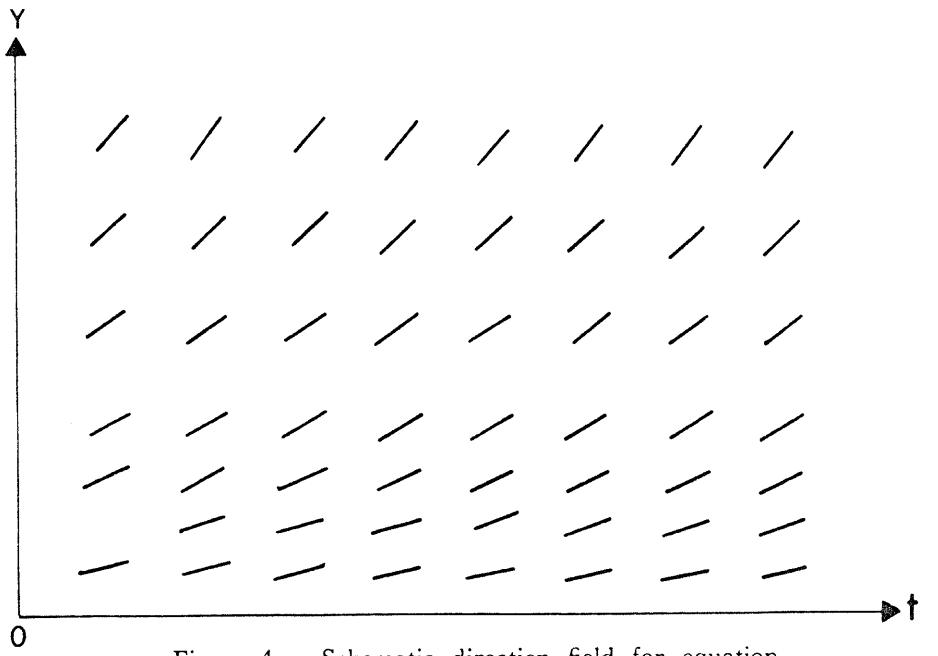


Figure 4. — Schematic direction field for equation (43).

Clearly, the form of the solution curves in figure 3 could be inferred from the direction field in figure 4.

Numeric Solutions

A numeric solution of a differential equation is a tabulated set of values with, in the case of equation (43), the values of Y and t at which it occurred. The table could, conceivably, consist of but two entries; i.e., Y_0, t_0 the beginning conditions, and Y_f, t_f, the final conditions. Rapid access to intermediate solution values and associated times, as well as initial and terminal conditions is a practical necessity for the methods of later sections, hence they are stored internally in the central memory of the computer.

The problem of developing algorithms, of which there are many, for solving differential equations numerically is indeed vast. The purpose at this point is to indicate that the single-step method of Runge-Kutta has been found satisfactory in the numerical solutions required in employing the quasilinearization method. Briefly, the Runge-Kutta (fourth order) method uses information about the solution and the slope of the solution at a single point (value of the independent variable) to project ahead a short distance, Δt, along the time axis, in the case of equation (43). The information concerning the solution and its slope at time t_1 is contained in the following constants:

$$m_1 = f(Y(t_1), t_1)\ \Delta t$$
$$m_2 = f(Y(t_1) + m_1/2, t_1 + \Delta t/2)\ \Delta t$$
$$m_3 = f(Y(t_1) + m_2/2, t_1 + \Delta t/2)\ \Delta t$$
$$m_4 = f(Y(t_1) + m_3, t_1 + \Delta t)\ \Delta t.$$

They are combined as follows to project the value of Y at t_2: $Y(t_2) = y(t_1) + 1/6\ (m_1 + 2m_2 + 2m_3 + m_4)$.

The above equations may be generalized directly to handle systems of differential equations (Conte 1965, Henrici 1964, Ralston 1965).

A fourth-order Runge-Kutta (RK) procedure was used in the examples of later sections. The routine, RKGS, from the Scientific Subroutine Package (IBM), contained a facility for doubling or halving the step size according to an error criterion specified by the user. This feature proved useful in detecting unstable systems, thus allowing the procedure to be terminated before excessive computing time was consumed.

On the surface, one might expect a multipoint method, such as a predictor-corrector, to be faster and cheaper, because it requires evaluation of the r.h.s. of the differential equation only twice per step, as opposed to four times for the RK method. However, predictor-corrector methods are not self-starting and it is common practice to use a fourth-order RK routine to compute the first four solution values, Y_{t_1}, Y_{t_2}, Y_{t_3}, Y_{t_4}, which the predictor-corrector then uses to project ahead to Y_{t_5}. As the routines are applied in later sections, four applications of the RK procedure traverses the entire integration interval, with the result that the predictor-corrector scheme is never implemented. Hence, a simple, self-starting, single-step, fourth-order RK method is used.

CHARACTERIZING THE SYSTEM STATE

Much time was spent examining how the system may be characterized in a manner that does not violate certain biological principles and how a system may be projected through time; i.e., by solving the governing equation. It was shown that at any point in time, t_k, the state of the system is given by the vector of state variables

$$[Y1(t_k)\quad Y2(t_k)\quad Y3(t_k)\quad \ldots\quad Yn(t_k)]^T.$$

This is a convenient means of expressing the system state; however, for our purposes it lacks sufficient conciseness for ease in mathematical analysis. A measure of the above vector that satisfies our needs is a vector norm.

Perhaps the most familiar vector norm is the Euclidean norm defined as (Faddeev and Faddeeva 1963),

$$||Y||_3 = [(Y1^2 + Y2^2 + Y3^2 + \ldots + Yn^2)]^{\frac{1}{2}}, \qquad (44)$$

or the distance of the point ($Y1\ Y2\ Y3\ \ldots\ Yn$) from the origin of our state space. There is no obvious meaning to the Euclidean norm for our purposes, hence, attention is directed at what is called the second vector norm, defined as

$$||Y||_2 = |Y1| + |Y2| + |Y3| + \ldots + |Yn|, \qquad (45)$$

where $|Y1|$ is, in our case, the absolute value of the standing crop in component $Y1$. Notice that if our governing equations possess the property of nonnegativity, equation (45) may be simplified to

$$||Y||_2' = Y1 + Y2 + Y3 + \ldots + Yn,$$

which is, of course, total standing crop (assuming the units of Yi are conformable for addition).[2] Likewise, given two adjacent points in state space corresponding to times t_1 and t_2,

$$||Y(t_2)||_2' - ||Y(t_1)||_2'$$

gives growth in total standing crop.

A geometric picture of system (stand) development through time may be obtained if the sequence of state vectors

$$\begin{bmatrix} Y1(t_1) \\ Y2(t_1) \\ Y3(t_1) \\ \cdot \\ \cdot \\ \cdot \\ Yn(t_1) \end{bmatrix}, \begin{bmatrix} Y1(t_2) \\ Y2(t_2) \\ Y3(t_2) \\ \cdot \\ \cdot \\ \cdot \\ Yn(t_2) \end{bmatrix}, \ldots, \begin{bmatrix} Y1(t_n) \\ Y2(t_n) \\ Y3(t_n) \\ \cdot \\ \cdot \\ \cdot \\ Yn(t_n) \end{bmatrix}$$

is interpreted in a geometric manner.

One such manner is to present these vectors as bar charts (Knuchel 1953). Or one might convert the bar charts to a series of polygons by connecting the midpoints for each class. Superimposing several of these polygons on the same coordinate system shows that in the case of even-aged forest stands development is characterized by the propagation of a "wave" across the size (d.b.h.) classes (fig. 5).[2]

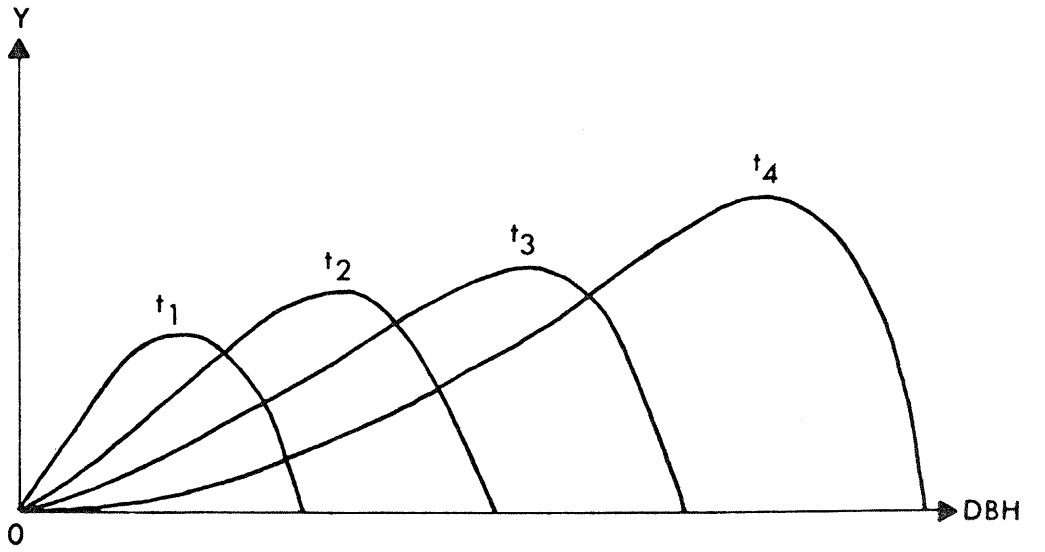


Figure 5. — Forest stand development in terms of wave propagation.

The interpretation employed in the remainder of this paper is that of the sequence of state vectors, each element of which locates a point in state space, tracing out a trajectory in

state space. A simple example for three-dimensional space is shown in figure 6. Clearly, given an initial condition (point in state space), the question of solving the governing equation(s) is equivalent to determining the trajectory the system will follow through time.

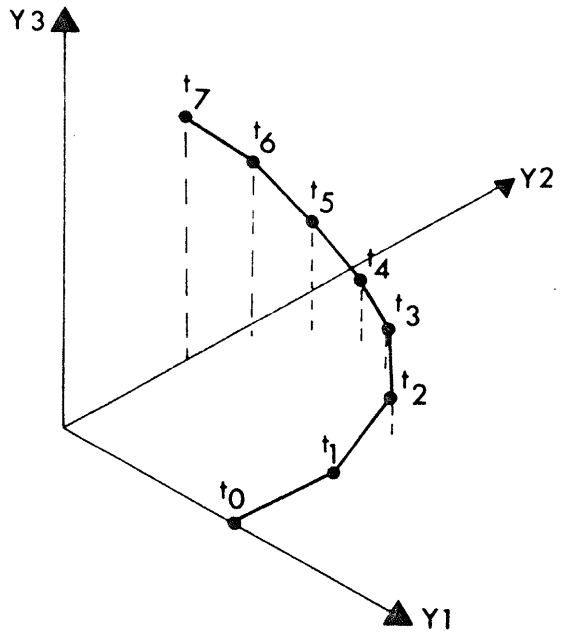


Figure 6. — Hypothetical forest stand development in terms of point transformations.

THE INVERSE PROBLEM

The problem central to constructing dynamic mathematical models of forest systems is the inverse of that just stated, that is, instead of asking "given certain governing equations for our system, what trajectory will the system follow?", one asks, "given observations on the system trajectory, what are the governing equations?" In its most basic form, the inverse problem may involve unknown forms of the governing equations. However, at this point the assumption is that the form of the governing equations is known, but numerical values of the parameters are not. Thus, our inverse problem may be rephrased as follows: "Given observations on the system trajectory, what are the numerical values for governing equation parameters that put the predicted system states in the desired conformity with the observed system states?"

What is the real world relevance of the inverse problem? Basically it rests on man's desire to exert control over his environment because of his preference for certain system states. In this case states take the form of stand structure, species composition, etc. As manipulators of forest systems, land managers may ask if leaving the system undisturbed will result in desired future states. Future states can be predicted by projecting the system into the future using past system states as a guide in determining governing equation parameters. If projected future states are not those most desired, a decision must be made as to how, given the resources available, a manager can alter the present state to insure the desired future states. This line of discussion leads naturally to the questions of multistage decision processes and optimal control, which are interesting but not part of our theme.

Boundary value problems are discussed in the next section, and we see that by using observations on the system trajectory as boundary conditions, the inverse problem may be solved. That is, the equations governing our system may be completely specified.

BOUNDARY CONDITION FORMULATION OF INVERSE PROBLEM

By specifying an initial condition for our governing equation, a particular element from a family of solutions may be isolated. See figure 7 for a schematic example using the generalized von Bertalanffy equation. For a further example, consider the following growth function similar to one suggested by Roston (1962):

$$dY/dt = c_1 Y - c_2 \int_o^t Y \, dt. \tag{46}$$

The growth of Y is positively related to the amount of Y present, but is inhibited by an accumulated proportion of Y. The inhibiting component may reflect an encroachment on available growing space (ecosystem resources) or an accumulation of toxic compounds that may be by-products of metabolism or environmental pollutants.

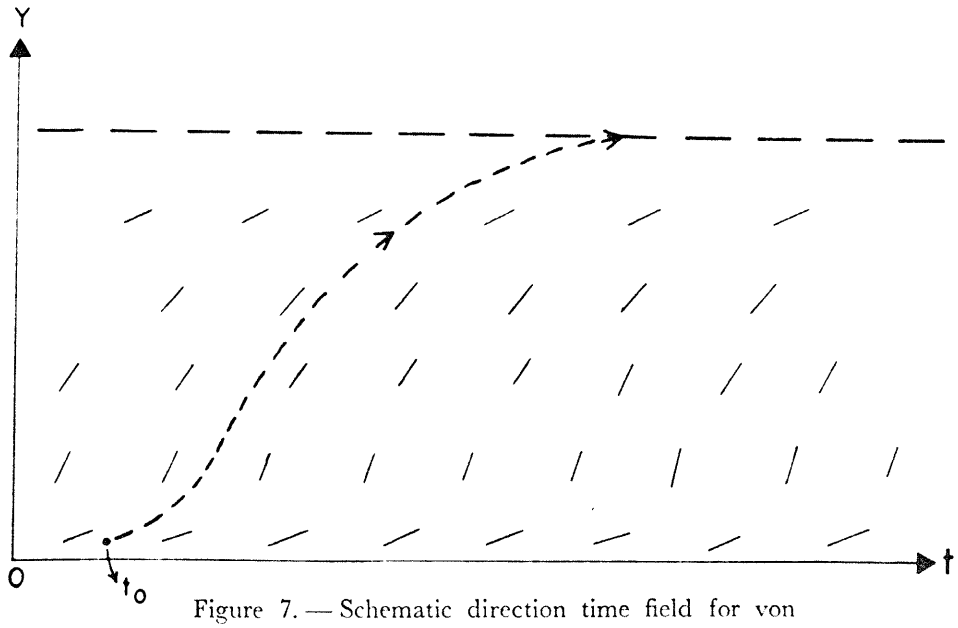


Figure 7. — Schematic direction time field for von Bertalanffy's equation showing role of initial condition.

Differentiation of equation (46) using Leibnitz's rule gives the linear homogeneous second-order differential equation

$$d^2Y/dt^2 - c_1 dY/dt + c_2 Y = 0. \tag{47}$$

Recall from the section on "Examining System Development" that such an equation has a solution comprised of a linear combination of homogeneous solutions of the form

$$Y = B_1 e^{r_1 t} + B_2 e^{r_2 t}, \tag{48}$$

where r_1 and r_2 are roots of the characteristic equation

$$z^2 - c_1 z + c_2 = 0.$$

In this case B_1 and B_2 are determined from initial conditions that involve Y and dY/dt; e.g.,

$$\begin{aligned} Y(0) &= a \\ dY/dt\Big|_{t=0} &= b. \end{aligned}$$

This is an initial value problem, because the constraints on Y and dY/dt are at time zero.

Consider now the following changes in constraints on the solution of equation (47). Instead of specifying both conditions at $t = 0$, one is specified at time $t = 0$ and the other at $t = t_f$. Furthermore, Y must meet both of these constraints instead of dY/dt meeting one as above. Thus, the problem is to solve

$$d^2Y/dt^2 - c_1 dY/dt + c_2 Y = 0, \tag{49}$$

with boundary conditions

$$\begin{aligned} Y(0) &= a \\ Y(t_f) &= b. \end{aligned} \tag{50}$$

Equations (49) and (50) constitute a linear two-point boundary value problem for which there are straightforward methods of solution.

However, many of the two-point boundary value problems that arise in the physical sciences, the calculus of variations, and control theory are nonlinear, and have no straightforward methods of solution. Some of these may, however, be solved by the quasilinearization method.

The solution of (47) may be placed in the framework of a system of differential equations by making the substitution, $Z = dY/dt$, and in place of equation (49) constructing the system of two first-order equations

$$\begin{aligned} dZ/dt &= c_1 Z - c_2 Y \\ dY/dt &= Z \end{aligned} \tag{51}$$

with the boundary conditions

$$\begin{aligned} Y(0) &= a \\ Y(t_f) &= b. \end{aligned} \tag{52}$$

Clearly, Y has two constraints and Z none. Successful numerical solution of (51) subject to (52) hinges on the ability to select a value for Z at $t = 0$ such that when $t = 0$, $Y = a$, and when $t = t_f$, $Y = b$.

If constraints on the solution exist at more than two points, it is referred to as a multipoint boundary value problem. As an example, consider the nonlinear first-order equation

$$dY/dt = aYe^{bY} \tag{53}$$

with boundary conditions

$$Y(t_1) = m_1, \quad Y(t_2) = m_2, \quad Y(t_3) = m_3. \tag{54}$$

In our applications, the constant parameters a and b in equation (53) are considered as unknowns and the following changes in the form of equation (53) are made:

(1) Consider a and b to be time-dependent variables; i.e., functions of time of the form

$$\begin{aligned} da/dt &= 0 \\ db/dt &= 0, \text{ and} \end{aligned}$$

(2) convert the single equation to a system of equations. Thus, a simple parameter estimation problem that will be solved in the following section is the multipoint boundary value problem

$$\begin{aligned} dY/dt &= aYe^{bY} \\ da/dt &= 0 \\ db/dt &= 0 \end{aligned} \tag{55}$$

with boundary conditions

$$Y(t_i) = m_i, \quad i = 1, 2, 3. \tag{56}$$

Here again, successful numerical solution rests on selecting values for $a(0)$ and $b(0)$ such that conditions (56) are met, where m_i are observed system states.

Likewise for systems of equations:

$$\begin{aligned} dY1/dt &= a\ Y1\ e^{b(Y1 + Y2 + Y3)} \\ dY2/dt &= c\ Y2\ e^{d(Y1 + Y3)} \\ dY3/dt &= f\ Y3\ e^{g(Y3)} \\ da/dt &= 0 \\ db/dt &= 0 \\ dc/dt &= 0 \\ dd/dt &= 0 \\ df/dt &= 0 \\ dg/dt &= 0 \end{aligned} \tag{57}$$

with boundary conditions

$$\begin{aligned} Y1(t_i) &= m_{1i}, \quad i = 1, 2, 3 \\ Y2(t_i) &= m_{2i}, \quad i = 1, 2, 3 \\ Y3(t_i) &= m_{3i}, \quad i = 1, 2, 3. \end{aligned} \tag{58}$$

Our approach will again be to select values for a, b, c, d, f, and g at $t = 0$, such that conditions (58) are met.

It is highly unlikely that values for a, b, c, d, f, and g at $t = 0$ could be selected such that $Y1\ (t_i) = m_{1,i}$, $Y2(t_i) = m_{2,i}$, $Y3(t_i) = m_{3,i}$ for all i. Hence, it is important that there is a means of making an initial estimate for a, b, c, d, f, and g, and, furthermore, a means of improving these estimates until a criterion of goodness of fit between $Y1(t_i)$, $Y2(t_i)$, and $Y3(t_i)$, and $m_{1,i}$, $m_{2,i}$, $m_{3,i}$, respectively, is met.

The quasilinearization method considered in the next section provides us with such a procedure. First, however, it is advisable to review some basic concepts concerning iteration functions.

Use of an iteration function (I.F.) involves making an initial guess at the answer, substituting it into an I.F., and getting a new approximation to the answer. The new approximation is then substituted into the I.F. and another approximation is obtained. This process is repeated until the value substituted into the I.F. is the same (to however many digits accuracy one desires) as the value given by the I.F. as the new approximation.

An I.F. is any function of the form $x^{(k+1)} = F(x^{(k)}, x^{(k-1)}, x^{(k-2)}, \ldots)$, where the superscript indicates the independent variable at the iteration. One seeks a "fixed point" of the I.F., which is that value of x, say α, such that $\alpha = F(\alpha)$. Thus, when $x = \alpha$, the iteration has converged.

There are many types of I.F.'s, the primary differences among them being (a) their interval and rate of convergence, (b) the number of previous iterations that are used, and (c) the number of initial approximations that must be specified to "start" the process. One might say that an ideal method would be one that is globally convergent, of at least second order (i.e., converges quadratically), and needs only one initial approximation to get started.

In what follows, the Newton-Raphson method is discussed in relation to (b) and (c) above. Using Traub's (1964) classification of I.F.'s, the Newton-Raphson method is a one-point function requiring evaluation of the derivative. "One-point" means the I.F. is of the form $x^{(k+1)} = F(x^{(k)})$, that is, only the previous iterate is used in obtaining a new approximation. It is well to note that most I.F.'s use information about the shape of the curve of the underlying function to speed convergence.

The form of the Newton-Raphson I.F. may be developed in the following way. Consider the function f shown in figure 8. The value of x at which $f(x) = 0$ is sought. A first approximation is obtained by selecting a value for x, $x^{(1)}$, and evaluating f; i.e., $f(x^{(1)})$. Using the point-slope formula we have

$$Y - f(x^{(1)}) = f'(x^{(1)})(x - x^{(1)}). \tag{59}$$

Setting $Y = 0$ and solving for x gives

$$x = x^{(1)} - f(x^{(1)})/f'(x^{(1)}). \tag{60}$$

If x is closer to α than $x^{(1)}$ — i.e., $|x - \alpha| < |x^{(1)} - \alpha|$ — we take x as our new estimate of the root and repeat the process. In general this gives

$$x^{(n)} = x^{(n-1)} - f(x^{(n-1)})/f'(x^{(n-1)}), \tag{61}$$

which is the Newton-Raphson iteration function, and is abbreviated

$$x^{(k)} = F(x^{(k-1)}),$$

where F corresponds to the right-hand side of equation (61).

In the next section it will be seen that the quasilinearization method is abstractly equivalent to the Newton-Raphson method. A fundamental difference, however, is that in the discussion above, the concern is with approximations (to the root α) in *value* space, whereas with the quasilinearization method the concern is with Newton-Raphson type approximations in *function* space. The concern, then, is with convergence to a function rather than to the value (root) of a function.

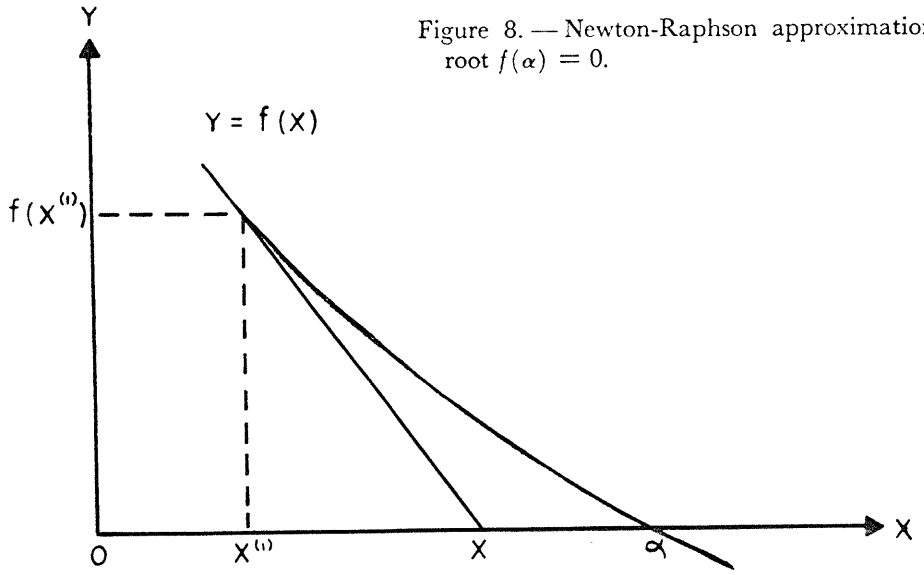


Figure 8. — Newton-Raphson approximation to the root $f(\alpha) = 0$.

SOLUTION OF THE INVERSE PROBLEM USING QUASILINEARIZATION

The approach in solving the following inverse problems is one of (a) converting nonlinear differential equations into approximately equivalent linear equations, and (b) solving the resulting linear boundary value problem using the method of complementary functions and standard numerical integration techniques. Why use this approach? It is because there are straightforward methods for solving linear boundary value problems with constant or variable coefficients. To see this, consider the following example of a linear nonhomogeneous second-order equation with variable coefficients.

$$d^2Y/dt^2 + P(t)\ dY/dt + Q(t)\ Y = R(t) \tag{62}$$

with boundary conditions

$$\begin{aligned} Y(0) &= a \\ Y(t_f) &= b. \end{aligned} \tag{63}$$

Linear equation theory dictates that the complete solution of (62) be of the form

$$Y(t) = y_p + c_1 y_{h1} + c_2 y_{h2}, \tag{64}$$

where y_p is the solution of the nonhomogeneous equation, and y_{h1} and y_{h2} are independent solutions of the homogeneous equation. How, then, do we form the particular and homog-

eneous solutions? First, one may obtain a particular solution by numerically integrating equation (62) subject to the initial conditions

$$\begin{aligned} Y(0) &= a \\ dY/dt\,|_{t=0} &= 0. \end{aligned}$$

Notice that the initial condition on Y corresponds to the first boundary condition in (63). Second one may obtain a homogeneous solution of equation (62) by numerically integrating

$$d^2Y/dt^2 + P(t)\ dY/dt + Q(t)\ Y = 0 \tag{65}$$

subject to any convenient initial conditions. As an example, one may use

$$\begin{aligned} Y(0) &= 1 \\ dY/dt\,|_{t=0} &= 0. \end{aligned} \tag{66a}$$

Two linearly independent homogeneous solutions of (62) may be required, so

$$\begin{aligned} Y(0) &= 0 \\ dY/dt\,|_{t=0} &= 1 \end{aligned} \tag{66b}$$

are used as initial conditions for the second.

The above are initial value problems. Because the first boundary condition has been satisfied by $y_p(t)$, our concern now is that the second boundary condition $y(t_f) = b$ is met. This is accomplished by forming the general solution at $t = 0$ and $t = t_f$, via superposition, which gives two linear algebraic equations in two unknowns, c_1 and c_2; i.e.,

$$\begin{aligned} Y(0) &= y_p(0) + c_1\, y_{h1}(0) + c_2\, y_{h2}(0) = a \\ Y(t_f) &= y_p(t_f) + c_1\, y_{h1}(t_f) + c_2\, y_{h2}(t_f) = b. \end{aligned} \tag{67}$$

What values should c_1 and c_2 take to insure that the boundary conditions (63) are met? By substituting the initial conditions into the first equation of (67)

$$a + c_1\ 1 + c_2\ 0 = a.$$

Clearly, c_1 must be zero. If $c_1 = 0$, then c_2 must equal

$$(b - y_p(t_f))/y_{h2}(t_f)$$

if the second boundary condition is to be satisfied.

Thus, this linear two-point boundary value problem is solved by computing one particular solution and two linearly independent homogeneous solutions. Then, a system of linear algebraic equations is solved for the integration constants so that the boundary conditions are satisfied. If we desire the complete solution, as we will later, we may use these integration constants in superposition.

If equation (62) is converted to a system of two first-order equations by making the substitution

$$dY/dt = Z, \tag{68}$$

the nonhomogeneous equation is

$$\begin{aligned} dY/dt &= Z \\ dZ/dt &= -P(t)\ Z - Q(t)\ Y + R(t) \end{aligned} \tag{69}$$

and the homogeneous equation is

$$\begin{aligned} dY/dt &= Z \\ dZ/dt &= -P(t)\ Z - Q(t)\ Y. \end{aligned} \tag{70}$$

The initial conditions for the solution of equation (69) (the particular solution) are

$$\begin{aligned} Y(0) &= a \\ Z(0) &= 0, \end{aligned} \tag{71}$$

and for the solution of equation (70) (the homogeneous solutions) are

$$\begin{aligned} Y(0) &= 1 \qquad Y(0) = 0 \\ &\quad \text{and} \\ Z(0) &= 0 \qquad Z(0) = 1. \end{aligned} \tag{72}$$

Because the boundary conditions involve Y and not Z, the integration constants c_1 and c_2 are again determined using equation (67). This approach is used in the following sections.

Thus, it is possible to solve linear boundary value problems, equations (62, 63), by using numerical integration techniques and methods of solving systems of linear algebraic equations. To see how it is possible to convert a nonlinear boundary value problem into a linearized form, consider the nonlinear second-order equation

$$d^2Y/dt^2 = f(Y(t),t) \tag{73}$$

with boundary conditions

$$\begin{aligned} Y(0) &= a \\ Y(t_f) &= b. \end{aligned}$$

Quasilinearization connotes replacing the nonlinear differential system by a "nearby" linear system. The nearby linear system is derived, in the case of equation (73), by expanding $f(Y,t)$ in a Taylor's series in Y about Y_0:

$$f(Y,t) = f(Y_0,t) + f_Y(Y_0,t)(Y-Y_0) + \frac{f_{YY}(Y_0,t)}{2!}(Y-Y_0)^2 + \ldots.$$

Retaining only the linear portion, the differential equation (73) takes the form

$$d^2Y/dt^2 = f(Y_0,t) + f_Y(Y_0,t)\ Y - f_Y(Y_0,t)\ Y_0, \tag{74}$$

where, if Y_0 is presumed a known function of t, the equation is linear in Y.

Starting with some initial guess function $Y_0(t)$, a sequence of functions is generated by means of the equation

$$\begin{aligned} d^2Y_{n+1}/dt^2 &= f_Y(Y_n(t),t)Y_{n+1} + f(Y_n(t),t) - \\ &\qquad f_Y(Y_n(t),t)Y_n(t) \\ \text{or}\quad d^2Y_{n+1}/dt^2 &= a(t)Y_{n+1} + b(t), \end{aligned} \tag{75}$$

where $a(t)$ and $b(t)$ are known (previously computed and stored internally in the computer) functions of t.

Clearly, this recurrence relation is analogous to a one-point I.F. requiring evaluation of the derivative; i.e., the Newton-Raphson I.F.

How does one know when convergence of the above recurrence relation has occurred? Roughly speaking it is when the function $Y_n(t)$, the solution at the n^{th} iteration, deviates less than some prescribed amount from the function $Y_{n+1}(t)$, the current solution.

Most of this work involves systems of first-order equations, hence, the analogous recurrence relation for such systems is

$$\frac{d\vec{Y}_{n+1}}{dt} = \vec{f}(\vec{Y}_n(t),t) + \vec{J}(\vec{Y}_n(t),t)\ (\vec{Y}_{n+1} - \vec{Y}_n(t)), \tag{76}$$

where $\longrightarrow$ indicates column vectors and f is the vector function of r.h.s. of the governing equations, and J is the Jacobian matrix defined as

$$J(Y_n) = \begin{bmatrix} \frac{\partial f_1}{\partial y_{1,n}} & \frac{\partial f_1}{\partial y_{2,n}} & \frac{\partial f_1}{\partial y_{3,n}} & \cdots & \frac{\partial f_1}{\partial y_{m,n}} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \frac{\partial f_m}{\partial y_{1,n}} & \frac{\partial f_m}{\partial y_{2,n}} & \frac{\partial f_m}{\partial y_{3,n}} & \cdots & \frac{\partial f_m}{\partial y_{m,n}} \end{bmatrix}. \tag{77}$$

Again, $Y_n(t)$ is a computationally known vector function on the interval $0 \leq t \leq t_f$. Because equation (76) is a linear nonhomogeneous first-order equation, superposition may be used to form the complete solution as

$$Y(t) = y_p(t) + c_1 y_{h1}(t) + c_2 y_{h2}(t) + c_3 y_{h3}(t) + \ldots + c_m y_{hm}(t).$$

The details of the above procedure are perhaps best seen through an example. Assume the existence of three observations on a system trajectory,

$$\begin{aligned} Y(0) &= m_1 \\ Y(t_1) &= m_2 \\ Y(t_f) &= m_3, \text{ and} \end{aligned} \tag{78}$$

the hypothesized governing equation

$$dY/dt = aYe^{bY}. \tag{79}$$

To completely determine the equation governing the system, numerical values for a and b must be determined such that conditions (78) are met. Converting a and b to functions of time and constructing a system of equations gives

$$\begin{aligned} dY/dt &= aYe^{bY} = f_1 \\ da/dt &= 0 \quad\ = f_2 \\ db/dt &= 0 \quad\ = f_3. \end{aligned} \tag{80}$$

In terms of equation (76), we have

$$\begin{bmatrix} \frac{dY_{n+1}}{dt} \\ \frac{da_{n+1}}{dt} \\ \frac{db_{n+1}}{dt} \end{bmatrix} = \begin{bmatrix} a_n Y_n(t) e^{b_n Y_n(t)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial Y_n} & \frac{\partial f_1}{\partial a_n} & \frac{\partial f_1}{\partial b_n} \\ \frac{\partial f_2}{\partial Y_n} & \frac{\partial f_2}{\partial a_n} & \frac{\partial f_2}{\partial b_n} \\ \frac{\partial f_3}{\partial Y_n} & \frac{\partial f_3}{\partial a_n} & \frac{\partial f_3}{\partial b_n} \end{bmatrix} \begin{bmatrix} Y_{n+1}-Y_n \\ a_{n+1}-a_n \\ b_{n+1}-b_n \end{bmatrix}. \tag{81}$$

Equation (81) reduces to

$$\frac{dY_{n+1}}{dt} = a_n Y_n e^{b_n Y_n} + \frac{\partial f_1}{\partial Y_n}(Y_{n+1}-Y_n) + \frac{\partial f_1}{\partial a_n}(a_{n+1}-a_n) + \frac{\partial f_1}{\partial b_n}(b_{n+1}-b_n) \qquad (82)$$

$$\frac{da_{n+1}}{dt} = 0$$

$$\frac{db_{n+1}}{dt} = 0$$

with boundary conditions $Y(0) = m_1$, $Y(t_1) = m_2$, and $Y(t_f) = m_3$. This, then, is a multipoint boundary value problem formulation of our inverse problem. Notice that equation (82) is a linear nonhomogeneous differential equation, and as such has a solution of the form

$$\begin{matrix} Y(t) = y_p(t) \\ a(t) = a_p(t) \\ b(t) = b_p(t) \end{matrix} + c_1 \begin{bmatrix} y_{h1}(t) \\ a_{h1}(t) \\ b_{h1}(t) \end{bmatrix} + c_2 \begin{bmatrix} y_{h2}(t) \\ a_{h2}(t) \\ b_{h2}(t) \end{bmatrix} + c_3 \begin{bmatrix} y_{h3}(t) \\ a_{h3}(t) \\ b_{h3}(t) \end{bmatrix}, \qquad (83)$$

where $[y_p\ a_p\ b_p]^T$ is the solution of the nonhomogeneous equation (82) and

$$\begin{bmatrix} y_{h1}(t) \\ a_{h1}(t) \\ b_{h1}(t) \end{bmatrix}, \quad \begin{bmatrix} y_{h2}(t) \\ a_{h2}(t) \\ b_{h2}(t) \end{bmatrix}, \text{ and } \begin{bmatrix} y_{h3}(t) \\ a_{h3}(t) \\ b_{h3}(t) \end{bmatrix}$$

are independent solutions of the homogeneous form of (82); i.e.,

$$dY_{n+1}/dt = \frac{\partial f_1}{\partial Y_n} Y_{n+1} + \frac{\partial f_1}{\partial a_n} a_{n+1} + \frac{\partial f_1}{\partial b_n} b_{n+1} \qquad (84)$$

$$da_{n+1}/dt = 0$$

$$db_{n+1}/dt = 0.$$

The logic in selecting initial conditions for particular and homogeneous solutions is entirely analogous to that used in equations (66a) and (66b). For the particular solution the first boundary condition is used as the initial condition on Y and $a(0) = b(0) = 0$; i.e.,

$$\begin{bmatrix} m_1 \\ 0 \\ 0 \end{bmatrix}.$$

To obtain three independent homogeneous solutions the convenient initial conditions

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

may be used. These, again, are initial value problems that may be handled in a direct manner using numerical integration techniques.

To insure that the second and third boundary conditions are met, we form the general solution, equation (83). Because the boundary conditions involve Y, not a or b, we consider the solution for Y at the three times 0, t_1 and t_f; i.e.,

$$\begin{aligned} Y(0) &= y_p(0) + c_1 y_{h1}(0) + c_2 y_{h2}(0) + c_3 y_{h3}(0) = m_1 \\ Y(t_1) &= y_p(t_1) + c_1 y_{h1}(t_1) + c_2 y_{h2}(t_1) + c_3 y_{h3}(t_1) = m_2 \\ Y(t_f) &= y_p(t_f) + c_1 y_{h1}(t_f) + c_2 y_{h2}(t_f) + c_3 y_{h3}(t_f) = m_3. \end{aligned}$$

This is similar to equation (67). Proceeding in an analogous manner here we have for the first equation

$$m_1 + c_1\,0 + c_2\,0 + c_3\,1 = m_1,$$

which dictates that c_3 be zero. The result is two equations in two unknowns,

$$\begin{aligned} y_p(t_1) + c_1 y_{h1}(t_1) + c_2 y_{h2}(t_1) &= m_2 \text{ and} \\ y_p(t_f) + c_1 y_{h1}(t_f) + c_2 y_{h2}(t_f) &= m_3, \end{aligned} \tag{85}$$

which may easily be solved for c_1 and c_2. Only two homogeneous solutions are required; i.e., so initial conditions $[0\ 1\ 0]^T$ and $[0\ 0\ 1]^T$ are used.

Once c_1, c_2, and c_3 are known, they are substituted into equation (83) and the solution vector function

$$\begin{bmatrix} Y_{n+1}(t) \\ a_{n+1}(t) \\ b_{n+1}(t) \end{bmatrix}$$

is formed. After a check for convergence, by making a comparison of

$$\begin{bmatrix} a_{n+1}(0) \\ b_{n+1}(0) \end{bmatrix} \quad \text{vs} \quad \begin{bmatrix} a_n(0) \\ b_n(0) \end{bmatrix},$$

one may, if need be, return to equations (81) and (82) and consider the vector function just determined as

$$\begin{bmatrix} Y_n(t) \\ a_n(t) \\ b_n(t) \end{bmatrix}.$$

In cases where many observations are available on a system trajectory, such as

$$Y(t_i) = m_i, \quad i = 1, 2, 3, \ldots, n,$$

we have an overdetermined system corresponding to equation (85). In such cases each boundary condition may not be met, so some measure of goodness of fit is used between

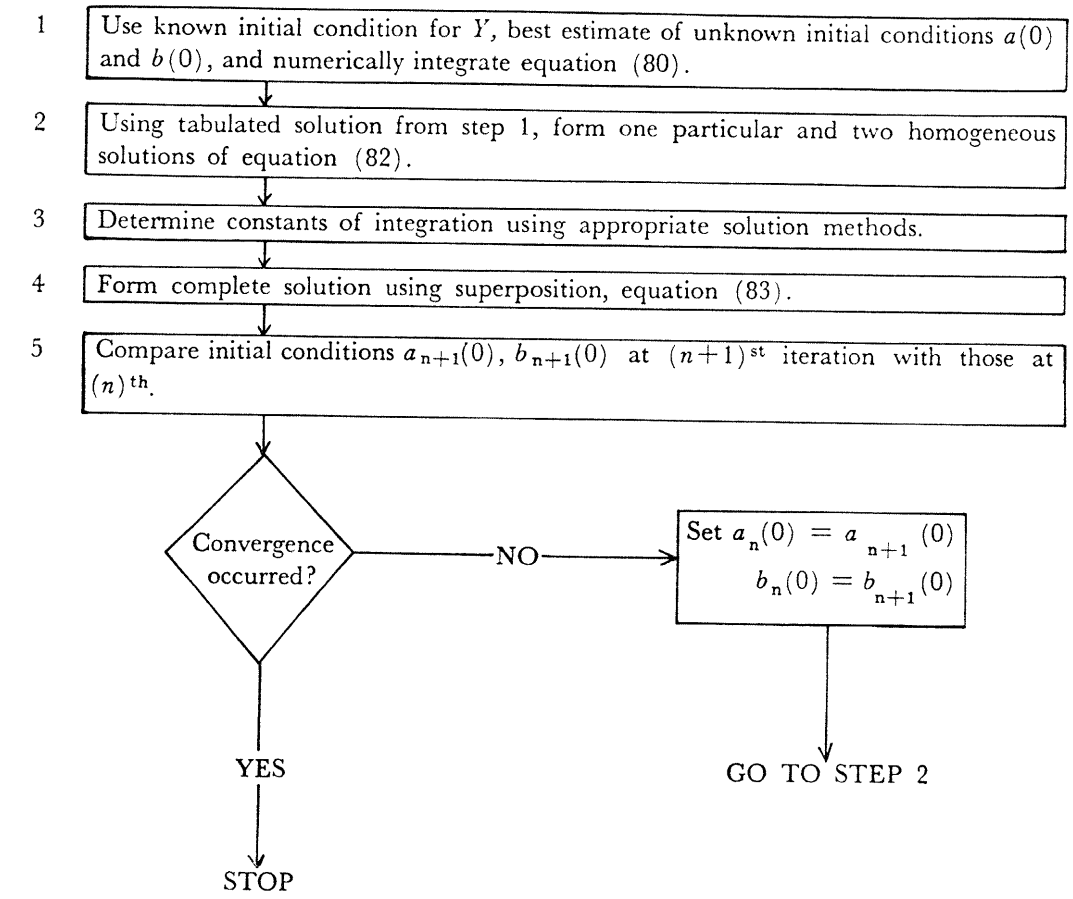
$$\begin{aligned} &Y(t_i) \text{ and } m_i; \text{ i.e.,} \\ &Y(t_i) \cong m_i \text{ for all } i. \end{aligned} \tag{86}$$

The criterion used in later sections is that of least squares. Thus, integration constants c_1 and c_2 are determined such that they minimize

$$S = \sum_i (Y(t_i) - m_i)^2. \tag{87}$$

There is some indication that other measures of goodness of fit may be more appropriate than least squares, for instance, minimax (Bellman 1964).

The following flow-chart conveys a more concise picture of the steps involved:



Example 1

For a first example consider the following observational data on one system component in a northern hardwood stand located on the Argonne Experimental Forest in Wisconsin. The component is defined as all sugar maple (*Acer saccharum*) trees that were living in 1968 and were present in the 4- to 8-inch diameter class in 1947. Summarizing the historical records for this component on a plot gives

Year	*Sum of tree diameters* (*Inches*)
1947	60.6
1949	63.9
1950	65.4
1951	66.9
1954	72.6
1958	78.6
1962	85.9
1967	92.5
1968	93.4

These form the boundary conditions that guide the solution of equation (80). Because only two homogeneous solutions are required, the system is overdetermined. To guide our solution, the least squares criterion for goodness of fit is used. In other words, we want to determine c_1 and c_2 in (85) in a manner that minimizes

$$S = \sum_i (Y(t_i) - m_i)^2, \text{ where}$$

$$S = \sum_i (y_p(t_i) + c_1 y_{h1}(t_i) + c_2 y_{h2}(t_i) - m_i)^2. \tag{88}$$

This may be accomplished by solving the simultaneous equations

$$\frac{\partial S}{\partial c_1} = 0 \quad \text{and} \quad \frac{\partial S}{\partial c_2} = 0$$

for c_1 and c_2.

Using a computer program written by the author, this nonlinear multipoint boundary value problem is solved by following the steps outlined in the flow-chart.

The initial parameter estimates were

$$a(0) = .05$$
$$b(0) = -.01.$$

The estimates of parameter values at each iteration were

Iteration	*a*	*b*
1	0.075469	−0.017809
2	.090660	− .019302
3	.091015	− .019167
4	.091041	− .019172
5	.091040	− .019172

The equation governing this system component is, therefore,

$$dY/dt = .0910\ Y\ e^{-.019172\ Y}. \tag{89}$$

Integrating this equation with known initial conditions on Y, $Y(0) = 60.6$, gives the following values:

Year	*Observed*	*Predicted*	*Difference*
1947	60.6	60.6	—
1949	63.9	64.0	−0.1
1950	65.4	65.7	− .3
1951	66.9	67.4	− .5
1954	72.6	72.4	.2
1958	78.6	78.8	− .2
1962	85.9	85.1	.8
1967	92.5	92.4	.1
1968	93.4	93.8	− .4

A similar trial using the same source and type of data, but for the 8- to 12-inch diameter class of sugar maple, gave the following results

Iteration	*a*	*b*
Initial estimate	0.050000	−0.010000
1	.075405	− .020263
2	.134159	− .030693
3	.161696	− .030475
4	.161992	− .030541
5	.161949	− .030537
6	.161951	− .030537

Thus, the equation governing this system component is

$$dY/dt = .1619\ Y\ e^{-.030537\ Y} \tag{90}$$

Integrating this equation with known initial conditions, $Y(0) = 63.2$, gives

Year	*Observed*	*Predicted*	*Difference*
1947	63.2	63.2	—
1949	66.1	66.1	0.0
1950	67.0	67.5	− .5
1951	68.2	68.8	− .6
1954	72.8	72.8	.0
1958	77.9	77.7	.2
1962	83.0	82.2	.8
1967	87.4	87.3	.1
1968	87.8	88.3	− .5

Thus, it is possible to use observed system states as boundary conditions, and by solving the associated multipoint boundary value problem, to determine equation parameters that produce close agreement between predicted and observed states. Because the system components are well behaved, it follows logically that a projection for a limited time beyond 1968 may result in acceptable estimates of future component values.

The above examples, although showing close agreement between predicted and observed states, are inadequate in that they do not treat the forest as an interactive system. Hence, consideration is given to a more inclusive model that reflects simplified assumptions concerning interactions among three system components.

Example 2

The concern here is with finding numerical values for the parameters that best relate the predicted system state, as determined by

$$\begin{aligned} dY1/dt &= a\ Y1\ e^{b(Y1 + Y2 + Y3)} \\ dY2/dt &= c\ Y2\ e^{d(Y2 + Y3)} \\ dY3/dt &= f\ Y3\ e^{g(Y3)}, \end{aligned} \tag{91}$$

to the observed state given below. In this case, six homogeneous solutions are required, one for each unknown initial condition (parameter) (see equation (57)). Given nine observations in time, an overdetermined system is again present, so the least squares criterion for goodness of fit is used. In other words, we minimize

$$S = \sum_i ((Y1(t_i) - m_{1,i})^2 + (Y2(t_i) - m_{2,i})^2 + (Y3(t_i) - m_{3,i})^2) \tag{92}$$

where $m_{1,i}$ is the i^{th} observation on the first state variable, and

$$Y1(t) = y_{p1} + c_1 y_{h1} + c_2 y_{h2} + c_3 y_{h3} + c_4 y_{h4} + c_5 y_{h5} + c_6 y_{h6}.$$

The normal equations are derived in a straightforward manner by taking the partials of (92) with respect to c_1, c_2, c_3, c_4, c_5, and c_6, and equating them to zero.

Using the data in table 1 as our boundary conditions gives the following results:

Iteration	Parameter					
	a	*b*	*c*	*d*	*f*	*g*
Initial estimate	0.09100	−0.00100	0.16190	−0.01000	0.18370	−0.08650
1	.03219	− .00056	.11380	− .01335	.11488	− .07443
2	.06551	− .00765	.14142	− .01948	.15608	− .08283
5	.12152	− .00958	.18381	− .02285	.15132	− .08029
7	.12145	− .00957	.18385	− .02285	.15132	− .08029

Table 1. — *Observed stand states for all sugar maple on a plot on the Argonne Experimental Forest*

Year	D.b.h. class (inches)		
	4 to 8	8 to 12	12 to 16
	(Sum of diameters in inches)		
1947	60.6	63.2	27.1
1949	63.9	66.1	28.2
1950	65.4	67.0	28.4
1951	66.9	68.2	28.9
1954	72.6	72.8	30.2
1958	78.6	77.9	31.6
1962	85.9	83.0	33.3
1967	92.5	87.4	34.8
1968	93.4	87.8	35.1

Hence, the complete governing equations are

$$\begin{aligned} dY1/dt &= .12145\ Y1\ e^{-.00957(Y1 + Y2 + Y3)} \\ dY2/dt &= .18385\ Y2\ e^{-.02285(Y2 + Y3)} \\ dY3/dt &= .15132\ Y3\ e^{-.08029(Y3)} \end{aligned} \qquad (93)$$

Using these equations and the known initial conditions

$$\begin{aligned} Y1(0) &= 60.6 \\ Y2(0) &= 63.2 \\ Y3(0) &= 27.1, \end{aligned}$$

the deviations of predicted system states from observed states are shown in table 2. Again, observed and predicted states agree well.

Interestingly, parameters *a* and *c* as determined from the system of equations (0.12145 and 0.18385, respectively) are both greater than when single equations are used as governing equations for their respective components (0.09104 and 0.1619). This indicates that components *Y*1 and *Y*2 have more capacity to grow than is indicated by treating them as isolated components, or not in the context of an interactive system. Clearly, component *Y*1 has observed states as shown earlier because it is comprised of the smallest measured trees in the stand, not because all 4- to 8-inch sugar maples grow as indicated by the boundary conditions. Equation (89)

$$dY1/dt = .09104\ Y1\ e^{-.01917\ Y1}$$

the single component equation, tells nothing about the relationship of trees in this class to trees in the larger size classes. On the other hand, the first equation of the system,

$$dY1/dt = .12145\ Y1\ e^{-.00957(Y1 + Y2 + Y3)} \qquad (94)$$

tells us that components $Y2$ and $Y3$ are inhibiting the growth of $Y1$. Furthermore, if all of $Y3$ were removed, one would expect a response in the growth of component $Y1$. The above equation will give a response because the inhibiting component will be smaller. Importantly, the amount of response is taken care of by the system itself.

Table 2. — *Deviations of predicted from observed sums of diameters on a plot on the Argonne Experimental Forest*

Year	D.b.h. class (inches)		
	4 to 8	8 to 12	12 to 16
	(Deviation in inches)		
1947	--	--	--
1949	0.1	0.0	-0.1
1950	.3	.5	.0
1951	.5	.7	.0
1954	- .1	.0	.0
1958	.3	- .1	.1
1962	- .8	- .7	- .1
1967	- .1	.0	.0
1968	.4	.4	.0

Although equation (93) is a realistic characterization of how forest system components interact, like all models, it involves simplifying assumptions. One could, at the expense of more computing time and a slight increase in data requirements, have a separate coefficient for each inhibiting component in each equation. Likewise, remeasurement data on other components of the forest system such as browse could easily be included as separate equations in such a multidimensional model. The upper limit of equations and unknown parameters appears determined only by available historical data and computing costs. (Determination of the six parameters in equation (93) required 7 seconds central processor time on a CDC 6600 computer.)

IMPLICATIONS

The preceding examples show how it is possible to obtain governing equation parameter estimates on the basis of observed system states. The approach used represents a constructive alternative to regression techniques for models expressed as differential equations. The attractiveness of the boundary value problem approach is clear in example 2, where parameters are determined in three nonlinear equations simultaneously.

The models and boundary conditions discussed here are of the simplest type. However, the same approach used in this paper is applicable to models involving time lags, time-dependent coefficients, partial differential equations and to boundary conditions involving linear and nonlinear relationships between the state variables.

In conclusion, this approach allows the scientist to more completely quantify his knowledge of forest development processes, to express his theories in terms of one or more governing differential equations, and to determine unknown numerical constants in the governing equations — three important steps in the study of forest dynamics.

LITERATURE CITED

Bellman, R. E. 1962. Orbit determination as a multi-point boundary value problem and quasilinearization. Nat. Acad. Sci. U.SA. Proc. 48: 1327-1329.

Bellman, R. E. 1964. Quasilinearization, boundary-value problems and linear programming. Rand Corp. Memo. 4284-PR, 7 p.

Bellman, R. E., and Kalaba, R. E. 1965. Quasilinearization and nonlinear boundary-value problems 206 p. New York: Elsevier.

Conte, S. D. 1965. Elementary numerical analysis. 278 p. New York: McGraw-Hill.

DeRusso, P. M., Ray, R. J., and Close, C. M. 1965. State variables for engineers. 608 p. New York: John Wiley and Sons, Inc.

Faddeev, D. K., and Faddeeva, V. N. 1963. Computational methods of linear algebra. 621 p. San Francisco: W. H Freeman.

Hall, A. D., and Fagen, R. E. 1956. Definition of a system — in general systems. Yearbook Soc. for Advance. Gen. Syst. Theory 1:18-28.

Henrici, P. 1964. Elements of numerical analysis. 328 p. New York: John Wiley and Sons, Inc.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor on Quantitative Biol. Symp. 22: 415-427.

Kaplan, W. 1964. Elements of differential equations. 270 p. New York: Addison-Wesley Publishing Co., Inc.

Knuchel, H. 1953. Planning and control in the managed forest. (Translated by Mark L. Anderson.) 360 p. London: Oliver and Boyd.

Lee, E. S. 1968. Quasilinearization and invariant imbedding. 329 p. New York: Academic Press.

Lotka, A. J. 1928. Elements of mathematical biology. 465 p. New York: Dover Publications, Inc.

MacKinney, A. L., Schumacher, F. X., and Chaiken, L. E. 1937. Construction of yield tables for nonnormal loblolly pine stands. J. Agr. Res. 54(7): 531-545.

Miller, R. S. 1967. Pattern and process in competition. Advances in Ecol. Res. 4: 1-74.

Moser, J. W., and Hall, O. F. 1969. Deriving growth and yield functions for uneven-aged forest stands. Forest Sci. 15(2): 183-188.

Perlis, S. 1952. Theory of matrices. 237 p. New York: Addison-Wesley Publishing Co., Inc.

Rainville, E. D. 1964. Elementary differential equations. Ed. 3, 521 p. New York: MacMillan.

Ralston, A. 1965. A first course in numerical analysis. 576 p. New York: McGraw-Hill.

Rosen, R. 1967. Optimality principles in biology. 198 p. New York: Plenum Press.

Roston, S. 1962. On biological growth. Bull. Math. Biophysics 24: 369-373.

Traub, J. F. 1964. Iterative methods for the solution of equations. 310 p. Englewood Cliffs: Prentice-Hall Inc.

SOME RECENT RESEARCH PAPERS
OF THE
NORTH CENTRAL FOREST EXPERIMENT STATION

Durable Rustic Wooden Signboards, by Glenn A. Cooper. USDA Forest Serv. Res. Pap. NC-29, 15 p., illus. 1969.

The Market for Wood Picnic Structures, by Jerry A. Sesco. USDA Forest Serv. Res. Pap. NC-30, 7 p., illus. 1969.

Oak Forests of the Lake States and Their Management, by John L. Arend and Harold F. Scholz. USDA Forest Serv. Res. Pap. NC-31, 36 p., illus. 1969.

Wildlife Esthetics and Auto Campers in the Superior National Forest, by David W. Lime and Charles T. Cushwa. USDA Forest Serv. Res. Pap. NC-32, 8 p., illus. 1969.

Standing Timber Coefficients for Indiana Walnut Log Production, by James E. Blyth, Edwin Kallio, and John C. Callahan. USDA Forest Serv. Res. Pap. NC-33, 9 p., illus. 1969.

Climatic Conditions Preceding Historically Great Fires in the North Central Region, by Donald A. Haines and Rodney W. Sando. USDA Forest Serv. Res. Pap. NC-34, 19 p., illus. 1969.

Initial Thinning in Red Pine Plantations, by John H. Cooley. USDA Forest Serv. Res. Pap. NC-35, 6 p., illus. 1969.

Biological Growth Functions Describe Published Site Index Curves for Lake States Timber Species, by Allen L. Lundgren and William A. Dolid. USDA Forest Serv. Res. Pap. NC-36, 9 p., illus. 1970.

Defect Indicators in Sugar Maple — A Photographic Guide, by Richard M. Marden, and Charles L. Stayton. USDA Forest Serv. Res. Pap. NC-37, 29 p., illus. 1970.

Estimating Sugar Maple Bark Thickness and Volume, by Charles L. Stayton and Michael Hoffman. USDA Forest Serv. Res. Pap. NC-38, 8 p., illus. 1970.

ABOUT THE FOREST SERVICE . . .

As our Nation grows, people expect and need more from their forests — more wood; more water, fish, and wildlife; more recreation and natural beauty; more special forest products and forage. The Forest Service of the U.S. Department of Agriculture helps to fulfill these expectations and needs through three major activities:

- Conducting forest and range research at over 75 locations ranging from Puerto Rico to Alaska to Hawaii.
- Participating with all State forestry agencies in cooperative programs to protect, improve, and wisely use our Country's 395 million acres of State, local, and private forest lands.
- Managing and protecting the 187-million acre National Forest System.

The Forest Service does this by encouraging use of the new knowledge that research scientists develop; by setting an example in managing, under sustained yield, the National Forests and Grasslands for multiple use purposes; and by cooperating with all States and with private citizens in their efforts to achieve better management, protection, and use of forest resources.

Traditionally, Forest Service people have been active members of the communities and towns in which they live and work. They strive to secure for all, continuous benefits from the Country's forest resources.

For more than 60 years, the Forest Service has been serving the Nation as a leading natural resource conservation agency.