

The Algebraic Difference Approach  
Improves Fixed Base-Age Site Models  
based on Chapman-Richard Function

Plantation Management Research Cooperative  
Warnell School of Forest Resources  
University of Georgia

PMRC Technical Report 1999-9

Prepared by

C.J. CIESZEWSKI

October 11, 1999

## Abstract

A scientist analyzing plant growth may want to summarize data using a mathematical model. The growth potential and shape of the growth curves representing various sites may vary due to differences in each site's productivity factors, which include nutrient and moisture availability, climate, etc. With respect to curve shape, in a simple situation, growth may increase proportionally, for example, with nutrient availability, while in more realistic situations, curves will have distinctly different shapes for different nutrient levels. Nonetheless, if changes between sites are consistent and continuous, one general model, rather than multiple, site-specific models, can be used to illustrate growth on all sites. Because site productivity is composed of many variables that are not measurable within practical limitations, it may be advantageous to use plant size as the measure of this productivity. This leads to the development of growth models based on implicit equations, in which the  $Y$  variable is a function of both variables  $X$  and  $Y$ . In practice, an implicit growth model describes the plant size ( $Y$ ) as a function of the prediction age ( $x$ ), another arbitrary age ( $x_0$ ), called a base age, and the plant size at the base age ( $y_0$ ), called the reference size. Hence  $Y = f(x, x_0, y_0)$ . The base age can be either implicit and fixed, or explicit and adjustable. For practical applications, the latter is more desirable. Also desirable is that the growth models are built through incorporating biological bases into the equations so that they are readily interpretable for their users. This paper discusses a simple, general methodology of derivation of such implicit growth models using a practical example of an existing model from the literature.

## 1 Introduction

Various types of models contribute to more efficient forest management by facilitating inventory updates and projections, growth and yield forecasting, and site productivity identification and stratification. Among the many kinds of models used in forest management, the most prevalent are the site dependent models based on implicit functions. These models describe panel data, (Fig. 1b to 1d), i.e., pooled cross-sectional and time-series, or longitudinal (Fig. 1a), or repeated measurements data. Because almost all dynamic processes in forestry depend on the cross-sectional aspect of forest dynamics relating to different ecological and productivity sites, I refer to these models as “site models” or “site equations”. Site equations model repeated measures data and, with this respect, they are similar to mixed effects models, e.g., Lindstrom and Bates (1990), random effects models for longitudinal data, e.g., Chi and Reinsel (1989), panel data models, e.g., Furnival *et al.* (1990), and self-referencing models, e.g., Northway (1985).

In forestry the earliest efforts of growth modeling concentrated on two-dimensional relationships (e.g., height over age). Both hand drawn curves and the earliest equations capable of consistently generating more intricate shapes approximated two-dimensional relations. (Fig. 1a). Peschel (1938) credits Spath in 1797 and Hosfeld in 1822 with the first efforts of expressing the two-dimensional growth and yield relationships in forestry with mathematical equations. These site models, at times, were developed separately for different sites or even individually for different stands. Until today, many users of site models prefer to consider them as two-dimensional relationships and rarely discuss them in the context of three-dimensional systems; sometimes sites are denoted as discrete quality or productivity classes, e.g., **A**, **B**, . . . , or **I**, **II**, . . . .

Historically site models were presented as graphs or tables representing yield for a discrete collection of sites or stands, e.g., Szymkiewicz (1971). In the USA the most popular are mathematical models with site represented by site index ( $S$ ), that is, a height at a given base-age ( $A_b$ ). An early algebraic inclusion of  $S$  into simple anamorphic (Fig. 1) equations was followed by increasing equation complexities necessary to describe numerous characteristics, such as, polymorphism (Fig. 2) and variable asymptotes. (Fig. 3 and 4). A more advanced and desirable approach to site dependent modeling is based on initial condition difference equations that, hereafter, I call dynamic equations. Bailey and Clutter (1974) introduced to forestry the concept of **base-age invariance**, in which a dynamic equation can compute predictions directly from any age-height pair without compromising consistency of the predictions. They applied a technique that is known in forestry as the algebraic difference approach (ADA) and consists, essentially, of replacing a parameter in a base-model with its initial condition solution. This approach has been applied to model many forest characteristics.

To complement the ADA, which is primarily suitable for derivation of either anamorphic or single-asymptote polymorphic models, Cieszewski (1994) (also Cieszewski and Bailey 2000) presents a generalization of the ADA, called hereafter Generalized Algebraic Difference Approach or GADA. This approach is suitable for derivation of polymorphic models with concurrently variable asymptotes. The GADA incorporates expanding a base model with theories about modeled processes prior to substitution of the initial condition solution. It allows for a derivation of very flexible dynamic equations that are truly base-age invariant, polymorphic, and have variable asymptotes, as well as, other desirable properties, such as a theoretical basis and  $S$  equal height at base-age, which could not all be derived with the ADA (Bailey and Clutter 1974). The GADA is suitable for various kinds of growth, yield, and decline or oscillation patterns, as well as, for improving existing fixed base-age site index models, some of which I present here.

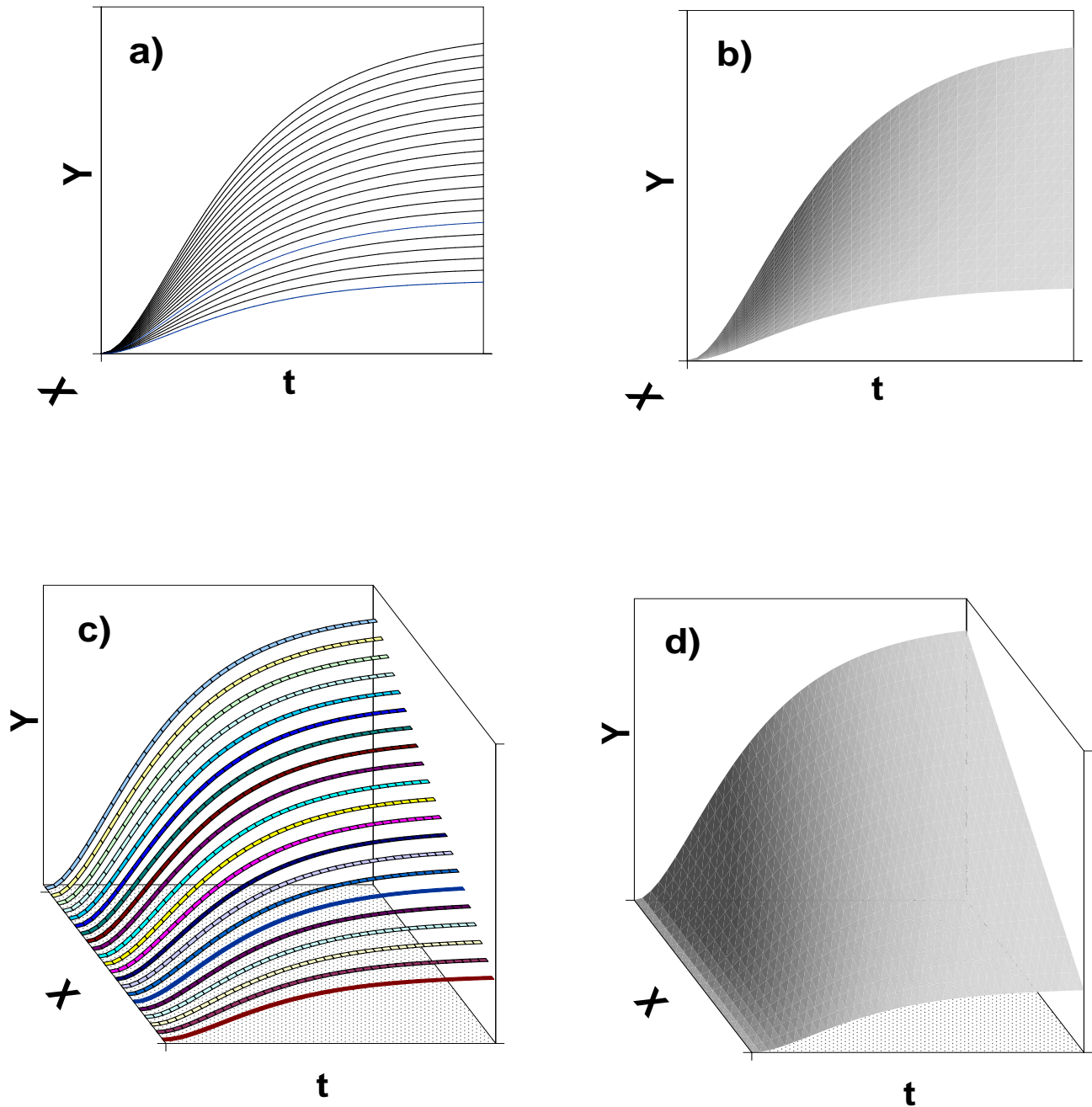


Figure 1: Anamorphic panel data model with variable asymptotes,  $t$  is time,  $Y$  is a response variable, and  $\mathcal{X}$  is a factor determining intensity; a) 0° rotation discrete representation; b) 0° rotation continuous representation; c) 30° rotation discrete representation; c) 30° rotation continuous representation;

## 2 The Method

The main step in the GADA is expanding base equations according to assumptions or theories about various growth characteristics. Examples of such characteristics are limiting size or asymptote, car-

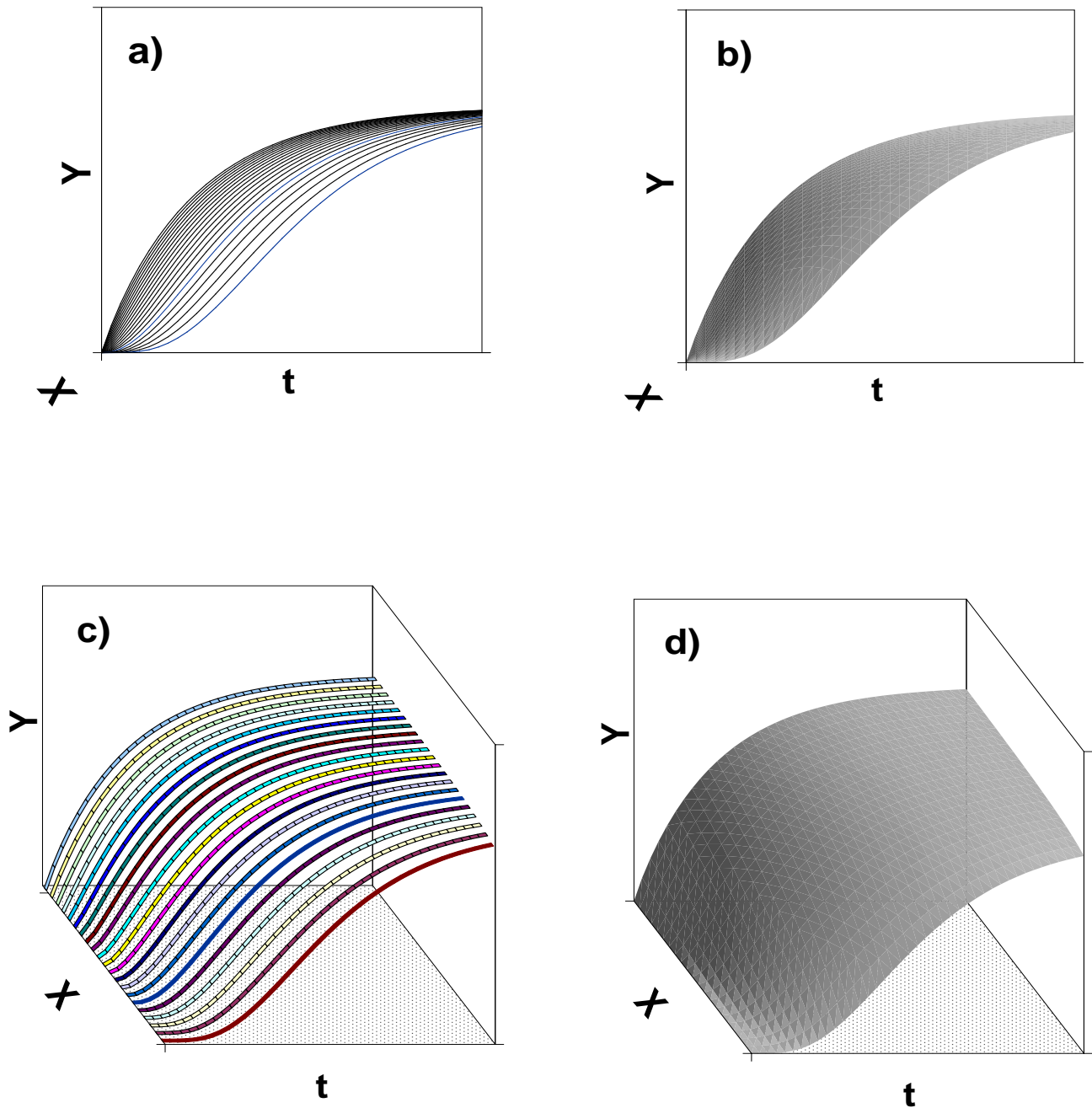


Figure 2: Polymorphic panel data model with single asymptote,  $t$  is time,  $Y$  is a response variable, and  $\mathcal{X}$  is a factor determining intensity; a) 0° rotation discrete representation; b) 0° rotation continuous representation; c) 30° rotation discrete representation; c) 30° rotation continuous representation;

rying capacity (a constant yield law), semi-saturation time, and other model parameters. Some of these characteristics may be subject to personal biases or preferences, but in the end, they all can be

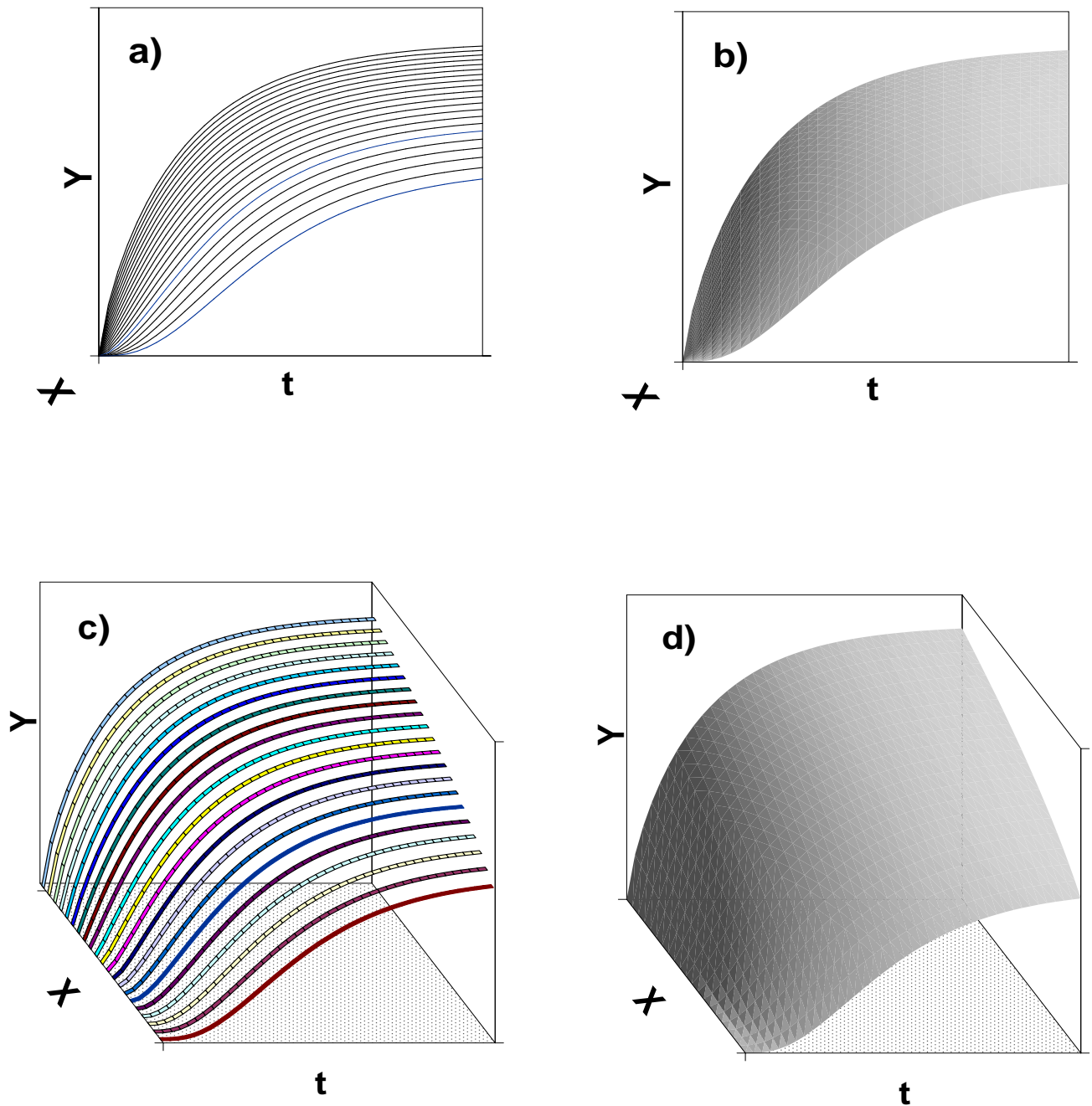


Figure 3: Polymorphic disjoint panel data model with variable asymptotes,  $t$  is time,  $Y$  is a response variable, and  $\mathcal{X}$  is a factor determining intensity; a)  $0^\circ$  rotation discrete representation; b)  $0^\circ$  rotation continuous representation; c)  $30^\circ$  rotation discrete representation; c)  $30^\circ$  rotation continuous representation;

tested with data. With particular respect to the subject of asymptotic growth, it is a rather controversial issue. Ricker (1979) contains a section titled “Asymptotic growth: is it real?”; Knight (1968)

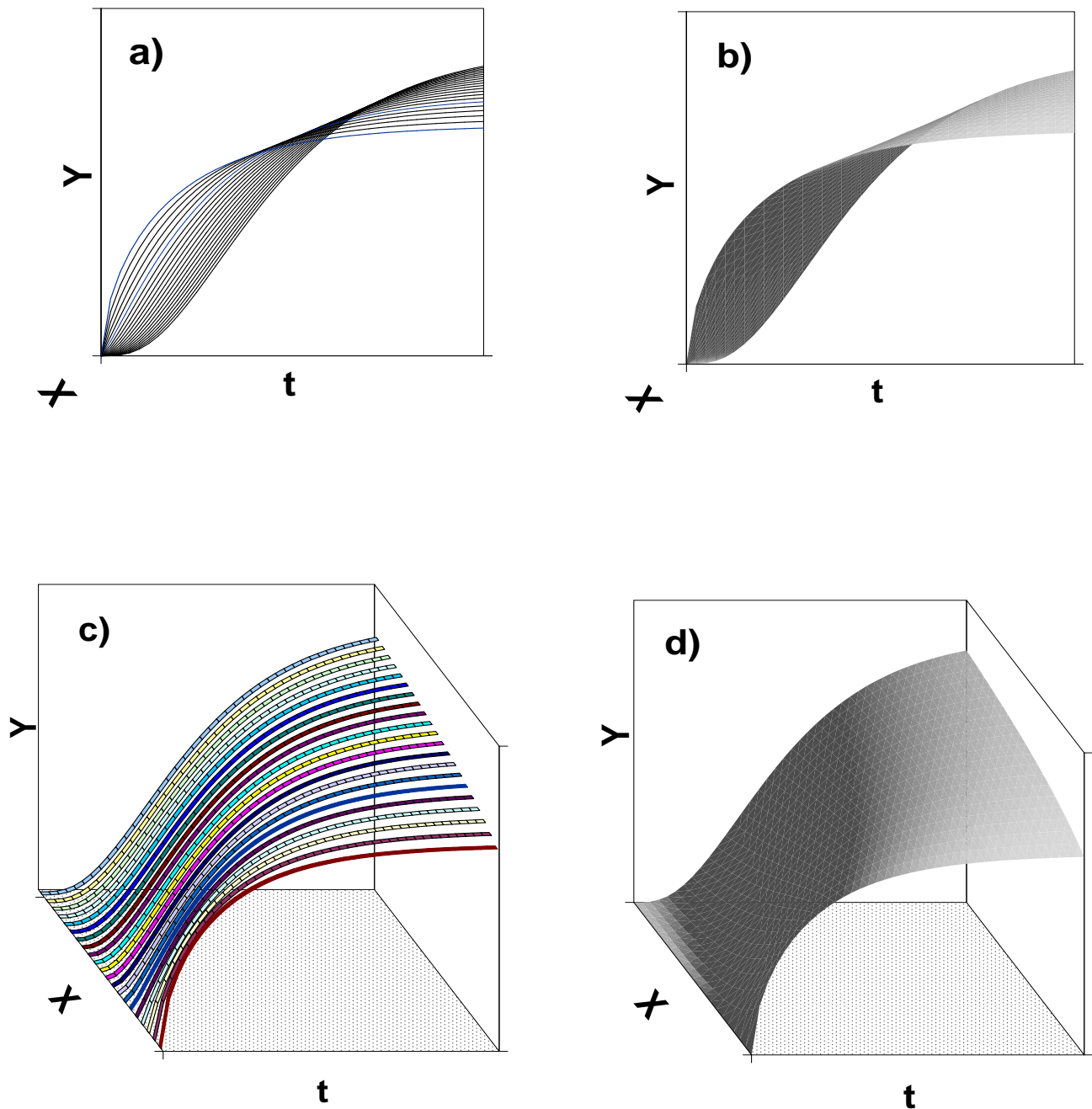


Figure 4: Polymorphic nondisjoint panel data model with variable asymptotes,  $t$  is time,  $Y$  is a response variable, and  $\mathcal{X}$  is a factor determining intensity; a)  $0^\circ$  rotation discrete representation; b)  $0^\circ$  rotation continuous representation; c)  $30^\circ$  rotation discrete representation; c)  $30^\circ$  rotation continuous representation;

calls it "nonsense disguised as mathematics" in his title; and Smith (1984) questions whether it's "fact or artifact." In addition, Biging (1985), Goudie (1984) and Anon. (1985) report difficulties in

estimating asymptotic parameters on data from young trees. As a solution, Schnute (1981) proposes deriving a generic equation, which can be either asymptotic or non-asymptotic, depending on the equation parameters and ultimately on the data. After noting the failure of asymptotic models, Bredenkamp and Gregoire (1988) apply this approach successfully and conclude that for their data an asymptotic model was inappropriate while a non-asymptotic model was appropriate. Cieszewski and Bella (1989 and 1991) also address a similar problem using the more generic forms of dynamic equations similar to those of Schnute (1981).

I consider asymptotes to be a desirable analytical tool used for model conceptualization and growth theory formulation. For example, the General Algebraic Difference Approach is based, in part, on building different theories about limiting growth and about the behavior of asymptotes. Furthermore, it is better to have more options than fewer, and usually more flexible generic models are favored. Yet, if the price of using the asymptotic model means not fitting the data at all, or very poorly, the situation warrants a non-asymptotic model.

For the purpose of the GADA, Cieszewski and Bailey (2000) identify a theoretical variable named the growth intensity factor  $\mathcal{X}$  and define it to be the quantification of those particular growth dynamics that are uniquely associated with site productivity and individual growth or survival capabilities.  $\mathcal{X}$  is used consistently in all equation formulations to describe the rules of changes in growth dynamics across different sites. It can be either a variable or a function of any number of variables affecting the growth. Since  $\mathcal{X}$  is only a theoretical variable and is practically unobtainable, it is eventually replaced with the initial conditions that are measurable, so that, the equation can be operationally useful for site modeling.

In this methodology (GADA), the implicit solution is applied to a site equation, which was explicitly modified to satisfactorily describe the combined longitudinal and cross-sectional changes in terms of two independent variables: the observable variable  $t$  and the unobservable variable  $\mathcal{X}$ . Thus, this three-dimensional site equation has to be defined prior to the substituting of the implicit solution, rather than during, e.g., (Bailey and Clutter 1974), or after it. This is the first characteristic distinguishing the GADA from all other approaches applied in forestry. The second one is that the initial condition solution is applied to an explicit unobservable variable in an equation with two independent variables ( $t$  and  $\mathcal{X}$ ). In the original ADA the solution was applied to a parameter in an equation with one independent variable. An equation with two independent variables represents a surface in a three-dimensional space (Fig. 1d), rather than a line on a two-dimensional plane. (Fig. 1a).

Similarly as in the ADA, the first step of the GADA is to select a base equation and to identify in it any desired number of site-specific parameters, that is

$Y(t) = f(t, \rho_1 \dots \rho_{n-1}, \rho_n) \tag{1}$
--

where  $\rho_1 \dots \rho_n$  are the equation parameters.

If in the base equation (1) a given site-specific parameter  $\rho_i$  is defined as a function  $g_i$  of  $\mathcal{X}$  and any number of  $j$  new parameters, viz.,  $\rho_i \equiv g_i(\mathcal{X}, \rho_{i_1} \dots \rho_{i_j})$ , the base equation (1) with multiple site-specific parameters is changed to the explicit three-dimensional site equation with two independent variables  $t$  and  $\mathcal{X}$ :



$$Y(t, \mathcal{X}) = f\left(t, \rho_1 \dots \rho_{m-1}, g_m(\mathcal{X}, \rho_{m_1} \dots \rho_{m_k}) \dots g_n(\mathcal{X}, \rho_{n_1} \dots \rho_{n_l})\right) \quad (2)$$

where  $Y(t, \mathcal{X})$  is a function of  $t$ ,  $\mathcal{X}$ , and  $m + k + l - 1$  parameters.

Given eq. (2) can be solved for  $\mathcal{X}$ , the right hand side of this solution, with initial condition values for  $t$  and  $Y$ , i.e.,  $\mathcal{X} = u(t, Y, \rho_1 \dots \rho_{n_l}) = u(t_0, Y_0, \rho_1 \dots \rho_{n_l})$  can be substituted in eq. (2) in place of  $\mathcal{X}$ , so the dynamic equation  $Y(t, t_0, Y_0) = f\left(t, \rho_1 \dots \rho_m, u(t_0, Y_0, \rho_1 \dots \rho_{n_l})\right)$ , after reformulation and elimination of redundant parameters, becomes the dynamic equation with an implicitly defined initial condition:

$$Y(t, t_0, Y_0) = f(t, t_0, Y_0, \rho_1 \dots \rho_w) \quad (3)$$

where

$$n - 1 \leq w \leq m + k + \dots + l - 1 \quad (4)$$

This method is primarily intended for derivation of new base-age invariant site models based on various hypotheses describing growth dynamics, but it may also be used to improve numerous existing site equations with respect to parsimony and equation generality. A prime example of two fixed base-age models that can benefit from the application of the GADA are the Chapman-Richards function (Richards 1959) based models by Hegy (1981) and Lundgren and Dolid (1970), that is,

$$H(t, S) = \alpha S (1 - e^{-\gamma t})^\delta \quad (5)$$

and by Biging (1985), which is,

$$H(t, S) = \alpha S^\beta (1 - e^{-\gamma t})^\delta \quad (6)$$

To see how the GADA lends itself to such applications, one can simply substitute  $\mathcal{X}$  for site index (i.e.,  $S$ ) in a base-age specific model and then follow through with the method illustrated above. This means that the element to be solved for and replaced by the initial conditions solution would be  $S$ . Not all equations are solvable for  $S$  and those that are not will not lend themselves to this approach. Nonetheless, many fixed-base-age models can be improved in this way.

Two examples of fixed base-age  $S$  models, which could benefit from the application of this method, are the equations (5) and (6). This can be demonstrated as follows.

	1)	and	2)
$Y(t, \mathcal{X})$	$\overbrace{\alpha \mathcal{X} (1 - e^{-\gamma t})^\delta}$	or:	$\overbrace{\alpha \mathcal{X}^\beta (1 - e^{-\gamma t})^\delta}$
where:			
$\mathcal{X}$	$Y / (\alpha (1 - e^{-\gamma t})^\delta)$	or:	$Y^{1/\beta} / (\alpha (1 - e^{-\gamma t})^\delta)^{1/\beta}$
	$= Y_0 / (\alpha (1 - e^{-\gamma t_0})^\delta)$	or:	$Y_0^{1/\beta} / (\alpha (1 - e^{-\gamma t_0})^\delta)^{1/\beta}$
so:			
$Y(t, t_0, Y_0)$	$\alpha Y_0 \frac{(1 - e^{-\gamma t})^\delta}{\alpha (1 - e^{-\gamma t_0})^\delta}$	or:	$\alpha (Y_0^{1/\beta})^\beta \frac{(1 - e^{-\gamma t})^\delta}{((\alpha (1 - e^{-\gamma t_0})^\delta)^{1/\beta})^\beta}$
	$= Y_0 \left( \frac{1 - e^{-\gamma t}}{1 - e^{-\gamma t_0}} \right)^\delta$	=	$Y_0 \left( \frac{1 - e^{-\gamma t}}{1 - e^{-\gamma t_0}} \right)^\delta$

Both parameters  $\alpha$  and  $\beta$  in the above equations are redundant and are self-canceling, while the models are conditioned to predict  $Y = S$  when  $t = A_b$ . After the GADA substitution and cancellation of redundant parameters, the resulting 2-parameter dynamic equation will be equivalent to the Clutter *et al.* (1983) equation:

$$Y(t, t_0, Y_0) = Y_0 \left( \frac{1 - e^{-\gamma t}}{1 - e^{-\gamma t_0}} \right)^\delta \quad (7)$$

Both of the above example equations and eq. (7) generate identical curves. In addition, the curves generated by eq. (7) predict  $Y = S$  when  $t = A_b$ . The equation is dynamic, parsimonious, and base-age invariant. Ironically, parameters  $\alpha$  and  $\beta$  in the two example equations only serve the function of creating a predicted  $Y$  that is unequal to  $S$  when  $t = A_b$ , which is undesirable.

### 3 Discussion

#### 3.1 Equation Flexibility

The Dynamic equations derived with the ADA or GADA are more flexible and general than corresponding fixed-base equations. A base-age specific model can be developed using a dynamic equation by simply fitting the dynamic equation to data the same way as any fixed-base-age equation would be fitted. Yet, dynamic equations allow for more flexibility in creating applications, conducting analysis, and generating curve shapes. Given an equation for a family of curves, usually any point  $(t_0; Y_0)$  on a curve can unequivocally define that curve. As  $t_0$  can assume any value,  $Y_0$  should not be considered a fixed-age  $S$ . The symbolic dynamic equation (3) represents anamorphic (Fig. 1) or polymorphic (Figures 2–4) height equations that directly use any age and height measurements. Furthermore, eq. (3) defines  $Y$  at age  $t$  as a function of another  $Y$  ( $Y_0$ ) at any other age  $t_0$ . When  $t = 25, 50, 75, 100$ , the associated  $Y$  values can be interpreted as  $S$ . The equation can generate site index curves:

- (i) as a function of height in specified age classes (if  $t_0$  is fixed and  $Y_0$  varies), and
- (ii) as a function of age in height classes (if  $Y_0$  is fixed and  $t_0$  varies),

If  $t_0 = 50$ ,  $Y_0$  can be interpreted as  $S$  and the equation will generate height curves:

- (iii) as a function of age in  $S$  classes (if  $S$  is fixed and  $t$  varies), and
- (iv) as a function of  $S$  in age classes (if  $t$  is fixed and  $S$  varies).

Because  $S$  is just another height at a point when  $t$  equals base-age, no equation reformulation is necessary to calculate  $S$  from height;  $S$  and any other heights are calculated from the same equation.

Parameter estimation, techniques used for regression analysis, data requirements and other possible considerations associated with the use of dynamic equations are independent of the method of equation derivation. Using the GADA to derive a dynamic equation does not preclude any of the various model-fitting approaches being used for parameter estimation. García (1983) reviews base-age-invariant parameter estimation with a fixed-base-age site equation using maximum likelihood methods. DuPlat and Tran-Ha (1986) describe a varying-parameter method of least squares non-linear regression analysis with an equation derived using the ADA, although they never fully exploit

the base-age-invariant aspect of the equation. Others have used base-age-invariant equations with base-age or multi-base-age specific parameter estimations. Whatever specific parameter estimation technique a researcher decides on, it may be applied to a dynamic equation.

Clearly, dynamic equations offer more flexibility in data requirements and handling than the base-age-specific equations. The parameters for dynamic equations may be estimated directly for any height over age data, whether from stem analysis or from permanent sample plots, without the need for heights at a particular age to be represented in the data. For example, the equation may be fitted to all possible combinations of data intervals (Borders *et al.* 1984, Furnival *et al.* 1990) or to non-overlapping measurement periods, i.e., fitting  $Y(t_n) = f(t_n, t_{n-1}, h_{n-1})$  with the data  $(t_0, Y_0; t, Y)$ ,  $(t, Y; t_3, h_3)$ ,  $(t_3, h_3; t_4, h_4)$  (Borders *et al.* 1984, Furnival *et al.* 1990).

### 3.2 Parsimony

The Generalized Algebraic Difference Approach is more parsimonious than most traditional approaches to site equation derivations or formulations and can derive more complex equations than the traditional Algebraic Difference Approach discussed by Bailey and Clutter (1974). In terms of the potential for final equation flexibility, the GADA exceeds the capabilities of both the ADA and the fixed-base-age modeling approaches. Although restricted by the need for explicit base site equation solvability for  $\mathcal{X}$ , this approach can, in various cases, produce more flexible equations with fewer parameters than their fixed-base-age counterparts.

The derivation defined by the GADA automatically reduces the number of parameters by canceling redundant terms during routine algebraic operations. The **strictly hypothetical definition** of  $\mathcal{X}$ , without commitment to any specific values, and the substituting into the base equation of its initial condition solution, are two key factors contributing to the extremely parsimonious nature of the GADA.

**The first factor** is that all such equations as eq. (5), eq. (6) and, for example,

$$Y(t, S) = \left( \left( \alpha_1 + \alpha_2^{\alpha_3 S^{\alpha_4 + \alpha_5 S^{\alpha_6}}} + \alpha_7 S^{\beta_1 S^{\beta_2 + S^{\beta_3}}} \right)^{\alpha_8 + \alpha_9 S} (1 - e^{-\gamma t}) \right)^\delta$$

can be defined, without any loss to the underlying capabilities of generating any specific curve shapes, as simply:

$$Y(t, \mathcal{X}) = \mathcal{X} (1 - e^{-\gamma t})^\delta$$

where  $\mathcal{X}$  could be as in equations (5) and (6), or the equation above:

$$\mathcal{X} = \begin{cases} \alpha S; & \text{or} \\ \alpha S^\beta; & \text{or} \\ \left( \left( \alpha_1 + \alpha_2^{\alpha_3 S^{\alpha_4 + \alpha_5 S^{\alpha_6}}} + \alpha_7 S^{\beta_1 S^{\beta_2 + S^{\beta_3}}} \right)^{\alpha_8 + \alpha_9 S} \right)^\delta; & \text{or} \\ \text{anything else — all the same as far as the GADA is concerned.} \end{cases}$$

**The second factor** is that  $\mathcal{X}$  disappears from the site equation without adding any new parameters. Consequently, the derived dynamic equation is base-age-invariant, and has many desirable properties, including that it always computes appropriate values at any base-ages.



In the Lake States curves (Lundgren and Dolid 1970), the  $\alpha$  values computed from eq. (8) (Table 2; col. 2) are different from the published values (col. 3). The heights predicted at base-age with this model for the three different site index classes in Table 2 are incorrect; these height curves do not go through appropriate heights at the base-age. Other models may have similar disparity problems. For example, eq. (6) will not generate curves that unconditionally go through appropriate heights at reference age unless the parameter  $\beta = 1$  and the parameter  $\alpha$  is equal to the values calculated by eq. (8) as in Table 1 (col. 2).

By setting  $Y(50, S) = S$  one may be able to calculate, for any  $\alpha$  and any  $\delta \neq 1$ , one unique  $S$  value, for which a height curve may in fact go through an appropriate height at reference age as

$$S(t, Y) = \left[ \alpha (1 - e^{-\gamma A_b})^\delta \right]^{\frac{1}{1-\beta}}$$

If, for some reason, it is desirable that height at base-age not equal site index, one can modify the equation using “nuisance parameters” by temporarily redefining  $Y_0$  with a simple scale parameter. The “nuisance parameters” are used only for the purpose of the regression analysis, and they are removed from the equation after completing the analysis.

## Acknowledgements

I am grateful to Dr. R.L. Bailey for his extensive help in editing earlier versions of this manuscript.

## Literature Cited

- Anonymous. 1985. Alberta Phase 3 Forest Inventory: Yield Tables for Unmanaged Stands. Alberta Dep. Energy and Nat. Res. Rep. 60.
- Bailey, R.L., and J.L. Clutter. 1974. Base-age Invariant Polymorphic Site Curves. *For. Sci.* 20: 155–159.
- Biging, G.S. 1985. Improved Estimates of Site-index Curves Using a Varying-parameter Model. *For. Sci.* 31: 248–259.
- Bredenkamp, B.V., and T.G. Gregoire. 1988. A Forestry Application of Schnute’s Generalized Growth Function. *For. Sci.* 34–3: 790–797.
- Cieszewski, C.J., and I.E. Bella. 1989. Polymorphic Height and Site-index Curves for Lodgepole Pine in Alberta. *Can. J. For. Res.* **19**: 1151-1160.
- Cieszewski, C.J., and I.E. Bella. 1991. Polymorphic Height and Site Index Curves for the Major Tree Species in Alberta. *For. Can., Northwest Reg., North. For. Cent., Edmonton, Alberta.* For. Manage. Note 51.
- Cieszewski, C.J. 1994. Development of a Variable Density Height-Growth-Model Through Defining Multidimensional Height Growth Spaces. Ph.D. Thesis, U. of A., Edmonton, AB, CA. 72p
- Goudie, J.W. 1984. Height and  $S$  curves for lodgepole pine and white spruce and interim managed stand yield tables for lodgepole pine in British Columbia. Internal report, FY-1983-84, Submitted to Research Branch, British Columbia Ministry of Forests.

- Knight, W. 1968. Asymptotic Growth: an Example of Nonsense Disguised as Mathematics. *J. Fish Res. Board Can.* 25: 1030–1307.
- Monserud, R.A. 1984. Height Growth and Site-index Curves for Inland Douglas-fir Based on Stem Analysis Data and Forest Habitat Type. *For. Sci.* 30: 943–965.
- Peschel, W. 1938. Die mathematischen Methoden zur Herleitung der Wachstumsgesetze von Baum und Bestand und die Ergebnisse ihrer Anwendung. *Tharandter Forstl. Jahrb.* 89, 169-247.
- Ricker, W.E. 1979. Growth Rates and Models. Chapter 11, *Fish Physiology*, Vol. 8, Academic Press, New York, NY.
- Schnute, J. 1981. A Versatile Growth Model with Statistically Stable Parameters. *Can. J. Fish Aquat. Sci.* 38(9): 1128–1140.
- Schumacher, F.X. 1939. A New Growth Curve and its Application to Timber Yield Studies. *J. For.* 37: 819–820.
- Smith, V.G. 1984. Asymptotic Site-index Curves, Fact or Artifact. *For. Chron.* 60: 150-156.