Three Approaches to Derivation of Inventory Projection Equations

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

PMRC Technical Report 1999–6

Prepared by

C.J. CIESZEWSKI

October 8, 1999

Abstract

Site equations compute values of a variable Y as a function of a variable t and a sample value of the variable Y at a sample $t = t_0$. For example, site equation may describe the plant size (Y) as a function of age (t) and a sample reference size (Y_0) at an arbitrary base age (t_0). The base-age can be explicit, as in dynamic equations (e.g., $Y = f(t, t_0, y_0)$), or implicit, as in fixed base-age equations (e.g., $Y = f(t, S)$). We compare here a fixed base-age height growth site equation with several dynamic equations derived through the traditional and the Generalized Algebraic Difference Approach, and draw conclusions about the methodologies and their expected outcomes. Both algebraic difference approaches are more parsimonious and robust than the fixed base age approaches. The generalized approach can derive more complex equations and allows for more flexible implementation of theoretical bases into model derivation than the traditional Algebraic Difference Approach.

Key words: model derivation and conditioning, nonlinear and biological models, site productivity, base age invariant equations, dynamic equations.

Background

The earliest efforts of growth modeling in forestry concentrated on two-dimensional relationships (e.g., height over age). Peschel (1938) credits Spath in 1797 and Hossfeld in 1822 with the first efforts to express growth and yield relationships in forestry as mathematical equations of two variables (e.g., $Y = f(t)$). However, both hand drawn curves and these earliest equations approximated only two-dimensional relations (Fig. 1a). In forestry, almost all dynamic processes are necessarily dependent on at least one more dimension of the different management practices and ecological and productivity sites—hence: site models. Historically, these models were developed separately for different sites or even individually for different stands. Till today, some users of site models prefer to consider them as two-dimensional relationships and sometimes even denote sites as independent discrete classes, e.g., **A**, **B**, ..., or **I**, **II**, ... (Fig. 1b). However, in the context of mathematical models, the site equations are better explained as describing three-dimensional relationships (Fig. 1c), or surfaces (Fig. 1d). These are defined by two observable variables and one unobservable variable, which symbolizes the site productivity and management (e.g., $Y = f(t, \mathcal{X})$), which is a function of many variables. To be practically useful these models are based on implicit functions using snapshot observations of the two observable variables Y and t instead the unobservable variable (e.g., $Y = f(t, t_0, Y_0)$). The unobservable variable defines the productivity sites is an unknown function of climate, soil, genetic, and management components and it cannot be reliably measured or even functionally defined.

Currently, in the USA, the most popular are mathematical models with site productivity represented by a fixed base-age site index (S), that is, a height at a fixed and implicit base-age (e.g., $Y = f(t, S)$). An early algebraic inclusion of S into simple anamorphic (Fig. 2) equations was followed by increasingly complex equations necessary to describe numerous desirable characteristics. Examples of such characteristics include concurrent polymorphism (Fig. 3) and variable asymptotes (Fig. 4), and, at times, even crossings of growth trajectories between different sites (Fig. 5).

More advanced approaches to site dependent modeling are based on initial condition difference equations of the form $Y = f(t, t_0, Y_0)$ that hereafter, We call dynamic equations. Bailey and Clutter (1974) introduced the concept of **base-age invariance**, according to 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), which essentially consists of replacing a base-model parameter with its initial condition solution. Many modelers have since applied this approach to model various forest characteristics, but none of these applications produced a polymorphic model with variable asymptotes; all models derived with ADA are either anamorphic or have single asymptotes.

Cieszewski and Bailey (2000)[1] presents a generalization of the ADA, Generalized Algebraic Difference Approach (GADA), founded on expanding the base model suitably to considered data and theories about the modeled phenomena. This approach allows for the derivation of more flexible dynamic equations than the ADA and is suitable for a broad range of growth, yield, and decline or oscillation models, as well as, for improving existing fixed base-age site index models.

I compare here example applications of the ADA and the GADA to modeling site dependent height growth simulated by a published fixed base-age site equation. Then, We compare the resulting equations against the published one.

[1] Also Cieszewski (1994).

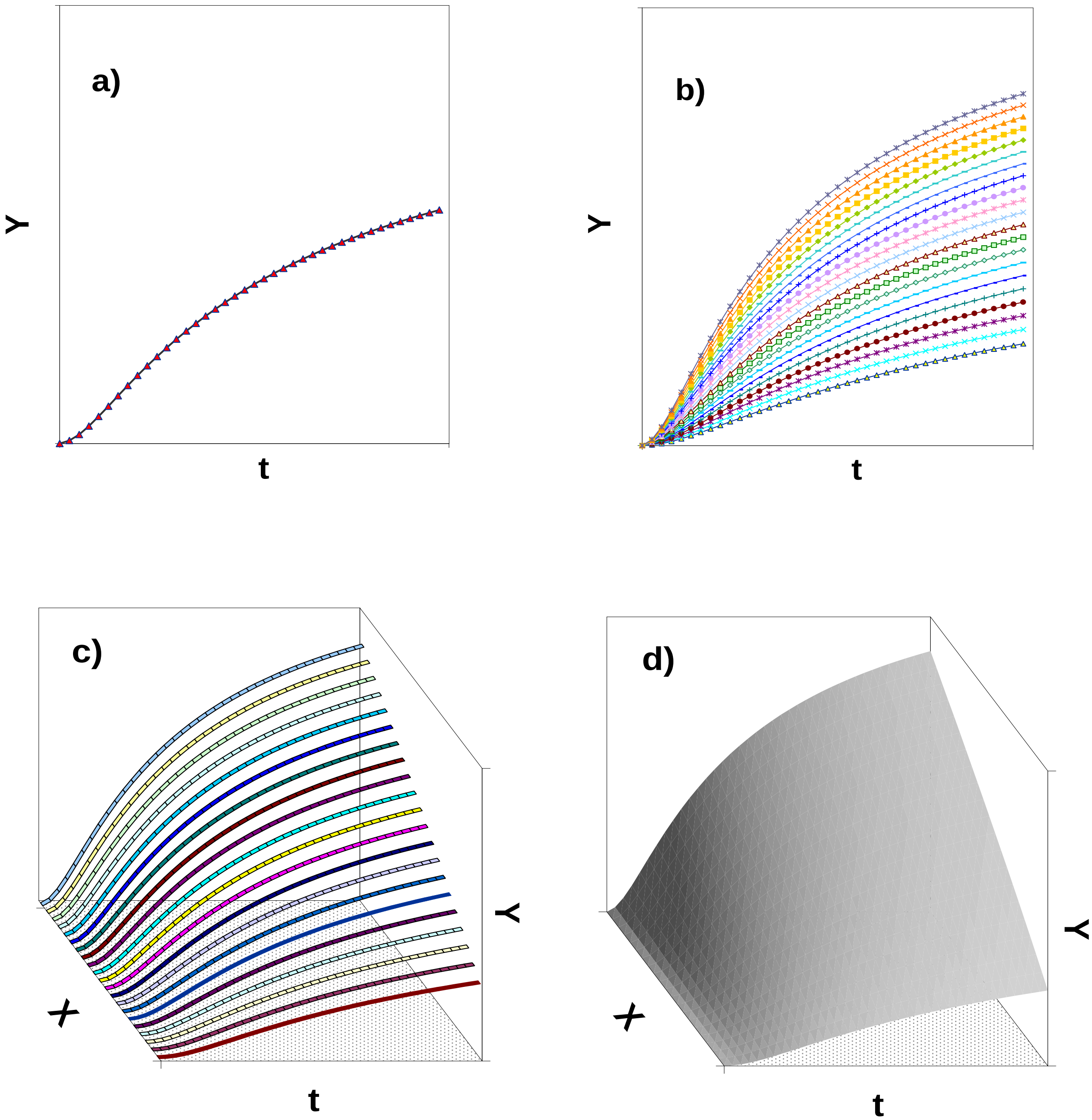

Figure 1: Geometric illustration of longitudinal data and model: (**a**); panel data and discrete model in two dimensions (**b**); panel data and discrete model in three dimensions (**c**); and all continuous three-dimentional panel data model (**d**); t is a longitudinal variable such as time, Y is the response variable and $\mathcal{X}$ is a third (cross-sectional) variable that may be unobservable.

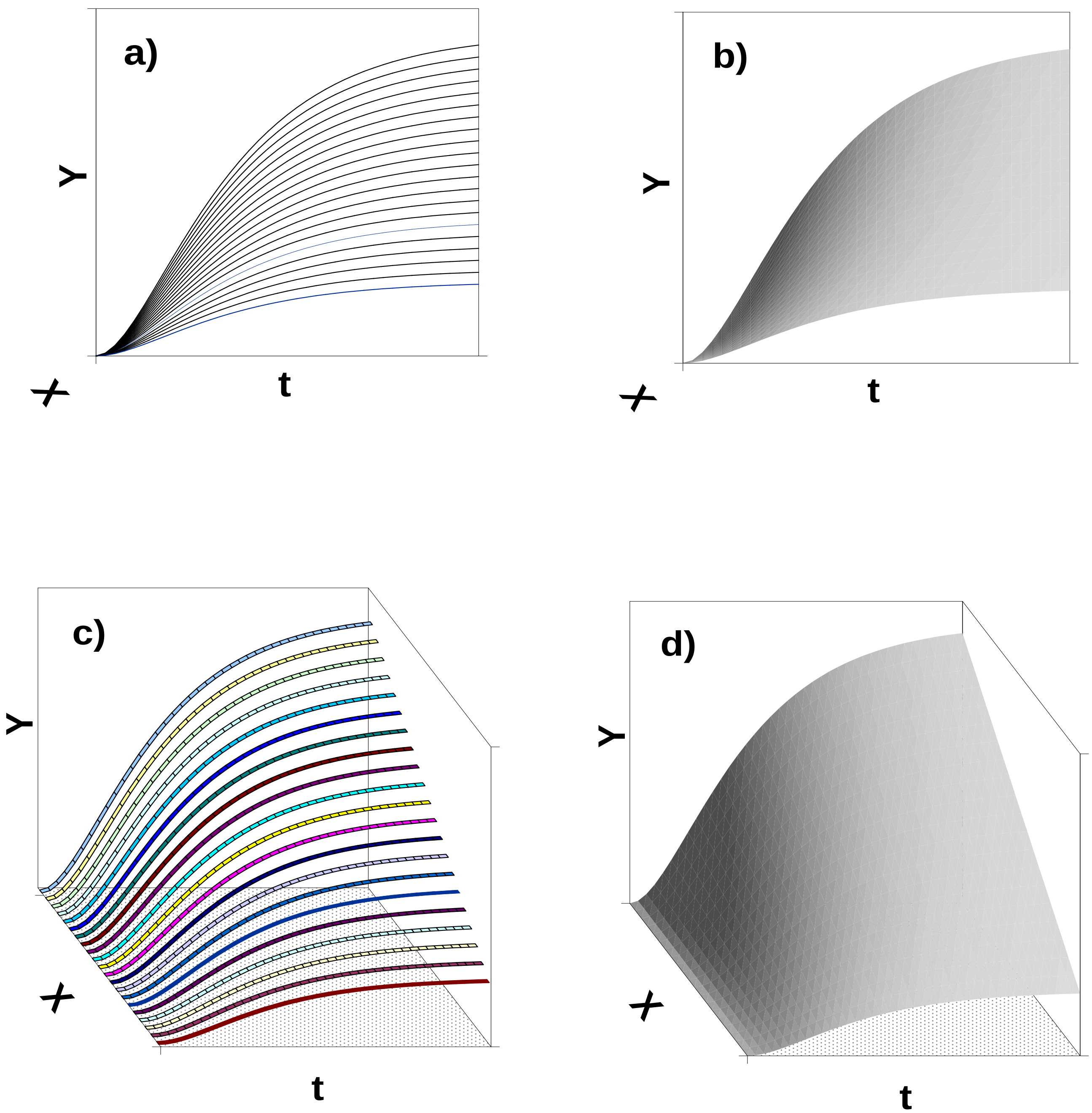

Figure 2: Anamorphic panel data model with variable asymptotes, t is time, Y is a response variable, and $\mathcal{X}$ is a factor determining intensity; a) 0^o rotation discrete representation; b) 0^o rotation continuous representation; c) 30^o rotation discrete representation; c) 30^o rotation continuous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

The Pseudo Data

We generated the pseudo-data with the following model (Monserud 1984):

$$H(t, S) = \frac{\alpha S^{\beta}}{1 + e^{\gamma + \delta \ln t + \zeta \ln S}} \qquad (1)$$

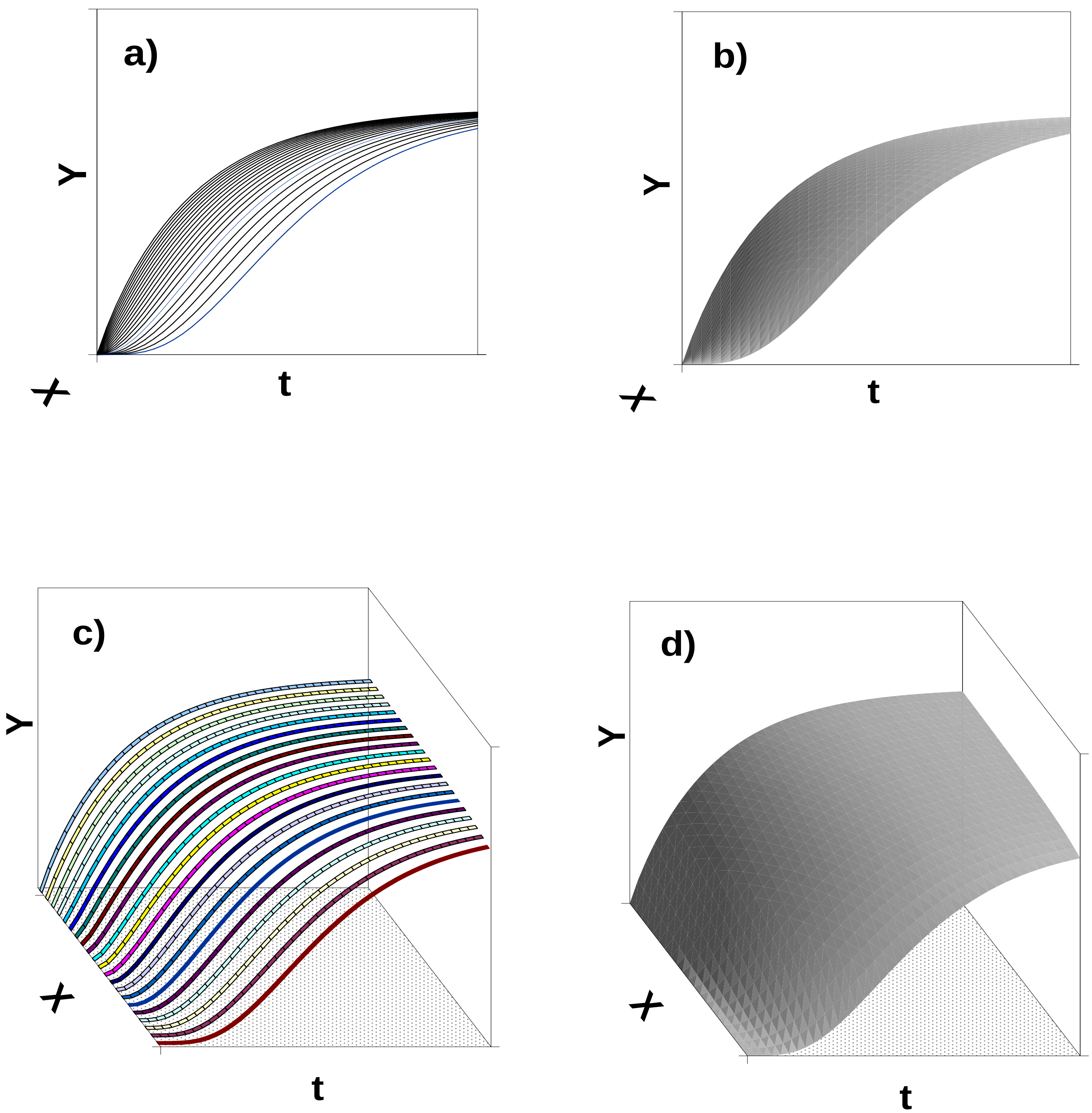

Figure 3: Polymorphic panel data model with single asymptote, t is time, Y is a response variable, and $\mathcal{X}$ is a factor determining intensity; a) 0^o rotation discrete representation; b) 0^o rotation continuous representation; c) 30^o rotation discrete representation; c) 30^o rotation continuous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

where: H is the height, t is age, S is site index, that is, height at age 50, and α, β, γ, δ, and ζ are the model parameters for the Douglas fir "general" model.

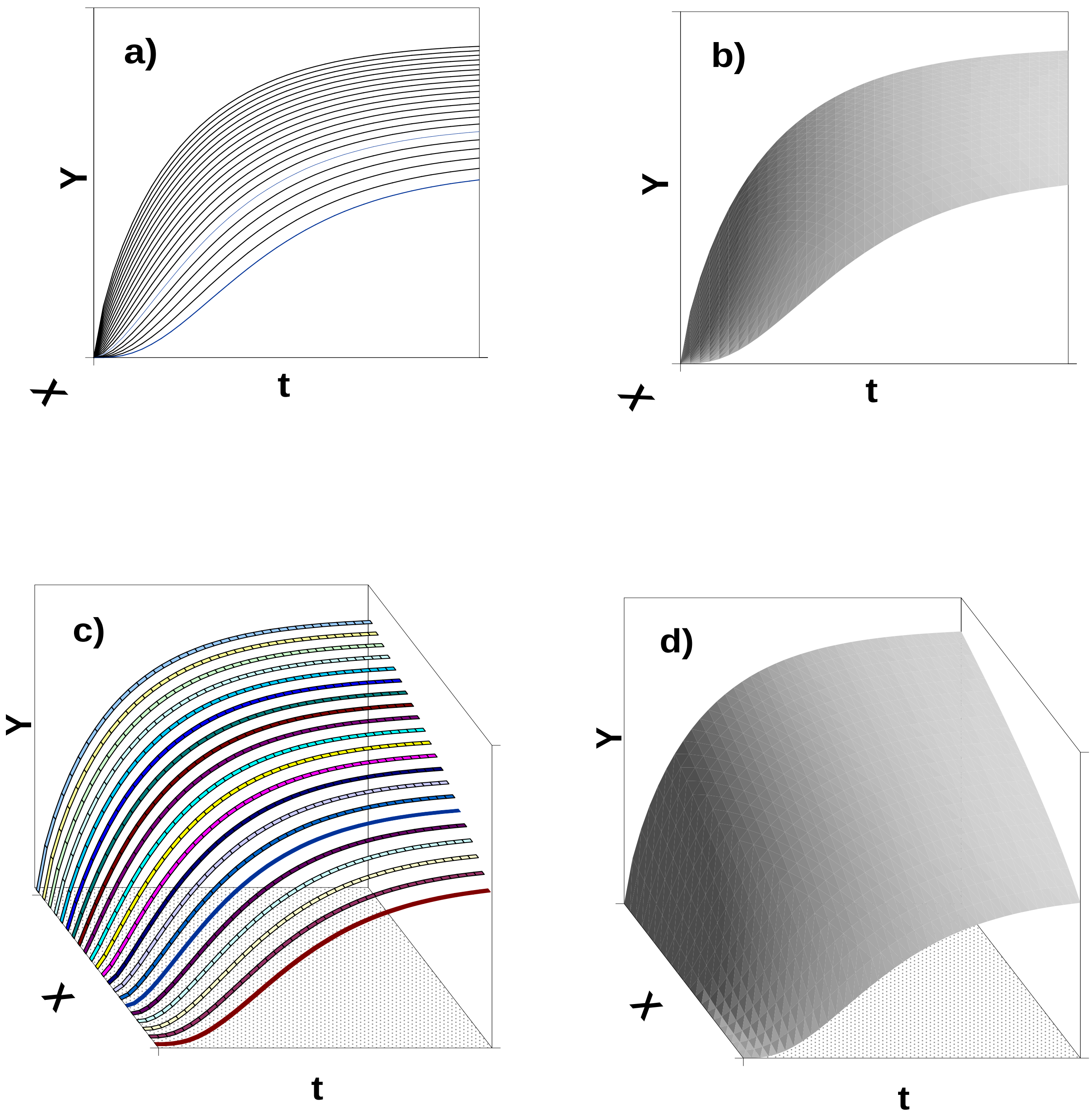

Figure 4: Polymorphic disjoint panel data model with variable asymptotes, t is time, Y is a response variable, and $\mathcal{X}$ is a factor determining intensity; a) 0^o rotation discrete representation; b) 0^o rotation continuous representation; c) 30^o rotation discrete representation; c) 30^o rotation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

The model was used to generate heights for decadal ages of 10 to 200 years for site indexes 12, 20 and 27 m at base-age 50 years as the applicable ranges represented by the original Inland Douglas

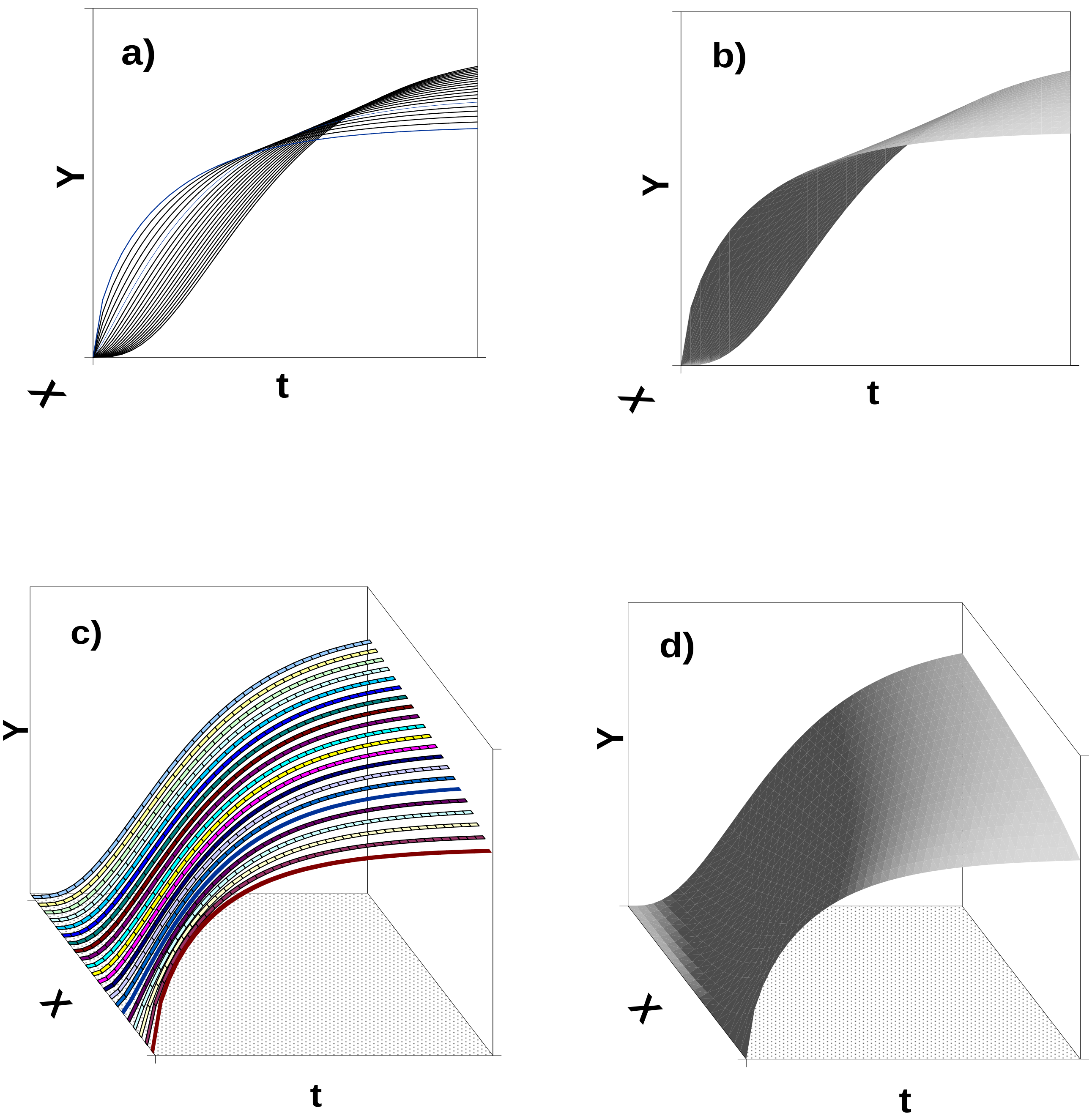

Figure 5: Polymorphic nondisjoint panel data model with variable asymptotes, t is time, Y is a response variable, and $\mathcal{X}$ is a factor determining intensity; a) 0^o rotation discrete representation; b) 0^o rotation continuous representation; c) 30^o rotation discrete representation; c) 30^o rotation continuous representation;

fir stem analysis data. A second set of predicted values, with pre-defined "noise" between -1 and 1 m with mean zero added to each prediction, was created with the help of a simple random number generator for the same decadal ages 10 to 200 years and the same site index values as above.

Methods

The main properties under consideration here are polymorphism and varying asymptotes. The ADA can derive only one of these properties at the time (e.g., Fig. 2 or Fig. 3). To compare equations derived with different approaches We use a common base equation to

i) derive different possible dynamic equations with the ADA,

ii) formulate a fixed base-age site index equation corresponding to one of the above dynamic equations, and,

iii) derive examples of different possible dynamic equations with the GADA;

iv) fit all the above equations to the pseudo-data; and

v) compare the results.

I start the derivations with the following log-logistic base equation,

$$Y(t) = \alpha/(1 + \beta e^{-t}) \tag{2}$$

where α is the asymptote parameter, β is the slope parameter, t is logarithmically transformed function of age such that, $t = \gamma \ln Age$, and all parameters are unique to equation (2) and its resultants in further derivations. In accordance with the Algebraic Difference Approach, one of the equations' parameters can be selected as a site specific parameter.

A-i. An anamorphic (Fig. 2) equation with variable asymptotes can be derived by solving for the parameter α as a function of initial conditions, i.e., $\alpha = Y_0\ (1 + \beta\, e^{-t_0})$, and substituting this solution to equation (2), which results in

$$Y(t, t_0, Y_0) = Y_0 \frac{1 + \beta\, e^{-t_0}}{1 + \beta\, e^{-t}} \tag{3}$$

A-ii. A polymorphic (Fig. 3) equation with a single asymptote can be derived by solving for β, i.e., $\beta = (Y_0 - \alpha)/(Y_0\, e^{-t_0})$, and substituting this solution to equation (2), which results in

$$Y(t, t_0, Y_0) = \alpha \left(1 + \frac{(Y_0 - \alpha)\, e^{-t}}{Y_0\, e^{-t_0}}\right)^{-1} \tag{4}$$

A-iii. Finally, a polymorphic (Fig. 3) equation with a single asymptote can also be derived with the ADA by solving and substituting for the parameter γ, but We do not consider this option any further.

The three dynamic equations above exhaust the possibilities for direct applications of the ADA to equation (2). The equations derived thereby can either have variable asymptotes and be anamorphic (Fig. 2) (A-i.) or have single asymptotes and be polymorphic (Fig. 3) (e.g., A-ii. or A-iii.). On the other hand, since the GADA is based on applying theories to formulating explicit site equations in terms of an explicit variable, it does not have similar limitations.

As described in Cieszewski and Bailey (2000), the main step in the GADA is expanding base equations according to assumptions or theories about various growth characteristics. Examples of such characteristics are the limiting size, or asymptote, carrying capacity (e.g., conforming to the 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 are tested with data.

Let $\mathcal{X}$ be the growth intensity factor, a theoretical variable, and the quantification of those particular growth dynamics that are uniquely associated with a site and individual characteristics of growth or survival capabilities. $\mathcal{X}$ is used 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 completely defined site equation, which can 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 entangling 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 the other approaches applied in forestry. The second characteristic distinguishing the GADA 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 single parameter of a one independent variable equation. 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).

If a given site-specific parameter, in the base equation, is defined as a function of $\mathcal{X}$ and any number of new parameters, the base equation with multiple site-specific parameters is changed to the explicit three-dimensional site equation with two independent variables t and $\mathcal{X}$. A specific form of the equation will depend on the assumptions made about functions of $\mathcal{X}$ replacing the site-specific model parameters. If the resulting equation can be solved for $\mathcal{X}$, the RHS of this solution with initial condition values of t and Y can be substituted for $\mathcal{X}$, which results in a dynamic equation.

The GADA can be used to derive all the models derived by the ADA. For example, to derive the polymorphic equation (4), one could formulate a response theory based on the parameter β being proportional to the unobservable variable $\mathcal{X}$, i.e., $\beta \equiv \beta'\mathcal{X}$, and then proceed with the methodology.

The GADA expands the concept of the ADA to multi-parameter site-specific responses. Various applications of the GADA to equation (2) can be based on different theories. Some examples follow.

G-i. ($\alpha \equiv \alpha'\mathcal{X}$ and $\beta \equiv \beta'\mathcal{X}$): the simplest theory is that both the growth limits and the slope are proportional to growth intensity, i.e.,

$$Y(t, \mathcal{X}) = \frac{\mathcal{X}}{1 + \beta\mathcal{X}e^{-t}}$$

which through a solution $\mathcal{X} = Y_0 \,/(1 - Y_0\beta e^{-t_0})$ leads to:

$$Y(t, t_0, Y_0) = \left(1/Y_0 - \beta\left(e^{-t_0} + e^{-t}\right)\right)^{-1} \tag{5}$$

G-ii. ($\alpha \equiv \alpha'\mathcal{X}$ and $\beta \equiv \beta'/\mathcal{X}$): since β in equation (2) sinifies a semi-saturation time, it can be considered as a slope- inverse. Therefore, a more plausible theory suggests that β is inversely-proportional to growth intensity, i.e.,

$$Y(t, \mathcal{X}) = \frac{\mathcal{X}}{1 + (\beta/\mathcal{X})e^{-t}} \tag{6}$$

and given a solution $\mathcal{X} = \frac{1}{2}(Y_0 \pm \mathcal{R})$, where $\mathcal{R} = \sqrt{{Y_0}^2 + 4\,Y_0\,\beta\,e^{-t_0}}$, the dynamic equation is:

$$Y(t, t_0, Y_0) = \frac{1}{2}\,\frac{(Y_0 + \mathcal{R})^2}{Y_0 + \mathcal{R} + 2\,\beta\,e^{-t}} \tag{7}$$

G-iii. ($\alpha \equiv \alpha' + \alpha''\mathcal{X}$ and $\beta \equiv \beta'/\mathcal{X}$): as asymptotes relate to theoretical values in infinity they may represent ranges of values that are offset from a direct proportion to growth intensity, i.e.,

$$Y(t, \mathcal{X}) = \frac{\alpha + \mathcal{X}}{1 + (\beta/\mathcal{X})e^{-t}}$$

and given a solution $\mathcal{X} = \frac{1}{2}(\phi \pm \mathcal{R})$ where: $\phi = (Y_0 - \alpha)$ and $\mathcal{R} = \sqrt{\phi^2 + 4\,Y_0\,\beta\,e^{-t_0}}$, the dynamic equation based on the theory of half-time inverse proportional to growth intensity with asymptotes that are linearly proportional but offset by a constant is:

$$Y(t, t_0, Y_0) = \frac{1}{2}\,\frac{(Y_0 + \mathcal{R})^2 - \alpha^2}{\phi + \mathcal{R} + 2\,\beta\,e^{-t}} \tag{8}$$

G-iv. ($\alpha \equiv \alpha'/\mathcal{X}$ and $\beta \equiv \beta'\mathcal{X}$): An alternative theory may be that the unobservable variable $\mathcal{X}$ that causes the differences between the longitudinal series is not anything like growth intensity but, for example, a level of ozone (O_3), length of dry season, or a depth of water table, etc., that has not been measured but varies among the sampled populations. Since ozone inhibits growth due to an excessive and harmful oxidization the response theories would have to be opposite to ones based on growth intensity. Thus, a plausible theory could be that the maximum yield (α) is inversely-proportional to the level of ozone and the half-saturation parameter β is proportional to the level of ozone, i.e.,

$$Y(t, \mathcal{X}) = \frac{\mathcal{X}^{-1}}{1 + (\beta\mathcal{X})e^{-t}}$$

and the solution is: $\mathcal{X} = 2(Y_0 \pm \mathcal{R})^{-1}$, where $\mathcal{R} = \sqrt{{Y_0}^2 + 4\,Y_0\,\beta\,e^{-t_0}}$. Not surprisingly, the dynamic equation based on the theory that asymptotes are inversely-proportional to ozone level and half-saturation times are proportional to ozone level is identical to the dynamic equation based on the theory that asymptotes are proportional to growth intensity and half-saturation times are inversely proportional to growth intensity (i.e., eq. (7)):

$$Y(t, t_0, Y_0) = \frac{1}{2}\,\frac{(Y_0 + \mathcal{R})^2}{Y_0 + \mathcal{R} + 2\,\beta\,e^{-t}} \tag{9}$$

The above are only five out of many, many possible formulations that would result from various theories concerning the biological limits and rates of growth expressed by the GADA. Finally, for the sake of comparison, We formulated one polymorphic fixed-base-age site index model using similar assumptions as the ones behind the polymorphic model. Hypothesizing similarly as in the previous case that the site responses are described by the parameter β and site index (S), instead of the unobservable variable ($\mathcal{X}$), i.e., $\beta \equiv \beta' S$, the polymorphic fixed-base-age site index equation is:

$$Y(t, S) = \alpha/(1 + \beta' S e^{-t}) \tag{10}$$

It may seem that this equation should be able to generate the same curves as equation (4) because it is based on almost identical assumptions. Indeed, the assumption: $\beta \equiv \beta' S$ seems very similar to the assumption: $\beta \equiv \beta'\mathcal{X}$, except for the fact that $\mathcal{X}$ can be anything while S can be only height at a certain age.

Statistical models based on the above equations have the following general form:

$$\hat{Y_{ij}} = f(t_{ij}, t_{i0}, \hat{Y}_{i0}, \Phi) \tag{11}$$

where: $\hat{Y_{ij}}$ is the estimated response variable of individual i at observation j, t_{ij} is the predictor variable, or age of individual i at observation j, t_{i0} is a constant specific to individual i, $\hat{Y}_{i0}$ is the estimated value of the response variable at the age t_{i0} and Φ is a vector of model global parameters common for all growth series. $\hat{Y}_{i0}$ can be estimated as a varying parameter unique to each growth series i.

The above statistical model formulation is a direct result of the algebraic derivation applied to a statistical base model of the form: $\hat{Y} = f(t, \mathcal{X}, \Phi)$, where $\mathcal{X} = f(t, \hat{Y}, \Phi)$, or $\hat{Y} = f(t, S, \Phi)$, where $S = f(t, \hat{Y}, \Phi)$. Accordingly, the values of $\hat{Y}_{0_i}$ should be estimated as site-specific parameters unique to each growth series; hence, the base-age independent parameter estimation procedures. However, it is common for site model developers to assume that $\hat{Y}_{0_i}$, or S, are "*error-free*" values equal to the height values observed in the data at the given base-age t_0; hence, the resulting base-age specific parameter estimation procedures. We explored both of these approaches to the estimation problem.

Results

After the dynamic equations (3) to (8) were fitted to the pseudo-data (see Table 1 for the results), as expected, equation (5) based on the theory that half-saturation times are proportional to growth intensity, produced unreasonable growth patterns (Fig. 6a). According to this theory, the achievement of different levels of yield at younger ages is delayed in proportion to the general level of growth intensity or site productivity. The curves produced by such a system merge close together at young ages and spread out at older ages. The model based on equation (5) was unable to simulate growth patterns even remotely similar to those observed in the data.

However, equation (9), based on an identical assumption that $\beta \equiv \beta' \mathcal{X}$ and an equally unreasonable assumption of $\alpha \equiv \alpha' / \mathcal{X}$, did very well in fitting the data (Fig. 6b). Similarly, so did equation (4) which is based on the same assumption of $\beta \equiv \beta' \mathcal{X}$. On the other hand, the fixed-base-age site index equation (10) built on a seemingly identical assumption to that for equation (4) ($\beta \equiv \beta' S$) did much worse than any other equation. Two questions that one may ask with respect to the above are:

i) Why did the one-parameter dynamic equation (4) fit the data so much better than the two-parameter fixed-base-age site index equation (10) if both of these equations are based on similar assumptions?

ii) Is it possible that the two equations describe different classes of growth patterns and that for some other data the two-parameter fixed-base-age site index equation could fit the data better than the dynamic equation?

The answers to these questions follow from a comparison of the derivations of equations (5) and (9). Consider that the unobservable variable $\mathcal{X}$ can be growth intensity or ozone level, or a function of both, or something else entirely. Whatever it is, it is replaced in the final model by the initial conditions. Thus, the dynamic equations represent generalized relationships that include the corresponding site index equations as special cases. It is inconceivable that the two-parameter fixed-base-age site index equation (10) could fit any particular data better than the one-parameter dynamic equation (4). The converse is likely. An example of a fit of the dynamic equation (4) to the fixed base-age equation (10) is illustrated in Figure 6d.

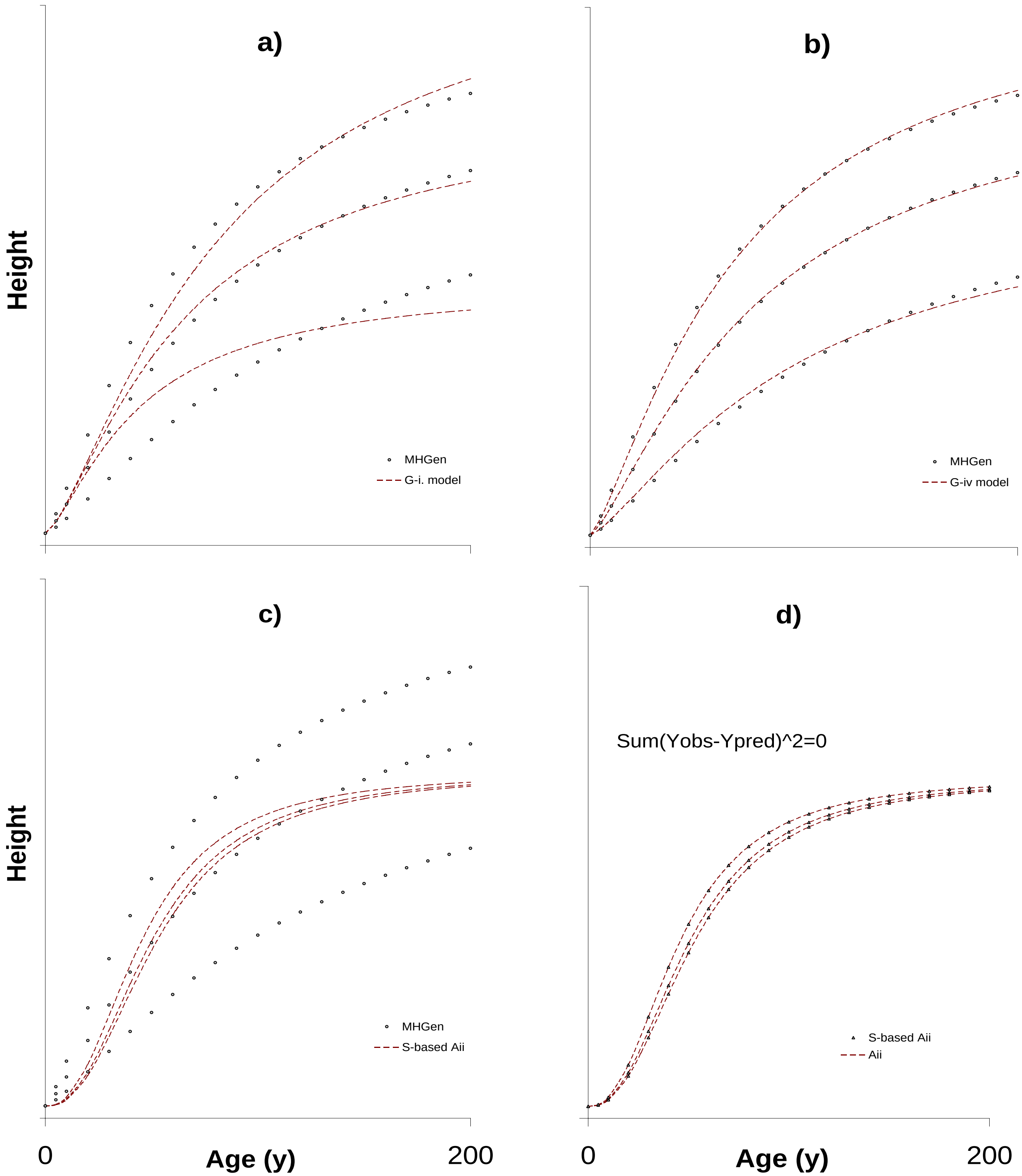

Figure 6: Fit of equations: **a**) dynamic equation G-i and the pseudo-data MHGen; **b**) dynamic equation G-iv and the pseudo-data MHGen; **c**) S-based equation S-i and the pseudo-data MHGen; and **d**) dynamic equation G-ii and pseudo-data from S-based equation S-i.

In what follows, only fittings of the models based on equation (3), (4), (7) and (8) are subsequently discussed (Fig. 7). Furthermore, equation (7) is a special case of equation (8), with $\alpha = 0$. Only the later equation was explored as the best demonstration of the advantage of polymorphism and variable asymptote (Fig. 4) equations derived with the GADA.

Table 1: Regression results for different models with varying properties of polymorphism and asymptotes (n=66, S = 65, 90, 40), fitted to pseudo-data generated with Monserud's (1984) model; ϵ denotes random noise added to the pseudo-data.

	Model Forms							
Statistic	A-i.	A-ii.	$S_{\sim A-ii.}$	A-ii$_{S_{\sim A-ii.}}$	G-i.	G-ii./G-iv.	G-iii.	G-iii.+ϵ
				Base age independent fitting				
Variance of residuals	12.159	3.744	732.049	0.000	51.028	2.501	0.021	3.531
Reg. Std. error	3.496	1.936	27.084	0.000	7.157	1.583	0.144	1.879
Mean residual	0.244	0.060	-1.214	0.000	0.447	0.078	-0.001	0.005
Minimum resudual	-6.984	-2.342	-44.185	0.000	-12.903	-3.585	-0.302	-3.464
Maximum residual	7.036	4.797	44.473	0.000	15.851	2.909	0.280	3.483
α	n/a	214.687	123.880	123.880	n/a	n/a	110.839	107.322
β	255.479	n/a	144.000	n/a	2.078	104466.7	39033.9	40605.0
γ	1.271	1.121	2.349	2.349	1.391	1.313	1.293	1.292
$\hat{Y}_{0_1}$	67.013	63.803	69.562	65.751	69.563	66.055	255.479	65.152
$\hat{Y}_{0_2}$	83.386	90.140	77.823	62.282	77.823	86.456	88.635	88.614
$\hat{Y}_{0_3}$	46.218	37.761	55.032	72.982	55.032	42.078	39.031	39.710
				Base age specific fitting (50y)				
Variance of residuals	78.817	4.909	n/a	n/a	n/a	n/a	0.080	6.904
Std. error of regression	9.146	2.243	n/a	n/a	n/a	n/a	0.283	2.629
Mean residual	-2.198	0.350	n/a	n/a	n/a	n/a	0.012	0.078
Minimum resudual	-22.738	-2.954	n/a	n/a	n/a	n/a	-0.609	-6.280
Maximum residual	14.220	5.449	n/a	n/a	n/a	n/a	0.475	5.214
α	n/a	228.278	n/a	n/a	n/a	n/a	111.870	92.448
β	249.8	n/a	n/a	n/a	n/a	n/a	38224.2	47377.8
γ	1.273	1.055	n/a	n/a	n/a	n/a	1.288	1.271
				Base age specific fitting (100y)				
Variance of residuals	18.180	4.374	n/a	n/a	n/a	n/a	0.026	6.557
Std. error of regression	4.276	2.576	n/a	n/a	n/a	n/a	0.163	2.581
Mean residual	-0.319	1.505	n/a	n/a	n/a	n/a	0.010	-0.322
Minimum resudual	-11.464	-0.792	n/a	n/a	n/a	n/a	-0.401	-5.470
Maximum residual	7.426	6.716	n/a	n/a	n/a	n/a	0.268	4.958
α	n/a	207.657	n/a	n/a	n/a	n/a	110.467	97.640
β	253.1	n/a	n/a	n/a	n/a	n/a	39188.5	41374.5
γ	1.270	1.146	n/a	n/a	n/a	n/a	1.292	1.209

Model fitting was based on a simple non-linear least squares regression. Both, base-age independent and base-age dependent, approaches produced similar results and identical conclusions (Table 1). The anamorphic version (Fig. 2, eq. (3)) was the worst (e.g., Fig. 7a and 8a), while the polymorphic model with variable asymptotes linearly proportional to growth intensity (i.e., eq. (8)) was the best (e.g., Fig. 7d and 8d). The difference between the fits of the two polymorphic models with variable asymptotes (eqs. (7) vs. (8), (Fig. 7c vs. 7d and 8c vs. 8d), was much smaller then the corresponding difference between the single asymptote polymorphic equation (4) (Fig. 3) and the polymorphic equation (7) with proportionally variable asymptotes, (Fig. 7b vs. 7c and 8b vs. 8c).

This, and the earlier example of equation (5) demonstrate the importance of adopting sound biological theories in equation derivation. Clearly, it is not sufficient to derive just any equation that is polymorphic and with variable asymptotes (Fig. 4). The specific responses of the polymorphism and the variation in asymptotes have to be consistent with biological reality and data. A prime example of this is the outstanding fit of equation (8) as derived with the GADA. The residual variance from fitting this equation is more than four orders of magnitude smaller (Table 1) than that from the best fit of the models derived with the ADA (i.e., eq. (4)).

The residual variance of the polymorphic (Fig. 3) model is over ten times smaller than that from the anamorphic (Fig. 2) model and over 2 times greater than the residual variance from the worst variable-asymptote polymorphic model (i.e., eq. (7)).

In principle, the model parameters and the shape of fitted curves depend partially on the method of fitting. However, this dependence is subject to different sources of errors and can exhibit different intensity. The greatest influence comes from the model's inadequacy to represent underlying growth patterns in the data (e.g., Fig. 9a vs. 9c). Smaller differences result from data measurement errors, and from time-dependent variation in growth conditions, such as weather patterns or years with varying precipitation (e.g., Fig. 9c vs. 9d). In addition, another source of such differences is inconsistency in overall growth patterns resulting from space-dependent variation in growth conditions and from variation in genetic growth predisposition. Such variation is present in many true data but cannot be realistically simulated with the pseudo-data used here. If, i) a model adequately reflects underlying growth patterns, ii) the growth series are consistent, and iii) there are no measurement errors in the data, all methods, the base-age invariant and the base-age variant, produce similar results (e.g., Fig. 9c). Such a situation is rather unlikely to happen. Yet, the consideration of the differences in model parameters and curve shapes created by different fitting techniques are secondary in importance to the identification of an appropriate model form that adequately reflects growth trends. As demonstrated above with the GADA application to different model derivations, an appropriate equation form and clean data will minimize the differences between the various outcomes of different fitting techniques. An appropriate equation form can make a much greater difference in minimizing the surface of square residuals than the differences resulting from applications of various fitting techniques; proper equation flexibility, adequately matching underlying data patterns, minimizes the differences between parameter estimates of different base-age dependent techniques.

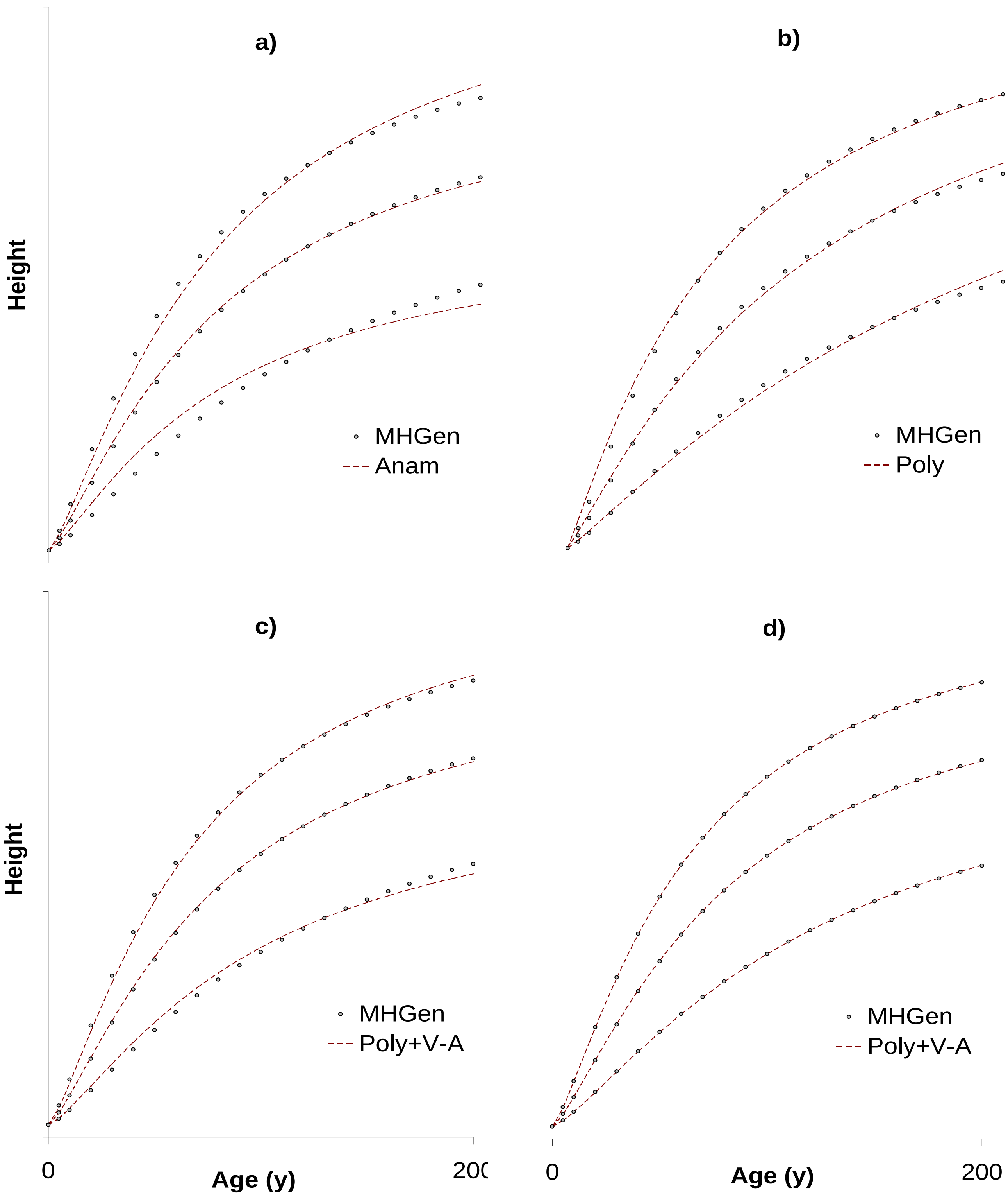

Figure 7: Impact of polymorphism and variable asymptotes of site models on base age invariant (BI) fits to the Monserud's (1984) pseudo-data: a) anamorphic model; b) polymorphic model with single asymptote; c) polymorphic model with directly proportional asymptotes; and d) polymorphic model with linearly proportional asymptotes;

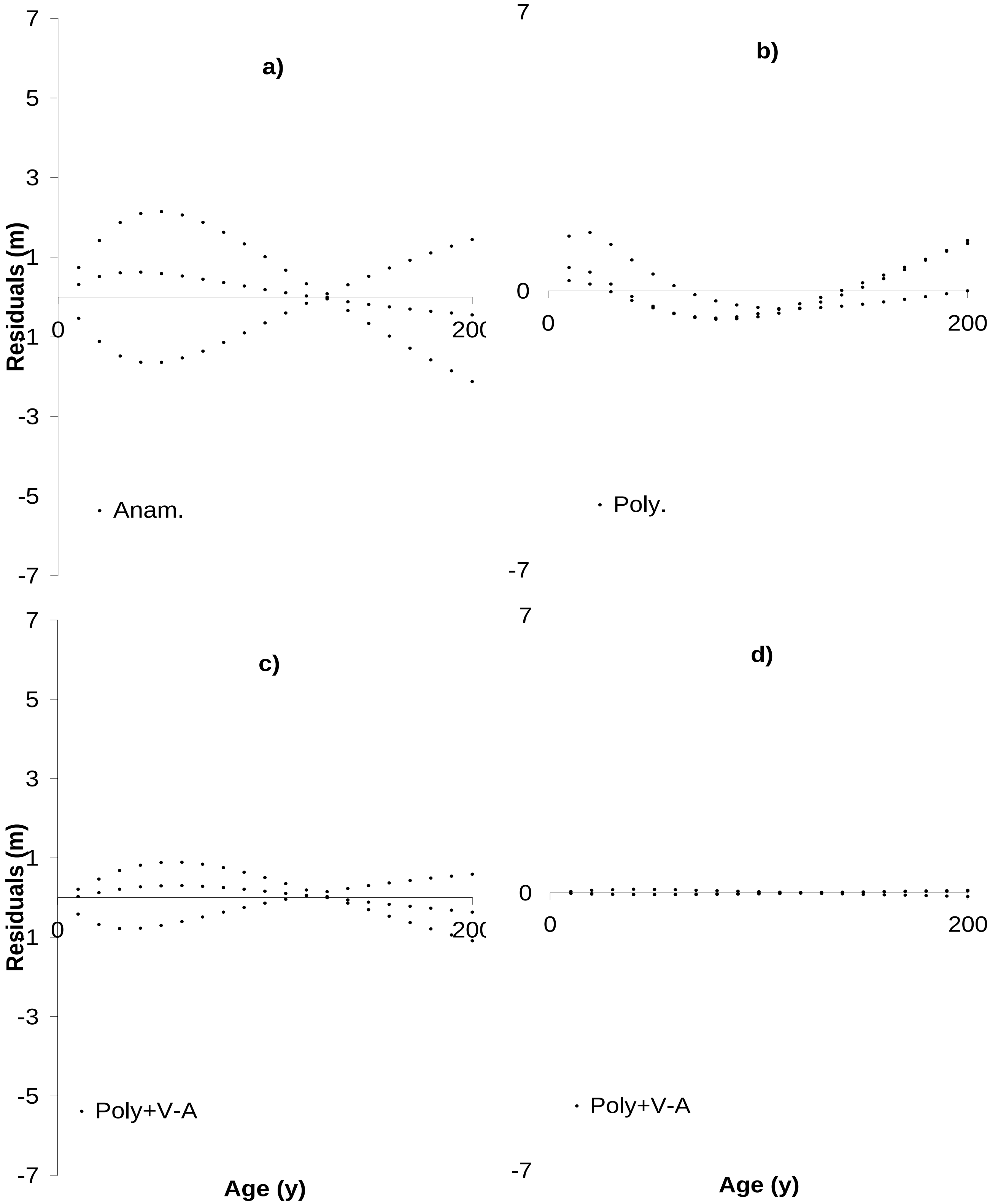

Figure 8: Residuals for the fitting of site models with different properties: a) anamorphic model; b) polymorphic model with single asymptote; c) polymorphic model with directly proportional asymptotes; and d) polymorphic model with linearly proportional asymptotes;

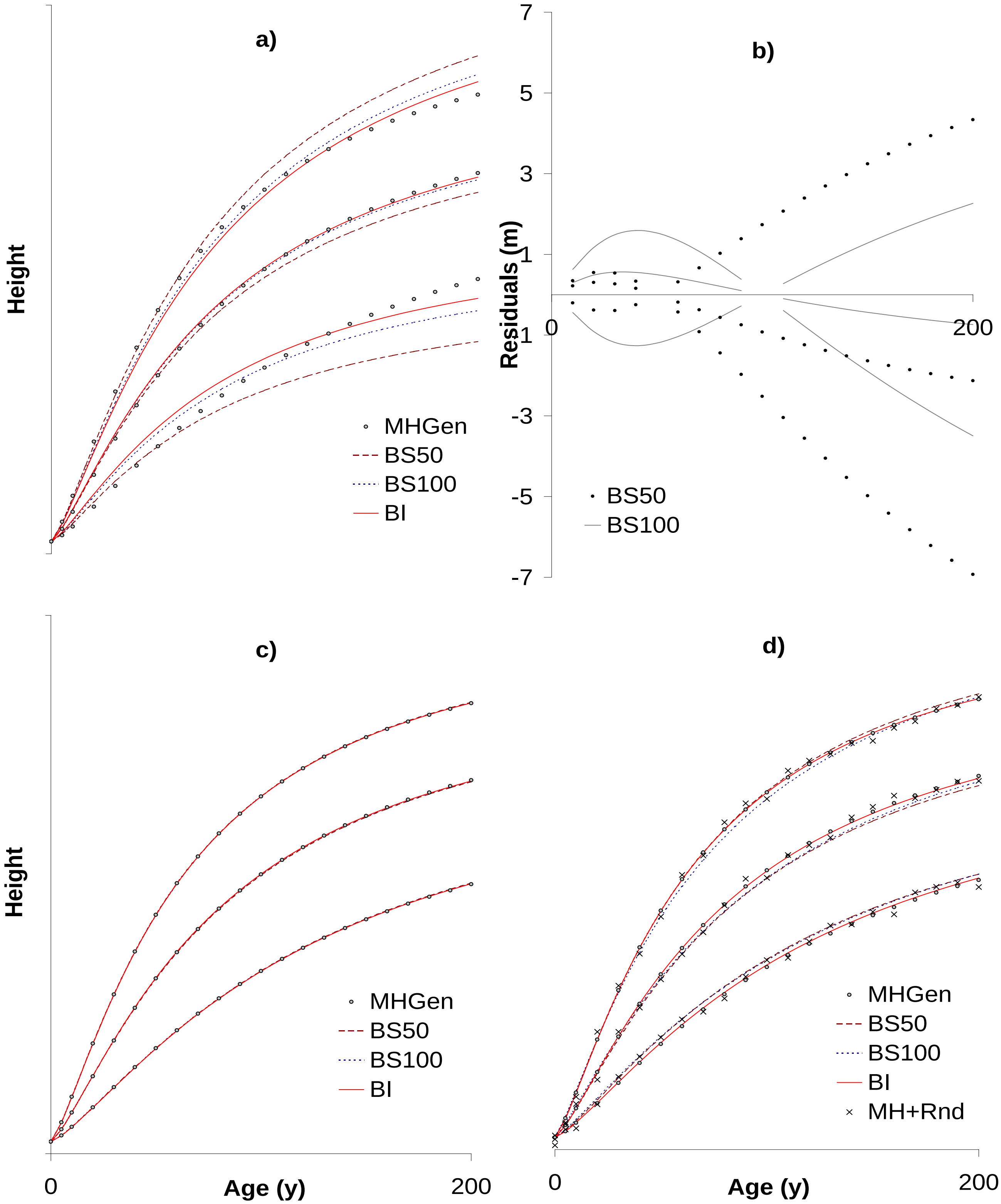

Figure 9: Impact of errors on base aga specific (BS) fitting of site models using base ages 50 and 100 years to the Monserud (1984) pseudo-data: a) anamorphic model error; b) residuals for the BS fitting of the anamorphic model; c) lack of model and measurement errors; and d) measurment error;

Discussion

Parsimony

The Generalized Algebraic Difference Approach is more parsimonious than most traditional approaches to site equation derivations or formulations. It can derive more complex equations than the Difference Approach (Bailey and Clutter 1974). In terms of the potential for final equation flexibility, it exceeds the capabilities of fixed-base-age modeling approaches. Although, restricted by the need for explicit base site equation solvability for $\mathcal{X}$ as illustrated in Figure 10, this new 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 cancellation of terms during routine algebraic operations. Such is not as likely to happen when dealing with fixed-base-age site index equations. An example can be the Schumacher (1939) equation with the asymptote $\alpha \propto \alpha' \mathcal{X}^\gamma$ and the slope $\beta \propto \beta' \mathcal{X}^\gamma$. Since $\mathcal{X}$ is an unobservable variable and, unlike site index, has only a theoretical meaning not intended for practical explicit use, it can be redefined as either $\mathcal{X}' = \alpha' \mathcal{X}^\gamma$ or $\mathcal{X}' = \beta' \mathcal{X}^\gamma$.

a) The modeler can start with a three-parameter explicit equation

$$Y(t, \mathcal{X}) = \alpha \mathcal{X}^\gamma + \beta \mathcal{X}^\gamma / t \tag{12}$$

that has the three-parameter solution $\mathcal{X} = (Y_0\, t_0/(\alpha\, t_0 + \beta))^{1/\gamma}$. The process of entangling this into the original explicit equation produces one of the following one-parameter[2] dynamic equations:

$$Y(t, t_0, Y_0) = Y_0 \frac{t_0\,(\alpha\, t + \beta)}{t\,(\alpha\, t_0 + \beta)} = \underbrace{\overbrace{Y_0 \frac{t_0\,(t + \beta/\alpha)}{t\,(t_0 + \beta/\alpha)}}^{\text{either,}} = \overbrace{Y_0 \frac{t_0\,((\alpha/\beta) t + 1)}{t\,((\alpha/\beta) t_0 + 1)}}^{\text{or}}}$$

b) The above is equivalent to starting with a one parameter explicit equation $Y(t, \mathcal{X}) = \mathcal{X} + \beta \mathcal{X}/t$ where the solution is $\mathcal{X} = Y_0\, t_0/(t_0 + \beta)$ and the dynamic equation is

$$Y(t, t_0, Y_0) = \overbrace{\underbrace{Y_0 \frac{t_0\,(t + \beta)}{t\,(t_0 + \beta)}} \equiv \underbrace{Y_0 \frac{t_0\,(t + \beta/\alpha)}{t\,(t_0 + \beta/\alpha)}}}^{\text{either,}}$$

c) Similarly, it is equivalent to starting with one parameter explicit equation $Y(t, \mathcal{X}) = \alpha \mathcal{X} + \mathcal{X}/t$ where $\mathcal{X} = Y_0\, t_0/(\alpha\, t_0 + 1)$ and the dynamic equation is

$$Y(t, t_0, Y_0) = \overbrace{\underbrace{Y_0 \frac{t_0\,(\alpha\, t + 1)}{t\,(\alpha\, t_0 + 1)}} \equiv \underbrace{Y_0 \frac{t_0\,((\alpha/\beta) t + 1)}{t\,((\alpha/\beta) t_0 + 1)}}}^{\text{or}}$$

d) On the contrary, if $\mathcal{X}$ is replaced by S_{50} in equation (12), a modeler is not very likely to notice that

$$Y(t, S_{50}) = \alpha S_{50}^\gamma + \beta S_{50}^\gamma / t$$

has three times as many parameters as necessarily for the given shapes of curves.

[2] Note that either β/α or its inverse is just a single parameter

$$Y(t,S) = \alpha S \left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}$$
where:
$$\mathcal{X} = \alpha S$$

$$Y(t,S) = \frac{\left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}}{\left(\alpha_0 + \alpha_1^{\alpha_2 S^{\beta_1}}\right)^{-\beta_2 S}}$$
where:
$$\mathcal{X} = \left(\alpha_0 + \alpha_1^{\alpha_2 S^{\beta_1}}\right)^{\beta_2 S}$$

$$Y(t,S) = \alpha S^{\beta} \left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}$$
where:
$$\mathcal{X} = \alpha S^{\beta}$$

$$Y(t,S) = \frac{\left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}}{\alpha S^{\beta}}$$
where:
$$\mathcal{X} = \frac{1}{\alpha S^{\beta}}$$

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

$$Y(t,S) = \frac{\left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}}{\alpha \sqrt[n]{\ln S^{\beta}}}$$
where:
$$\mathcal{X} = \frac{1}{\alpha \sqrt[n]{\ln S^{\beta}}}$$

$$\mathcal{X} = \frac{1}{\alpha S}$$
where:
$$Y(t,S) = \frac{\left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}}{\alpha S}$$

$$\mathcal{X} = Y(t,S) = \alpha \sqrt[n]{\ln S^{\beta}}$$
where:
$$Y(t,S) = \alpha \sqrt[n]{\ln S^{\beta}} \left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}$$

$$\mathcal{X} = \frac{1}{\left(\alpha_0 + \alpha_1^{\alpha_2 S^{\beta_1}}\right)^{\beta_2 S}}$$
where:
$$Y(t,S) = \frac{\left(1 - e^{-\gamma(t-\tau)}\right)^{\delta}}{\left(\alpha_0 + \alpha_1^{\alpha_2 S^{\beta_1}}\right)^{\beta_2 S}}$$

Figure 10: Three-parameter dynamic equation (central location) with 8 examples of its special cases in forms of multi-parameter fixed base age S-equations (outer locations); the dynamic equation can generate identical curves shapes as those from the 8 fixed-base-age equations.

In short, the **strictly hypothetical definition** of $\mathcal{X}$, without commitment to any values, and the **replacement of $\mathcal{X}$ with its initial condition solution**, are two key factors contributing to the extremely parsimonious nature of the GADA.

Theory Generalization and Robustness

The fact that the theoretical variable $\mathcal{X}$ has no restrictions in interpretation is eliminated during the derivations determines a high degree of model generalization and robustness. Not only are the dynamic equations derived with the GADA generalizations of many functional forms of the unobservable variable, but they are also generalizations of many, at times contradictory, theories that can rationalize the model (Fig. 11). The final representation of the winning theory behind the dynamic equation is actually decided by the data not the modeler. Thus, for example, an intended theory that a maximum yield is proportional to the unobservable variable will result in a model describes also a theory that, for example, the maximum yield is inversely proportional to the unobservable variable and many other theories in which the dependence on the unobservable variable could be in any positive or negative and direct or inverse functional relationship, (see Figure 11 for more examples).

Limiting size (α) is proportional to an exponential function of site index (S) and the rate of change (β) is proportional to the inverse of this function, e.g.,

$$\alpha = \alpha' \left(1 - e^{\gamma S}\right), \ \beta = \frac{\beta'}{1 - e^{\gamma S}}$$

Limiting size is proportional to an inverse power function of S and the rate of change is proportional to the power function of S, e.g.,

$$\alpha = \frac{1}{\alpha' S^{\gamma}}, \ \beta = \beta' S^{\gamma}$$

Limiting size is proportional to an inverse exponential function of S and the rate of change is proportional to the exponential function of S, e.g.,

$$\alpha = \frac{\alpha'}{1 - e^{\gamma S}}, \ \beta = \beta' \left(1 - e^{\gamma S}\right)$$

Limiting size is proportional to a power function of S and the rate of change is proportional to the inverse of this function of S, e.g.,

$$\alpha = \alpha' S^{\gamma}, \ \beta = \beta'/S^{\gamma}$$

Theory Generalization

$$\alpha_i = \beta'/\beta_i$$

or for convenience:

$$\alpha = \mathcal{X} \quad \text{and} \quad \beta = \beta'/\mathcal{X}$$

Limiting size is proportional to a power function of the ozone level (O_3) and the rate of change is proportional to the inverse power function of the ozone level, e.g.,

$$\alpha = \alpha' O_3^{\gamma}, \ \beta = \beta'/O_3^{\gamma}$$

Limiting size is proportional to an exponential function of the ozone level and the rate of change is proportional to the inverse exponential function of the ozone level, e.g.,

$$\alpha = \alpha' \left(1 - e^{\gamma O_3}\right), \ \beta = \frac{\beta'}{1 - e^{\gamma O_3}}$$

Limiting size is proportional to an inverse power function of the ozone level and the rate of change is proportional to the power function of the ozone level, e.g.,

$$\alpha = \frac{1}{\alpha' O_3^{\gamma}}, \ \beta = \beta' O_3^{\gamma}$$

Limiting size is proportional to an inverse exponential function of the ozone level (O_3) and the rate of change is proportional to the exponential function of the ozone level, e.g.,

$$\alpha = \frac{\alpha'}{1 - e^{\gamma O_3}}, \ \beta = \beta' \left(1 - e^{\gamma O_3}\right)$$

Figure 11: Examples of the GADA consolidation of seemingly different theories into a common model due to the fact that in this approach the only detrimental factor is the relative relationship between the affected parameters (all the nine theories would result in an identical dynamic equation).

Similarly, the examples of the dynamic equations (4) (Fig. 7b) and equation (9) (Fig. 6b) demonstrate a consolidation of multiple theories on the relationship between the half-saturation time and the unobservable variable. These dynamic equations were derived from a common theory that the half-saturation time is proportional to the growth intensity. However, they actually fitted the data using an unspoken theory of a corresponding inverse-proportional relationship. This is easily inferred from the comparison of the fit of these dynamic equations with the fit of the fixed-base-age site index equation (10) to the same data (Fig. 6c). In the latter equation, the proportional relationship theory is rigidly enforced by the explicit role of site index in the final equation. The result is a lack of fit — demonstrating the deficiency of the applied theory and the lack of flexibility in its implementation. The data-dependent generality in theory of the dynamic equation (4) is demonstrated by the exact fit of this equation to the pseudo-data generated by the equation (10). In this fitting the intended proportionality theory discussed above has materialized.

An algebraic proof of the GADA theory generalization is demonstrated by derivations of the two dynamic equations (5) and (9). They were derived from two contradictory theories. One assumed that both the asymptote and the half-saturation time were proportional to $\mathcal{X}$ and the other assumed that the asymptote and the half-saturation time were inversely proportional to $\mathcal{X}$. They both lead to an identical dynamic equation. This simply means that the resulting dynamic equation was a generalization of both of these theories. This generalization is a direct consequence of the algebraic methodology defined by the GADA approach and the algebraic properties of this approach as discussed in the previous section.

In essence, any specific theories describing the functional form of the unobservable variable and the complexity of its representation in the base equation are not relevant to the final model performance. The critical factors are the relative differences among the hypothesized responses of the different parameters in the base equation. The GADA's greatest advantage is that the derivation process generalizes the modeler's theories to much wider ranges of theories and interpretations of the unobservable variable. The theoretical interpretations within the dynamic equations are very robust and flexible. Yet, they are consistent and maximize payoffs from the modeler's efforts to formulate flexible, but parsimonious, dynamic equations.

Other Considerations

The asymptotic growth is a good example of a controversial issue in growth and yield modeling. The various opinions about asymptotic growth may be controversial. Ricker (1979) includes a section titled "Asymptotic growth: is it real?"; Knight (1968) 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 alike to those of Schnute (1981).

I have considered here asymptotes to be a desirable analytical tool used for model conceptualization and growth theory formulation. The methodology discussed here is based, in part, on building different theories about limiting growth and the behavior of asymptotes. Yet, if the price paid means not fitting the data at all or very poorly, the situation warrants a non-asymptotic model. Furthermore, it is better to have more options than fewer and usually more flexible generic models are favored. Derivation of the dynamic equations allows resolving the controversy about the asymptotes by producing models that are generic enough to include both asymptotic and non-asymptotic models in one common equation.

Another issue may be the incompatibility between input S and height predictions at base-age that causes height model inaccuracy and inconsistency near the base-age. This may be a minor problem with small differences, or a major problem with large ones. The direct use of age-height measurements precludes existence of this problem and is more realistic than the use of S, since S is almost never directly measured, but is estimated from direct height measurements at various ages.

Acknowledgements

I am grateful to Dr. R.L. Bailey and Mr. R.I. Elle for their 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.