Dynamical implications of the variability representation in site-index modelling

Oscar García

Received: 16 April 2010/Revised: 12 August 2010/Accepted: 23 November 2010/Published online: 21 December 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Issues in the development and formulation of forest site-index models are examined, linking the forestry terminology and methods to standard mathematical concepts. Variability complicates interpretation. Three sources of variation are distinguished: between sites, within sites, and observation error, with the article focusing mainly on the second one. Two site-index definitions arising from different views about the variability are contrasted. Modelling based on algebraic difference equations (ADE's) is analyzed in detail, relating it to concepts of state space flows used in modern dynamical systems theory. It is shown that, given a stand current state, ADE's predict growth rates that are independent of site quality.

Keywords Forest growth and yield · Site productivity · Algebraic difference equations · ADA · GADA · Differential equations · Dynamical systems

Introduction

Site-index models relate height, age, and site quality (potential productivity) in even-aged single-species stands. They are used for predicting stand height development and for assessing site quality. The principles can be traced back to the 18th Century (Batho and García 2006), and various approaches are described in books such as Belyea (1931); Spurr (1952); Assmann (1970); Clutter et al. (1983); von

Gadow and Hui (1999); Pretzsch (2009). General reviews have been published by Jones (1969); Carmean (1975); Hägglund (1981); Ortega and Montero (1988); Grey (1989). Some of the details are subtle. Assumptions and interpretations are not always clear and explicit, leading to misunderstandings and controversy. We focus here on some implications of natural variability of stand development. Mathematical and modelling aspects are stressed, but detailed statistical procedures are beyond the scope of this article.

Three sources of variability may be identified: (a) between sites, (b) within sites, and (c) observation error. These are illustrated in Fig. 1. The between-sites variation gives rise to a family of site curves, each representing a height–age trajectory for a certain site quality. It is discussed briefly in the next Section. For a given site, individual stands will deviate from the nominal site curve due to weather fluctuation and other factors. The implications of this within-site variability are the main topic here. In addition, height observations (and sometimes age) are subject to sampling and measurement error. These errors can be important in devising appropriate estimation and statistical inference procedures, but will not be discussed in detail.

Over time, site-index modelling has developed its own methods and terminology. The article makes an effort to link these to standard mathematical concepts. It is hoped that tapping into a wider pool of knowledge may facilitate future progress.

Communicated by G. Kändler.

O. García (✉)
University of Northern British Columbia, 3333 University Way,
Prince George, BC V2N 4Z9, Canada
e-mail: garcia@unbc.ca

Variation between sites

Differences in height growth across sites are the basis of the site-index concept. Ignoring some of the natural

Fig. 1 Nominal site-index curve (dashed), actual trajectory (solid), and observations (isolated dots). A and B indicate different site-index definitions (see text)

variation, the main ideas are not difficult to understand. It is assumed that stands follow height–age trajectories characteristic of each site quality, the site-index curves (or *site curves*) drawn with dots in Fig. 1. The curves do not intersect, except possibly at the origin, and greater heights at any particular age indicate higher site quality.

Equation-based models usually start with a growth curve function $H = g(A)$, where A is stand age and H is some measure of stand height. The growth curve is made to vary with site quality by including a site-dependent parameter, say q , so that $H = g_q(A) = g(A, q)$. For instance, with the Schumacher (1939) function $H = a \exp(-b/A)$, typically one of the parameters a or b is taken as site-dependent (or *local*), while the other is assumed to be common to all sites and stands (*global*). Thus, one may have a model where the site curves differ by an H -scale factor, $H = q \exp(-b/A)$ (called *anamorphic*), or by an A -scale factor, $H = a \exp(-q/A)$. More generally, both original parameters might be assumed to be site-dependent, as in $H = \alpha \beta^q \exp(-q/A)$, where q is local, and α and β are new global parameters.

The local parameter q serves only as a label for the individual site curves and can be chosen in different ways. Curves may be labeled simply by discrete quality classes, often with roman numerals. The most common continuous labeling scheme uses a *site index*, S , defined as the curve height at some reference *base age* A_b . The site index is related to any other site-dependent parameter q through $S = g_q(A_b)$.

Variation within sites

Clearly, real stands will inevitably deviate from the curve specified by any deterministic model. The model does not

necessarily ignore this and the curve may be interpreted as a point estimate, a predicted, expected, or most likely height–age trajectory. We shall not be specific about the differences among these (mean, mode, median, etc.) and will say “predicted” or “nominal”.

Which site index?

A first consequence of this within-site variability is the existence of different definitions of site index. Some authors have explicitly or implicitly defined site index as the actual height reached by a particular stand at the base age. This is a property of the stand and is different from the definition based on predicted height given in the previous Section, which is a property of the site. “Stand site index” corresponds to point A in Fig. 1 and “site site index” to point B.

Definitions cannot be right or wrong, but the proper statistical treatment differs, and lack of clarity on this point can (and has) lead to misunderstanding and controversy. Under the *stand site index* view, models appropriate for predicting height may differ from those for assessing site quality, and statistical analysis typically involves error-in-variables situations (Curtis et al. 1974; Goelz and Burk 1992). The *site site index* approach is more abstract, although it may be closer to the original idea.

Focusing instead on local and global parameters side-steps these issues.

Dynamics

The nominal model height–age trajectories correspond to a prediction “at birth”. By analogy to a person’s life expectancy, and for similar reasons, a future height predicted at birth can be expected to be different from that predicted later in life. Generally, for a stand growing in site quality q , one knows (or has an estimate of) its height H_1 at age A_1 and wants to predict the height H_2 at some other age A_2 . Writing $t = A_2 - A_1$ for the prediction interval, the predicted height is some function

$$H_2 = F_q(A_1, H_1, t). \quad (1)$$

It is assumed that this function is continuous and smooth (differentiable). These assumptions are implicit in the use of growth functions like those in Sect. 2, even though the model does not usually pretend to reproduce the seasonal growth fluctuations within a year (but see García 1979, 1999, for one way of doing this).

The function (1) is special in that it must satisfy two consistency conditions: prediction over a zero-length interval must return the starting value, i.e. $F_q(A_1, H_1, 0) = H_1$, and two predictions over consecutive intervals must give the

same result as a single prediction over the whole: $F_q[A_1 + t, F_q(A_1, H_1, t), u] = F_q(A_1, H_1, t + u)$ (Sullivan and Clutter 1972; Clutter et al. 1983, p. 123; García 1979, 1994)¹. Such a function, or more precisely (1) together with the obvious age “prediction” function $A_2 = A_1 + t$, is known as a (global) *transition function* (Padulo and Arbib 1974; García 1994, and references therein) or a *flow* (Arnold 1973, Chap. 1) (a semi-flow if predictions back in time, $t < 0$, are not allowed). In forestry, these have been called also difference equations (Clutter et al. 1983), algebraic difference equations (ADE, Cieszewski and Bailey 2000), or self-referencing functions (Northway 1985), at least when independent of q (see below).

Any flow (1) can be obtained as the solution of an ordinary differential equation (ODE)

$$\frac{dH}{dt} = f_q(A, H) \quad (2)$$

(and $dA/dt = 1$), which is sometimes more convenient (Arnold 1973). In fact, modelling often starts with an ODE formulation, and the transition function arises through integration. The reversal of this classical view by the Russian school of Anosov, Arnold, and others, starting from flows as the more primitive concept, is attractive because it is often argued that ODE’s are not appropriate for forest modelling. It is also closely related to the ADE ideas introduced by Clutter (1963) and Bailey and Clutter (1974).

Alternatively, flows may be described by an *invariant*, or first integral, an expression that remains constant over trajectories of the flow. A trajectory is the curve generated by varying t in (1), for fixed A_1 and H_1 . This relates also to ADE’s, as shown below. For multivariate generalizations, see García (2010).

Site curves and ADE’s

The algebraic difference approach (ADA/GADA)² produces a flow or ADE compatible with a given site equation $H = g(A, q)$ (Bailey and Clutter 1974; Clutter et al. 1983; Cieszewski and Bailey 2000). The usual procedure consists of solving for the local parameter, $q = \varphi(A, H)$, and equating the value at two height–age points:

$$\varphi(A_1, H_1) = \varphi(A_2, H_2).$$

¹ In forestry, the property has been called *path invariance*. Mathematically, this is a one-parameter (t) continuous group of transformations. An example of Lie group, named after the Norwegian mathematician Sophus Lie (pronounced “Lee”).

² This terminology should not be confused with the standard mathematical meanings. In mathematics, difference equations deal with sequences of uniformly spaced values. An algebraic equation may contain elementary operations and rational exponents, excluding exponentials and other transcendental functions (James, 1992).

Solving for H_2 , one obtains the flow equation (ADE). For example, with the anamorphic Schumacher model $H = q \exp(-b/A)$; $q = H \exp(b/A)$, $H_1 \exp(b/A_1) = H_2 \exp(b/A_2)$, and $H_2 = H_1 \exp(b/A_1) / \exp(b/A_2) = H_1 \exp[b(1/A_1 - 1/A_2)]$.

Note that $\varphi(A, H)$ is an invariant, constant for all the points on a trajectory. Alternatively, one may differentiate the invariant and obtain the ODE for the flow:

$$\begin{aligned} \frac{dH \exp(b/A)}{dA} &= 0 \\ \frac{dH}{dA} \exp(b/A) + H \exp(b/A)(-b/A^2) &= 0, \end{aligned}$$

from where,

$$\frac{dH}{dA} = b \frac{H}{A^2}.$$

This ODE was actually the starting point of Schumacher (1939).

The ADE is a relationship between any two points lying on the same site curve. The corresponding ODE is called the ODE of a one-parameter family of curves (e.g. Agnew 1960, Chap. 4); in this case, the family of site curves parameterized by q .

The ADE, or its ODE, predicts that any stand of site quality q that currently lies on the site- q curve will continue to follow that curve, as one might expect. However, that is not the only flow and ODE with this property. One could use any other invariant in the derivation, not just the one that corresponds to the local parameter. A few possibilities for the anamorphic Schumacher are shown in Table 1. The ADE, first row, is the (only) flow that does not depend on site quality. The one on the second row is obtained through solving for b ; the others cannot be obtained by the method of equating parameters. The ODE in row 3 is a function of A , and that in 4 is a function of H . There is an infinity of ODE’s depending on both A and H (and q) that produce the same growth curve, starting from a point on the curve. In other words, a site equation $H = g(A, q)$ is not sufficient by itself to predict future dynamics when the stand deviates from the nominal curve; hypotheses about the growth rate are needed.

Prediction

Consider a stand for which we know, or have estimates of, the current age A_1 , height H_1 , and the site quality q . In the absence of within-site variability, (A_1, H_1) would be on the site- q curve and, as just explained, any of the flows or ODE’s discussed in the previous section then give the same predictions, following the nominal site curve. In reality, the stand would almost certainly have deviated from the

Table 1 Some flows and ODES that generate $H = q \exp(-b/A)$

	Invariant	Flow equation	ODE
1	$H \exp(b/A)$	$H_2 = H_1 \exp(b/A_1 - b/A_2)$	$dH/dA = bH/A^2$
2	$A \ln(H/q)$	$H_2 = q(H_1/q)^{A_1/A_2}$	$dH/dA = H \ln(q/H)/A$
3	$H - q \exp(-b/A)$	$H_2 = H_1 + q[\exp(-b/A_2) - \exp(-b/A_1)]$	$dH/dA = bq \exp(-b/A)/A^2$
4	$A + b/\ln(H/q)$	$H_2 = q \exp[b/(A_1 - A_2 + b/\ln(H_1/q))]$	$dH/dA = H \ln(H/q)^2/b$
5	$q/H - \exp(b/A)$	$H_2 = q/[\exp(b/A_2) - \exp(b/A_1) + q/H_1]$	$dH/dA = (b/q)H^2 \exp(b/A)/A^2$

nominal curve, and then the different flow and ODE equations produce different predictions.

This is illustrated in Fig. 2, for an anamorphic Schumacher site model with $b = 18$. Representative site curves are drawn with dashes and labeled with the site index for base age 25. The Clutter-Bailey ADE always follows the site curve passing through the current point, regardless of site quality. On the other hand, the solid curves show the trajectories predicted by the flow on the forth row of Table 1 for stands of site index 20 that have deviated above or below the nominal curve.

ADE's describe site curves well and, by eliminating local parameters, make it possible to fit site models using standard nonlinear regression packages. But their representation of growth dynamics in the presence of within-site variation is questionable. A site-index 20 stand that happens to be on the site-index 18 curve can be expected to grow faster than a site-index 18 stand that so far has stayed the course.

Among the alternatives, there are good biological and other reasons to prefer the one where the growth rate (2) depends on size (H), but not on age (e.g. Peñuelas 2005).

For the Schumacher, it is the one on row 4 of Table 1 and in Fig. 2.

Figure 3 displays similar predicted trajectories from the model of Hu and García (2010). It is based on an age-invariant Bertalanffy-Richards ODE. For another example, see the *EasySDE User Guide* at <http://forestgrowth.unbc.ca/sde>.

One situation where the various predictions coincide is where the only information available about a stand is the current age and height, without any knowledge of site quality or previous measurements. Then q must be inferred from the starting point, and it can be seen that the resulting trajectories will be the same. This might be a common occurrence when applying these systems in practice. During model development, however, there are usually multiple measurements in a sample plot, and ADA essentially derives site quality separately for each pair of measurements, ignoring the restriction of a common value for the plot. The question of if discarding some information is compensated by avoiding the complications of having to deal with local parameters would depend on the data and other considerations, and seems difficult to answer in general.

Fig. 2 Dashed: Site-index curves and ADE trajectories, anamorphic Schumacher model with $b = 18$. Continuous: predicted trajectories for site index 20 according to the fourth row of Table 1. Site indices for base age 25

Fig. 3 From model by Hu and García (2010). Dashed: site-index curves for sites 19, 20, 21. Solid: projected heights for stands growing in site quality 20, currently at the dot locations. ADE predictions follow the dashed curves

Conclusions

The site modelling methods pioneered by Clutter and others deviated from the classical ODE-centered approaches to dynamical systems common in other fields. Although they might have seemed ad hoc, it is remarkable how they actually paralleled to some extent more recent developments in Mathematics (Arnold 1973; Anosov et al. 1997).

Given current stand conditions, the growth rate in ADE/ADA/GADA models is independent of site quality. This might be seen as a conceptual flaw. Deviations from the nominal curves, however, are unlikely to be large, so the practical implications are not entirely clear. Certainly, ADA made feasible the development of good site models at a time when computing resources were limited. The computational and flexibility advantages of ADA/GADA techniques might well carry into the future.

Only essentially deterministic aspects of growth forecasting have been discussed, dealing with predicted or nominal trajectories. A superimposed stochastic structure is important for hypothesis testing and in the search for good estimators. One natural extension is to include environmental perturbations in the ODE's, possibly adding also observation errors (Hotelling 1927; Seber and Wild 2003, Sect. 7.5). It is tempting to use hierarchical modelling for the sources of variation, treating the local parameter as “random” (Snijders 2003; Hall and Bailey 2001); but it should be remembered that height–age data is rarely a random sample from the target population, and the effects of violating this assumption are unclear.

Acknowledgments This work owes much to extensive discussions over the years with Keith Rennolls, and with Chris Cieszewski. Encouragement and suggestions from Mike Strub are gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Agnew RP (1960) Differential equations, 2nd edn. McGraw-Hill, New York
- Anosov DV, Aranson SK, Arnold VI, Bronshstein IU, Grines VZ, Il'yashenko YS (1997) Ordinary differential equations and smooth dynamical systems. Springer, New York
- Arnold VI (1973) Ordinary differential equations. The MIT Press, Cambridge, MA
- Assmann E (1970) The principles of forest yield study. Pergamon Press, Oxford, England
- Bailey RL, Clutter JL (1974) Base-age invariant polymorphic site curves. For Sci 20:155–159
- Batho A, García O (2006) De Perthuis and the origins of site index: a historical note. For Biom Model Inform Sci 1:1–10
- Belyea HC (1931) Forest measurement. Wiley, New York
- Carmean WH (1975) Forest site quality evaluation in the United States. In: Brady NC (eds) Advances in agronomy, vol 27. Academic Press, New York, pp 209–269
- Cieszewski CJ, Bailey RL (2000) Generalized algebraic difference approach: theory based derivation of dynamic site equations with polymorphism and variable asymptotes. For Sci 46(1):116–126
- Clutter JL (1963) Compatible growth and yield models for loblolly pine. For Sci 9:354–371
- Clutter JL, Fortson JC, Pienaar LV, RL Bailey GHB (1983) Timber management: a quantitative approach. Wiley, New York
- Curtis RO, DeMars DJ, Herman FR (1974) Which dependent variable in site index–height–age regressions? For Sci 20:74–80
- García O (1979) Modelling stand development with stochastic differential equations. In: Elliott DE (ed) Mensuration systems for forest management planning. New Zealand Forest Service, Forest Research Institute Symposium No. 20, pp 315–334. <http://web.unbc.ca/~garcia/publ/sym20.pdf>
- García O (1994) The state-space approach in growth modelling. Can J For Res 24:1894–1903
- García O (1999) Height growth of *Pinus radiata* in New Zealand. N Z J For Sci 29(1):131–145
- García O (2010) A parsimonious dynamic stand model for interior spruce in British Columbia. For Sci (to appear)
- Goelz JCG, Burk TE (1992) Development of a well-behaved site-index equation: Jack pine in north central Ontario. Can J For Res 22:776–784
- Grey DC (1989) Site index—a review. S Afr For J 148:28–32
- Hägglund B (1981) Evaluation of forest site productivity. For Abstr 42:515–527
- Hall DB, Bailey RL (2001) Modelling and prediction of forest growth variables based on multilevel nonlinear mixed models. For Sci 47:311–321
- Hotelling H (1927) Differential equations subject to error, and population estimates. J Am Stat Assoc 22:283–314
- Hu Z, García O (2010) A height-growth and site-index model for interior spruce in the sub-boreal spruce biogeoclimatic zone of British Columbia. Can J For Res 40(6):1175–1183
- James RC (1992) Mathematics dictionary. 5th edn. Chapman & Hall, New York
- Jones JR (1969) Review and comparison of site evaluation methods. Research Paper RM-51, USDA Forest Service, 27 p
- Northway SM (1985) Notes: fitting site index equations and other self-referencing functions. For Sci 31:233–235
- Ortega A, Montero G (1988) Evaluación de la calidad de las estaciones forestales. Revisión bibliográfica. Ecología 2:151–184
- Padulo L, Arbib MA (1974) System theory. Hemisphere Pub. Co, Washington, DC
- Peñuelas J (2005) Plant physiology: a big issue for trees. Nature 437(7061):965–966. doi:10.1038/437965a
- Pretzsch H (2009) Forest dynamics, growth and yield: from measurement to model. Springer, Berlin
- Schumacher FX (1939) A new growth curve and its application to timber-yield studies. J For 37:819–820
- Seber GAF, Wild CJ (2003) Nonlinear regression. Wiley-Interscience, New York
- Snijders TAB (2003) Multilevel analysis. In: Lewis-Beck M, Bryman A, Liao T (eds) The SAGE encyclopedia of social science research methods, vol II. Sage, Thousand Oaks, CA, pp 673–677
- Spurr SH (1952) Forest inventory. Ronald Press, New York
- Sullivan AD, Clutter JL (1972) A simultaneous growth and yield model for loblolly pine. For Sci 18(1):76–86
- von Gadow K, Hui G (1999) Modelling forest development. Forestry sciences, vol 57. Kluwer Academic Publishers, Dordrecht, The Netherlands