

Using segmented regression to model the density–size relationship in direct-seeded slash pine stands

Quang V. Cao ^{*}, Thomas J. Dean

School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

Received 21 June 2007; received in revised form 2 October 2007; accepted 3 October 2007

Abstract

This study investigated the application of segmented regression techniques to modeling the trajectory of tree density and quadratic mean diameter through time for individual stands. The full model contains three segments for characterizing the three different stages of the trajectory on a log–log scale. The first segment represents precanopy conditions where no mortality is expected after what normally occurs before initial tree establishment. The next two segments are quadratic functions to accommodate two distinct mortality rates if present in the trajectory. Quadratic functions were selected based on published trajectories for loblolly pine (*Pinus taeda* L.) and slash pine (*Pinus elliottii* var. *elliottii* Engelm.) in southeastern U.S. A reduced model is formed by simply joining two quadratic segments. Data from direct-seeded slash pine stands used in this study fit the reduced model. The two-segment model requires a join point, which was predicted as a power function of the logarithm of initial stand density. The resulting trajectories matched observed trends in the data reasonably well. The segmented regression approach is appealing for modeling very complicated functional forms such as the density–size relationship presented in this paper.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Size–density trajectory; *Pinus elliottii*; Self-thinning; Mortality rate

1. Introduction

Depending on the objective, modelers have used various approaches to quantitatively describe the growth trajectory of an even-age monoculture in terms of either average stand diameter or mean stem volume and number of trees per hectare under the constraint of a limiting relationship between mean size and tree number (e.g., Smith and Hann, 1984; Puettmann et al., 1993; Tang et al., 1994). The constraint originates from observations of even-age monocultures asymptotically approaching a linear boundary between the log of mean tree size and the log of spatial density. This observation is not universal, however, as other data suggests a curvilinear trajectory throughout the lifespan of a stand, at least for loblolly and slash pine (Zeide, 1987; Cao et al., 2000).

One of the incentives for modeling the size–density trajectories of individual stands is that it eliminates the need to censor data to investigate the properties of the upper boundary of the tree size and tree density relationship; the

boundary emerges from overlapping trajectories of individual stands (Smith and Hann, 1986). Another advantage of individual trajectories is their incorporation into stand growth models (e.g., Tang et al., 1994). To fully benefit from this technique the models need to accurately describe the life-long developmental trajectories of the stand. Even-age monocultures move through several stages of development, each with characteristic mortality rates which influence accuracy of quantitative descriptions of the trajectory. Modelers have assumed a monotonic increase in mortality with increasing proximity to the size–density boundary after canopy closure. In addition, Long and Smith (1984) proposed an intervening developmental stage between canopy closure and self-thinning called full-site occupancy that is characterized by hastened size differentiation among trees in the stand and skewness in size distribution. Pre- and postcanopy closure mortality rates suggest that a complete description of a stand's developmental trajectory requires at least two distinct curve segments; an intervening developmental stage may indicate the need for a third segment. The objective of this study was to apply segmented regression techniques to modeling the trajectory of stand density and quadratic mean diameter of individual stands through time.

^{*} Corresponding author. Tel.: +1 225 578 4218; fax: +1 225 578 4227.

E-mail addresses: qciao@lsu.edu (Q.V. Cao), fwdean@lsu.edu (T.J. Dean).

Table 1

Distribution of 147 permanent plots in direct-seeded slash pine stands, by measurement ages

Stand age (years)		Number of plots
First measurement	Last measurement	
8	17	1
8	19	18
8	20	10
8	23	25
8	27	24
9	15	1
9	18	3
9	20	3
9	28	8
10	25	20
11	22	17
12	27	12
13	28	5

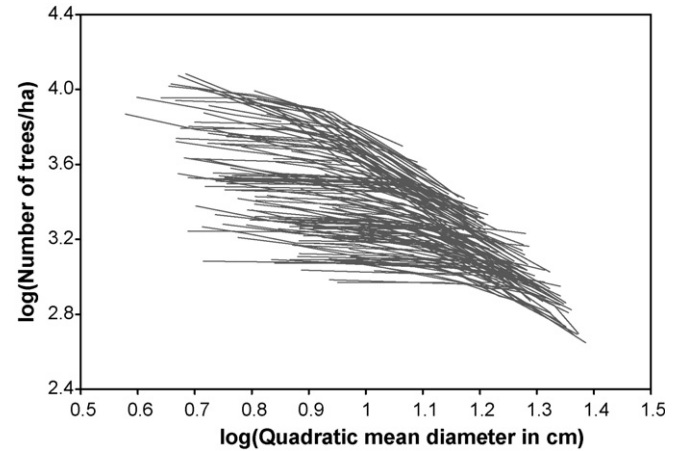


Fig. 1. Density–size trajectories observed from 147 permanent plots in direct-seeded slash pine stands.

2. Data

Data available for this study were from 147 permanent plots from direct-seeded slash pine (*Pinus elliottii* var. *elliottii* Engelm.) stands on cutover sites located in Natchitoches and Rapides parishes (central Louisiana) and in Washington parish (southeast Louisiana). A detailed description of the data can be found in Baldwin (1985) and Lohrey (1987). Plot size ranged from 0.040 to 0.048 ha. Some plots were precommercially thinned at age 3 or 4 years. Stand age ranged from 8 to 28 years, stand density from 445 to 12,108 trees/ha, basal area from 2.6 to 52.6 m²/ha, and site index (base age 25 years) from 9 to 23 m. Distribution of plots by measurement ages is presented in Table 1. Each plot was measured from three to six times, at 3–10 years apart, resulting in a total of 615 measurements encompassing 468 growth periods. Table 2 shows the mean and standard deviation of number of trees per hectare and quadratic

mean diameter for each measurement age. The trajectories of stand density and quadratic mean diameter for these measurements are shown in Fig. 1.

3. Methods

Segmented regression models have been employed to describe complicated functional forms such as tree taper (Max and Burkhardt, 1976; Fang et al., 2000; Coble and Hilpp, 2006) and the height–age relationship (Devan and Burkhardt, 1982; Borders et al., 1984). Similar techniques were recently used to determine what observations of size–density trajectories were within particular stages and phases of stand development (VanderSchaaf and Burkhardt, in press). In this study, segmented regression techniques were applied to model trajectories of stand density and quadratic mean diameter through time.

Table 2

Mean (and S.D.) of number of trees per ha and quadratic mean diameter measured through time from 147 permanent plots in direct-seeded slash pine stands

Age	Number of plots	Stand density (number/ha)	Quadratic mean diameter (cm)
8	78	3715 (2733)	6.7 (1.5)
9	15	2112 (996)	6.7 (1.5)
10	20	2409 (860)	9.1 (1.0)
11	53	3201 (2410)	9.2 (3.0)
12	35	4583 (2657)	8.5 (1.8)
13	34	3251 (2099)	9.9 (2.2)
14	53	2736 (2078)	11.3 (3.4)
15	24	1604 (500)	13.8 (1.4)
17	71	3161 (2165)	12.0 (3.5)
18	34	2493 (1610)	12.8 (2.7)
19	18	1848 (786)	15.0 (2.2)
20	20	1250 (662)	17.9 (2.6)
22	53	2794 (1343)	12.8 (2.7)
23	38	1735 (1316)	16.5 (3.6)
25	20	799 (220)	21.3 (1.5)
27	36	2087 (624)	16.0 (2.3)
28	13	2420 (1269)	15.1 (2.7)

3.1. A three-segment regression model

A linear–quadratic–quadratic segmented model for the general density–size relationship has the following form:

$$y = y_0 + \beta_1 I_1(x - a_1)^2 + \beta_2 I_2(x - a_2)^2 + \varepsilon \quad (1)$$

where $y = \log(N)$; N = stand density (number of trees/ha); $y_0 = \log(N_0)$; N_0 = initial stand density (at age 0); $x = \log(Dq)$; $a_1 = \alpha_{11}y_0^{\alpha_{12}}$ = first join point; $a_2 = \alpha_{21}y_0^{\alpha_{22}}$ = second join point; β_j and α_{jk} = regression coefficients, j and $k = 1, 2$; $I_j = 1$ if $x > a_j$, 0 otherwise, $j = 1, 2$; ε = error term; and $\log(\cdot)$ = logarithm base 10.

This model uses three segments joined together at two join points to describe the three stages of the density–size trajectory. The first segment, $y = y_0$ for $x < a_1$, is a horizontal line, representing the early stages of stand development when no density-dependent mortality is expected after what normally occurs before initial tree establishment. Both Zeide (1987) and Cao et al. (2000) have noted life-long curvilinear trajectories between average tree size and tree density in loblolly and slash pine. The next two segments, $y = y_0 + \beta_1(x - a_1)^2$ for $a_1 < x < a_2$ and $y = y_0 + \beta_1(x - a_1)^2 + \beta_2(x - a_2)^2$ for $x > a_2$ are quadratic functions that would account for both a curvilinear trajectory and distinct changes in mortality rate after canopy closure. The two join points (a_1 and a_2) are separate functions of logarithm of initial density (y_0).

Eq. (1), which expresses current stand density as a function of stand density at age 0, can be reformulated to express current stand density in terms of past stand density. The following equation allows estimation of parameters from the segmented model based on measurements from consecutive growth periods:

$$y_2 = y_1 + \beta_1 [I_{12}(x_2 - a_1)^2 - I_{11}(x_1 - a_1)^2] + \beta_2 [I_{22}(x_2 - a_2)^2 - I_{21}(x_1 - a_2)^2] + \varepsilon \quad (2)$$

where y_i = logarithm of stand density at time $i = 1, 2$ (beginning and end of the growth period); x_i = logarithm of quadratic mean diameter at time i ; $I_{ij} = 1$ if $x_i > a_j$, 0 otherwise, for time i ($i = 1, 2$) and join point j ($j = 1, 2$).

3.2. A two-segment regression model

If a data set contains remeasurements from stands that have already closed canopy and are beginning to experience mortality, the following quadratic–quadratic segmented model might be adequate:

$$y = y_0 + \beta_1(x - a)^2 + \beta_2 I(x - a)^2 + \varepsilon \quad (3)$$

where $a = \alpha_1 y_0^{\alpha_2}$ = the only join point; β_j and α_j = regression coefficients, $j = 1, 2$; $I = 1$ if $x > a$, 0 otherwise.

Parameter estimates of (3) can be obtained from

$$y_2 = y_1 + \beta_1 [(x_2 - a)^2 - (x_1 - a)^2] + \beta_2 [I_2(x_2 - a)^2 - I_1(x_1 - a)^2] + \varepsilon, \quad (4)$$

Table 3
Parameter estimates for Eqs. (2) and (4)

Equation	Parameter	Estimate	S.E.	p-Level
(2)	β_1	−0.3333	0.2786	0.2322
	β_2	−2.6604	0.2804	0.0001
	α_{11}	18.0963	32.2419	0.5749
	α_{12}	−2.7903	1.7352	0.1085
	α_{21}	7.1533	0.8235	0.0001
	α_{22}	−1.6326	0.0975	0.0001
(4)	β_1	−0.0868	0.0149	0.0001
	β_2	−2.8364	0.1078	0.0001
	α_1	7.9689	0.5394	0.0001
	α_2	−1.7370	0.0567	0.0001

using measurements from the beginning and end of consecutive growth periods.

Eqs. (2) and (4) are used for estimation of parameters. Once the parameters were estimated, these equations will be useful for predicting the trajectory of a stand from a measured pair of N and Dq .

4. Results and discussion

The stand density at the first measurement for each plot was assumed to be initial density (N_0). Both Eqs. (2) and (4) were fitted to the data using NLIN, a SAS nonlinear regression procedure (SAS institute Inc., 1999). Parameters β_1 , α_{11} and α_{12} were not significant at the 5% level (Table 3), indicating that the three-segment model (Eqs. (1) and (2)) was overparameterized. The data evidently did not contain sufficient information for estimating the first join point. On the other hand, all parameters of the two-segment model were highly significant (Table 3). This model explained 97.8% of the variation. The two-segment model (Eqs. (3) and (4)) is therefore deemed appropriate for this data set.

4.1. Three-segment versus two-segment models

The three-segment model (Eq. (1)) is able to completely describe the density–size trajectory of a stand. However, in cases where mortality is evident in the first measured period from many plots, the two-segment model (Eq. (3)) might be sufficient. The three- and two-segment models can therefore be considered full and reduced models, respectively, from a regression viewpoint. An F -test can be easily performed to determine which of the two models is appropriate for a given data set. The F -test was not necessary in this case because all parameters of the three-segment model were not significant at the 5% level, indicating that this model was not suitable for the data.

The data used in this study certainly fit the profile for the two-segment model. Mortality occurred in most plots after the first measurement at ages ranging from 8 to 13 years nullifying the initial horizontal segment in Eq. (1). The significant coefficients for the join point indicated a predictable change in the mortality rate for the individual stand trajectories. Averaged across initial tree densities typical of managed stands, the join

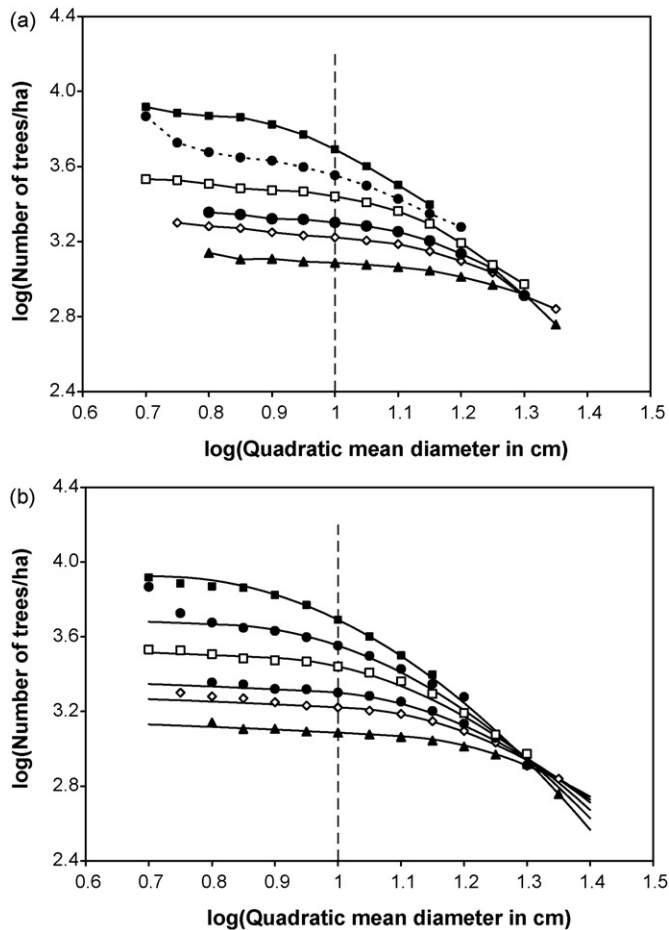


Fig. 2. Average density–size trajectories for six stand density groups (a), and plus curves generated from the two-segment regression model (b).

points correspond to a relative density characteristic of full-site occupancy, which according to Long (1985) begins at 35% of maximum density. Due to the constraint of smooth transition from one segment to the other, the join point might be near, but not exactly where the change in shape of the trajectory occurs. Nevertheless, the relative density corresponding to the join points for the two-segment model for $N < 2500$ averages 42% for these slash pine stands. At this stage of development, stand growth is near maximum for some species (Smith and Hann, 1986; Marshall et al., 1992). If growth translates to proportional consumption of resources as surmised by Miller (1995), a change in mortality rates might be a logical consequence of the intraspecific competition associated with near-maximum growth.

4.2. Trend curves

For each plot (trajectory), values of $\log(N)$ were computed by use of linear interpolation for values of $\log(Dq)$ from 0.7 to 1.35 at 0.05 intervals. Trajectories were then classified into six groups (each containing an approximately equal number of plots) based on $\log(N)$ values evaluated at a standard Dq of 10 cm. These values are similar in concept to stand density

index (Reineke, 1933), which is stand density at an index diameter of 25.4 cm. Averages of the $\log(N)$ values for each of the $\log(Dq)$ values were computed for each group and are shown in Fig. 2a.

Curves were generated from the two-segment regression model (Eqs. (3) and (4)) such that they passed through the six average points at $Dq = 10$ cm (Fig. 2b). The curves explained 98.8% of the variation of the average points. While an R^2 based on averages is not statistically sound and not a valid quantitative measure of goodness of fit, it is presented here only to show that the regression curves adequately represented trends in the data.

The above process was repeated to construct 2, 4, 8 and 10 groups of similar density at $Dq = 10$ cm. Results from these cases were similar to those obtained in the case of six groups discussed in detail above.

Since the trend lines are the results of two quadratic curves, they are curvilinear throughout their length. As a consequence of this shape, total basal area per hectare increases with increasing values of Dq to a maximum, then decreasing thereafter. During the period of basal area accretion, the slope between $\log(Dq)$ and $\log(N)$ will be greater than -2.0 (or absolute value less than 2.0); after basal area peaks, the slope will be less than -2.0 . Estimating slopes beyond the maximum values of Dq will obviously produce unrealistic gradients given the simple quadratic models fit to these data.

Fig. 2a shows that three of the trend curves cross at $\log(Dq) \geq 1.3$ (or $Dq \geq 20$ cm). Since the segmented curves were fit to these individual trajectories, the fitted curves also cross at these values of Dq (Fig. 2b). Crossing trajectories violates the assumption that individual stands asymptote to a common upper size–density limit. While this assumption may simplify curve fitting, individual trajectories frequently cross each other (e.g., Drew and Flewelling, 1979; Hibbs, 1987; Dean and Baldwin, 1993; Sharma and Zhang, 2007). Predicting the trajectory of individual stands is different than trying to discern an emergent property among self-thinning stands; consequently, curves fit to individual stands should reflect their individual patterns, allowing crossovers if present in the data set.

References

- Baldwin Jr., V.C., 1985. Survival curves for unthinned and early-thinned direct-seeded slash pine stands, Proc. of the Third Biennial South. Silv. Res. Conf. USDA For. Serv. Gen. Tech. Rep. SO-54, pp. 460–465.
- Borders, B.E., Bailey, R.L., Ware, K.D., 1984. Slash pine site index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. For. Sci. 30, 411–423.
- Cao, Q.V., Dean, T.J., Baldwin Jr., V.C., 2000. Modeling the size–density relationship in direct-seeded slash pine stands. For. Sci. 46, 317–321.
- Coble, D.W., Hilpp, K., 2006. Compatible cubic-foot stem volume and upper-stem diameter equations for semi-intensive plantation grown loblolly pine trees in East Texas. South. J. Appl. For. 30, 132–141.
- Dean, T.J., Baldwin Jr., V.C., 1993. Using a density-management diagram to develop thinning schedules for loblolly pine plantations. USDA Forest Service Southern Forest Experiment Station Research Paper SO-275, p. 7.
- Devan, J.S., Burkhart, H.E., 1982. Polymorphic site index equations for loblolly pine based on a segmented polynomial differential model. For. Sci. 28, 544–555.

- Drew, T.J., Flewelling, J.W., 1979. Stand density management: an alternative approach and its application to Douglas-fir plantations. *For. Sci.* 25, 518–532.
- Fang, Z., Borders, B.E., Bailey, R.L., 2000. Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *For. Sci.* 46, 1–12.
- Hibbs, D.E., 1987. The self-thinning rule and red alder management. *For. Ecol. Manage.* 18, 273–281.
- Lohrey, R.E., 1987. Site index curves for direct-seeded slash pines in Louisiana. *South. J. Appl. For.* 11, 15–17.
- Long, J.N., 1985. A practical approach to density management. *For. Chron.* 61, 23–27.
- Long, J.N., Smith, F.W., 1984. Relation between size and density in developing stands: a description and possible mechanism. *For. Ecol. Manage.* 7, 191–206.
- Marshall, D.D., Bell, J.F., Tappeiner, J.C., 1992. Levels-of-growing-stock cooperative studies in Douglas-fir – report no. 10 – the Hoskins Study, 1963–83. USDA Forest Service Pacific Northwest Research Station Research Paper U1.
- Max, T.A., Burkhart, H.E., 1976. Segmented polynomial regression applied to taper equations. *For. Sci.* 22, 283–289.
- Miller, H.G., 1995. The influence of stand development on nutrient demand, growth and allocation. *Plant Soil* 169, 225–232.
- Puettmann, K.J., Hann, D.W., Hibbs, D.E., 1993. Evaluation of the size–density relationships for pure red alder and Douglas-fir stands. *For. Sci.* 39, 7–27.
- Reineke, L.H., 1933. Perfecting a stand-density index for even-aged forests. *J. Agric. Res.* 46, 627–638.
- SAS institute Inc., 1999. SAS/STAT user's guide, version 8. SAS Institute Inc., Cary, N.C., 3864 p.
- Sharma, M., Zhang, S.Y., 2007. Stand density management diagram for jack pine stands in eastern Canada. *N. J. Appl. For.* 24, 22–29.
- Smith, N.J., Hann, D.W., 1984. A new analytical model based on the $-3/2$ power rule of self-thinning. *Can. J. For. Res.* 14, 605–609.
- Smith, N.J., Hann, D.W., 1986. A growth model based on the self-thinning rule. *Can. J. For. Res.* 16, 330–334.
- Tang, S.Z., Meng, C.H., Meng, F.R., Wang, Y.H., 1994. A growth and self-thinning model for pure even-age stands: theory and applications. *For. Ecol. Manage.* 70, 67–73.
- VanderSchaaf, C.L., Burkhart, H.E., in press. Using segmented regression to estimate stages and phases of stand development. *Forest Sci.*
- Zeide, B., 1987. Analysis of the $3/2$ power law of self-thinning. *For. Sci.* 33, 517–537.