

---

# Projecting Diameter Growth in Tropical Trees: A New Modeling Approach

Martin Ricker and Rafael del Río

---

**ABSTRACT.** An important and heretofore unresolved challenge in forestry has been how to project long-term tree growth (i.e., decades to hundreds of years) from short-term measurements (here 1 year) for trees that do not present annual growth rings in their trunk wood. Such a method is crucial in the lowland tropics, where few long-term growth measurements have been taken, and where frequently trees lack reliable annual growth rings because of the lack of winters or highly seasonal dry periods. The new piecewise linear (PL) growth model, developed in this article, relates logarithmic relative growth to trunk diameter. Having obtained the coefficients from piecewise linear regression, the long-term age-diameter curve is calculated, i.e., the expected average growth curve of a statistical population of individual trees. The model is applied to the following five tree species without annual growth rings from the tropical rainforest in Los Tuxtlas (Veracruz, Mexico): *Aspidosperma megalocarpon*, *Cordia alliodora*, *Dialium guianense*, *Guarea grandifolia*, and *Persea schiedeana*. Using the tools of multiple linear regression, the PL model is highly flexible to derive sigmoid, exponential, and over-exponential growth separately for different diameter segments. FOR. SCI. 50(2):213–224.

**Key Words:** Bootstrap, exponential integral, multiple linear regression with fixed intercept, piecewise linear regression carried out as multiple linear regression.

---

FOR MOST TREE SPECIES AND SITES IN THE LOWLAND TROPICS, no previous scientific information on growth and yield exists, and one has to rely on questionable estimates from local people. Without cold winters or strong climatic seasonality, trees do not form reliably distinct annual growth rings, and it is not possible to determine ages by counting growth rings (Jacoby 1989, Pilar-Ibarra 2000). Consequently, yield projections exist only for some sites of a few timber species, such as *Swietenia macrophylla* King (Mahogany) or *Cedrela odorata* L. (Spanish Cedar). This in turn leads to an almost complete lack of species-

and site-specific cost-benefit analysis, and subsequently to the virtual nonexistence of truly informed forest management on a landscape-wide scale in most of the tropics.

This article presents a new modeling approach to derive the long-term age-diameter growth curve from species- and site-specific data obtained in the short-term (minimum 1 year). To understand conceptually the basic idea, consider the following hypothetical scenario: A number of different-aged but genetically identical trees grow under perfectly homogeneous environmental conditions. Then the old

---

Martin Ricker, Estación de Biología Tropical "Los Tuxtlas," Universidad Nacional Autónoma de México (UNAM), Apartado postal 94, San Andrés Tuxtla, Veracruz 95701, Mexico—Phone: 52-29494-26623; mrickr@servidor.unam.mx, martin\_tuxtlas@yahoo.com.mx. Rafael del Río, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), UNAM, Apartado postal 20-726, Administración 20, Delegación Álvaro Obregón, México D.F. 01000, Mexico—Phone: 52-55-5622-3595; delrio@servidor.unam.mx, delrio@leibniz.iimas.unam.mx.

**Acknowledgments:** We are grateful to Miguel-Angel Sinaca and Eladio Velasco (Laguna Escondida, Mexico) who carried out much of the fieldwork in Los Tuxtlas. Apolinar Calderón (IIMAS, UNAM) gave advice on the application of Runge-Kutta methods. We also thank Federico O'Reilly and Carlos Diaz (IIMAS, UNAM) for useful conversations about statistical applications. Douglas C. Daly (The New York Botanical Garden, USA) kindly commented on a manuscript draft. The review process, involving several drafts, was long but enlightening: We are grateful to over five reviewers and (associate) editors of *Forest Science* who provided constructive and important critique. The research was partially supported by the Consejo Nacional de Ciencia y Tecnología (Mexico City, CONACYT project 37,444-E).

---

Manuscript received August 21, 2001, accepted August 22, 2003.

Copyright © 2004 by the Society of American Foresters



(large) trees will tell exactly how the young (small) trees will grow once they reach the age of the old trees, without having to wait until the young trees reveal that information by themselves. This approach is used to develop a regression model that relates relative growth with diameter. From this relationship, the long-term age-diameter curve is derived, i.e., the expected average growth curve of a statistical population of individual trees.

While the presented steps of the method can be carried out with commercial software, such a procedure would be tedious. Therefore, the first author wrote the necessary software, which is available from him via e-mail.

## Model Development

Ricker (1998) suggested the following piecewise linear relationship as the basis of a new tree growth model:

$$\ln[dD/dAge)/D] = a_i + b_i \cdot D \quad \{i = 1, 2, \dots \text{segment}, D_{\text{Min},i} \leq D \leq D_{\text{Max},i}\} \quad (1)$$

In this equation,  $D$  is the trunk diameter,  $D_{\text{Min},i}$  and  $D_{\text{Max},i}$  the corresponding minimum and maximum diameters for a given segment  $i$ , and  $a_i$  and  $b_i$  are the segment's regression coefficients.

While Equation 1 employs instantaneous diameter increment ( $dD/dAge$ ), which technically is the slope of the tangent of the growth function at given diameter, field measurements can only estimate the periodic diameter increment (here over 1 year). The slope of the measurable annual increment over a year will be a reasonable estimate for the slope of the tangent only if one uses the mean diameter between the diameters corresponding to the first and last measurement on a given tree [ $D_{\text{Mean}} = D_{\text{Initial}} + (D_{\text{Final}} - D_{\text{Initial}})/2$ ].

Accurate estimation is important to apply the model correctly: The trunk diameter increment should be calculated from measurement of the perimeter over bark, above buttresses at a marked height ( $D = \text{perimeter}/\pi$ ). The height should ideally be the same for all trees (at breast height), although deviations are not severe given that diameters and their increments do not vary strongly with tree height variation of some decimeters. The bark should be cleaned from moss, and lianas separated from the trunk. On large trees, one has to mark several points around the perimeter (a tree of 1 m diameter has a perimeter of 3.14 m!). Depending on the species' bark, one can use paint, but the bark of some species is peeling away. More reliable are aluminum tree nails. It may be useful to measure the perimeter parallel on two to three heights (20–50 cm apart) to get two to three comparable diameter increments after a year (also if marks disappear over the year). After the second measurement one can then discard measurements that are obviously "off," or take the average of the two to three increments.

Taking the second measurement is best after exactly a year, because it represents one complete growth cycle, with all dry and wet seasons. Deviation in measurement time from a year can be interpolated linearly to 365 days. Depending on the trees' growth phases over the year, however,

the higher the deviation, the more inaccurate the annual increment estimate may become. For taking the second measurement, one should have the data from the first measurement at hand in the field; with very few exceptions of dehydration in dry forests, a tree will never decrease in diameter increment. Therefore, a second perimeter measurement that is smaller than the first measurement a year ago is an error measurement, and should be verified immediately in the field. Also one should have a rough idea of the maximum possible diameter increment of the species in the region (e.g., by asking local people). Some trees may have to be discarded after the year from the study because they died, were cut, or the bark got damaged, but selecting initially 100 trees of a wide diameter range, randomly from the chosen habitat, represents a good statistical number.

Returning to the mathematical model, Equation 1 says in words that, on a logarithmic scale, instantaneous relative growth of the trunk diameter is linearly related to the diameter itself. To model this function over the whole diameter range of a tree, several piecewise segments have to be used, as shown graphically for five species in Figures 1–3 on the left. Why does this function represent a good tree growth model?

1. With increasing age or diameter, relative growth in plants  $[(dD/dAge)/D]$  goes always from high values (frequently over 100%) to low values (close to zero). Tree seedlings with diameters of a few millimeters grow a few millimeters per year, when they establish themselves on a site, while adult trees of many decimeters diameter still grow only millimeters (or centimeters at best). Therefore, any function that models relative growth of tree diameter will slope downwards. A combination of highly flexible piecewise linear segments is one possibility to model the downward path.
2. Using linear segments as interpolating splines has great advantages: The whole set of tools from multiple linear regression techniques becomes available.
3. Taking the logarithm of relative growth makes the data homoscedastic (small diameters present larger variance of relative growth than do large diameters) and avoids an impossible trajectory into negative relative growth.

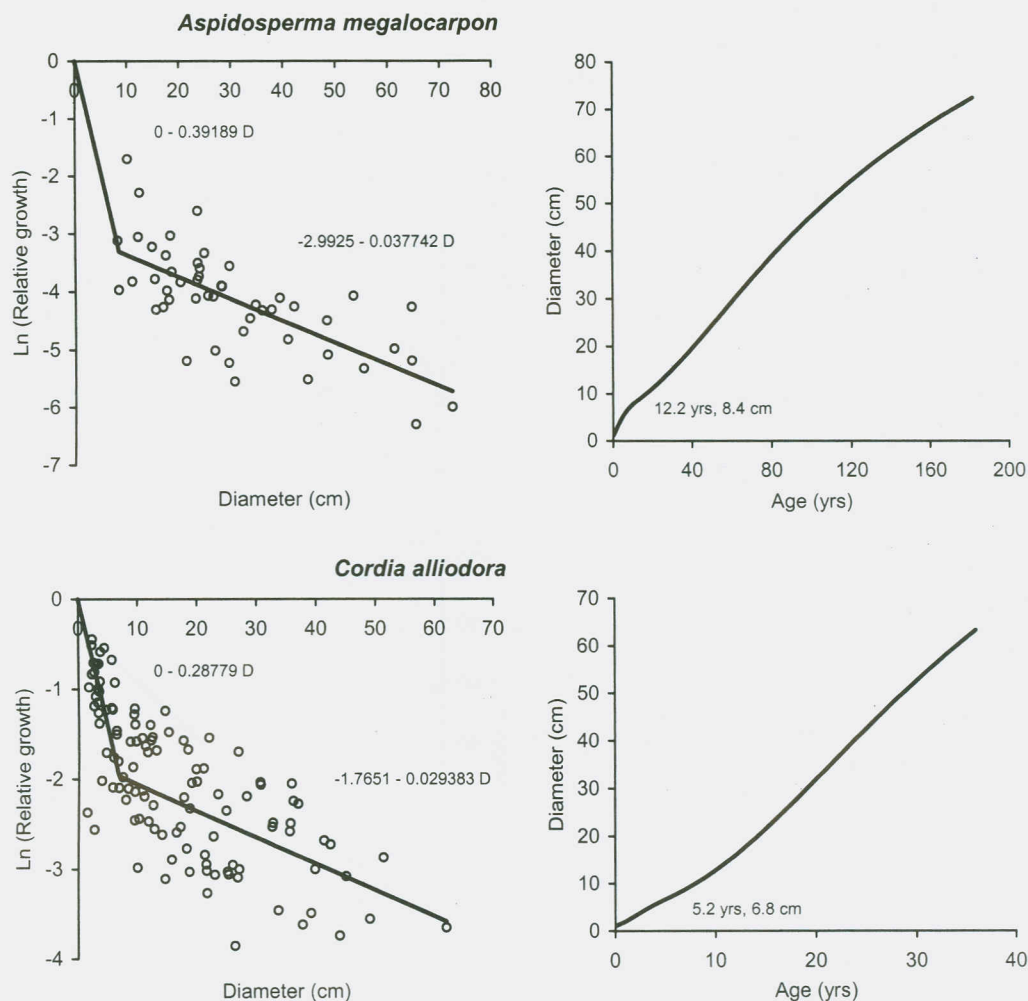
Next we turn to the statistical aspects of how to carry out the piecewise linear regression.

## Piecewise Linear Regression with the PL Model

Before the age-diameter curve can be derived, piecewise linear regression has to be carried out to model the relationship between diameter and logarithmic relative growth. The regression model in the case of three segments is the following:

$$\begin{aligned} \ln(RG) &= a_1 + b_1 \cdot D \text{ for } D_{\text{Min}} \leq D \leq KD_1 \\ \ln(RG) &= a_2 + b_2 \cdot D \text{ for } KD_1 \leq D \leq KD_2 \\ \ln(RG) &= a_3 + b_3 \cdot D \text{ for } KD_2 \leq D \leq D_{\text{Max}} \end{aligned} \quad (2)$$





**Figure 1.** Piecewise linear regression with growth data from *Aspidosperma megalocarpon* and *Cordia alliodora* from the tropical rainforest in Los Tuxtlas (Veracruz, Mexico). For each segment, the resulting regression equation is given in the graphs on the left ( $D$  = diameter). Residuals are normally distributed, homoscedastic, and serially independent. For the graphs on the right, the corresponding age-diameter curves were derived with Equation 4. The kink ages and kink diameters from the piecewise linear regressions are given. The calibrating point is 1 cm at 0 years for both species. The kinks are hardly recognizable in the age-diameter curves. Note the different scales for the species' growth curves.

with constraints:  $a_2 = a_1 + KD_1 \cdot (b_1 - b_2)$

$$a_3 = a_1 + KD_1 \cdot (b_1 - b_2) + KD_2(b_2 - b_3)$$

In this system of equations,  $RG$  is relative growth,  $D$  is the trunk diameter,  $D_{Min}$  and  $D_{Max}$  the minimum and maximum measured diameters of the data set, and  $KD_1$  and  $KD_2$  are the two kink diameters (also known as "join points," "break-points," or "knots"). The two constraints fix  $a_2$  and  $a_3$ , once the kink diameters are chosen; otherwise the piecewise linear segments would not connect at the kink diameters.

Piecewise linear regression can be carried out with nonlinear regression algorithms. Equation 2 would be converted to the regression model  $Ln(RG) = a_1 + b_1 \cdot (D_{Min} \leq D < KD_1) + b_2 \cdot [(KD_1 \leq D < KD_2) - KD_1] + b_3 \cdot [(KD_2 \leq D \leq D_{Max}) - KD_2]$ , which can be specified in nonlinear regression software, such as SAS or SYSTAT. There is,

however, a serious disadvantages of the nonlinear regression approach: Depending on the data and the starting values, the iterative nonlinear regression algorithms may fail to find the regression coefficients, or present suboptimal results. In contrast, in linear regression the computation is straightforward (noniterative). Therefore we present in Appendix 1 a method to carry out piecewise linear regression as multiple linear regression, with each segment taken as a separate dimension.

The parameters  $a_1$ ,  $a_2$ , and  $a_3$  represent the  $Y$ -intercepts of the three piecewise linear segments. In particular,  $a_1$  corresponds to the logarithmic relative growth rate infinitesimally close to a diameter of  $D = 0$  (see Equation 1). A vertical shift up or downward of the regression line changes the relative growth rate at each point of the segment in a multiplicative way, given the logarithmic  $Y$ -axis. For example, an upward shift of  $Ln(2) = 0.693$  means that relative growth at each diameter becomes twice its former value.

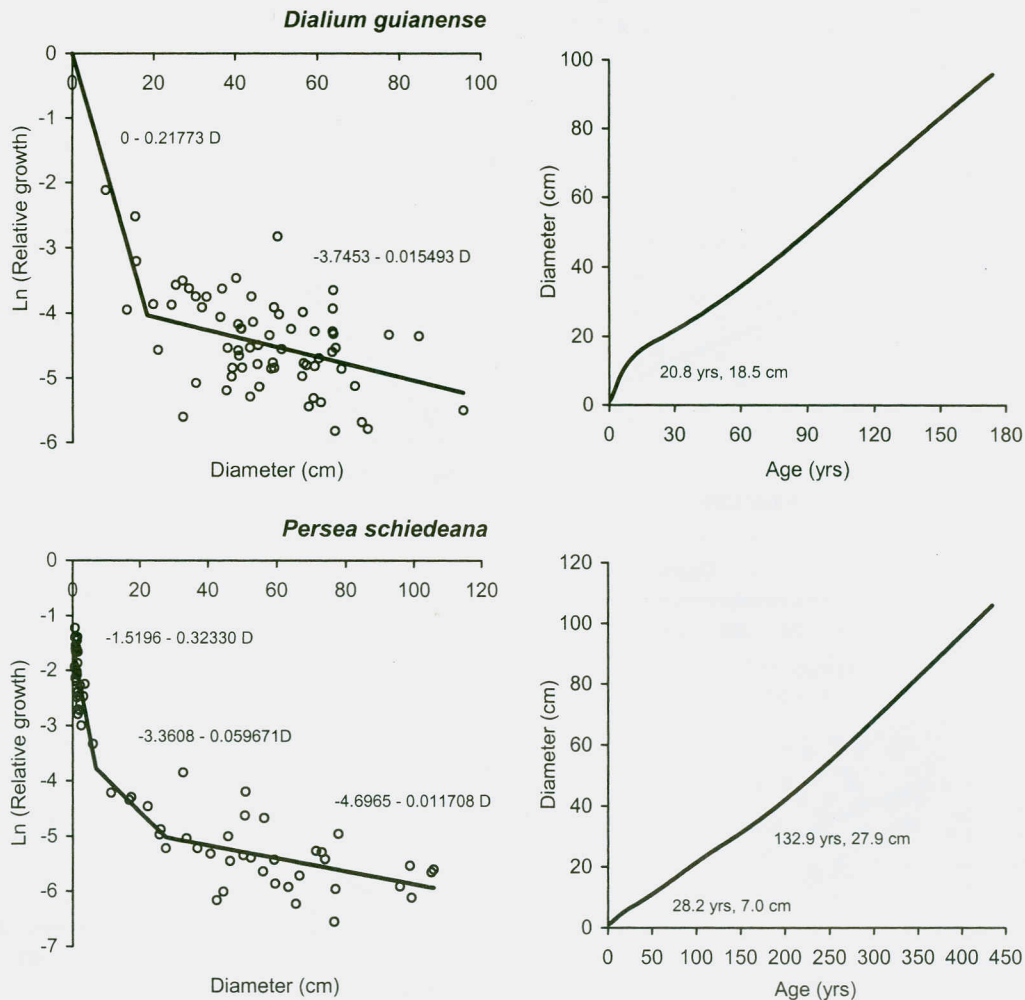


Figure 2. Piecewise linear regression with growth data from *Dialium guianense* and *Persea schiedeana* (as in Figure 1). For *Persea schiedeana*, the Y-intercept was not fixed at zero because there were many data points for small diameters.

Therefore, the parameter  $a$  determines the growth rates at different diameters, and changing  $a_i$  with fixed  $b_i$  affects the different growth rates over the whole diameter range of a given segment in a uniform way.

Depending on the data, the parameter  $a_1$  can take on any value: A test to examine if  $a_1$  is significantly different from zero is of no particular interest, as it would mean only that initial relative growth could be 100% [ $\ln(100\%) + b \cdot 0 = 0$  in Equation 1]. If the initial relative growth [i.e.,  $\exp(a_1)$ ] is known from a seedling experiment or supposed to take on a certain value, then the multiple linear regression should be carried out with a fixed Y-intercept. Appendix 2 derives the statistical formulas to calculate the regression coefficients in that case (as used in four cases in Figures 1–3 on the left).

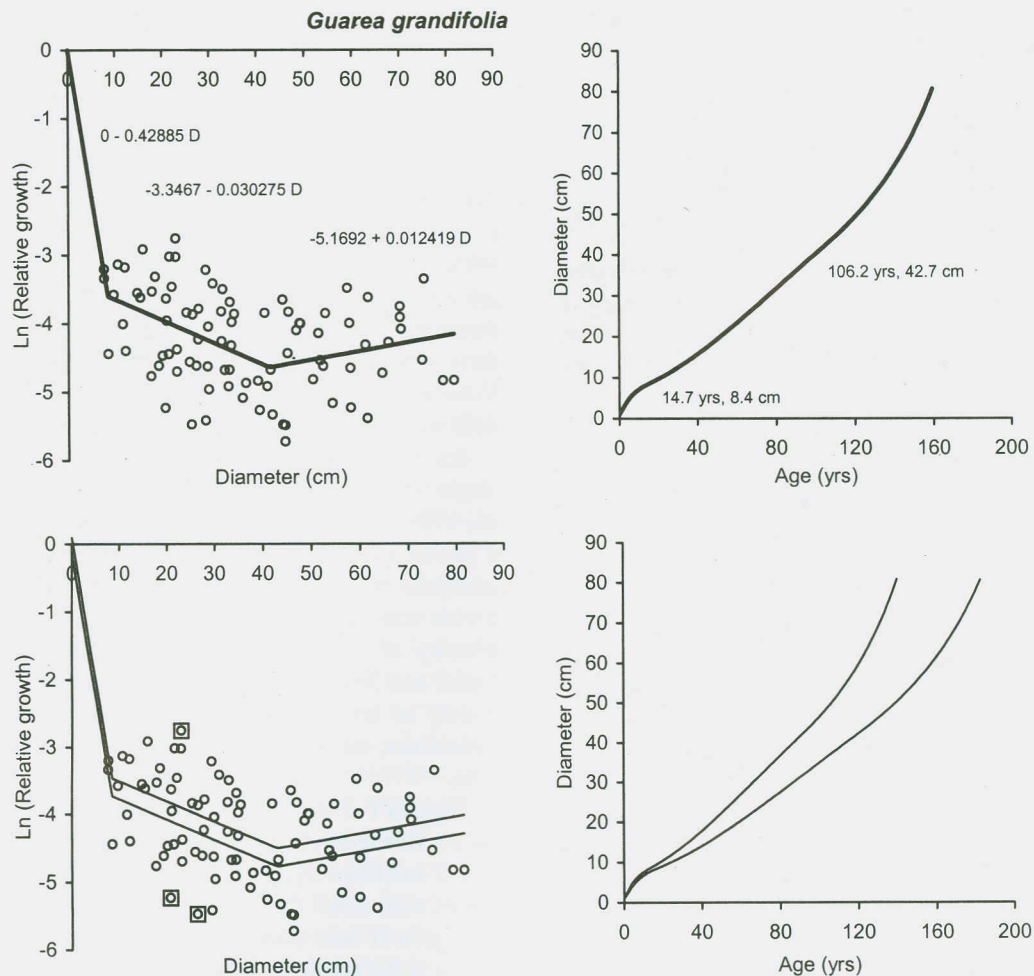
The number of segments may vary, depending on the diameter range considered and the trees' growth history (two to three segments in the examples of this article). There are two, each other opposing criteria to select the number of segments: On the one hand, more segments will lead to a better fit of the piecewise linear regression line to the data cloud. On the other hand, the difference between the  $b$ -slopes of any two segments has to be statistically significant

to support objectively the inclusion of the intermediate kink. If the slopes of two neighboring segments are not significantly different, the kink should be eliminated and the regression should be carried out with one segment less. Therefore, after a visual inspection of the graphed data, one will start with the highest reasonable number of segments (say three), and reduce the number of segments stepwise, until all pairwise tests of differences between  $b$ -slopes are statistically significant, or only one segment is left.

Note in this context that the usual significance test in regression analysis to examine  $b_i = 0$  is irrelevant here. That would only mean to test whether the growth could be exponential, as  $b_i = 0$  implies exponential growth (see Equation 5 below), but not whether the growth curve should be calculated with one segment less.

The comparison of the slopes  $b_1$  with  $b_2$ ,  $b_1$  with  $b_3$ , and  $b_2$  with  $b_3$  (in the case of three segments) is carried out with a  $t$ -test. The formula  $t_{\text{Sample}} = |b_1 - b_2| / [s_Y \cdot (g_{11} + g_{22} - 2 \cdot g_{12})^{0.5}]$ , given by Sokal and Rohlf (1995) to compare slopes in multiple linear regression, involves the Gaussian multipliers (called  $g$ ) that have to be provided by a computer program. The  $t$ -value has  $(n - k - 1)$  degrees of





**Figure 3.** Piecewise linear regression with growth data from *Guarea grandifolia* (as in Figure 1). The bottom graphs show in addition the regression curves and age-diameter curves that correspond to the 95% confidence interval for the regression's Y-mean. Three statistical outliers are indicated with squares.

freedom for a free  $Y$ -intercept and  $(n - k)$  degrees of freedom for a fixed  $Y$ -intercept, where  $n$  is the number of data points and  $k$  the number of piecewise linear segments. These are the usual number of degrees of freedom as employed in multiple linear regression (Sokal and Rohlf 1995), where in our case each piecewise linear segment corresponds to an independent  $X$ -variable (see Appendix 1).

### Derivation of the Age-Diameter Function

Having completed the piecewise linear regression, the objective is to convert the function between diameter and relative growth (Figures 1–3, left) into a mathematically equivalent function between age and diameter (Figures 1–3, right). In this section, a single segment will be considered, so that we can use  $a$  and  $b$  instead of  $a_i$  and  $b_i$ .

One can transform Equation 1 into the following equation of instantaneous increment as a function of diameter:

$$dD/dAge = D \cdot \text{Exp}(a + b \cdot D) \quad (3)$$

Next, Equation 3 is converted into  $dAge/dD = \text{Exp}(-a - b \cdot D)/D$ . Integrating both sides of this equation with respect to  $D$ , and  $usbD_1$  at  $Age_1$  being any point on the growth curve, yields the

function  $Age_2 = Age_1 + \text{Exp}(-a) \int_{D_1}^{D_2} [\text{Exp}(-b \cdot D)/D] dD$ . Defining the exponential integral  $Ei(-b \cdot D)$  as in Appendix 3 (A3.1), and applying Equation A3.2 (with  $c = -b$  and  $y = D$ ), one gets  $Age_2$  as a function of  $D_2$ :

$$Age_2 = Age_1 + \text{Exp}(-a) \cdot [Ei(-b \cdot D_2) - Ei(-b \cdot D_1)] \quad (4)$$

In Equation 4, both diameters  $D_1$  and  $D_2$  have to be larger than zero, different from the regression of logarithmic growth as a function of diameter where the diameter can numerically be zero. The underlying reason is that growth is modeled as a multiplicative process, and for an absolute growth curve there must be some initial diameter that can be proportionally increased.

For  $b = 0$ ,  $\int_{D_1}^{D_2} [\text{Exp}(-b \cdot D)/D] dD$  is simply  $\text{Ln}(D_2/D_1)$ . For  $b \neq 0$ , the exponential integral  $Ei(-b \cdot D)$  converts into an infinite polynomial (Jeffrey 1995) and therefore has to be approximated with an acceptable degree of accuracy. Appendix 3 provides a new and easy-to-use method to calculate the exponential integral  $Ei(x)$  relatively quickly, for example with a spreadsheet program.



With Equation 4, one can calculate  $Age_2$  corresponding to the diameter  $D_2$ , given  $Age_1$ ,  $D_1$ ,  $a$ , and  $b$ . However, usually in a growth curve one wishes to determine a diameter  $D_2$  for given  $Age_2$ , not the other way around. The easy case is for  $b = 0$ , where Equation 4 converts into an exponential growth function:

$$D_2 = D_1 \cdot \text{Exp}[(Age_2 - Age_1) \cdot \text{Exp}(a)] \quad \{\text{for } b = 0\} \quad (5)$$

For  $b$  not being zero, there are two mathematical options. The mathematically simple way consists of employing a converging trial-and-error algorithm that determines  $D_2$  in Equation 4 in ever-smaller steps for given  $Age_2$ . The more sophisticated way is to take Equation 3 and apply a Runge-Kutta method for solving ordinary differential equations (Gerald and Wheatley 1994).

Depending on the shape parameter  $b$ , the piecewise linear (PL) model derives sigmoid, exponential, or over-exponential growth of the age-diameter curve (but note that it cannot model negative growth). With negative  $b$ ,  $D_2(Age_2)$  is a sigmoid function with an accelerating beginning and a decelerating end. The resulting formula for the turning point diameter (TPD) between the two parts is especially simple:

$$TPD = -1/b \quad \{\text{for } b < 0\} \quad (6)$$

It is obtained by taking the second derivative of Equation 4 with respect to  $D_2$ , setting the derivative equal to zero, and solving for  $D_2$ . The corresponding turning point age (TPA) is:

$$TPA = Age_1 + \text{Exp}(-a) \cdot [Ei(1) - Ei(-b \cdot D_1)] \quad \{\text{for } b < 0\} \quad (7)$$

Interestingly, the turning point diameter is dependent exclusively on the slope of the segment's regression line in the relationship of logarithmic relative growth over diameter. All other parameters ( $a$ ,  $Age_1$ ,  $D_1$ ) influence the turning point age (i.e., the time it takes to reach the turning point diameter) but not the turning point diameter itself.

With  $b = 0$ , the model simplifies to the exponential curve given already with Equation 5. Positive  $b$  results in an over-exponential age-diameter curve with ever more increasing relative growth rate; Equation 1 makes this obvious.

Equations 4, 5, and 7, respectively, require one age-diameter point  $D_1$  at  $Age_1$ . This point calibrates the position of the growth curve  $D_2(Age_2)$ . In Equation 4, one can move the age-diameter curve freely to the left or right by manipulating  $Age_1$  at given  $a$ ,  $b$ , and  $D_1$  (mathematically,  $Age_1$  can also be negative). Similarly, at given  $a$ ,  $b$ , and  $Age_1$  one can manipulate  $D_1$ . If a change in  $Age_1$  is opposite to a change in  $\text{Exp}(-a) \cdot Ei(-b \cdot D_1)$ , then the age-diameter curve remains exactly in the same position.

In the next section, we will apply the PL model to empirical field data.

## Model Application

Figures 1–3 show the application of the PL model with growth data from five tree species of the Los Tuxtlas

rainforest (Veracruz, Mexico). Site descriptions of this tropical rainforest can be found in Bongers et al. (1988) and Ibarra-Manríquez et al. (1997). The five species are *Aspidosperma megalocarpon* Müll. Arg. (common name "nazareno," plant family Apocynaceae), *Cordia alliodora* (Ruiz et Pav.) Oken ("súchil acahualero," Boraginaceae), *Dialium guianense* (Aubl.) Sandwith ("paque," Caesalpiniaceae), *Guarea grandifolia* ("sabino," Meliaceae), and *Persea schiedeana* Nees ("chinine," Lauraceae). These tree species are well-known from a floristic inventory of the Mexican Autonomous National University's (UNAM) 644-ha reserve (Ibarra-Manríquez and Sinaca 1995, 1996a, 1996b). Voucher specimens for each species are deposited in UNAM's field station herbarium.

Each data point in Figures 1–3 on the left presents a single tree, measured twice between Mar. 3, 1997 and Apr. 26, 1999 (depending on the species and individual tree over a period of 345–633 days, but increments were linearly interpolated to 365 days). Trees were selected to come from a wide range of trunk diameters, within an area of 575 ha of primary and secondary forest, as well as cattle pastures. Initial and final trunk diameters were determined approximately at breast height (or above buttresses) from their perimeters, measured at a marked height with a measuring tape.

Figures 1–3 show on the left the resulting regression line. Of the five species, only *P. schiedeana* was modeled with a free  $Y$ -intercept  $a_1$ , given the existence of numerous data points with small diameters. In the other cases, initial relative growth was assumed to be 100% (which makes the origin in the graphs to be the  $Y$ -intercept). *A. megalocarpon*, *C. alliodora*, and *D. guianense* present two piecewise linear segments, while *G. grandifolia* and *P. schiedeana* present three segments.

Applying Equation 4 for the five species, the resulting age-diameter curves are shown in Figures 1–3 on the right. The calibrating point  $D_1$  at  $Age_1$  is 1 cm at 0 years as a possible but arbitrary point for comparing the age-diameter curves of different species. There are obvious differences in growth rate and shape between the species. One way to compare growth rates is calculating the expected average annual increment over the first 30 years of growth. *C. alliodora* has an estimated average annual diameter increment of 1.75 cm during the first 30 years, *D. guianense* 0.72 cm (41% of *C. alliodora*'s increment), *A. megalocarpon* 0.50 cm (29%), *G. grandifolia* 0.42 cm (24%), and *P. schiedeana* 0.24 cm (14%). Its relatively fast growth makes *C. alliodora* a popular timber species in plantations: The largest encountered tree of 62 cm trunk diameter was estimated to be only 36 years old. On the other hand, *P. schiedeana* according to this analysis is a slowly growing tree species in Los Tuxtlas, requiring 413 years to reach 100 cm trunk diameter at breast height. For this comparison between species, one has to take into account that the sampled trees of the tree species were not found on the same site, so that different site conditions between species may also play in.

There are also differences in the growth curves' shape



between species. *A. megalocarpon* has a clearly pronounced sigmoid shape of the age-diameter curve for the second segment (turning point diameter  $TPD_2 = -1/-0.037742 = 26.5$  cm). The age-diameter curve for *C. alliadora* is more linear ( $TPD_2 = 34.0$  cm). *D. guianense* has a pronounced sigmoid age-diameter curve for the first segment ( $TPD_1 = 4.6$  cm), but a relatively linear one for the second segment ( $TPD_2 = 64.5$  cm). *P. schiedeana* was modeled with three segments, which are notably difficult to distinguish in the age-diameter curve ( $KD_1 = 7.0$  cm,  $KD_2 = 27.9$  cm). Finally, *G. grandifolia* is interesting because it presents a positive slope in its third regression segment ( $b_3 = +0.012419$  cm<sup>-1</sup>).

To be useful, the PL model should be able to provide a measure of reliability of the growth projections. Figure 3 (bottom) shows for the data of *G. grandifolia* the regression curves and resulting age-diameter curves that correspond to the 95% confidence interval of the mean logarithmic relative growth. A simultaneous move up or down of all piecewise linear segments together is considered. This confidence interval is based on the regression's standard error of the sampled *Y*-mean, i.e., the average logarithmic relative growth. The standard error of the sampled *Y*-mean is  $(s^2/n)^{0.5}$ . The unexplained variance  $s^2$  is  $[(1 - R^2) \cdot \Sigma y^2] / (n - k - 1)$ , where  $R^2$  is the coefficient of multiple determination of the piecewise linear regression and  $\Sigma y^2$  the total sum of squares. The confidence intervals are then calculated with a *t*-distribution (for details see Sokal and Rohlf 1995).

The 95% confidence interval of the mean logarithmic relative growth reflects the overall uncertainty about the growth rate and implies that all  $a_i$  (of *i* segments) have to be moved up or down together by the same interval. Calculating confidence curves in this way is based on four assumptions:

1. The sampled *Y*-mean is normally distributed, which is automatically true if the residuals are normally distributed. Note that given the central limit theorem, depending on the number of data points, *Y*-mean approaches a normal distribution even if the residuals are not normally distributed (Sokal and Rohlf 1990).
2. The variance of the residuals is homogeneous along the regression line and the residuals are serially independent. This assures a homogeneous uncertainty over the whole diameter range (see Figure 3, bottom left).
3. There is no uncertainty about the calibrating point  $D_1$  at  $Age_1$ .

4. There is no uncertainty about the slopes  $b_i$  of the *i* segments. This, in accordance with Equation 6, implies that there is no doubt about the turning point diameters.

Normality of the residuals was analyzed with the Kolmogorov-Smirnov test, homoscedasticity with Bartlett's test, and serial independence according to von Neumann (Sokal and Rohlf 1995). Points 1 and 2 are fulfilled in the *G. grandifolia* example of Figure 3. This allows also to detect statistical outliers by calculating standardized residuals and comparing them against the corresponding *t*-value (Sokal and Rohlf 1995). Three such outliers are indicated in Figure 3 (bottom left) with squares.

To find confidence intervals for the slopes of each segment, bootstrapping was carried out for *G. grandifolia* (see Chernick 1999). The original statistical sample population in Figure 3 (left) was considered as a pseudoparametric population, from which 1,000 sample populations of the same size ( $n = 88$ ) were randomly derived. For each pseudosample population, the PL model was applied ( $a_1 = 0$  always), and 1,000 combinations of  $b_1$ ,  $KD_1$ ,  $a_2$ ,  $b_2$ ,  $KD_2$ ,  $a_3$ , and  $b_3$  derived. Table 1 gives the following information for each of the seven variables: (1) the original value of each parameter as given in Figure 3 (top), (2) the 26th entry after ordering the 1,000 data points (separately for each variable) from the smallest to largest value as the lower 95% confidence limits, (3) the corresponding percent deviation from the original value, (4) the 975th entry as the upper 95% confidence limit, and (5) again the corresponding percent deviation from the original value.

The percent deviations from the original value of six of the seven variables in Table 1 range from -231.6% to 168.4%. Only the slope of the second segment  $b_2$  deviates by a much higher percentage into the positive range (1,675.8%), but this segment is also constrained to connect the first and the second segment.

The most interesting aspect in this example is that the 95% confidence interval of the slope of the third segment ( $b_3 = +0.012419$  cm<sup>-1</sup> in Figure 3, bottom left) ranges from a negative value (-0.016345), corresponding to sigmoid growth, to a positive value (+0.033338), corresponding to over-exponential growth. While we do know that over-exponential age-diameter curves can exist at low growth rates in trees (unpublished data from a tree with annual growth rings of *Pinus teocote* Schiede ex Schltdl. &

**Table 1. Bootstrap parameters of the regression coefficients for *Guarea grandifolia*.<sup>a</sup>**

	$b_1$	$KD_1$	$a_2$	$b_2$	$KD_2$	$a_3$	$b_3$
Original (Figure 3)	-0.42885 cm <sup>-1</sup>	8.4 cm	-3.3467	-0.030275 cm <sup>-1</sup>	42.7 cm	-5.1692	0.012419 cm <sup>-1</sup>
Lower 95% CL	-0.49922 cm <sup>-1</sup>	7.6 cm	-9.0080	-0.089081 cm <sup>-1</sup>	9.7 cm	-6.3538	-0.016345 cm <sup>-1</sup>
Deviation	-16.4%	-9.5%	-169.2%	-194.2%	-77.3%	-22.9%	-231.6%
Upper 95% CL	-0.32413 cm <sup>-1</sup>	11.6 cm	-2.4015	0.477068 cm <sup>-1</sup>	45.9 cm	-3.5659	0.033338 cm <sup>-1</sup>
Deviation	24.4%	38.1%	28.2%	1675.8%	7.5%	31.0%	168.4%

<sup>a</sup> Notes: 1,000 bootstrap iterations, fixed  $a_1 = 0$ ,  $KD$  = kink diameter, CL = confidence limit, Deviation = percent deviation from the original estimates given in Figure 3 (top).



Cham.), in this *G. grandifolia* data there remains uncertainty about it.

## Discussion

Modeling tree growth is fundamental to forest science and has been the subject of numerous articles (see summaries in Zeide 1993, Liu and Ashton 1995, Vanclay 1994, Vanclay 1995). Our article focuses on an individual-tree model for trunk diameter growth, i.e., the expected average growth curve of a statistical population of individual trees. The trunk diameter has the advantage that annual increments can be measured relatively accurately, and it is possible to relate other variables such as trunk volume allometrically to the trunk diameter.

The simplest approach to diameter growth modeling consists of choosing a growth function for nonlinear regression, with empirical data of age-diameter points from trees of a given species and site. The result of this nonlinear regression is an average growth path of the long-term age-diameter curve over the past decades. The most widely used growth model in forestry has been the von-Bertalanffy-Richards-Chapman (BRC) model, often called the Chapman-Richards model (Bredenkamp and Gregoire 1988, Zeide 1993). It was introduced to forestry by Pienaar and Turnbull (1973) and has the basic functional form  $D = D_{\text{Max}} \cdot [1 - \text{Exp}(-a \cdot \text{Age})]^b$ , where  $D$  is the diameter and  $a$ ,  $b$ , and  $D_{\text{Max}}$  are regression coefficients.

In the lowland tropics, however, this simplest approach is usually impossible, because without strong climatic seasonality, tree ages generally cannot be determined from annual growth rings in the trunk (Jacoby 1989, Pilar-Ibarra 2000). The only alternatives are expensive carbon dating (Chambers et al. 1998) or indirect methods, where ages are inferred from other variables such as annual increment (see Martínez-Ramos and Álvarez-Buylla 1998). In this article, we contribute a novel indirect method based on annual trunk increments. In contrast to the few existing indirect approaches, most notably the simulation approach by Lieberman and Lieberman (1985), the new PL model is apparently the first one that takes advantage of linear regression with all its tools in an indirect approach to derive the long-term age-diameter curve from short-term measurements.

The approach to convert the relationship between diameter and logarithmic relative growth mathematically into the relationship between age and diameter is not limited exclusively to the PL model. Ricker et al. (1999) convert the above-mentioned equation of the BRC model into the corresponding equation  $\text{Ln}(RG) = \text{Ln}\{a \cdot b \cdot [(D/D_{\text{Max}})^{-1/b} - 1]\}$  and find the three coefficients  $a$ ,  $b$ , and  $D_{\text{Max}}$  via nonlinear regression. Putting the coefficients back into the original equation provides the projected long-term age-diameter curve. The same approach is possible for some other nonlinear models. All these nonlinear models, however, have severe drawbacks in comparison with the PL model:

1. The PL model with an a priori undetermined number of regression coefficients is more flexible and can adapt better to different types of growth curves than

the BRC model and nonlinear models with three regression coefficients. In contrast to the BRC model, the PL model is able to model sigmoid ( $b < 0$ ), exponential ( $b = 0$ ), and over-exponential growth ( $b > 0$ ), as well as combinations between these growth types, within a tree's overall growth curve.

2. For the PL model, the parameters have a much clearer interpretation than for the BRC and other nonlinear models: The  $Y$ -intercepts of the segments ( $a_i$ ) determine the relative growth rates at different diameters, the slopes are responsible for the turning point diameter ( $TPD = -1/b_i$ ), and the calibrating point ( $D_1$  at  $\text{Age}_1$ ) determines the absolute position of the growth curve.
3. In nonlinear regression modeling, increasing the number of regression coefficients (at given number of explaining variables, here diameter) causes eventually "endless" iterative searches because of overfitting. In contrast, the PL model has a straightforward criterion in case of overfitting: two neighboring segments will have slopes that are not significantly different and one of the two segments should be eliminated.
4. In linear regression, ordinary least squares (OLS) estimators are automatically the "Best Linear Unbiased Estimators" (BLUE), provided that the assumptions of homoscedasticity, independence, and zero mean of the residual errors are met. If in addition the data are normally distributed, then OLS also provide minimum variance unbiased estimation (MVUA) among all linear and nonlinear unbiased estimators (Maddala 1992).
5. Using the tools of multiple linear regression, the PL model can include a theoretically unlimited number of explanatory genetic or environmental variables, such as leaf nutrients or competition indices, and the long-term growth curve can be derived as a function of these variables. This application will be the topic of future research.

The basic idea of avoiding the necessity of taking measurements over the lifetime of a tree without annual growth rings is that existing large trees can tell us how currently small trees will grow on average in the future, when the increments measured on the large trees are representative for the increments of the small trees once they reach that large size. It implies that different-sized trees are selected at random within a given environment. For example, it is assumed that large trees on the average are not from a superior site, when small trees on the average are from an inferior site.

Furthermore, the increments measured in the short-term are assumed to be representative for the long-term. Importantly, the climate of the measurement year(s) has to be representative of the long-term average climate. If temperature and precipitation during the measurement year(s) are not representative for the long-term, growth projection can be strongly biased. For the data of the five species in Figures



1–3, the 1952–1998 climate was analyzed from the Coyame meteorological station, approximately 15 km air-distance away from the study site. The monthly mean maximum daily temperature was on average 28.0°C, and the average annual total precipitation was 4,543 mm. For the 2 years from 1 Apr. 1997 to 31 Mar. 1999, the corresponding values were 27.2°C and 3,287 mm. Therefore, the climate during the measurement period was slightly cooler (3%) and considerably drier (28%) than the “long-term” climate 1952–1998. This probably causes the derived long-term growth curves in Figures 1–3 to underestimate the true growth curves, as higher temperature and more precipitation are expected to increase the growth rate.

Growth projections are fundamental to any forest management plan, as well as to cost-benefit analysis. The tropics contain a large diversity of tree species in distinct environments: We hope that this new modeling tool for projecting long-term tree growth will contribute to overcome the widespread lack of informed forest management.

### Literature Cited

- BONGERS, F., J. POPMA, J. MEAVE DEL CASTILLO, AND J. CARABIAS. 1988. Structure and floristic composition of the lowland rain forest of Los Tuxtlas, Mexico. *Vegetatio* 74:55–80.
- BREDEKAMP, B.V., AND T.G. GREGOIRE. 1988. A forestry application of Schnute's generalized growth function. *For. Sci.* 34:790–797.
- CHAMBERS, J.Q., N. HIGUCHI, AND J.P. SCHIMEL. 1998. Ancient trees in Amazonia. *Nature* 391:135–136.
- CHERNICK, M.R. 1999. *Bootstrap methods: A practitioner's guide*. John Wiley & Sons, New York, NY. 264 p.
- CODY, W.J., AND H.C. THACHER. 1969. Chebyshev approximations for the exponential integral  $Ei(x)$ . *Math. Comp.* 23:289–303.
- GAUTSCHI, W., AND W.F. CAHILL. 1965. Exponential integrals and related functions. P. 228–251 in *Handbook of mathematical functions*, Abramowitz, M., and I.A. Stegun (eds.). Dover Publications, New York, NY.
- GERALD, C.F., AND P.O. WHEATLEY. 1994. *Applied numerical analysis*. Addison-Wesley Publishing Company, Reading, MA. 748 p.
- HARRIS, F.E. 1957. Tables of the exponential integral  $Ei(x)$ . *MTAC* 11:9–16.
- HASTINGS JR., C., J.T. HAYWARD, AND J.P. WONG, JR. 1955. *Approximations for digital computers*. Princeton University Press, Princeton, NJ. 201 p.
- IBARRA-MANRÍQUEZ, G., AND S. SINACA C. 1995. Lista florística comentada de la Estación de Biología Tropical “Los Tuxtlas,” Veracruz, México. *Rev. Biol. Trop.* 43(1–3):75–115.
- IBARRA-MANRÍQUEZ, G., AND S. SINACA C. 1996a. Estación de Biología Tropical “Los Tuxtlas,” Veracruz, México: Lista florística comentada (Mimosaceae a Verbenaceae). *Rev. Biol. Trop.* 44(1):41–60.
- IBARRA-MANRÍQUEZ, G., AND S. SINACA C. 1996b. Lista comentada de la Estación de Biología Tropical “Los Tuxtlas,” Veracruz, México (Violaceae – Zingiberaceae). *Rev. Biol. Trop.* 44(2):427–447.
- IBARRA-MANRÍQUEZ, G., M. RICKER, G. ÁNGELES, S. SINACA C., AND M.A. SINACA C. 1997. Useful plants of the Los Tuxtlas rain forest (Veracruz, Mexico): Considerations of their market potential. *Econ. Bot.* 51:362–376.
- JACOBY, G.C. 1989. Overview of tree-ring analysis in tropical regions. *IAWA Bull. NS* 10:99–108.
- JEFFREY, A. 1995. *Handbook of mathematical formulas and integrals*. Academic Press, San Diego, CA. 410 p.
- LIEBERMAN, M., AND D. LIEBERMAN. 1985. Simulation of growth curves from periodic increment data. *Ecology* 66:632–635.
- LIU, J., AND P.S. ASHTON. 1995. Individual-based simulation models for forest succession and management. *For. Ecol. Manage.* 73:157–175.
- MADDALA, G.S. 1992. *Introduction to econometrics*. MacMillan Publishing Company, New York, NY. 631 p.
- MARTÍNEZ-RAMOS, M., AND E.R. ÁLVAREZ-BUYLLA. 1998. How old are tropical rain forest trees? *Trends Pl. Sci.* 3:400–405.
- MILLER, J., AND R.P. HURST. 1958. Simplified calculation of the exponential integral. *MTAC* 12:187–193.
- PIENAAR, L.V., AND K.J. TURNBULL. 1973. The Chapman-Richards generalization of von Bertalanffy's growth model for basal area growth and yield in even-aged stands. *For. Sci.* 19:2–22.
- PILAR-IBARRA, R. 2000. Estudio del crecimiento diametral del tallo de *Guarea grandifolia* en la selva de Los Tuxtlas, Veracruz. Tesis de licenciatura en biología, Escuela Nacional de Estudios Profesionales Iztacala, Universidad Nacional Autónoma de México, México D.F., Mexico. 104 p.
- RICKER, M. 1998. Enriching the tropical rain forest with native fruit trees: A biological and economic analysis in Los Tuxtlas (Veracruz, Mexico). Ph. D. thesis, Yale University, School of Forestry and Environmental Studies, and Graduate School, New Haven, CT. UMI Dissertation Services, Ann Arbor, MI. 262 p.
- RICKER, M., R.O. MENDELSON, D.C. DALY, AND G. ÁNGELES. 1999. Enriching the rainforest with native fruit trees: an ecological and economic analysis in Los Tuxtlas (Veracruz, Mexico). *Ecol. Econ.* 31:439–448.
- SOKAL, R.R., AND F.J. ROHLF. 1995. *Biometry*. W.H. Freeman and Company, New York, NY. 887 p. [Together with: Rohlf, F.J., and R.R. Sokal, 1995. *Statistical Tables*. W.H. Freeman and Company, New York, NY. 199 p.]
- VANCLAY, J.K. 1994. *Modelling forest growth and yield: Applications to mixed tropical forests*. CAB International, Wallingford, Oxon, United Kingdom. 312 p.
- VANCLAY, J.K. 1995. Growth models for tropical forests: A synthesis of models and methods. *For. Sci.* 41:7–42.
- VAN DER LAAN, C.G., AND N.M. TEMME. 1984. Calculation of special functions: The gamma function, the exponential integrals and error-like functions. *Centrum voor Wiskunde en Informatica (CWI)*, Amsterdam, The Netherlands. 231 p.



### APPENDIX 1: Piecewise Linear Regression in Three Segments, Carried Out as Multiple Linear Regression

Piecewise linear regression can be calculated with the tool of multiple linear regression, with each segment taken as a distinct dimension. For example in a three-dimensional space, in multiple linear regression with two explanatory variables, the joint point corresponds to the corner of the rectangular plane that is adjusted to the data cloud.

The regression model for three segments in this approach is  $Ln(RG) = a_0 + b_1 \cdot D_{\text{First\_segment}} + b_2 \cdot D_{\text{Second\_segment}} + b_3 \cdot D_{\text{Third\_segment}}$  (see Sokal and Rohlf 1995 for the mathematical procedures). The two kink diameters have to be chosen before preparing the input matrix. The data input, as applicable in any commercial multiple linear regression program, has to be carried out as follows, when the data are ordered according to increasing diameter:

1. segment	2. segment	3. segment	Y-variable
$D_1$	<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	$Ln(RG_1)$
$D_2$	<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	$Ln(RG_2)$
...	<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	...
$D_h$	<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	$Ln(RG_h)$
<b><math>KD_1</math></b>	$D_{h+1}$	<b><math>KD_2</math></b>	$Ln(RG_{h+1})$
<b><math>KD_1</math></b>	$D_{h+2}$	<b><math>KD_2</math></b>	$Ln(RG_{h+2})$
...	...	<b><math>KD_2</math></b>	...
<b><math>KD_1</math></b>	$D_{h+j}$	<b><math>KD_2</math></b>	$Ln(RG_{h+j})$
<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	$D_{h+j+1}$	$Ln(RG_{h+j+1})$
<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	$D_{h+j+2}$	$Ln(RG_{h+j+2})$
<b><math>KD_1</math></b>	<b><math>KD_2</math></b>	...	...

The abbreviations mean:  $D$  = trunk diameter,  $KD$  = kink diameter (emphasized in boldface here),  $RG$  = annual relative growth,  $h$  = the  $h$ th data point, and  $j$  = the  $j$ th data point. The data entry scheme assures that any two segments join exactly at the chosen kink diameters. To understand this, think of adjusting a plane to the data in the case of two segments: The kink is one corner of that plane.

The first three columns correspond to the diameter data of the three segments. Given that the data are ordered according to increasing diameter, in the first column ("1. segment") the data for the measured diameters ( $D_1$  to  $D_h$ ) are entered up to the first kink diameter ( $KD_1$ ). For the rest of the column, the first kink diameter is repeated. In the second column ("2. segment"), again the first kink diameter is repeated up to that row, where in the first column the last diameter before the first kink diameter was entered ( $D_h$ ). The following row in the second column contains the first diameter that is larger than the first kink diameter ( $D_{h+1}$ ). The subsequent diameters are entered in this column up to the last diameter that is smaller than the second kink diameter ( $D_{h+j}$ ). For the rest of the second column, the second kink diameter is repeated. The third column ("3. segment") again repeats the second kink diameter up to the row where in the second column the last diameter before the second kink diameter was entered ( $D_{h+j}$ ). Then, the remaining

diameters are entered in the third column. The fourth column contains the logarithmic relative growth data that correspond to the diameters given in any of the first three columns. Note that while here the data are ordered according to increasing diameter, this does not have to be so, as long as the relationships within each row are respected.

The resulting  $Y$ -intercept  $a_0$  of the multiple linear regression is not the  $Y$ -intercept  $a_1$  of the first segment (as defined in Equation 2). The three  $Y$ -intercepts, corresponding to the three segments, have to be calculated with the equation system A1.1. This calculation reconverts the segments from a four-dimensional space back to the two-dimensional space:

$$\begin{aligned} a_1 &= a_0 + b_2 \cdot KD_1 + b_3 \cdot KD_2 \\ a_2 &= a_0 + b_1 \cdot KD_1 + b_3 \cdot KD_2 \\ a_3 &= a_0 + b_1 \cdot KD_1 + b_2 \cdot KD_2 \end{aligned} \quad (A1.1)$$

For significance analyses, there are two less degrees of freedom compared to simple linear regression ( $n - 4$  instead of  $n - 2$  in the case of three segments). The four degrees of freedom that are rested correspond to the estimation of  $a_0$ ,  $b_1$ ,  $b_2$ , and  $b_3$  from the data. Note in this approach that we do not rest degrees of freedom neither for the determination of the kink diameters  $KD_1$  and  $KD_2$  nor for the determination of the  $Y$ -intercepts  $a_1$ ,  $a_2$  and  $a_3$ . The reason is that these parameters are not estimated statistically from the data: The kink diameters are predetermined in the data entry scheme, and the  $Y$ -intercepts are calculated with the system of Equations A1.1 from the other parameters.

To find the most likely growth path, the kink diameters  $KD_1$  and  $KD_2$  have to be optimized so as to minimize the unexplained sum of squares. This can be achieved by an iterative search-algorithm that tries data input matrices with combinations of two kink diameters in ever-smaller steps (jumping forth and back over the final kink diameters), always starting again with those new kink diameters that result in a lower unexplained sum of squares.

### APPENDIX 2: Multiple Linear Regression with the PL Model and Fixed Initial Relative Growth for Three Segments

The  $Y$ -intercept  $a_1$  is the logarithm of the initial relative growth. Fixing  $a_1$  at a desired value implies that the slope of the first segment  $b_1$  is not free, but rather has to be found as a function of  $a_1$ . This appendix shows the derivation of the formulas that are needed in a computer algorithm to calculate piecewise linear regression for three segments with fixed  $Y$ -intercept.

The usual  $Y$ -intercept  $a_0$  is calculated from the averages of the data entry columns ( $X_{i,Avg}$ ) as follows:

$$a_0 = Y_{Avg} - b_1 \cdot X_{1,Avg} - b_2 \cdot X_{2,Avg} - b_3 \cdot X_{3,Avg} \quad (A2.1)$$

The actual  $Y$ -intercept of the first segment in the PL model, according to Equation A1.1, is:

$$a_1 = a_0 + b_2 \cdot KD_1 + b_3 \cdot KD_2 \quad (A2.2)$$



Replacing  $a_0$  in Equation A2.2 with A2.1:

$$a_1 = Y_{Avg} - b_1 \cdot X_{1,Avg} - b_2 \cdot X_{2,Avg} - b_3 \cdot X_{3,Avg} + b_2 \cdot KD_1 + b_3 \cdot KD_2 \quad (A2.3)$$

Bringing  $b_1$  in Equation A2.3 on one side, the slope of the first segment can be calculated as a function of all remaining variables:

$$b_1 = (Y_{Avg} - a_1)/X_{1,Avg} - b_2 \cdot (X_{2,Avg} - KD_1)/X_{1,Avg} - b_3 \cdot (X_{3,Avg} - KD_2)/X_{1,Avg} \quad (A2.4)$$

To solve in general the multiple linear regression with three segments in the PL model, one starts off with the following system of equations of sums of squares and sums of products (Sokal and Rohlf 1995):

$$\begin{aligned} b_1 \cdot \Sigma(x_1^2) + b_2 \cdot \Sigma(x_1 \cdot x_2) + b_3 \cdot \Sigma(x_1 \cdot x_3) &= \Sigma(x_1 \cdot y) \\ b_1 \cdot \Sigma(x_1 \cdot x_2) + b_2 \cdot \Sigma(x_2^2) + b_3 \cdot \Sigma(x_2 \cdot x_3) &= \Sigma(x_2 \cdot y) \\ \dots & \\ b_1 \cdot \Sigma(x_1 \cdot x_k) + b_2 \cdot \Sigma(x_2 \cdot x_k) &+ b_3 \cdot \Sigma(x_k \cdot x_3) = \Sigma(x_k \cdot y) \end{aligned} \quad (A2.5)$$

Without a fixed  $Y$ -intercept, and with given sums-of-squares and sums-of-products, the matrix is solved for  $b_1$ ,  $b_2$ , and  $b_3$ . With one or two segments, the procedure is the same, as the sums corresponding to  $b_2$  and  $b_3$  are simply set to zero.

Replacing in the system of Equations A2.5 the variable  $b_1$  with the function given in A2.4 and applying some calculus provides the final matrix for multiple linear regression with fixed  $Y$ -intercept  $a_1$ :

$$\begin{aligned} b_2 \cdot [\Sigma(x_1 \cdot x_2) - \Sigma(x_1^2) \cdot (X_{2,Avg} - KD_1)/X_{1,Avg}] &+ b_3 \cdot [\Sigma(x_1 \cdot x_3) - \Sigma(x_1^2) \cdot (X_{3,Avg} - KD_2)/X_{1,Avg}] \\ &= \Sigma(x_1 \cdot y) - \Sigma(x_1^2) \cdot (Y_{Avg} - a_1)/X_{1,Avg} \\ b_2 \cdot [\Sigma(x_2^2) - \Sigma(x_1 \cdot x_2) \cdot (X_{2,Avg} - KD_1)/X_{1,Avg}] &+ b_3 \cdot [\Sigma(x_2 \cdot x_3) - \Sigma(x_1 \cdot x_2) \cdot (X_{3,Avg} - KD_2)/X_{1,Avg}] \\ &= \Sigma(x_2 \cdot y) - \Sigma(x_1 \cdot x_2) \cdot (Y_{Avg} - a_1)/X_{1,Avg} \end{aligned} \quad (A2.6)$$

This matrix of two equations with two unknowns is solved for  $b_2$  and  $b_3$ , using a method for solving systems of simultaneous linear equations (most notably Gaussian elimination, see for example Gerald and Wheatley 1994). Subsequently,  $b_1$  is calculated with Equation A2.4. The degrees of freedom are augmented by one, compared to the case where initial relative growth is found by the regression from the data, resulting in  $n - 3$  in the case of three segments (corresponding to finding  $b_1$ ,  $b_2$ , and  $b_3$  from the data).

### APPENDIX 3: Numerical Evaluation of the Exponential Integral $Ei(x)$

The exponential integral  $Ei(x)$  is defined as follows (van de Laan and Temme 1984):

$$\text{For } x < 0: \quad Ei(x) = \int_{-\infty}^x [Exp(y)/y]dy \quad (A3.1)$$

$$\text{For } x > 0: \quad Ei(x) = \lim_{\epsilon \downarrow 0} \{ \int_{-\infty}^{-\epsilon} [Exp(y)/y]dy + \int_{\epsilon}^x [Exp(y)/y]dy \}$$

For  $y = 0$  the function  $f(y) = Exp(y)/y$  is singular, and  $Ei(0) = -\infty$ . Therefore, for  $x > 0$  one uses the Cauchy principal value to define  $Ei(x)$ . In this way, the to-be-integrated interval can include zero (for example:  $\int_{-2}^{+3} [Exp(y)/y]dy = Ei(3) - Ei(-2) = 9.983$ ).

For  $\int_{x_1}^{x_2} [Exp(c \cdot y)/y]dy$ , with  $c$  being a real number and applying the calculus rule of "change of variables" ( $c \cdot y = z$ ), we obtain:

$$\int_{x_1}^{x_2} [Exp(c \cdot y)/y]dy = Ei(c \cdot x_2) - Ei(c \cdot x_1) \quad \{\text{for } x_1 > 0, x_2 > 0\} \quad (A3.2)$$

All computational methods calculate the exponential integral as an approximation with a desired degree of accuracy (Hastings et al. 1955, Harris 1957, Miller and Hurst 1958, Gautschi and Cahill 1965, Cody and Thacher 1969, van der Laan and Temme 1984, Jeffrey 1995). Here we present a new method of numerical evaluation of  $Ei(x)$  that works particularly well for the new growth model of this article, because there are no constraints for choosing  $x$ , the desired accuracy can be chosen freely, and it can be calculated even with a spreadsheet program.

$Ei(x)$  can be expressed as a Taylor expansion (an infinite polynomial), where Euler's constant  $\gamma = 0.577,215,664,901,533$  appears (van de Laan and Temme 1984, Jeffrey 1995). The addition of  $\gamma$ , however, is irrelevant when calculating the integral between two integration bounds, because it cancels out. With  $n$  being any positive integer number:

$$Ei(x) = \gamma + Ln|x| + x^1/(1! \cdot 1) + \dots + x^n/(n! \cdot n) + \dots \quad \{x \neq 0\} \quad (A3.3)$$

Equation A3.3 can be reformulated when truncating the polynomial so that only  $i$  polynomial terms remain (where  $i$  is a positive integer number):

$$Ei(x) = \gamma + Ln|x| + x^1/(1! \cdot 1) + \dots + x^i/(i! \cdot i) + Remainder(x,i) \quad (A3.4)$$

By deriving a bound formula for the  $Remainder(x,i)$ , a criterion is provided for the number of terms  $i$  that have to be added for a desired accuracy of the numerical result:

$$Remainder(x,i) = x^{i+1}/[(i+1)! \cdot (i+1)] + \dots + x^{i+1+n}/[(i+1+n)! \cdot (i+1+n)] + \dots \quad (A3.5)$$

$$Remainder(x,i) < \{ x^{i+1}/[(i+1)! \cdot (i+1)] \cdot \{ 1 + x/(i+2) + x^2/[(i+3) \cdot (i+2)] + \dots + x^n/[(i+1+n)!/(i+1)!] + \dots \}$$



Applying the formula for a geometric series:

$$\sum_{n=0}^{\infty} [x/(i+2)]^n = 1/[1 - x/(i+2)]$$

$$\{\text{for } |x/(i+2) < 1\}$$

The latter condition is equivalent to  $i > (|x| - 2)$  polynomial terms.

$$\text{Remainder}(x,i) < \{x^{i+1}/[(i+1)! \cdot (i+1)]\}$$

$$/[1 - x/(i+2)]$$

$$\text{Remainder}(x,i) < x^{i+1}/\{i! \cdot (i+1)^2\}$$

$$\cdot [1 - x/(i+2)] \quad (\text{A3.6})$$

Inequality A3.6 works for positive as well as negative  $x$ . Note that in the case of negative  $x$ , the remainder is positive for uneven  $i$  and negative for even  $i$ . As an example, take  $x = -5.2$  and a desired accuracy of nine decimals. The smallest possible  $i > |x| - 2$  is four polynomial terms, but

the remainder as calculated with Equation A3.6 is still large (between  $-3.4$  and  $0$ ). With increasing  $i$ , the remainder converges toward zero. With  $i = 25$  polynomial terms, the exponential integral is calculated with Equation A.3.2 as  $Ei(-5.2) \approx \gamma + Ln|-5.2| + (-5.2)^1/(1! \cdot 1) + \dots + (-5.2)^{25}/(25! \cdot 25) = -0.000,908,621,9$ , and the remainder is between  $0$  and  $3 \cdot 10^{-10}$ . Consequently the true value of  $Ei(-5.2)$  is between  $-0.000,908,621,9$  and  $-0.000,908,621,6$  ( $-0.000,908,621,9 + 3 \cdot 10^{-10}$ ). Rounding one decimal results for both cases in  $-0.000,908,622$ , i.e., the desired accuracy of nine decimals has been reached.

This method works for any  $|x| \neq 0$ . For very large  $x$ , however, the condition  $i > x - 2$  makes the method slow or impractical. For negative values of  $x$ ,  $Ei(x)$  approaches zero. When the method is employed with the accuracy of a spreadsheet program, one may encounter rounding error problems for  $x < -15$ , in which case one can set  $Ei(x) \approx 0$ . These problems, however, are irrelevant for data from real-world trees in our growth model.