

Crown age estimation of a monocotyledonous tree species *Dracaena cinnabari* using logistic regression

Radim Adolt · Hana Habrova · Petr Madera

Received: 28 February 2011 / Revised: 19 November 2011 / Accepted: 20 February 2012 / Published online: 29 March 2012
© Springer-Verlag 2012

Abstract Unique woodlands of *Dracaena cinnabari* (DC) are at risk throughout most of their range (Socotra Island, Yemen) as a result of missing regeneration and overmaturity. Effective conservation measures depend on reliable predictions of future population dynamics, which depend on accurate data on current age structure. However, age determination of *Dracaena* sp. has long been a scientific challenge, because the common method of tree ring counts cannot be applied to this or to most other monocotyledonous trees. In the present study, the indirect method for crown age estimation proposed by Adolt and Pavlis (Trees 18:43–53, 2004) was further developed using a more appropriate statistical technique and an intuitive model formulation. This new technique is based on the relationship between the number of branching orders and the number of flowering events that result from a specific growth pattern. We used logistic regression to directly model annual flowering probability, the reciprocal value of which corresponds to the length of the interval between flowering events. Our methodology was applied to data sets collected at two ecologically distinct sites. In Firmihin, the

time between flowering events decreases from 28 years between the first and second event to 10 years between the 25th and 26th event. The length of time between flower events in Skant, however, was estimated to be a constant value of 6.5 years. We propose the application of generalised mixed-effects models and methods of survey sampling to improve the accuracy of crown age estimation in DC. Our methodology may also be useful for age estimations of other tree species with similar growth patterns, such as *Dracaena draco* and *Aloe dichotoma*.

Keywords *Dracaena cinnabari* · Tree age · Age estimation · Socotra · Flowering probability · Generalised linear model · Generalised linear mixed model · Logistic regression

Introduction

Populations of endemic *Dracaena cinnabari* (DC) trees on Socotra Island, Yemen, are among the most endangered in the world, which is likely a consequence of inadequate regeneration in most stands. Several areas on the island are extensively forested with these monocotyledonous trees; however, over much of the original range of this species, only remnant populations remain. Without accurate information on the age structure of DC, reliable estimates of future population dynamics are impossible. Nonetheless, such estimates are necessary in the development of effective habitat and species conservation strategies.

The genus *Dracaena* is exceptional among monocotyledonous plants, because of the secondary thickening of its stems and roots and the tree-like physiognomy of various species (Razdorskij 1954). Adolt and Pavlis (2004) reviewed sources on *Dracaena* sp. including taxonomy,

Communicated by A. Braeuning.

R. Adolt (✉)
NFI Methodology and Analysis,
Forest Management Institute Brandys nad Labem,
Namesti Miru 497, 767 01 Kromeriz, Czech Republic
e-mail: adolt.radim@uhul.cz

H. Habrova · P. Madera
Faculty of Forestry and Wood Technology,
Mendel University in Brno, Zemedelska 3,
613 00 Brno, Czech Republic
e-mail: xhabrova@mendelu.cz

P. Madera
e-mail: xmadera@mendelu.cz

physiognomy and former and current occurrence of several tree species.

Age estimations of monocotyledonous trees usually cannot be based on tree ring counts, such as those typically used with temperate or boreal species. In monocotyledonous trees, either the tree rings are not developed or they do not correspond to annual growth patterns (Razdorskij 1954). According to Worbes (2002), seasonal radial growth and the development of associated annual growth rings have been confirmed in numerous tropical tree species, but only for one monocotyledonous species (*Echinodorus andreuxii*).

Few analyses dedicated to age estimation in monocotyledonous trees have been published, Symon (1974) and Magdefrau (1975) published growth data from a long-term direct observation of *D. draco* in an artificial environment. However, the information presented in these papers is only descriptive, as no statistically sound method has been proposed regarding the estimation of crown or tree age for other species and populations.

Recently, an analytical approach of age estimation for two *Cecropia* species was used. Heuret et al. (2002) described synchronicity between flowering and branching events of a tropical dicotyledonous tree *Cecropia obtusa*. They found that subsequent branch tiers occurred with high probability every 35th node. The same frequency of occurrence but with shifted phase was observed in case of scars after inflorescences developed. Assuming flowering takes place once a year, the frequency of branch tier creation is also yearly. This information can be used to determine the age of particular trees. Zalamea et al. (2008) came to practically identical findings for *Cecropia sciadophylla*. The only difference was that branch tiers and inflorescence scars occurred for every 25 or 23 nodes depending on the location of study. The results of the previous two publications cannot be transferred to age estimation of *Dracaena cinnabari* because of the different growth of this species.

In the case of *Dracaena* sp., the creation of new branch segments is induced by flowering events. In addition, only a small portion of terminal buds flowers each year (Zimmermann and Tomlinson 1969).

Indirect methods of age estimation using age–diameter curves were tested for temperate European tree species *Fagus sylvatica* and *Quercus robur* (Rozas 2003). However in a given forest stand, individuals with quite different growth rates often coexist (Stewart 1986), so a considerable variation in the age of trees belonging to the same size class is usual (Norton and Ogden 1990; Lusk and Smith 1998). Each point in the age–diameter scatter diagram is surrounded by an area of uncertainty which prevents the parametrisation of faithful predictive models (Ogden 1985).

Rozendaal and Zuidema (2011) emphasise the effect of temporal autocorrelation in tree growth rates (fast-growing trees keep growing fast for a longer time span, and, analogously, slow-growing trees tend to continue growing slowly), which leads to an increased variation of age at a given tree dimension. With reference to Brienens and Zuidema (2006a, b), the authors report a high temporal autocorrelation of growth rates in dry tropical forests. This should hold true even for small trees, which in normal conditions experience several periods of suppression and release resulting in weak temporal autocorrelation. In dry tropical forests, different growth rates among trees most likely reflect spatial patterns of water availability (Brienens et al. 2010).

According to Rozendaal and Zuidema (2011), the well-known ^{14}C method is not suitable for age determination in live trees as the atmospheric concentration of ^{14}C had several peaks between 1650 and 1940, making a correct age estimation difficult.

To our knowledge, the first attempt at an analytical approach to age determination in *Dracaena* sp. was described by Adolt and Pavlis (2004). This age determination technique is based on the fact that, in *Dracaena* sp., the growth of a branch segment is interrupted by a flowering event, after which one or more new branches are formed (Zimmermann and Tomlinson 1969). Thus, the number of branching orders, i.e. the number of branch segments between a terminal bud and the base of crown (the top of stem) is related to the number of flowering events. Crown age can be estimated as the total duration of the estimated time intervals between all flowering events. The present study is based on this original concept, and the objective is to carry the technique a step further by applying it to an extended data set from two ecologically different localities—Firmihin Plateau and Skant. We have emphasised the use of a statistical approach that is appropriate based on the probability distribution of the data.

Adolt and Pavlis (2004) used ordinary least squares (OLS) regression to parameterise the necessary model. The use of OLS regression is justified only if the conditional probability distribution of the response variable is normal. This requirement is not satisfied by our data. As a consequence, statistical inference may be misleading, and estimated crown age may be biased. From a statistical standpoint, a generalised linear model adapted to a particular probability distribution can provide a more accurate estimate. Within this framework, new, easy-to-use and versatile equations for estimating crown age can be theoretically determined.

Our technique focuses on an estimation of crown age only. Tree age is obtained as the sum of crown age and the stem age when the tree flowers for the first time. In stands where nearly all of the trees have a developed crown and

no regeneration is present or expected; an unknown stem age is not a limitation to predicting population development patterns. According to our own observations and personal communication with colleagues, nearly all DC populations on Socotra have developed crowns and lack regeneration, with the exception of a few hardly accessible stands in the Hagher Mountains.

Materials and methods

Study area

Socotra Island is situated between 12°19'–12°42'N latitude and 53°18'–54°32'E longitude in a dry tropical climatic zone with two monsoon periods (Habrova 2007; Scholte and De Geest 2010). Climatological characteristics of selected localities are included in Table 1. Populations of DC are extant only on the eastern half of the island (Kral and Pavlis 2006). This species generally ranges from 300 to 1,500 m a.s.l., preferring a limestone-based subsoil. The majority of these trees grow in the third altitudinal vegetation zone, where they are a dominant species of the tree layer (Habrova 2004). The largest populations are found in the Hagher Mountains, in Firmihin and in Hamadero. Figure 1 shows a DC-dominated woodland in the northern part of Firmihin.

Field data

Field data were collected in 2001, 2003 and 2004 in Firmihin and Skant. The numbers of measured sample plots, trees and analysed branches are shown in Table 2. The data set collected in 2001 had been previously used by Adolt and Pavlis (2004) to derive the first model for age estimation of DC. Sample trees were measured on subjectively located plots of irregular shape and variable total area. No probability sampling, in the sense that it is described in the survey sampling literature, was used. From each sample tree, the following variables were obtained: (1) total tree height, (2) the height of the stem (not

Table 1 Climatic conditions of Skant and Firmihin: Socotra Island (Habrova 2007)

Locality	Skant	Firmihin
Geographic coordinates	54°02'E 12°35'N	54°01'E 12°28'N
Altitude a. s. l. (m)	1,450	440
Mean annual temperature (°C)	17.9	23.4
Maximum temperature (°C)	32.0	36.2
Minimum temperature (°C)	8.2	14.35
Mean annual humidity (%)	80.0	71.8



Fig. 1 Woodland dominated by *D. cinnabari* (umbrella-shaped trees) in the northern part of Firmihin

Table 2 The number of sample trees and plots by location and measurement year

Locality	Measurement year	Number of plots	Number of sample trees	Number of branches
Firmihin	2001	1	31	1,869
Firmihin	2003	1	144	3,212
Skant	2003	2	73	1,976
Firmihin	2004	1	71	2,192

measured in 2001), (3) diameter at breast height, (4) the number of terminal buds on the crown periphery, (5) the number of inflorescences and (6) the mean number of branching orders on these terminal buds.

Modelling yearly flowering probability by logistic regression

A generalised linear model (GLM) extends the ordinary linear regression model to encompass a non-normal probability distribution of a response variable (Agresti 2002). This is achieved by a transformation of the response variable via a link function, which is used to relate a linear combination of explanatory variables, called linear predictors, to the response variable. For a binomial response variable, the most common GLM variant is the logistic regression.

Using logistic regression, we can model the flowering probability π_i for a terminal bud of tree i in a direct way. In this framework, the number of blooming buds on the crown periphery of a sample tree i is considered to be a binomial random variable $Y_i \sim \mathcal{B}(n_i, \pi_i)$, where n_i is the total number of buds on the crown periphery of tree i . The observed ratios of blooming terminal buds (plus the number of fruit panicles or their residues from a given year) to the total number of terminal buds on the crown periphery are used as blooming probability estimates for particular sample

trees. Logistic regression uses a specific link function—the logit transformation of the response:

$$l_{\pi_i} = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (1)$$

This is then modelled by a linear function:

$$l_{\pi_i} = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_k \cdot X_{ik}, \quad (2)$$

where $\beta_0 \dots \beta_k$ are the model parameters to be estimated, and $X_{i1} \dots X_{ik}$ are the values of k explanatory variables for tree i . An explanatory variable can also be categorical, with $k + 1$ possible values. In that case, $X_{i1} \dots X_{ik}$ are k indicator or ‘dummy’ variables with 0 or 1 values. An example of such a variable is the locality or site category. The estimated yearly flowering probability is expressed by a logistic function:

$$\pi_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_k \cdot X_{ik})}} \quad (3)$$

recalling the name of the technique itself.

Estimation of crown age

The time to the next flowering event on a particular terminal δ_{t_p} corresponds to the reciprocal value of a probability π_p that a terminal bud on a crown with p branching orders will bloom in a given year:

$$\delta_{t_p} = \frac{1}{\pi_p} \quad (4)$$

Exactly, one flowering event corresponds to a one unit increase in the mean number of branching orders. Provided the flowering probabilities are known for each moment during the crown’s existence, we can sum the lengths δ_{t_p} of the intervals between flowering events to determine the actual age of the crown:

$$\hat{t}_c = \sum_{p=1}^{\text{int}(P)} \frac{1}{\pi_p} + \frac{P - \text{int}(P)}{\pi_{[\text{int}(P)+1]}}, \quad (5)$$

where P is the mean number of branching orders -0.5 . Subtracting 0.5 is necessary, because it is not the number of branching orders but the number of swollen nodes between branch segments that corresponds to the number of flowering events. For example, a tree with only one branch segment, the crown of which has not flowered yet, has flowered just once, with the inflorescence occurring on its stem. We do not know exactly how long it will take until the next flowering event on a particular branch segment. Thus, we approximate the number of flowering events as 0.5 because other alternatives, such as 0 or 1, would result in biased estimates of crown age. The number of branching orders for an actively flowering terminal bud is evaluated as the number of branch segments on the path below it $+0.5$.

The mean number of branching orders can be calculated over a total of either peripheral branches or all branches of a given tree. The decision on which of these totals to use must correspond with the choice of explanatory variables during the model-fitting stage. The term $\text{int}(P)$ represents an integer of P . The second part of the formula refers to a proportion of the time interval to the next flowering event corresponding to a fraction of P (e.g. for $P = 12.6$ it is 0.6 times the length of the interval between the 12th and 13th flowering event).

Estimates of crown age using this method must be based on explanatory variables, the values of which can be deduced for any given point of time during the crown’s existence. It is not necessary for values of these variables to vary over time; a constant value fulfills this condition as well. There are two variables with such properties. Intuitively, the most promising is the number of branching orders. The second potentially useful variable is stem length, which remains constant after a tree has flowered for the first time. Formula (5) represents a discrete approximation of a more general relationship, which is based on a definite integral (7) of a function (6) describing the length of the interval between two consequent flowering events.

$$\hat{\delta}_{t_p} = 1 + e^{-(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2)} \quad (6)$$

This function is reciprocal to a function describing yearly flowering probability. Only two relevant explanatory variables are considered: the mean number of branching orders (X_1) and the stem length (X_2). Values of β_0 , β_1 and β_2 are estimated using logistic regression (see “[Modelling yearly flowering probability by logistic regression](#)” and “[Parametrisation of models and hypothesis testing](#)” for details). Solving the definite integral (7)

$$\hat{t}_c = \int_1^P \left[1 + e^{-(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2)} \right] dX_1 + \left[1 + e^{-(\beta_0 + \beta_1 + \beta_2 \cdot X_2)} \right] \quad (7)$$

and after a necessary algebraic manipulation, we obtain (8) as an estimate of crown age for trees with a mean number of branching orders of ≥ 1.5 . The second term, after the integral itself, is an estimate of the length of time between the tree’s first flowering event (the inflorescence formed at the top of the stem) and the time at which the mean number of branching segments reaches 1.5 (with a P value of 1).

$$\hat{t}_c = P + e^{-(\beta_0 + \beta_1 + \beta_2 \cdot X_2)} \left[\frac{e^{-\beta_1 \cdot (P-1)} - 1}{-\beta_1} + 1 \right] \quad (8)$$

Equation (9) is applicable for trees with a mean number of branching orders of ≥ 1 but ≤ 1.5 (with a corresponding P value of between 0.5 and 1).

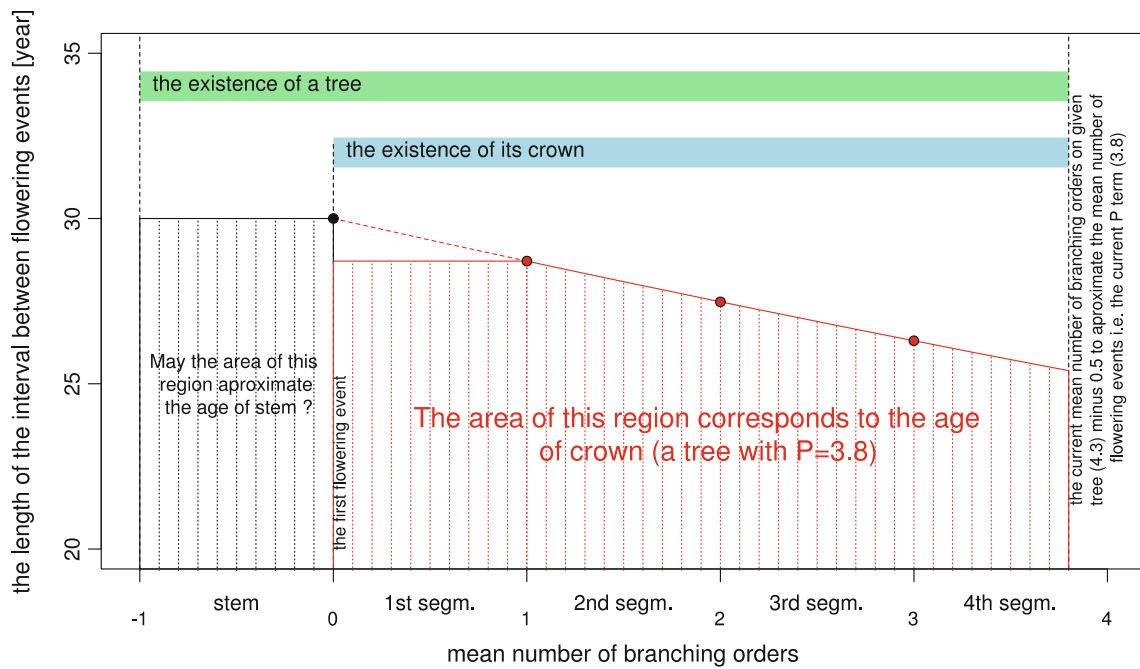


Fig. 2 A graph showing an estimation of the crown age of a tree with a mean number of branching orders equal to 4.3. To construct this plot, parameters of the *firmihin.03q* model were used. A detailed explanation of the technique is included in “[Estimation of crown age](#)”

$$\hat{i}_c = P \cdot \left[1 + e^{-(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2)} \right] \tag{9}$$

If the interval between flowering events does not depend on the mean number of branching orders, i.e. the parameter β_1 is not significantly different from zero, we must use Eq. (10) instead of (8) and (9):

$$\hat{i}_c = P \cdot \left[1 + e^{-(\beta_0 + \beta_2 \cdot X_2)} \right] \tag{10}$$

If the stem age is not considered significant in a particular model, the corresponding terms (the product $\beta_2 \cdot X_2$) can be easily left out of (6), (7), (8), (9) and (10).

Notice that the lower integration limit in (7) equals one instead of zero as might be expected. For theoretical reasons, no trees without a crown were included in the analysis. The probability of observing an inflorescence on these trees cannot be used to describe their growth rates; instead, this value reflects the actual age structure in a population of such trees and is greatly influenced by regeneration and mortality processes. Evaluating flowering probability for a tree with zero branching orders, i.e. for a tree without a crown, is a pure extrapolation. One should not extrapolate the model below one if there is no additional information supporting such an approach. In addition, the P term representing the mean number of branching orders -0.5 cannot be <0.5 , because the mean number of branching orders never drops below 1. This holds true at least for our sample trees and for trees whose crown age can be reasonably predicted using this method. If this constraint is discarded, stem age may also be evaluated. However,

because of a lack of theoretical justifications, such an approach cannot guarantee unbiased tree age estimates. A graphical representation of the age determination principle is shown in Fig. 2.

Parametrisation of models and hypothesis testing

The search for suitable models was driven by an educated guess. We did not apply stepwise regression or any similar technique of automatic variable selection. Because there were few variables of potential use, we could easily parameterise and test all reasonable model configurations. The significance of individual predictors was tested with a classical Wald test (W test) and a likelihood-ratio test (LR test). In cases of contradictory results, preference was given to the LR test because it uses more information than the W test and is therefore more versatile. According to Agresti (2002), the LR test is more reliable for small and moderate sample sizes. For large sample sizes, the properties of these two tests are asymptotically equivalent.

Special attention was paid to overdispersion diagnostics. The term overdispersion indicates a higher variability in the ratios of flowering buds than would be expected in a binomial distribution. Whenever the null deviance was markedly higher than the number of residual degrees of freedom, the model was refitted using a quasibinomial family rather than a binomial family of link functions. In such models, LR tests with F , rather than χ^2 , distribution are recommended (Smilauer 2007).

We used an analysis of statistical power to infer the possible effect size (a measure of the practical importance of an explanatory variable) of predictors that were not found to be statistically significant. Statistical power analyses can be helpful in determining whether different sets of significant predictors are caused by chance only (a type II error) or reflect true differences between particular populations. Details regarding the calculation of statistical power are included in “[Calculation of statistical power](#)”.

Statistical analyses were conducted using the R statistical software (R Development Core Team 2009). Logistic regression models were fitted using the ‘glm’ function with the logit transformation. Likelihood-ratio tests were obtained using the ‘anova’ function with the distribution parameter set to F or χ^2 . The analysis of statistical power was conducted using our self-coded functions implementing equations from “[Calculation of statistical power](#)”.

Confidence intervals for crown age predictions

Variance of any linear combination of estimated model parameters $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ can be analytically expressed provided the covariance matrix of the parameter estimates is known (Sarndal et al. 2003, chapter 5.4, page 169). Unfortunately, the functions for crown age prediction as derived in “[Estimation of crown age](#)” are non-linear with respect to parameter estimates. Consequently, the variance of age prediction must be estimated either by linear approximation (Sarndal et al. 2003, chapter 5.5, p. 172) or through simulations.

In this analysis, we simulated 10,000 estimates of the true parameter vector for each of the resulting models (*firmihin.03q* and *skant.03q*). Parameter estimates were drawn by random selections from multivariate normal distributions using estimated model parameters and their covariance matrices. The simulations were implemented using the MASS package available in R (Venables and Ripley 2002). Crown age was evaluated for each simulated vector of parameters with a mean number of branching orders of between 1 and 30. Equations (8), (9) and (10) were used selectively depending on the locality, the mean number of branching orders and the set of significant predictors. For a given mean number of branching orders, the resulting confidence intervals were obtained as the corresponding $1 - \alpha/2$ (two-sided confidence intervals) quantiles of the simulated distribution of the age estimates.

Calculation of statistical power

Statistical tests can generate two types of errors. A type I error occurs when a null hypothesis is rejected, but is in fact true. The probability of this error occurring is

evaluated within the testing procedure itself as the well-known P value. This probability is then compared to the value of α representing the maximum acceptable probability of a type I error. Whenever P is smaller than α , we reject the null hypothesis that a parameter is not significant. A type II error occurs when a null hypothesis is accepted when it is in fact false. The probability of this error, termed β , cannot be evaluated in a general case because it depends on a true, but unknown, alternative hypothesis. Finally, the statistical power evaluated as $1 - \beta$ represents the probability of rejecting a false null hypothesis. With respect to parameter testing, the statistical power is the probability of detecting the true significance of an explanatory variable. For a more comprehensive and general treatment of this topic, see Tabachnick and Fidell (2007).

Agresti (2002) provides the equations for determining the necessary sample size in one-sided tests as a function of the required power, effect size and α . A simple algebraic manipulation leads to an explicit expression of statistical power:

$$1 - \beta = \Phi(Z_\beta) \quad (11)$$

$$Z_\beta = \left[\sqrt{\frac{n\hat{\pi}\tau^2}{(1 + 2\hat{\pi}\delta) \cdot (1 - R^2)}} - Z_\alpha \right] e^{\frac{\delta}{4}} \quad (12)$$

$$\delta = \left[1 + (1 + \tau^2)e^{\frac{\delta}{4}} \right] / \left[1 + e^{\frac{\delta}{4}} \right] \quad (13)$$

$$\tau = \ln \left\{ \frac{\pi_{(\bar{x}+S_{x_k})}[1 - \pi_{\bar{x}}]}{[1 - \pi_{(\bar{x}+S_{x_k})}]\pi_{\bar{x}}} \right\} \quad (14)$$

The term $1 - \beta$ in (11) is the resulting statistical power. Z_α and Z_β are quantiles of the standardised normal distribution for the probabilities of type I and type II errors, respectively. The sample size is denoted by n , and it corresponds to the total number of analysed branches. $\hat{\pi} = \pi_{\bar{x}}$ is the flowering probability for \bar{x} , i.e. the mean vector of explanatory variables. $\pi_{(\bar{x}+S_{x_k})}$ is the flowering probability for the mean vector of explanatory variables excluding variable x_k , which is set to its mean plus one standard deviation. τ is the natural logarithm of the ratio of the odds corresponding to $\pi_{(\bar{x}+S_{x_k})}$ and $\pi_{\bar{x}}$ probabilities, and it represents a measure of effect size. Φ represents the normal cumulative probability distribution function. R is the multiple correlation coefficient between the x_k and the remaining explanatory variables, and δ is a term used to make the formulas in Eq. (12) simpler. If the model includes only one explanatory variable, the term $(1 - R^2)$ in the denominator within Eq. (12) must be left out.

Our analysis was focused on Δ_{δ_p} , the difference between the lengths of the interval between flowering events for the mean vector of explanatory variables \bar{x} and for $\bar{x} + S_{x_k}$:

$$\Delta_{\delta_{ip}} = \frac{1}{\pi_{\bar{x}}} - \frac{1}{\pi_{(\bar{x} + S_{x_k})}} \tag{15}$$

For our purposes, $\Delta_{\delta_{ip}}$ is a more informative measure of effect size than τ , as defined in (14). To evaluate the statistical power, we express τ using $\Delta_{\delta_{ip}}$ and $\pi_{\bar{x}}$:

$$\tau = \ln \left[\frac{1 - \pi_{\bar{x}}}{1 - \pi_{\bar{x}}(\Delta_{\delta_{ip}} - 1)} \right] \tag{16}$$

According to Agresti (2002), the above outlined formulas for power calculations are valid provided the unconditional distribution of explanatory variables is normal. We also assume that the models do not contain overdispersion, and observations should not be correlated. The obvious, and to this data inherent, autocorrelation decreases the number of independent observations, i.e. the effective sample size n in Eq. (12) does not exactly correspond to the number of analysed branches (for more details, see “An extension using generalised mixed models”). Overdispersion was present in all fitted models. Under such circumstances, the true power is expected to be lower than calculated.

Results

The logistic model for the 2001 data set

This model was fitted to the data set already used by Adolt and Pavlis (2004) to derive the first model for crown age estimation of DC. The properties of all resulting models are summarised in Table 3. Parameters β_0 and β_1 correspond to the intercept and linear terms, respectively, for the mean number of branching orders. Standard errors of parameter estimates are abbreviated as σ_{β_0} and σ_{β_1} . The terms P_{β_0} and P_{β_1} are the P values in the Wald significance tests, and Φ is an estimated dispersion parameter—a factor by which the observed variability of the ratio of blooming terminal buds is greater than expected for a binomial random variable

Table 3 Parameters and characteristics of the resulting models

Model	<i>firmihin.00q</i>	<i>firmihin.03q</i>	<i>skant.03q</i>
Locality	Firmihin	Firmihin	Skant
Measurement year(s)	2001	2001, 2003, 2004	2003
β_0	-2.8521	-3.3674	-1.711
β_1	NA	0.0455	NA
σ_{β_0}	0.1795	0.2338	0.09647
σ_{β_1}	NA	0.0157	NA
P_{β_0}	3.78e-16	2e-16	2e-16
P_{β_1}	NA	0.0041	NA
Φ	3.108	2.890	2.372

(a present overdispersion, see “Parametrisation of models and hypothesis testing” for details).

The mean number of branching orders was not found to be a significant explanatory variable (LR test, $F = 0.0314$, $P = 0.8605$).

Stem length was not tested as a possible predictor of flowering probability because it was not available in the data set. Substituting the resulting estimate for an intercept term into Eq. (6), we arrive at 18.3 years as an estimate of the interval between flowering events with 95% confidence limits from 13.7 to 27.5 years.

The logistic models for the two locations

The resulting models for the Firmihin and Skant locations contain different sets of explanatory variables. For complete information regarding model parameters, see Table 3. The concordance between the modelled and observed ratios of flowering terminals is shown in Fig. 3. The graph also shows the observed ratios of blooming buds (blue triangles and red squares) for sample trees grouped by their numbers of branching orders; numeric labels indicate the number of trees in each group.

In Firmihin, the yearly flowering probability increases with the mean number of branching orders (LR test, $F = 8.4344$, $P = 0.0040$). However, in Skant, this explanatory variable was not found to be significant (LR test, $F = 1.2628$, $P = 0.2649$).

Stem length is not a significant predictor of flowering probability in Firmihin (LR test, $F = 0.8162$, $P = 0.3673$). This was also the case in Skant, but the P value in the particular LR test ($F = 2.9899$, $P = 0.0881$) approaches the conventional significance level ($\alpha = 0.05$).

In Firmihin, the interval between flowering events decreases from approximately 28 years between the first and second event to approximately 10 years between the 25th and 26th flowering event. However, in Skant, this

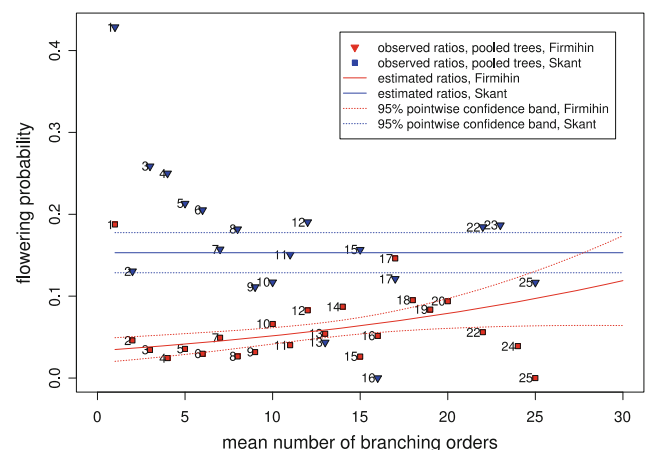


Fig. 3 Flowering probability, *firmihin.03q*, *skant.03q* models

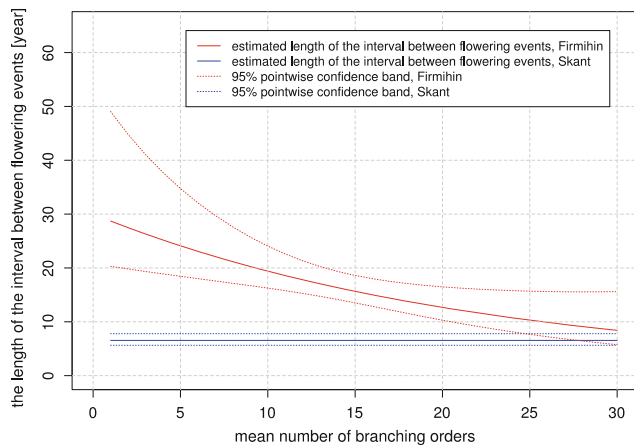


Fig. 4 The length of the interval between flowering events, *firmin.03q*, *skant.03q* models

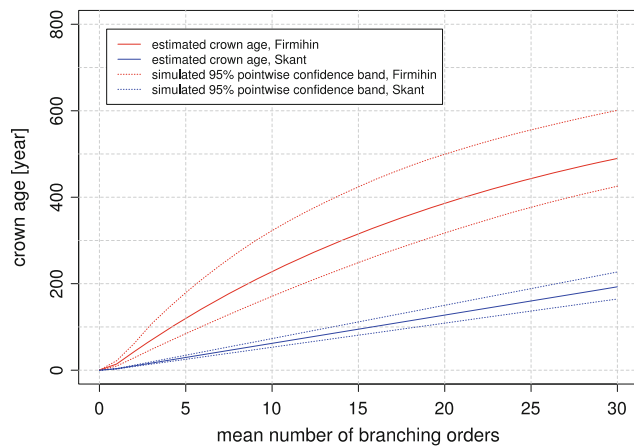


Fig. 5 Crown age, *firmin.03q*, *skant.03q* models

interval has a constant value of 6.5 years. Figures 4 and 5 show the predictions of the intervals between flowering events and the current crown age for the two localities. Table 4 shows the estimated crown age in years and the associated 95% confidence intervals as a function of the mean number of branching orders.

Discussion

Interpretation and comparison of the resulting models

Using the data set from Firmihin collected in 2001, Adolt and Pavlis (2004) established an expected length of the interval between flowering events of 18.7 years with a 95% confidence interval from 13.7 to 29.6 years. They concluded that flowering probability is independent of the number of branching orders. Our results, applying a new methodology to the same data, are in close agreement with this earlier analysis. However, results for the same locality using a

Table 4 Crown age predictions and their simulated confidence intervals

Model Locality	<i>firmin.03q</i>			<i>skant.03q</i>		
	Firmihin	Lower	Upper	Skant	Lower	Upper
X_1	Fit	Lower	Upper	Fit	Lower	Upper
1	14	10	22	3	3	4
2	41	28	61	10	8	12
3	70	48	106	16	14	19
4	97	66	144	23	20	27
5	122	84	179	29	25	35
6	146	102	212	36	31	42
7	169	120	243	42	36	50
8	191	137	272	49	42	58
9	212	154	298	56	47	65
10	233	171	323	62	53	73
11	252	187	346	69	59	81
12	271	203	367	75	64	88
13	288	219	387	82	70	96
14	305	234	406	88	75	104
15	322	249	424	95	81	112
16	337	264	442	101	86	119
17	352	277	457	108	92	127
18	367	291	472	114	98	135
19	381	304	487	121	103	142
20	394	317	499	127	109	150
21	406	330	512	134	114	158
22	419	342	524	140	120	166
23	430	354	535	147	126	173
24	442	366	546	154	131	181
25	452	377	556	160	137	189
26	462	387	565	167	142	196
27	472	397	575	173	148	204
28	482	407	584	180	154	212
29	491	416	592	186	159	219
30	500	426	601	193	165	227

X_1 represents the mean number of branching orders

pooled data set collected in 2001, 2003 and 2004 show that the interval between flowering events decreases with the mean number of branching orders. This does not necessarily contradict the results found by Adolt and Pavlis (2004). The interval between flowering events may in fact decrease with the increasing age of a tree even if a significance test did not reject the null hypothesis. The power of the LR test for an extended *firmin.00q* model lies below 0.75 (for $\alpha = 0.05$ and $\Delta_{\delta_{ip}} = 3.9$ years; see the intersection of horizontal and vertical red dashed lines in Fig. 7). Hence, the probability of a type II error is still relatively high, and no firm conclusions can be drawn as to whether the mean number of branching orders has no effect on flowering probability. Whereas in Firmihin the mean number of branching orders was found to

be a significant predictor, in Skant this parameter was not significant. The graph depicting the statistical power of the LR test for the extended Skant model (graph 8) shows that for $\alpha = 0.05$ and effect sizes $\Delta_{\delta_{ip}}$ above 1.2 years, the power exceeds 0.95. Thus, in Skant, no effect size of either predictor (the mean number of branching orders or stem length) is expected to exceed 1.2 years.

Stem length was not found to be a significant predictor in Firmihin either. In this case, however, the power exceeds 0.95 for effect sizes above 1.8 years (see graph 9). Therefore, an effect below this rather elevated value may be present. Unfortunately, the values of statistical power are likely lower than those calculated as a result of the non-normal distribution of the number of branching orders and stem lengths, overdispersion in the models and autocorrelation among branches of the same tree. Hence, the effect sizes of both the above discussed predictors are likely higher at the 0.95 power threshold.

The selection of sample plot locations and their size was subjective, so the data are representative of the particular plots only. Strictly speaking, no generalisation to any population except the plots themselves is possible. In this sense, we cannot compare the resulting models for the two different occasions or localities, because the data were not collected on the same plots and trees. In Firmihin, we found that the length of the interval between flowering events decreases from approximately 28 years between the first and second event to approximately 10 years between the 25th and 26th event. However, in Skant, this interval remains constant (6.5 years) or increases slowly from approximately 5–8 years between the first and 25th flowering occasion.

Marked differences in the flowering intervals of each of the two localities may be induced by very different site conditions. In addition, the two populations may represent two different taxa, which could be anticipated based on the

very different growth habits of the trees present at each site. Unfortunately, these speculations cannot be proven by our analysis and with the available data.

The crown age histograms constructed from our sample trees differ in shape as well as mean (Fig. 6). In Skant, the histogram shows a pattern very common in biological populations, i.e. a left-skewed distribution with a prevalence of young individuals. The histogram for Firmihin indicates roughly equal percentages of younger and older trees. However, this symmetry (in line with direct observations at this locality) indicates the absence of regeneration.

Habrova et al. (2009) predicted the number of DC trees that would develop at a location in Skant during the next 100 years based on the constant flowering interval of 19 years as it was published by Adolt and Pavlis (2004). As a result, the findings of these authors may change considering the potentially much shorter length of the interval between flowering events in Skant arrived at in the present study.

Hubalkova (2011) conducted an analysis similar to the one of Habrova et al. (2009). Hubalkova used the same, constant length of the flowering interval (19 years) to predict the future development of DC stand on a study site in Firmihin. Her results may also change due to the non-constant length of the flowering interval in Firmihin, as proposed by our current study.

None of the above-mentioned differences between the two sites can be reliably explained until the models are parameterised using data collected according to survey sampling principles and improved by following the recommendations presented here.

An extension using generalised mixed models

In our data set there are two obvious levels of data clustering: locations with several sample plots and sample plots with several sample trees. Less obvious, but very important clusters, are made up of individual trees that contain several terminal buds within their crowns. The number of flowering terminal buds as a modelled binomial variable, $Y_i \sim \mathcal{B}(n_i, \pi_i)$, is defined as the sum of Bernoulli trials (Agresti 2002) with 0 (no flowers or fruits) or 1 (flowers or fruits present) possible outcome. Bernoulli variables should not be correlated when a binomial distribution of their sums is considered. This circumstance does not occur in the present case, because the presence of flowers on the terminal buds of a tree depends on factors inherent to the whole tree. For example, among DC, individual trees tend to have either several blooming terminal buds or only a few (or none), which typically depends on a given tree's flowering history in previous years. Because our analysis ignored the clustered nature of the data, a positive correlation can also be expected for estimation errors. For example, in a particular locality, the errors of crown age

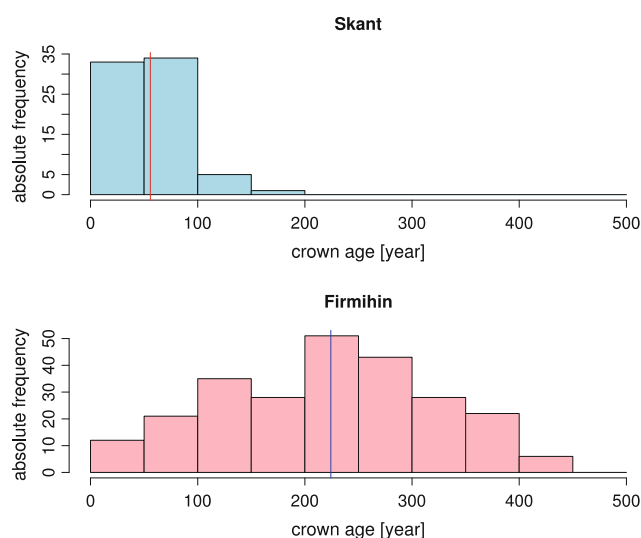


Fig. 6 Crown age histograms constructed using sample trees from Firmihin (pink) and Skant (blue)

estimation for individual trees would be predominantly positive or predominantly negative, i.e. crown age would be systematically overestimated or underestimated. This is the case with an OLS regression model, which only includes parameters describing the effect of a particular predictor (e.g. the number of branching orders) on the values of a response variable (e.g. the probability of blooming in a given year). These parameters, called fixed effects, are considered constant for all observations regardless of which cluster, i.e. locality, plot or tree, they refer to.

One possible solution is to include cluster-specific terms, i.e. taking the random effects into account in the analysis. Random effects are unobserved and take the same value for each observation within a cluster but different values for different clusters. Because the resulting models contain fixed as well as random effects, they are called mixed-effects models. A generalised linear mixed model (GLMM) incorporates fixed as well as random effects into the linear predictor of a generalised linear model (see “[Modelling yearly flowering probability by logistic regression](#)”). If we consider random effects for all model parameters, the linear predictor of a logistic GLM model becomes:

$$l_{\pi_{ij}} = \beta_0 + u_{0j} + (\beta_1 + u_{1j}) \cdot X_{i1} + (\beta_2 + u_{2j}) \cdot X_{i2} + \dots + (\beta_k + u_{kj}) \cdot X_{ik} \quad (17)$$

The additional terms $u_{0j}, u_{1j}, u_{2j}, \dots, u_{kj}$ represent random effects values specific to cluster j . In this formulation, only one level of data grouping is considered; however, more complicated hierarchical GLMMs can be formulated with several levels of groupings (plots nested into localities, trees nested into plots, branches nested into trees and other nested groupings). Equation (17) clearly illustrates that random effects act as parameter corrections for particular clusters and that, given the values of random effects, any GLMM leads to a GLM.

Although the concept of generalised mixed effects models was well developed and appeared to be suitable, it was not used in the present analysis for several reasons. We could not account for the clustering of the terminal buds at the level of individual trees, because the number of branching orders was not recorded for individual terminal buds but for the whole crown periphery. There were also too few localities and sample plots to include random effects at these levels. This became apparent when we attempted to parameterise the model and is also confirmed by the recommendations of other authors (Bolker et al. 2008).

One advantage of using GLMMs rather than ordinary GLMs is the ability to adjust model parameters for new localities with a minimum number of additional measurements. This can be done with a calibration technique that predicts the most probable values of random effects using prior knowledge of marginal model parameters and the variance–covariance matrix of random effects. Calama and

Montero (2005) published a mixed effects and calibration technique to model and predict stem diameter increments. Calama and Montero (2006) and Trincado and Burkhart (2006) used a mixed effects approach for stem-taper functions. The application of this technique to the present study would be a statistically sound inventory of the DC population followed by simulations of its future state.

Improvements to data collection

From a methodological perspective, the purposeful (subjective) collection of field data should be replaced by a theoretically well-founded sampling technique. Because the data must be representative of a certain population, it must first be defined; an obvious way to accomplish this is to delineate a geographic region and consider DC specimens within it. It is also important to define a timescale to which the parameters to be estimated refer. In addition, for each population element included in the sample, the probability of its selection should be known or deducible in order to describe the population using a general Horwitz–Thompson approach to parameter estimation (Cochran 1977). Fixed-area, circular plots represent an easily applicable sampling design. The way in which plot locations are chosen must not rely on an expert’s guess or a similarly subjective procedure. Instead, we recommend the use of one of the spatial sampling patterns reviewed by Barabesi (2003).

An additional important aspect of data collection is the choice of elements defining the population itself. In our study, the population elements were trees on which the number of total and flowering terminal buds on the periphery of the crown were observed. However, modelling the number of flowering terminal buds summed over trees as a binomial variable violates the assumption of independence (see “[An extension using generalised mixed models](#)”). To avoid this drawback, flowering data should be collected at the level of individual terminal buds. The attributes of particular terminals (e.g. the number of branching orders on a path to the crown base) and trees (e.g. the length of the stems) could be used as explanatory variables at appropriate levels of grouping in a mixed effects approach. Of course, collecting such data requires a sampling protocol at that level of detail. Gregoire and Valentine (2008) describe a technique called randomised branch sampling that can be used to estimate properties of populations defined as paths within branched objects, as is typical for tree crowns. Based on our experience, a relatively long time is needed to sample just a few branches on one DC tree, suggesting that an adaptation of the technique would be necessary.

Our analysis was cross-sectional; although we used data collected over 4 years, flowering status was observed on a different set of sample trees on each occasion of measurement. Consequently, the true development of flowering

probability as a function of tree age may be different than that determined by our models, and the estimated tree age could be biased. A longitudinal analysis based on flowering status observed over several years on the same terminal buds would rule out this potential source of bias. This technique would also prevent bias resulting from the possible existence of seed years. Several flowering periods within 1 year may also occur. Thus, there should be additional observations within each year to detect each flowering event on sampled terminal buds. In addition to flowers, we recommend collecting data on fruit panicles. Fruit panicles remain present on terminal buds for a longer time period. Thus, flowering events can be detected several weeks or even months later.

Conclusion

We have presented a statistically sound methodology for an indirect age estimation of *Dracaena cinnabari*, a monocotyledonous tree species native to Socotra Island. An original idea published by (Adolt and Pavlis 2004) has been further developed here, suggesting a more suitable statistical technique for model formulation and parameterisation. Using the data set from (Adolt and Pavlis 2004), our new technique did not affect the form of the model or the expected length of the interval between flowering events, reinforcing the theoretical equality of both approaches.

In Firmihin, the length of the flowering interval decreases from approximately 28 years between the first and second flowering event to approximately 10 years between the 25th and 26th flowering event. However, in Skant, this interval remained constant at 6.5 years. The age of the crown of an individual tree can be predicted using Eqs. (8), (9) or (10).

For future analyses, we recommend a data collection procedure based on a well-founded sampling technique that guarantees the design-unbiased estimation of model parameters and other population figures. Data collection should concentrate on the flowering status of particular terminal buds with repeated observations over several years. With this approach, generalised mixed effects models could be parameterised to include individual trees as the first level of data grouping. Despite its increased complexity, the mixed models approach delivers unbiased age estimations that can be adjusted for particular locations, plots or even individual trees.

Acknowledgments This project was supported by the Internal Granting Agency of the Faculty of Forestry and Wood Technology at Mendel University in Brno (Project 12/2010) and by the Ministry of Education of the Czech Republic (Project MSM 6215648902).

Appendix: Nomograms for statistical power of LR tests

See Figs. 7, 8 and 9.

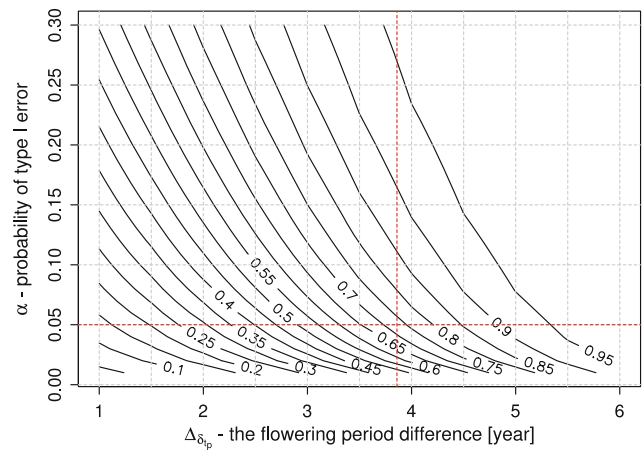


Fig. 7 The power of an LR test for one predictor in a logistic regression, an extended *firmihin.00q* model, $n = 1,869$, $\pi_{\bar{x}} = 0.0524$

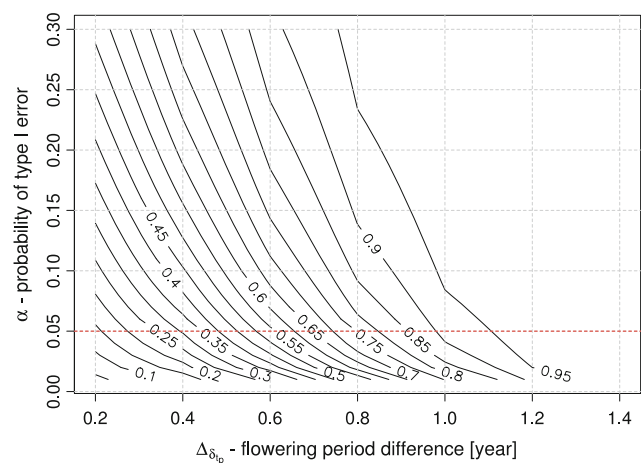


Fig. 8 The power of an LR test for one predictor in a logistic regression, an extended *skant.03q* model, $n = 1,967$, $\pi_{\bar{x}} = 0.1530$

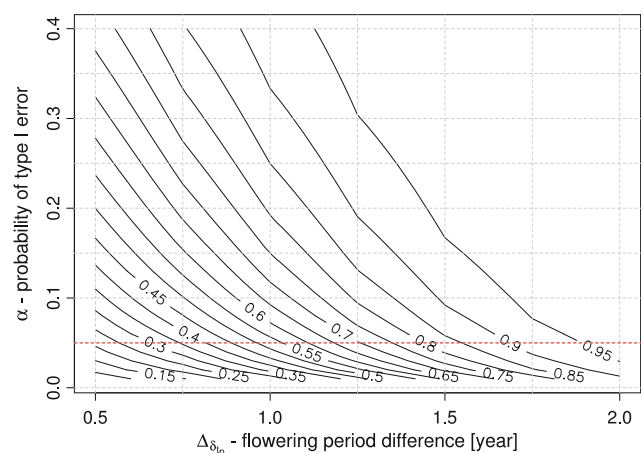


Fig. 9 The power of an LR test for two predictors in a logistic regression, an extended *firmihin.03* model, $R = -0.34478$, $n = 5,404$, $\pi_{\bar{x}} = 0.0524$

References

- Adolt R, Pavlis J (2004) Age structure and growth of *Dracaena cinnabari* populations on Socotra. *Trees* 18:43–53
- Agresti A (2002) *Categorical data analysis*. Wiley, Hoboken
- Barabesi L (2003) A Monte Carlo integration approach to Horvitz–Thompson estimation in replicated environmental designs. *Int J Stat LXI*:355–374
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Henry M, Stevens H, White JS (2008) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24(3):127–135
- Brienen RJW, Zuidema PA (2006) Lifetime growth patterns and ages of Bolivian rain forest trees obtained by tree ring analysis. *J Ecol* 94:481–493
- Brienen RJW, Zuidema PA (2006) The use of tree rings in tropical forest management: projecting timber yields of four Bolivian tree species. *For Ecol Manag* 226:256–267
- Brienen RJW, Zuidema PA, Martinez-Ramos M (2010) Attaining the canopy in dry and moist tropical forests: strong differences in tree growth trajectories reflect variation in growing conditions. *Oecologia* 163:485–496
- Calama R, Montero G (2005) Multilevel linear mixed model for tree diameter increment in stone pine (*Pinus pinea*): a calibrating approach. *Silva Fenn* 39(1):37–54
- Calama R, Montero G (2006) Stand and tree-level variability on stem form and tree volume in *Pinus pinea* L.: a multilevel random components approach. *Sist Recur For* 15:24–41
- Cochran GW (1977) *Sampling Techniques*. Wiley, Hoboken
- Gregoire GT, Valentine TH (2008) *Sampling strategies for natural resources and the environment*. Chapman and Hall/CRC, Boca Raton. ISBN 978-1-58488-370
- Habrova H (2004) Geobiocoenological differentiation as a tool for sustainable land-use of Socotra Island. *Ekologia* 23:47–57
- Habrova H (2007) Climate of Socotra with a view to rains. *Tayf Soqotra Newsllett* 4:15–16
- Habrova H, Cermak Z, Pavlis J (2009) Dragon's blood tree threatened by overmaturity, not by extinction: dynamics of a *Dracaena cinnabari* woodland in the mountains of Socotra. *Biol Conserv* 142:772–778
- Heuret P, Barthelemy D, Guedon Y, Columier J, Tancre X (2002) Synchronization of growth, branching and flowering processes in the South American tropical tree *Cecropia obtusa* (cecropiaceae). *Am J Bot* 89(7):1180–1187
- Hubalkova I (2011) Prediction of dragon's blood tree (*Dracaena cinnabari* balf.) stand sample density on Socotra Island (key study). *J Landsc Ecol* 4:5–20
- Kral K, Pavlis J (2006) The first detailed land cover map of Socotra Island by Landsat/ETM+ data. *Int J Remote Sens* 27(15):3239–3250
- Lusk HC, Smith B (1998) Life history differences and tree species coexistence in an old-growth New Zealand rain forest. *Ecology* 79:795–806
- Magdefrau K (1975) The age of large *Dracaena* specimens of Tenerife. *Flora* 164:347–357
- Norton AD, Ogden J (1990) Problems with the use of tree rings in the study of forest population dynamics. In: Cook ER, Kairiukstis LA (eds) *Methods of dendrochronology: applications in the environmental science*. Kluwer, Dordrecht
- Ogden J (1985) Past, present and future: studies on population dynamics of some long-lived trees. In: White J (eds) *Studies on plant demography: a festschrift for John L Harper*. Academic Press, London
- R Development Core Team (2009) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria, 2009. <http://www.R-project.org>. ISBN 3-900051-07-0
- Razdorskij VF (1954) *Anatomie rostlin*, 1st edn. Nakladatelství Československé Akademie Věd, Praha
- Rozas V (2003) Tree age estimates in *Fagus sylvatica* and *Quercus robur*: testing previous and improved methods. *Plant Ecol* 167:193–212
- Rozendaal DMA, Zuidema PA (2011) Dendroecology in the tropics: a review. *Trees Struct Funct* 25:3–16
- Sarndal CE, Swensson B, Wretman J (2003) *Model assisted survey sampling*. Springer, Berlin
- Scholte P, De Geest P (2010) The climate of Socotra Island (Yemen): a first-time assessment of the timing of the monsoon wind reversal and its influence on precipitation and vegetation patterns. *J Arid Environ* 74(11):1507–1515. doi:10.1016/j.jaridenv.2010.05.017
- Smilauer R (2007) *Moderní regresní metody*. Biologická Fakulta JU, Ceske Budejovice
- Stewart HG (1986) Population dynamics of a montane conifer forest, western Cascade Range, Oregon, USA. *Ecology* 67:543–544
- Symon ED (1974) The growth of *Dracaena draco*—dragon's blood tree. *J Arnold Arbor* 55:51–58
- Tabachnick BG, Fidell LS (2007) *Using multivariate statistics*. Allyn and Bacon, Boston
- Trincado G, Burkhart E (2006) A generalized approach for modeling and localizing stem profile curves. *For Sci* 52:670–681
- Venables NW, Ripley DB (2002) *Modern applied statistics with S*, 4th edn. Springer, New York. <http://www.stats.ox.ac.uk/pub/MASS>. ISBN 0-387-95457-0
- Worbes M (2002) One hundred years of tree-ring research in the tropics a brief history and an outlook to future challenges. *Dendrochronologia* 20(1–2):217–231
- Zalamea CP, Stevenson PR, Madrinan S, Aubert PM, Heuret P (2008) Growth pattern and age determination for *Cecropia sciadophylla* (urticaceae). *Am J Bot* 95(3):263–271
- Zimmermann HM, Tomlinson BP (1969) The vascular system in the axis of *Dracaena fragrans* (Agavaceae). *J Arnold Arbor* 50:370–383