ON THE STATISTICAL ANALYSIS OF REPRODUCTIVE SIZE THRESHOLDS IN DIPTEROCARP FORESTS

S.C. Thomas

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Ave., Cambridge MA 02138, United States of America

&

S. Appanah

Forest Research Institute Malaysia, Kepong, 52109 Kuala Lumpur, Malaysia

Received March 1993

THOMAS, S.C. & APPANAH, S. 1995. On the statistical analysis of reproductive size thresholds in dipterocarp forests. In a recent publication, evidence was presented that smaller trees can fruit in logged dipterocarp forests in Peninsular Malaysia. Here we point out an important statistical flaw in the paper: namely, that the mean size of fruiting trees in logged forest may be depressed simply due to the removal of larger individuals during harvesting. An improved analysis is presented, based on estimation of the reproductive size thresholds of trees using a modified form of logistic regression. We conclude that there is insufficient evidence to strongly support the idea that trees fruit at smaller sizes in logged forest; however, there is a suggestive trend in this direction. Guidelines are given for further data collection and analysis to definitively answer this important question.

Key words: Dipterocarps - fruiting - logging - reproductive size

THOMAS, S.C. & APPANAH, S. 1995. Statiskal analisa mengenai had saiz kritikal pembiakan dalam hutan dipterokap. Dalam penerbitan yang terkini, terdapat bukti yang menunjukkan bahawa pokok yang berukuran kecil berkeupayaan untuk berbuah di kawasan hutan dipterokap yang telah dibalak. Dalam kertas kerja ini, kami *menunjukkan terdapat kesalahan dalam analisa tersebut iaitu purata kebesaran pokok* berbuah dalam hutan yang telah dibalak mungkin berkurangan disebabkan tebangan pokok-pokok yang besar semasa pembalakan. Analisa yang lebih baik ditunjukkan, berdasarkan kepada anggaran had saiz kritikal pembiakan pokok-pokok menggunakan bentuk regresi logistik. Kami membuat kesimpulan yang mana tiada bukti yang cukup untuk menyokong pendapat yang mengatakan pokok kecil berbuah di hutan yang telah dibalak. Walau bagaimanapun terdapat tanda-tanda kebenaran pendapat tersebut. Satu garis panduan untuk kutipan dan analisis data yang selanjutnya diberikan untuk menjawab dengan pasti persoalan yang penting ini.

Introduction

Future timber harvests in dipterocarp forests are dependent on adequate seedling stocks of commercially valuable species. It is therefore of critical importance to understand how seedling stocks are generated, and particularly how the generation of important timber species may be affected by harvesting activities. Recently Appanah and Manaf (1990) presented evidence that dipterocarp trees reproduce at relatively smaller sizes in logged forest than in primary forest. The purpose of this paper is twofold: (1) to point out an important flaw in the statistical analysis presented by Appanah and Manaf; and (2) to conduct an improved statistical analysis, one that directly examines differences in reproductive size thresholds among dipterocarp trees studied in disturbed versus primary forests.

The problem

The previous analysis by Appanah and Manaf (1990) is based on a comparison of the mean size (specifically dbh, or diameter at 1.3 m) of flowering individuals observed in primary forest versus logged forest. Such a comparison is valid only if the size distribution for all trees (reproductive and non-reproductive) is the same for both populations. However, in logged forests the largest trees have generally been removed. The effects of this can most easily be seen by considering size distribution for two forest plots in which all trees above a certain size always reproduce (Figure 1). If trees above a certain dbh are removed by logging, then the mean dbh of reproductive individuals will be lower in that plot even if there is no increase in reproduction among smaller trees. Thus, there is a strong bias toward a smaller mean size for fruiting trees in logged forest that is simply the result of an absence of the largest trees in the surviving population.

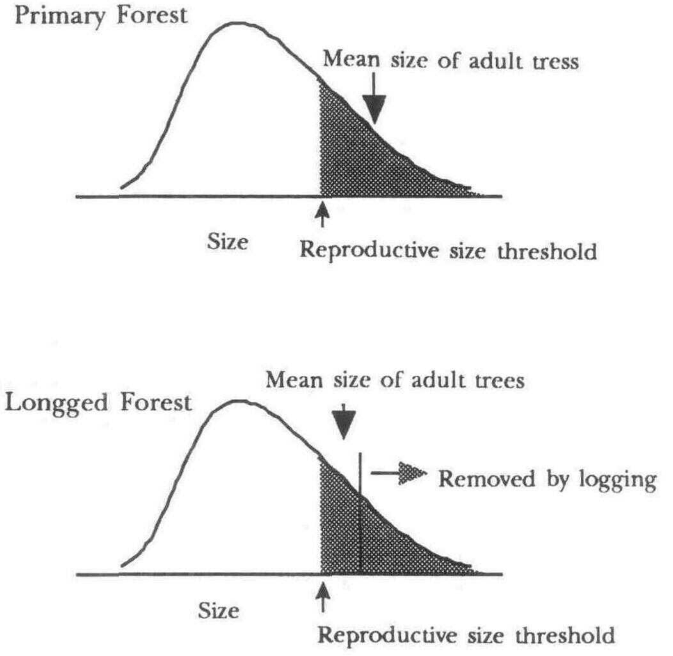

Figure 1. The statistical problem: the mean size of adult trees in logged forest will generally be lower than that observed in undisturbed forest simply because the largest trees in a population are generally removed by logging

Materials and methods

The mean size of reproductive individuals is not actually of particular significance to the problem at hand. Rather, to best determine if smaller trees can fruit in logged forest, one should estimate the lower size limit for reproduction, or "reproductive size threshold". One method by which to estimate size at reproductive onset is based on examining the probability of reproduction as a function of tree size. If there is a perfect size-cutoff between reproductive and non-reproductive trees (as in Figure 1), then a step function would describe this relationship, and size cutoff could simply be noted by inspection. However, in most tree populations one finds some large trees that do not reproduce as well as small trees that do not. By analogy to the case of a perfect size cutoff, the inflection point of a continuous function describing the probability of reproduction as a function of tree size may be used as a statistical estimate of the reproductive size threshold. Theoretically, this inflection point corresponds to the mode (or peak value) of the distribution of tree size at reproductive onset that would be determined if trees were repeatedly censused through time (this is strictly the case under the assumption that after reproductive onset trees consistently fruit in subsequent reproductive episodes: see Thomas 1995). The following mathematical function (a modified form of the logistic regression equation: e.g. Hosmer & Lemeshow 1989) is particularly tractable for quantifying the relationship between size and the probability of reproduction:

$$P = k\left(\frac{e^{a+b\ln S}}{1 + e^{a+b\ln S}}\right) \qquad \text{Equation 1}$$

In this equation P is the probability of reproduction, S is a size metric (i.e. dbh), e is the base of natural logarithms, ln is the natural logarithm function, and k, a and b are constants. The parameter k describes the probability of reproduction among large trees; if the largest trees in the population are all reproductive, k will equal 1.0. The parameter b describes the sharpness of the size cutoff; if there is no trend of reproductive probability with plant size, then b = 0; for a perfect size cutoff b approaches infinity. The parameter a is a scaling factor that is primarily of interest in calculating the inflection point of this function, which estimates the mode for size at first reproduction in the population. The inflection point is calculated as:

$$S_{crit} = e^{\,b^{-1}\left[\ln\frac{(b-1)}{(b+1)} - a\right]} \qquad \text{Equation 2}$$

Here S_{crit} is the modal size at reproductive onset, and other variables are as defined above. These equations have previously been found to provide good estimates of reproductive probability as a function of stem diameter in 37 non-dipterocarp tree species studied at Pasoh Forest Reserve, Negeri Sembilan, Malaysia (Thomas 1993, 1995).

Equation 1 was fitted to the data presented by Appanah and Manaf (1990) using maximal likelihood methods that make direct use of the binary data (i.e. reproduction for each tree is scored as 0 for non-fruiting and 1 for fruiting individuals). The Nonlin procedure in SYSTAT (Wilkinson 1989) was used to estimate the function, and to calculate confidence intervals for the parameters. It is difficult to directly calculate confidence intervals for size threshold values (S_{crit}). We instead compare values for the parameter a and b between primary and logged forests, using standard error estimates to conduct approximate t-tests (see Sokal & Rohlf 1981). One or both of these parameters must be significantly different between sites in order for there to be a significant difference between size threshold values. Running means of ranked stem diameter values and corresponding reproductive state values are used to qualitatively examine the shape of the observed relationships. An approximate procedure for calculating reproductive size thresholds using simple manipulations of tabulated data is presented in the appendix.

Results and discussion

Plots of running means do suggest substantial differences in reproductive size-dependence between trees in logged vs. primary forest (Figure 2). Equation 1 provides a reasonable description of the quantitative trends, although in the case of the primary forest data a single small reproductive tree (of 21.9 cm dbh) has a very large effect, and as a result the estimated function is somewhat flattened. If this tree is omitted from the analysis, the qualitative fit for the remaining data is much better (see Figure 2). For the primary forest population, we estimate a reproductive size threshold of 60.3 cm dbh. This value is much larger than the thresholds estimated for the arboretum population (38.4 cm) and the logged forest population (33.8 cm). The combined data sets for these populations (as shown in Figure 2) yield an estimated size threshold of 36.5 cm. There also appears to be a trend toward a greater proportion of large trees in flower in logged than in primary forest. A slight difference in this direction is indicated by estimated values for the parameter k in equation 1 ($k = 660$ vs. $k = 628$ in logged vs. primary forests respectively).

Although the calculated difference between the size threshold values is quite large, it does not reach statistical significance, as indicated by tests of differences between the parameters in equation 1. The parameter b estimates the "distinctness" of the reproductive size threshold. Although there is a trend toward a more distinct size threshold in logged than in primary forests ($b = 7.85$ vs. $b = 3.40$ respectively), this difference is not significant ($p = 0.195$). Similarly, there are no statistically significant differences for the parameters a and k in logged vs. primary forests (the calculated values for a are 28.487 vs 14.527 in logged vs. primary forests respectively).

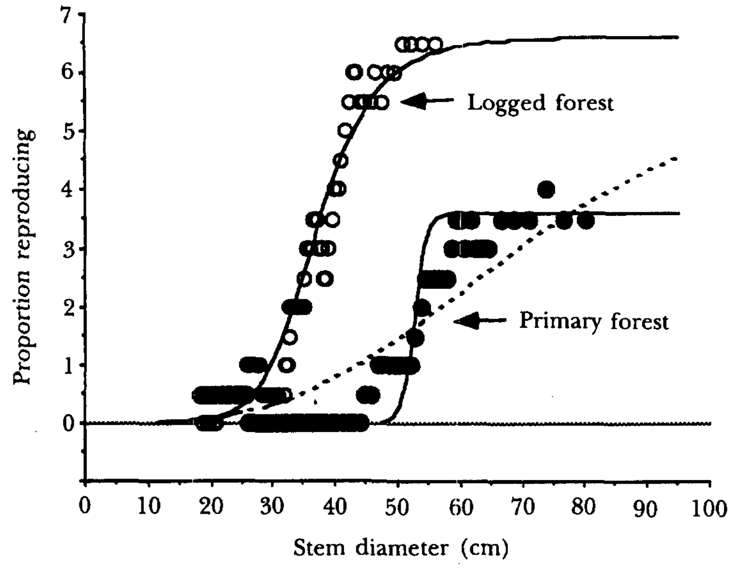

Figure 2. Reproductive size thresholds may be calculated as the inflection points for continuous functions that relate the probability of reproduction to stem diameter for the logged and undisturbed forests. The data of Appanah and Manaf (1990) are shown as points representing the running mean values for ranked stem diameter values vs. reproductive state (scored as 1 for fruiting trees, and 0 for non-fruiting trees). The curves fit equation 1 directly to the binary data, using maximal likelihood methods. The dotted line for the undisturbed forest data includes all trees, while the solid line excludes one small outlying tree from this analysis

The statistical pitfall encountered in Appanah and Manaf's study is a very common one. Very few studies have adequately defined or presented statistical tests of reproductive thresholds, in spite of their importance from a management perspective. We hope the present analysis will aid in guiding others away from some particularly hazardous statistical shoals. We recognize that access to computers and appropriate statistical software may be problematic in some areas. An approximate calculation method is given in the appendix which gives results very similar to the logistic regression analysis used here.

The data presented by Appanah and Manaf (1990) provide a preliminary basis for calculating reproductive size thresholds of dipterocarp trees in logged vs. primary forest. However, when the appropriate analyses are conducted, it is apparent that the sample sizes are insufficient to strongly support the claim that smaller trees can in fact fruit in logged dipterocarp forest. Additional samples that include a good size range of trees in several samples of both logged and primary dipterocarp forests are necessary. As a rule of thumb, such samples should include at least 25 and preferably 50 reproductive trees, and similar or greater number of non-reproductives. There is clearly a need for more data to address the general issue of reproduction by trees left after logging operations. As previously emphasized (Appanah & Manaf 1990), information on crown sizes, individual tree responses to crown exposure, and species-specific differences are also of considerable importance from a management perspective.

References

APPANAH, S. & MANAF, M. R. A. 1990. Smaller trees can fruit in logged dipterocarp forests. *Journal of Tropical Forest Sciences* 3 : 80 - 87.

HOSMER, D.W. & LEMESHOW, S. 1989. *Applied Logistic Regression.* Wiley, New York.

SOKAL, R. R. & ROHLF, F. J. 1981. *Biometry.* 2nd edition. W.H. Freeman and Co., New York.

THOMAS, S. C. 1993. Interspecific allometry in Malaysian rain forest trees. Ph.D. dissertation, Harvard University, Cambridge, Massachusetts, U.S.A.

THOMAS, S.C. 1995. Relative size at onset of maturity in rain forest trees: a comparative analysis of 37 Malaysian species. Submitted to *Oikos.*

WILKINSON, L. 1989. *SYSTAT: The System for Statistics.* SYSTAT, Inc., Evanston, Il.

Appendix

The following is an approximate method for calculating reproductive size thresholds using simple manipulations of tabulated data:

(1) Create a table of all dbh values in ascending order, and mark all reproductive individuals.
(2) Calculate the total sample size divided by 10, rounding to the nearest integer. This is X.
(3) Examine the set of the largest X trees in the data set, and tally the number of reproductive trees in this set.
(4) Divide this value by 2, rounding up to the nearest integer. This is Y.
(5) Find the smallest reproductive tree in the data set. This is tree Z.
(6) Examine the adjacent set of X trees in the data table, beginning with tree Z.
(7a) If this set of X trees includes Y or more reproductives, then the dbh of Z is taken as the size threshold.
(7b) If not, then find the next largest reproductive tree in the data set. Call this the new tree Z, and repeat from step (6).

Y/X is an estimate of k in equation 1; the diameter of tree Z is an estimate of the reproductive size threshold. This approximate method assumes that the largest 10% of trees sampled are well above the reproductive size threshold, and that there is a good sample of trees near the reproductive size threshold. The approximation was tested on a large data set of 37 Malaysian tree species for which the logistic regression analysis had previously been conducted (Thomas 1993, 1995). The two methods yield very similar values: in a regression analysis, the r^2 between the two estimates was 0.95, and the slope was not significantly different from 1.0.

Example

For the following data set (Pasoh Plot 1 from Appanah & Manaf 1990), X = 10, Y = 2, and the final tree Z measures 53.8 cm dbh.

12.5	13.1	13.4	14.5	14.7	15.0	17.1	17.4	17.6	18.7	19.5
20.0	21.0	21.5	21.5	21.9	22.0	22.1	22.3	22.7	23.3	23.5
23.5	23.8	25.0	25.5	26.0	26.6	27.4	28.3	28.5	28.5	29.9
30.4	31.5	31.5	31.7	31.8	31.9	32.0	32.2	33.4	33.5	34.2
34.3	35.5	35.8	36.6	36.9	37.1	37.5	39.0	40.2	41.2	41.3
42.1	42.9	43.4	46.0	49.5	50.0	50.4	52.0	52.6	53.0	53.8
53.8	54.3	54.7	56.3	56.8	57.0	57.0	58.0	58.0	58.5	60.5
60.6	60.8	62.8	63.0	63.0	64.8	65.8	66.8	68.9	68.9	71.2
72.0	78.3	78.5	93.0	101.2	101.5	110.0	120.5	125.0		