

Modelling regeneration and recruitment in a tropical rain forest

JEROME K. VANCLAY

Royal Veterinary and Agricultural University, Thorvaldsensvej 57, DK-1871 Frederiksberg C, Denmark

Received July 18, 1991

Accepted February 14, 1992

VANCLAY, J.K. 1992. Modelling regeneration and recruitment in a tropical rain forest. *Can. J. For. Res.* **22**: 1235–1248.

A two-stage model predicts the recruitment (i.e., the number of stems reaching or exceeding 10 cm DBH) of the 100 species that account for 97% of all the recruitment observed on 217 permanent sample plots in the tropical rain forest of north Queensland. The first stage predicts the probability of the occurrence of any recruitment from stand basal area and the presence of that species in the existing stand. These probabilities can be implemented stochastically, or deterministically by summing the probabilities and initiating recruitment on unity. The second stage indicates the expected amount of recruitment, given that it is known to occur, and employs stand basal area, the relative number of trees of that species in the stand, and site quality. This approach is easily implemented in growth models and planning systems.

VANCLAY, J.K. 1992. Modelling regeneration and recruitment in a tropical rain forest. *Can. J. For. Res.* **22** : 1235–1248.

Un modèle à deux niveaux prédit le recrutement, c'est-à-dire le nombre de tiges de 10 cm ou plus au DHP, des 100 espèces qui comptent pour 97% de tout le recrutement tel qu'observé dans 217 parcelles échantillons permanentes dans la forêt ombrophile tropicale du nord du Queensland. Le premier niveau prédit la probabilité de recrutement d'une espèce à partir de la surface terrière du peuplement et de la présence de cette espèce dans le peuplement existant. Ces probabilités peuvent être traitées de façon stochastique ou déterministe en additionnant les probabilités et en commençant le recrutement lorsque la valeur de un est atteinte. Le second niveau indique la quantité attendue de recrutement, sachant qu'il y a du recrutement, et utilise la surface terrière du peuplement, le nombre relatif d'individus d'une espèce dans le peuplement et la qualité de station. Cette approche est facilement utilisable dans des modèles de croissance et des systèmes de prise de décision.

[Traduit par la rédaction]

Introduction

Growth models may neglect the simulation of regeneration because of a lack of data or difficulty in modelling, or because it is considered unnecessary because silviculture involves clear-felling and replanting. However, natural regeneration forms an essential component of selection harvesting systems used in rain-forest management, and long-term yield forecasts must take account of the presence and amount of this regeneration.

A review of the literature indicates two approaches to predicting regeneration and recruitment. Regeneration models predict the development of trees from seed or seedlings. As suitable data for modelling regeneration are difficult to obtain, many models predict recruitment rather than regeneration. Recruitment models predict the number of stems reaching or exceeding some specified nominal size limit (e.g., 1.3 m height, 3 m height, 10 cm DBH over bark etc.). Recruitment models may employ a static approach that predicts a constant amount each year irrespective of stand condition, or may be dynamic and respond to stand condition.

Static approaches to the prediction of recruitment assume that the amount of recruitment observed during the period of data collection reflects the long-term average, and that this amount will not vary greatly between projection periods predicted by the model. Such assumptions of static recruitment are common in stand table projection and matrix approaches. Usher's (1966) matrix model for Scots pine (*Pinus sylvestris* L.) predicted recruitment as a static proportion of the number of trees in the larger size classes, and thus recruitment increased as stand density increased. More realistic matrix approaches may predict recruitment diminishing with increasing stand density (e.g., Buongiorno and Michie 1980) or only after the death of another tree (e.g., Bosch 1971). Although these approaches are rather empirical, they may provide useful estimates of recruitment for stands that do not differ greatly from the source stands used for model development.

Other approaches attempt to predict the number of stems recruited as a function of stand condition. These vary from the highly empirical to those that model some simple biological hypothesis. Letourneau (1979) used an empirical approach with 33 estimated parameters to predict numbers of stems, and accounted for time between remeasures in estimating size of recruits. Landford and Cunia (1977) predicted total number of recruits (at 4 in. (10 cm) diameter) deterministically, but the size and species stochastically. Both these models used estimates of sapling density (numbers of stems in the 1, 2, and 3 in. (2.5, 5.1, and 7.6 cm) DBH size classes) in their equations, and this estimate was assumed to remain constant through time; as saplings were recruited, new regeneration was assumed to take their place. This limits the utility of the model for extrapolating inventory data, as such data (sapling density) may not be recorded during operational inventory.

Hann (1980) predicted recruitment with an exponential function of site index, stand basal area, and basal area in the smallest size class. His simulation cycle was 5 years, and this ensured a realistic 5-year lag in the appearance of recruitment stimulated by the reduction in stand basal area following logging.

Vanclay (1989a) predicted the total amount of recruitment at 20 cm diameter in tropical moist forests in north Queensland as a linear function of stand basal area and site quality. The composition of this recruitment was determined by predicting the proportion in each of five species groups and standardizing the proportions. The proportion for each species group was predicted from the stand basal area, the site quality, and the basal area of that species group (e.g., for the large, fast-growing species):

$$P_1 = [1 + \exp(2.407 + 0.005\ 608\ BA - 0.011\ 05\ B_1 - 0.004\ 64\ B_1 SQ)]^{-1}$$

where BA is total stand basal area (m^2/ha), B_i is the basal area of the i th species group (m^2/ha), and SQ is the growth index. These proportions were standardized to ensure they summed to unity:

$$P'_i = \frac{P_i}{\sum P}$$

Botkin *et al.* (1972) established an ecological approach to modelling recruitment at 0.5 cm diameter on their 10×10 m plots. They assumed that a seed source was available for each of the major species considered by their model and compiled a list of possible species for the plot being modelled, on the basis of shade tolerance, growing season, and soil moisture. If the plot leaf area index (LAI) was less than a specified threshold, 60–75 cherry trees were recruited on the plot. If the plot LAI exceeded the first threshold, but was less than a second larger threshold, some (0–13) birches were recruited. If the plot LAI exceeded both these thresholds, a random choice of the remaining suitable shade-tolerant species was made, and a random number (0, 1, or 2) of each species was recruited. Shugart and West (1977) followed a similar approach, but identified the requirements of each species for mineral soil or leaf litter, introduced stochastic elements of variable weather and animal browsing, and also modelled sprouting from dead trees. They recruited trees when they reached breast height. Similar succession models exist for subtropical rain forests in Australia (Shugart *et al.* 1980) and tropical rain forests in Central America (Doyle 1981). Reed (1980) followed an approach somewhat similar to that of Botkin *et al.* (1972), but introduced alternate seed years and “off years” and imposed a maximum stocking of 2500 stems/ha, irrespective of size, above which no recruitment could occur.

One of the difficulties in modelling recruitment is the great variability in regeneration. Stand condition accounts for some of this variation, as do periodicity of mast years and prevailing climate, but regeneration remains a rather stochastic process, providing difficulties for efficient model estimation. Much of the variability associated with regeneration is due to the fact that during any period some regeneration may or may not occur, and that if the data are partitioned into a two-state system, the ability to predict the amount of regeneration is greatly enhanced. With a two-state approach, the first equation estimates the probability that some regeneration or recruitment will occur, and can be estimated using logistic regression with presence–absence as the dependent variable. The second stage is a conditional function to predict the amount of recruitment, given that some is known to occur, and can be estimated using ordinary linear regression. Hamilton and Brickell (1983) gave an example of such a two-stage approach applied to the prediction of defective volume in standing trees, which can be applied equally well to recruitment modelling.

Stage and Ferguson (1982) and Ferguson *et al.* (1986) used a two-stage approach to predict recruitment in the PROGNOSIS model (Stage 1973; Wykoff *et al.* 1982; Wykoff 1986). They used a stochastic procedure to predict the regeneration on 50 subplots, each 1/300 acre (about 0.001 ha), and these data were aggregated into the main PROGNOSIS model at 10 and 20 years after disturbance. They predicted the probability that some regeneration would occur using environmental variables (habitat, slope, aspect, elevation), distance to seed source, residual basal area, and time since disturbance. Given that

regeneration is known to occur, the expected number of trees is determined using pseudorandom numbers, and it determines the number of cohorts for that subplot. The number of species present and the identity of these species were also stochastically determined. Height of regeneration is deterministically predicted. The model considers three categories of regeneration: “best” trees, comprising the two tallest trees per plot regardless of species, the tallest tree of each additional species present, and the four tallest of any remaining trees. Best trees were assumed to be advance growth of shade-tolerant species if established 3 years prior to disturbance. Regeneration within each category was predicted independently and differed considerably in composition (e.g., advance growth was more likely to comprise shade-tolerant species).

Some models simulate the growth of trees from seedlings to breast height within a separate “regeneration model,” or “understory operation” (Ek and Monserud 1974a, 1974b; Dudek and Ek 1980). The approach is sufficiently flexible that almost any size may be used as the criterion for recruitment into the main stand. Ek and Brodie (1975) simulated only suckers developing after logging (predicted from stand basal area before and after logging, site index, and treatment), but many models simulate the development of regeneration throughout the development of the stand. Modelling may start at any of a number of stages. Leak (1968) modelled regeneration from the stage of flower development; Ek and Monserud (1974a, 1974b) modelled it from seedfall. Germination could equally well be considered, and Vanclay (1988) started with “establishment” when the seedling had survived its 1st year after germination. An advantage of the approach is that it realistically models the time taken for regeneration to be recruited following a reduction in stand density due to logging. In contrast, recruitment models may need to employ an explicit time lag. Recruitment models which employ an expression of stand density may lead to overestimates of recruitment in the projection period immediately following logging.

Vanclay (1988) predicted the amount of established 1-year-old regeneration in a *Callitris* forest from stand basal area and site quality. Regeneration was modelled as cohorts representing height classes until it reached breast height, when it was recruited to the main model. A maximum of 10 cohorts were employed. Under ideal conditions (good sites with low stocking), these cohorts represented annual flushes of regeneration. Where regeneration exhibited slower growth and took more than 10 years to reach breast height, the most similar cohorts were combined to ensure that the limit of 10 cohorts was not exceeded.

Ek and Monserud (1974a, 1974b) adopted a more detailed approach to predict recruitment into their stochastic distance-dependent individual tree model. The regeneration model used cohorts representing the number of stems for each species and age in a number of subplots within the main plot being simulated. A Monte Carlo approach selected good, moderate, and poor seed years according to the observed frequency for each species. Seed and sprout production were estimated for each overstory tree as a function of its size and the threshold age, and were distributed across the subplots according to the parent tree's position, height, and crown width. Germination was predicted as a stochastic function of microsite and canopy cover conditions. Each year, a germinant or tree in the understory may die or survive and attain

some height increment (function of cover, species, and age). When a tree reached breast height it was recruited into the main model. If a tree did not attain this height within a specified time (e.g., 25 years for black spruce), it was assumed to have died.

Monserud and Ek (1977) refined this approach, improving the efficiency by reducing the number of cohorts to be modelled. They assumed that understory tree size was more relevant than tree age and modelled the development of trees to 7.6 m height using five height cohorts of varying size, using movement ratios to predict upgrowth. The height increment of the mean tree was predicted from the potential height increment (a function of height and site), overstory competition index (a relative size-distance index), shade tolerance (a function of species and height), and stand density. Monserud and Ek determined that five cohorts were required to model the understory without compromising accuracy. Such detailed approaches (Ek and Monserud 1974a, 1974b; Monserud and Ek 1977) may not be warranted for yield prediction models, but may be relevant for models that analyze the silvicultural alternatives for intensively managed stands.

One difficulty with regeneration models is ensuring compatibility with inventory data when the model is used for yield prediction. Inventory data frequently sample only the larger stems (e.g., ≥ 10 cm DBH), and smaller stems may remain unsampled. Thus there may be some censorship of data. Such problems are more common for regeneration models (which predict regeneration at very small sizes) than for recruitment models, which predict recruits at larger diameters (e.g., 10 cm). To avoid this censorship, it is necessary either for the inventory to provide a count of the smaller stems or for a model to predict the likely incidence of such stems from overstory stocking. Augmenting such censored data with an "average" small tree distribution for the forest type is preferable to using the unadjusted censored data (Randall *et al.* 1988).

Difficulties in obtaining uncensored data during an operational inventory limit the utility of regeneration models for yield forecasting. Data concerning regeneration are often not available, or may be unreliable owing to inability to reliably identify species of seedlings, whereas recruitment data are always available from permanent sample plots. The germination and initial survival of seedlings in the rain forest is an uncertain phenomenon; vast numbers of individuals and species may germinate but never attain a significant size. The longer term survival and continued growth of these seedlings is much more under environmental control and thus recruitment is more predictable than regeneration. Thus modelling recruitment at some nominal size represents a viable alternative.

Data

The present study concerns the tropical rain forests of north-east Queensland. These forests have been managed for conservation and timber production for more than 80 years (Just 1991), and prior to their recent inclusion on the *World Heritage List*, provided a sustained yield of 60 000 m³ of veneer and sawlogs annually through conservative selection harvesting (Preston and Vanclay 1988; Vanclay 1991a). The Queensland Department of Forestry (1983) research programme provided a data base of 250 permanent sample plots with a measurement history of up to 40 years (Vanclay 1990). All trees exceeding 10 cm DBH (diameter over bark at breast

height (1.3 m) or above buttressing) were numbered, tagged, and measured for diameter. The plots sample virgin, logged, and silviculturally treated forests.

Pairs of consecutive remeasurements (i.e., all nonoverlapping intervals) were selected from the data base and formatted to provide a data file suitable for analysis. Site quality (Vanclay 1989b) could not be estimated for some plots, and the omission of these plots left 217 plots for the present study. Since plots were remeasured, these 217 plots provided 791 observations of the incidence and amount of recruitment (at 10 cm DBH) for each species. The file also included details of stand and site variables such as basal area, site quality, and soil type.

Correct species identity was recorded in the data base using a three-character mnemonic. However, species identification is often difficult in these forests, and routine resource inventory procedures record only the standard trade name (Standards Association of Australia 1983), using a set of mnemonics known as the harvesting and marketing (H & M) codes. Although the H & M code retains the correct identity of most species, several members of a genus may share a common code, as may members of more than one genus with similar timber characteristics. Some 100 noncommercial species and trees of undetermined identity may be labelled as miscellaneous. As the present study was to develop a recruitment model for operational yield prediction purposes (e.g., Vanclay and Preston 1989), it was appropriate to use the H & M codes. The data file used in the present study included 239 such H & M codes (including miscellaneous).

Method

The present data posed three difficulties: a range of plot sizes and measurement intervals, the large number of species characteristic of tropical rain forests, and the variable nature of recruitment. The data suggested an excessively heteroscedastic pattern of recruitment, until partitioned for the two-stage modelling approach, with one equation to predict the probability that any recruitment of a given species occurs and another to predict the amount, given that recruitment is known to occur. Once so partitioned, regression analyses were possible.

The data file included data drawn from a range of plot sizes and measurement intervals, both of which have an influence on the likelihood of recruitment occurring. Longer measurement intervals have a higher probability of recruitment, and this can be accommodated by weighting the regression analysis (e.g., Hamilton and Edwards 1976). The amount of recruitment can easily be adjusted for plot size and measurement interval by converting to stems per hectare per annum, but the influence of plot size on the probability of recruitment is more complex. Plot sizes varied from 0.03 to 0.5 ha; about half of them were half-acre (0.2023-ha) plots. The present study assumes that the probability of recruitment is determined primarily by stand condition (and measurement interval), and that the influence of plot size in the present study is negligible. This assumption is tenable largely because the presence of a species is a key explanatory variable in predictions, and it compensates for plot size. Recruitment of a species is more likely if that species is already present on the plot; a species is more likely to be present on a bigger plot, and recruitment of any species is more likely on a bigger plot. Thus this assumption regarding plot size is reasonable given that presence of the species is one of the explanatory variables.

Number of species to model

Of the 239 H & M species groups occurring on the permanent sample plots, 213 were observed to occur as recruitment on one or more occasions. However, the contribution of individual species to the total recruitment varied greatly. Sixty species accounted for 90%, 80 species accounted for 95%, and 100 species accounted for 97% of all recruitment observed. The remaining 3% of recruitment comprised 113 species, all of which offered insufficient data for meaningful analyses of regeneration characteristics. Accordingly, only the more prevalent 100 species were included in the recruitment model. These species were those that were observed as recruitment on nine or more occasions; they included a reasonable number of commercial and noncommercial species and light-demanding and shade-tolerant species, and they should provide a reasonable representation of the forest. Simulating the recruitment of these 100 species should provide sufficient precision for yield prediction purposes, provided that simulated ecological consequences are interpreted carefully.

Two options exist for recruitment of the remaining 113 species: they may be aggregated with the miscellaneous group, or they may be ignored. Both options have disadvantages. The less abundant species may have ecological characteristics unlike those dominating the miscellaneous group, and aggregating these could lead to bias. Similarly, ignoring this recruitment also leads to bias. In the present study, they were ignored since they represent such a small component of the total recruitment.

Aggregating species

It is impractical to develop recruitment models for each of these 100 individual tree species because of the large number of functions that would be required and because the paucity of data for many species inhibits the development of reliable relationships. Thus for efficient estimation of recruitment, it is desirable to aggregate these species into several groups. This reduces the number of functions required to a more manageable number, and avoids the requirement for specific equations for species with few data. Such groupings need not form the basis for growth modelling, as simulation models can retain the individual identity of all species (Vanclay and Preston 1989), but are necessary for the estimation of increment, mortality, and recruitment functions. Ideally, species should be grouped on *a priori* grounds, and tests performed to justify the validity of such groupings. This may be possible in temperate forests where there are few species with well-documented ecological characteristics, but is unrealistic for the 100 species in the present study. Taxonomy (family or genus) may not provide a good guide to the regeneration strategy, and other methods (e.g., Swaine and Whitmore 1988) based on successional status, seed morphology, etc. may be rather subjective. Regeneration is dependent upon stand density and other factors, so a grouping based on average recruitment may be specific to the set used. Not only is it difficult to resolve which species to combine, but it is not clear how many groups are required.

Meldahl *et al.* (1985), Leech *et al.* (1991), and Vanclay (1991b) have examined procedures to resolve these questions. Meldahl *et al.* (1985) argued that the grouping should reflect the dynamics of growth, and this could be best expressed through the coefficients of a regression equation on diameter increment. They attempted cluster analysis on these coefficients, but found that reasonable results could be obtained

only when the regression analysis was constrained to a single explanatory variable. Their best results were obtained using the basal area of trees larger than the current tree as the explanatory variable. The cluster analysis was weighted by the inverse of the significance level of slope parameter, and it provided 20 clusters from 110 species-type equations. The amount of data assigned to each cluster varied greatly, and the outcome was subjectively adjusted to provide the final grouping. The adequacy of final groups was tested by fitting a multiparameter linear function and examining the total (across clusters) residual sums of squares, on the assumption that a better grouping would result in a better fit. Whilst the method did provide a grouping of similar elements, it did not provide a unique solution.

Leech *et al.* (1991) used a Behrens-Fisher analogue of Hotelling's T^2 to group 27 species for fitting volume equations. They used a polynomial equation to predict tree volume (V) from tree diameter (D) for tree i :

$$V_i = \beta_{0i} + \beta_{1i}D + \beta_{2i}D^2 + \dots + \beta_{ni}D^n$$

Then, using the vector of coefficients

$$u'_i = [\beta_{0i}, \beta_{1i}, \beta_{2i}, \dots, \beta_{ni}]$$

Hotelling's T^2 between two species i and j can be defined as

$$d_{ij}^2 = (u_i - u_j)'S^{-1}(u_i - u_j)$$

where S^{-1} is the combined covariance matrix of regression coefficients for species i and j . By calculating all possible combinations, a symmetric matrix with zero diagonal elements can be formed. Principal coordinate analysis (Gower 1966) was used to group species on the basis of this matrix. Leech *et al.* (1991) concluded that the technique should only be used when the order of the polynomial and the sign of the highest term are the same for each of the two individual species equations. The method was also computationally intensive.

Vanclay (1991b) devised an objective means to aggregate 237 species into 41 groups to enable efficient estimation of diameter increment functions for a growth model of tropical rain forests in north Queensland. This approach involved the following:

- (1) Ranking species in order of increasing number of observations.
- (2) Assigning the species of highest rank as the founding species of group 1.
- (3) For each species in decreasing order of rank, conducting pairwise F -tests with all founding species of higher rank. If the incoming species exhibited a significantly different ($P < 0.01$) increment pattern from all existing founding species, it became the founding species of a new group. Species not significantly different from all founding species remained ungrouped.
- (4) After identifying the founding species, comparing the ungrouped species, in order of rank, with all existing groups, and combining each with the most similar group. Similarity was determined as the grouping which led to the smallest increase in residual sum of squares when the incoming species was incorporated in the group. These comparisons were made with the whole group, not just with the founding species.