Compatible Deterministic and Stochastic Predictions by Probabilistic Modeling of Individual Trees

Jerome K. Vanclay

ABSTRACT. A single growth model can provide both deterministic and stochastic predictions which are compatible. Change may be expressed using probabilistic functions which can represent proportions of populations or probabilities for individuals. The former represents determinism while the latter enables the stochastic implementation. The same functional relationships may thus be used to generate compatible deterministic and stochastic predictions. All components of forest growth and change, including diameter increment, can be expressed as probabilistic functions, enabling construction of a single model which provides compatible stochastic and deterministic outcomes. Users may specify the minimum expansion factor corresponding to the simulated plot size and thus control the granularity of predictions. Such a model may facilitate numerical estimation of confidence intervals about yield forecasts and sustained yield estimates.
FOR. SCI. 37(6):1656-1663.
ADDITIONAL KEY WORDS. Logistic equations, Monte Carlo simulation, movement ratios.

One of the important milestones in growth modeling in the 1960s was the realization that growth and yield models must be compatible (Buckman 1962, Clutter 1963). Forest managers had a need for both growth and yield models (or tables), and it was important that these guides provided compatible results. There is also a need for compatible models of different resolution, and Daniels and Burkhart (1988) have demonstrated a model that produces compatible results at the stand, class and tree levels.

Precisely the same requirements may be extended to deterministic and stochastic models. Stochastic models offer great potential for risk analyses and better appraisal of investment options, but predictions from such models must be compatible with the deterministic models used in other aspects of planning. Computing capacity and operations research methodology do not yet enable stochastic models to be used for all planning and forecasting purposes; many applications demand the efficiency of a deterministic model. However, it is possible to design a compatible model which can operate deterministically or stochastically, the expectation or many-run average of the latter providing results identical to the former.

This concept of compatibility differs from the work of Buckman (1962) and Clutter (1963), who were concerned that yield estimates derived by summing growth forecasts should be compatible with direct predictions of yield (i.e., that growth and yield should be predicted from the differential and integral form of the same equation respectively). Several studies have confirmed the compatibility of forecasts made using alternative stochastic and deterministic implementations of the same mortality function (e.g., Ek 1980, Weber et al. 1986), but these studies have not been extended to models where stochastic increment and regeneration functions are included.

The Prognosis model (Stage 1973, Wykoff 1986) preserves heteroscedastic variation by incrementing individual trees stochastically while the aggregate stand increment is assumed to be deterministic. STEMS (Hahn and Brand 1979, Brand et al. 1988) accommodates random error through the mortality function and offers the user the choice of deterministic or stochastic mortality (Belcher et al. 1982). Gertner (1987) modified STEMS to enable stochastic predictions by adding conditional multivariate normal variates to the crown ratio and diameter increment predictions. Gertner (1987) also illustrated the use of error propagation methods to estimate precision of growth and yield simulations. The error propagation approach is more efficient than Monte Carlo methods for estimating precision, but may not satisfy all the possible uses for stochastic models.

DESIGN CONCEPTS

Deterministic growth models can be converted to stochastic models by adding normal variates to the predictions (e.g., Gertner 1987). However, the use of probabilistic functions offers simplicity, and eliminates a number of subjective assumptions inherent in other approaches. The principle that allows the construction of a compatible deterministic/ stochastic growth model is the realization that a prediction from a probabilistic function can be regarded as a proportion of a population (the expected outcome providing the deterministic case), or as a probability concerning an individual (the stochastic case). Since all components of growth and change in a forest stand can be expressed in terms of probabilities, a compatible deterministic/stochastic growth model can be created.

Mortality functions are routinely expressed as logistic functions (Hamilton 1974, 1980, Buchman 1979, Buchman et al. 1983) predicting fractions which may be used as proportions or as probabilities. Models such as STEMS (Belcher et al. 1982) exploit this in enabling the user to select either deterministic or stochastic mortality prediction. For stochastic mortality, a random number (between 0 and 1) is drawn, and if the predicted probability of death exceeds this random number the tree dies and the expansion factor is set to zero (otherwise the expansion factor remains unchanged). For deterministic mortality, the expansion factor is reduced by the proportion predicted by the mortality function. The same mortality function is used for both implementations.

Diameter increment (e.g., Lowell and Mitchell 1987), deteriorating merchantability (e.g., Vanclay 1991b) and regeneration (e.g., Ferguson et al. 1986) may also be predicted using probabilistic functions. Probabilistic prediction of diameter increment is equivalent to movement ratios which indicate the proportion of trees moving to the next size class in stand table projection methods of growth forecasting (e.g., Wahlenberg 1941, Davis and Johnson 1987). Lowell and Mitchell (1987) have demonstrated that diameter increment and mortality of individual trees can be estimated simultaneously using logistic functions. There is, however, a disadvantage with this simultaneous approach, as either the remeasurement interval of the data must correspond with the projection interval of the model, or the data must be interpolated accordingly. Hamilton (1974) argues the advantages of predicting survival rather than mortality, in that the n-year probability of survival is the nth power of the annual probability (providing that stand conditions remain constant), and this enables data with various remeasurement intervals to be used without interpolation. The important principle here is that the transient state (i.e., survival) rather than the absorbing state (mortality) should be modeled. The elegant simplicity of Hamilton's (1974) approach derives from the fact that there is a single transient state. Lowell and Mitchell's (1987) approach poses a more complex problem, as there is more than one transient state. One cannot simply modify it by modeling survival in same class, survival in the same or next class, …, and survival in any class, using cumulative probabilities and expressing mortality as the 99th state. An example indicates the bias in such an approach. Suppose the median increment is 0.5 cm per annum, the maximum 2 cm, the class size 1 cm, and there is no mortality. Then for any year, the probability of moving not more than 5 classes should equal 1, as no trees could grow more than 5 cm. But over 10 years, the probability of moving not more than 5 classes should be 0.5, and this implies an annual probability of 0.933 (i.e., $0.933^{10} = 0.50$). Thus it may be expedient to model movement and survival independently.

Modeling movement remains a challenging problem because of the several transient states required, and several options exist. Increment data can be interpolated to annual movements and summarized as a Markov matrix (e.g., Usher 1966, Harrison and Michie 1985) to be used as a lookup table in the model. Logistic equations based on annual movements could also be developed using Lowell and Mitchell's (1987) approach, but without including mortality. Another alternative is to use the actual observed movements, and use the states "remains in same class," "moves not more than one class each year," "moves not more than two classes each year," and so on. This formulation does not suffer the bias illustrated in the example above, but utilizes only

part of the information contained in the data. The best approach should utilize the fact a tree which moves one class in an n-year period, moves in one of those years, and remains in the remaining n - 1 years. This approach becomes quite complex for long intervals, especially if the class size is chosen such that trees can move more than one class during a single interval. In view of the complexity of the latter, the most viable approach may be interpolation to annual (or other suitable inter-al) movements, similar to the approach of Lowell and Mitchell (1987). Implementation is simplified if the interval and class size are chosen (sufficiently short and wide respectively) to ensure that the probability of trees moving more than one class during a single period is insignificant.

These ideas can be implemented in diameter class models (e.g., Ek 1974) and individual tree or tree list models (e.g., STEMS, Prognosis). Implementation in diameter class models simply involves replacing the movement ratio. For individual tree and tree list models. each record is considered a class, and implementation is analogous to the diameter class model. Instead of incrementing tree diameter through fractions of a millimeter, movement by standard (e.g., one centimeter) increments can be modeled probabilistically. As inventory data have finite precision (e.g., operational resource inventory in Queensland records diameters only to the nearest centimeter), there need be no loss of precision, and the heteroscedacity evident in the growth data would be preserved. Thus, in deterministic mode, a proportion of trees comprising each record would increment one centimeter each year (i.e.. the record would split into two new records). In stochastic mode, one or several random numbers could be drawn to see if the entire or corresponding several parts of the record would increment one centimeter. Where individual tree and tree list models involve few individual tree characteristics (e.g., species and diameter only), implementation is easy, and "housekeeping" to recombine similar records allows computational efficiency. However. models employing many individual tree characteristics (e.g., species, diameter, tree height, height to crown base, etc.) become computationally intensive as the number of tree records increases, and recombining of records is impeded by different combinations of these tree variables.

The use of probabilistic equations overcomes the subjective implementation of swindles such as record tripling. The Prognosis model may employ a swindle which effectively estimates the average of many replications without actually conducting the replications. The Prognosis swindle triples each tree record, the new records having 15, 60, and 25% of the original expansion factor, and an increment of $\mu-1.549\sigma$, $\mu-0.1423\sigma$ and $\mu+1.271\sigma$ respectively (Stage 1973). The share of the increment assigned to each new tree record is derived from the normal $N(\mu, \sigma)$ distribution, but the decision to triple (rather than to double or quadruple) and the 15:60:25 apportionment is rather subjective. Other models double records, apportioning 50:50 with 0.8 and 1.2 times the predicted diameter increment for *Callitris* forest (Vanclay 1988), and 75:25 with 0.9 and 1.3 times the predicted diameter increment for tropical rainforest (Vanclay 1989a). While the relative increments were gauged from the data, the apportionment (50:50 and 75:25) was subjective. Probabilistic modeling of diameter increment avoids these subjective decisions.

Other subjective decisions are inherent in the simulated plot size and in various housekeeping routines used in many deterministic growth models. Many growth models simulate on a per hectare or per acre basis, irrespective of the plot size of the user's data. Models may invoke housekeeping routines to reduce the number of tree records (e.g., the COMPRESS option in Prognosis, Wykoff 1986), and these may or may not be under user control. Other models avoid the proliferation of tree records by modeling increment deterministically (without doubling or tripling) and mortality stochastically (e.g., STEMS when expansion factor <1/ac, Belcher et al. 1982). These inherent plot sizes and housekeeping routines may significantly influence predicted outcomes, yet are often inaccessible to the user. Kimmins et al. (1990) argued that such calibration "twiddle knobs" should be kept to a minimum, and where inevitable should be controlled by the user, not hidden in the code. The approach outlined below offers one way to give that control to the user.

IMPLEMENTATION

By allowing the user to specify two parameters, a single per hectare growth model becomes flexible enough to efficiently produce deterministic and stochastic results consistent with any simulated plot size. These can be specified by the user at run time, so that consecutive simulations can alternately be stochastic and deterministic, with varying plot sizes, using the same model with identical functions to ensure compatible results.

One number (m) specified by the user indicates the minimum expansion factor, and thus reflects the plot size being simulated (or the resolution required), and controls the proliferation of minute proportions during deterministic operation. Thus for high resolution large plot simulations, m should be small, while for lower resolution or smaller plots, m may be large.

The other number specified (n) determines the maximum number of random numbers that will be drawn for any tree record. If the expansion factor exceeds $n \times m$, the simulation will deterministically employ the predicted proportion of the class rather than taking repeated stochastic events. In Prognosis, Stage (1973) assumed that the stand-level aggregate of many stochastic projections of individual trees would be deterministic. The converse is used here: there are computational advantages in using a deterministic outcome rather than taking repeated stochastic outcomes where a large number of trees is represented by a single record.

In determining the action to take for each cohort, the expansion factor is compared with the user's specified values of m and $n \times m$. If the expansion factor exceeds $n \times m$, the predicted proportion of the expansion factor will advance, die or deteriorate (except that when $n = 0$ and expansion factor $<m$, movement probabilities are accumulated until 1.0 is attained, when the tree size is incremented, to avoid further splitting tree records with expansion factors $<m$). If the expansion factor is less than m and $n>0$, a random number will be drawn, and the fate of all the trees in that cohort depend upon the relativity between the random number and the predicted probability. If the expansion factor lies between m and $n \times m$ $(n > 0)$, then several (expansion factor/m, rounded up) random numbers will be drawn, and each m (or part thereof) trees in the cohort will be treated stochastically.

Regeneration may also be predicted stochastically or deterministically using probabilistic equations. If the model is in deterministic mode $(n = 0)$, probabilities are accumulated until unity is reached and regeneration initiated. In stochastic model $(n > 0)$, random numbers are drawn to determine the incidence of regeneration in any year.

When $n = 0$, the model will behave deterministically for any minimum expansion factor (m). With $n = 1$ and m small, results will also be near deterministic with high resolution, similar to some existing deterministic models (e.g., Prognosis and STEMS). When $n > 0$, the model draws a random number for each m (or part thereof) trees represented in each class and takes the stochastic outcome. When m is large, the resolution of the predicted outcome is more granular, as is appropriate when simulating a small plot. Parameters m and n may take any positive value. The choice of m and n influence computational efficiency, and for stochastic simulations, n may be as small as 5, as little increase in variability is observed as n increases.

The serial and seasonal correlation of tree growth is well established and should be accommodated when drawing random numbers during stochastic simulations. Stage (1973) preserved random variables associated with each tree record, so that appropriate serial correlation could be maintained during the next simulation cycle. Within the present framework, appropriate serial correlation can be achieved by assigning a random number r_{i*} to each record for the duration of the simulation, drawing a random number r_{*j} for each cycle, and taking a third random number r_{ij} for the interaction of individual and season. The final variable for determining the stochastic action to be taken should be a weighted average of these three random numbers, weighted appropriately for the species and location. Preliminary analyses for natural forests in Queensland suggest that appropriate weights may be 0.6, 0.2, and 0.2 for tree, year, and interaction respectively.

EXAMPLE

The Queensland rainforest growth model (Vanclay 1989a) has been extensively modified to enhance its accuracy, flexibility, and scope (Vanclay and Preston 1989), and is now being further modified to accommodate the proposals outlined above. Previous versions have employed a tree list approach with each tree record indicating species, size (cm diameter), expansion factor (number/ha), and merchantability (veneer log, sawlog, useless). The new version is similar, except that size is expressed as an integer. The new growth function is a logistic function predicting movement to bigger size classes and replaces the previous diameter increment function. Mortality, deterioration, and regeneration were predicted using logistic functions which could be employed without alteration. Serial and seasonal correlation have not yet been implemented in the model.

The model uses 41 equations to predict the diameter 'increment of the 400-odd species recognized on inventory plots in north Queensland. The equations are illustrated for *Flindersia pimenteliana,* the most abundant species in the permanent sample plot database used to estimate the equations. The diameter increment function previously used to simulate the growth of this species was (Vanclay 1991a):

$$Log(DI + 0.02) = -0.2354 -0.06056DBH +0.9673Log(DBH)$$
$$+0.08851SQLog(DBH) -0.9366Log(SBA)$$
$$- 0.02684OBA + 0.1415PS \qquad (1)$$

where DI is diameter increment (cm/ann), DBH is tree diameter (cm), SQ is site quality (Vanclay 1989b), SBA is stand basal area (m^2/ha), OBA is overtopping basal area (m^2/ha) defined as the basal area of stems whose diameter exceeds that of the subject tree, and PS is a dummy (0, 1) variable which, for this species, takes the value 1 if the plot is located on soils derived from recent alluvial, volcanic, or granitic parent materials, and zero for soils derived from sedimentary or metamorphic parent materials.

The same attributes are useful for predicting probability of movement. The following logistic equation was fitted using maximum likelihood estimation:.

$$Z = +0.7378 -0.1079DBH +1.987Log(DBH)$$
$$+0.1455SQLog(DBH) -1.994Log(SBA)$$
$$-0.03548OBA +0.4221PS \qquad (2)$$

and

$$P\{Int[DBH_0 + (DBH_n - DBH_o)/n] > Int[DBH_0]\} = (1 + e^{-z})^{-1}$$

where P is the probability that a tree completes a centimeter of growth within that year (i.e., grows from $<x$ to $\geq x$ cm diameter, where x is an integer number of centimeters), DBH_0 is the initial diameter, DBH_n is the final diameter, n is the number of years, and Int rounds down to an integer value. One further advantage of this approach is that the logistic equation is more robust in the presence of outliers and decrements in the data, enabling data to be used in estimating Equation (2), which Vanclay (1991a) had withheld from linear regression analyses for Equation (1) (e.g., $DI \leq -0.02$).

TABLE 1. Effect of varying values of *m* and *n*.

Minimum expansion factor (m)	Stand basal area (m^2/ha) of Plot 01/88/0079 after 100-year simulation. Mean (and standard deviation) of 100 simulations with							
	n= 0		*n = 1*		*n= 5*		*n = 20*	
100.	42.2	(0.0)	41.5	(6.0)	-†		-	
10.	42.2	(0.0)	41.9	(3.7)	43.0	(3.7)	-	
1.	42.3	(0.0)	42.7	(1.3)	43.0	(1.7)	42.5	(1.4)
0.1	42.1	(0.0)	42.8	(0.5)	42.6	(0.7)	42.6	(0.7)
0.01	42.1	(0.0)	42.6	(0.6)	42.7	(0.5)	42.8	(0.6)

† These combinations ineffective as maximum expansion factor <50 during simulations. Previous deterministic tree list model (Vanclay and Preston 1989) prediction: 42.3
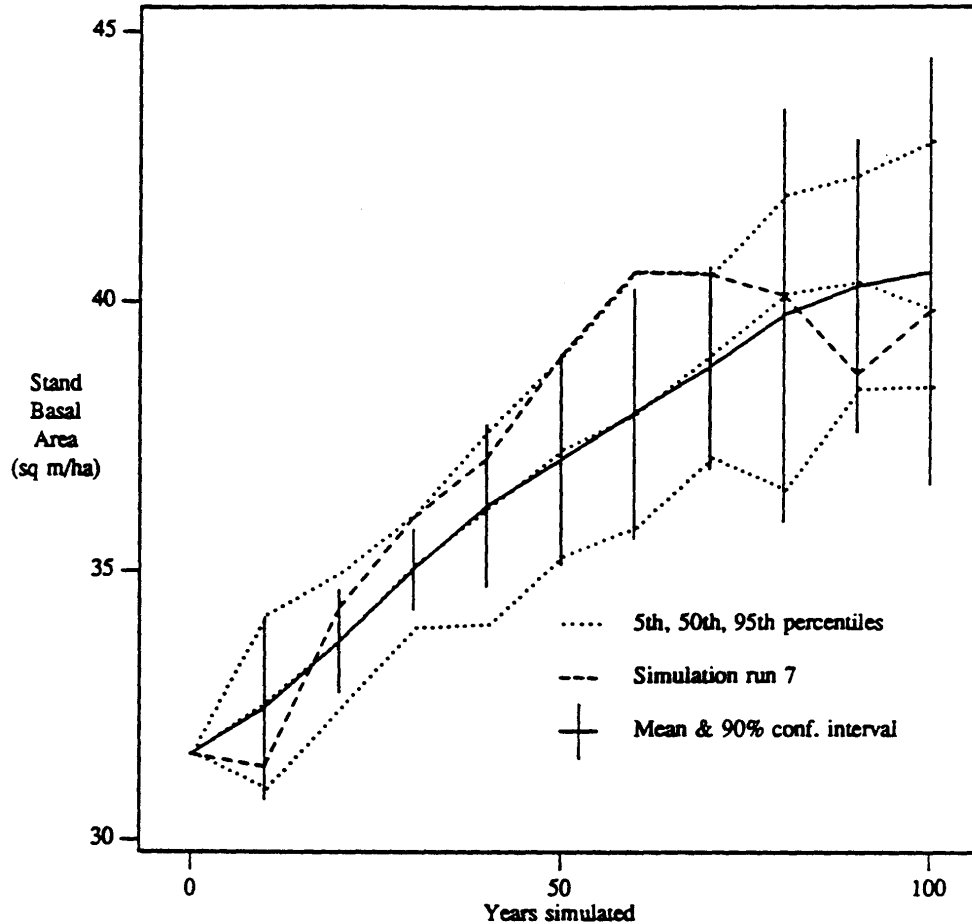
FIGURE 1. One-hundred-year simulations for plot 01/88/0079

Table 1 shows the effects on simulated future stand basal areas, of various values of parameters *m* and *n*. The plot (01/88/0079) used for this example was a 100 meters square (1.0 ha) and measured all trees over 40 cm diameter. In addition, five point samples (2 m^2/ha BAF) were used to sample trees over 10 cm (but less than 40 cm) diameter. A total of 108 trees of 33 species were sampled. Mean predictions are quite similar for a range of *m* and *n,* and compare favorably with predictions made from a previous deterministic model (Vanclay and Preston 1989). The minimum expansion factor (*m*) has a major influence on the variability of predictions, while the number of random numbers drawn for any tree record (*n*) has relatively little impact on predictions (Table 1). Standard deviations are likely to be underestimated due to the absence of seasonal and serial correlations in the present model.

Figure 1 shows 100-year predictions for this plot, made with *m* = 1 and n = 5. The 5th, 50th, and 95th percentiles are the smallest, median, and largest of 19 simulations and give a nonparametric 90% confidence interval for predictions. These were obtained by ranking outcomes at each 10-year interval, and individual simulation runs may vary between these limits (e.g., run 7, illustrated with the dashed line). The mean and 90% confidence intervals derived from the standard deviation are also shown. The parametric and nonparametric confidence intervals are in reasonably close agreement.

LITERATURE CITED

BELCHER, D.M., M.R. HOLDAWAY, and G.J. BRAND. 1982. A description of STEMS. USDA For. Serv. Gen. Tech. Rep. NC-79. 18 p.

BRAND, G.J., M.R. HOLDAWAY, and S.R. SHIFLEY. 1988. A description of the TWIGS and STEMS individual tree-based growth simulation models and their applications. P. 950-960 in Forest Growth Modelling and Prediction, Ek, A.R., S.R. Shifley, and T.E. Burk (eds.). Proc. IUFRO Conference. USDA For. Serv. Gen. Tech. Rep. NC-120. 1149 p.

BUCHMAN, R.G. 1979. Mortality functions. P. 47-55 in A generalized forest growth projection system applied to the Lake States region. USDA For. Serv. Gen. Tech. Rep. NC-49. 96 p.

BUCHMAN, R.G., S.P. PEDERSON, and N.R. WALTERS. 1983. A tree survival model with application to species of the Great Lakes region. Can. J. For. Res. 13:601-608.

BUCKMAN, R.E. 1962. Growth and yield of red pine in Minnesota. USDA For. Serv. Tech. Bull. 1272.

CLUTTER, J.L. 1963. Compatible growth and yield models for loblolly pine. For. Sci. 9(3):354-371.

DANIELS, R.F., and H.E. BURKHART. 1988. An integrated system of forest stand models. For. Ecol. Manage. 23:159-177.

DAVIS, L.S., and K.N. JOHNSON. 1987. Forest management. Ed. 3. McGraw-Hill, New York. 790 p.

EK, A.R. 1974. Nonlinear models for stand table projection in northern hardwoods stands. Can. J. For. Res. 4(1):23-27.

EK, A.R. 1980. A preliminary trial of alternative methods for treating mortality in the multi-purpose forest projection system (MFPS) model. Univ. Minn. Coll. For. Dep. For. Resour. Staff Pap. Ser. No. 8. 6 p.

FERGUSON, D.E., A.R. STAGE, and R.J. BOYD. 1986. Predicting regeneration in the grand fir-cedar-hemlock ecosystem of the northern rocky mountains. For. Sci. Monogr. 26. 41 p.

GERTNER, G. 1987. Approximating precision in simulation projections: An efficient alternative to Monte Carlo methods. For. Sci. 33(1):230-239.

HAHN, J.T., and G.J. BRAND. 1979. An implementation program. P. 61-78 in A generalized forest growth projection system applied to the Lake States region. USDA For. Serv. Gen. Tech. Rep. NC-49. 96 p.

HAMILTON, D.A. 1974. Event probabilities estimated by regression. USDA For. Serv. Res. Pap. INT-152. 18 p.

HAMILTON, D.A. 1980. Modelling mortality: A component of growth and yield modelling. P. 82-99 in Forecasting forest stand dynamics. Brown, K.M., and F.R. Clarke (eds.). Proc. Workshop. Sch. For., Lakehead Univ., Thunder Bay, Ont.

HARRISON, T.P., and B.R. MICHIE. 1985. A generalized approach to the use of matrix growth models. For. Sci. 31(4):850-856.

KIMMINS, J.P., K.A. SCOULLAR, M.J. APPS and W.A. KURZ. 1990. The FORCYTE experience: A decade of model development. P. 60-67 in Forest modelling symposium, Boughton, B.J., and J.K. Samoil (eds.). Proc. Symp. Forestry Canada Inf. Rep. NOR-X-308. 180 p.

LOWELL, K.E., and RJ. MITCHELL. 1987. Stand growth projection: Simultaneous estimation of growth and mortality using a single probabilistic function. Can. J. For. Res. 17:1466-1470.

STAGE, A.R. 1973. Prognosis model for stand development. USDA For. Serv. Res. Pap. INT-137. 32 p.

USHER, M.B. 1966. A matrix approach to the management of renewable resources, with special reference to selection forests. J. Appl. Ecol. 3:355-367.

VANCLAY, J.K. 1988. A stand growth model for cypress pine. P. 310-332 in Modelling trees, stands and forests, Leech, J.W., et al. (eds.). Univ. of Melbourne, Sch. of Forestry, Bull. No. 5. 559 p.

VANCLAY, J.K. 1989a. A growth model for north Queensland rainforests. For. Ecol. Manage. 27:245-271.

VANCLAY, J.K. 1989b. Site productivity assessment in rainforests: An objective approach using indicator species. P. 225-241 *in* Proc. Seminar on Growth and Yield in Tropical Mixed/Moist Forests. Wan Razali Mohd., H.T. Chan, and S. Appanah (eds.). For. Res. Inst. Malaysia. 250 p.

VANCLAY, J.K. 1991a. Aggregating species to develop diameter increment functions for tropical rainforests. For. Ecol. Manage. 42:143-168.

VANCLAY, J.K. 1991b. Modelling changes in the merchantability of individual trees in tropical rainforest. Commonw. For. Rev. 70(3): 105-111.

VANCLAY, J.K., and R.A. PRESTON. 1989. Sustainable timber harvesting in the rainforests of northern Queensland. P. 181-191 *in* Forest planning for people. Proc. 13th bienn. conf. of the Institute of Foresters of Australia. 232 p.

WEBER, L.A., A.R. Ex, and T.D. DROESSLER. 1986. Comparison of stochastic and deterministic mortality estimation in an individual tree based stand growth model. Can. J. For. Res. 16:11391141.

WAHLENBERG, W.G. 1941. Methods for forecasting timber growth in irregular stands. USDA Tech. Bull. 796. 55 p.

WYKOFF, W.R. 1986. Supplement to the User's guide for the stand prognosis model--version 5.0. USDA For. Serv. Gen. Tech. Rep. INT-208. 36 p.

AUTHOR AND ACKNOWLEDGMENTS