VALIDATION OF GROWTH MODELS USED IN FOREST MANAGEMENT

C. J. GOULDING*

ABSTRACT

The concept of validation is discussed in the context of models used for simulating the growth of trees. Validation attempts to increase confidence in a model's ability to provide useful and correct inferences about the growth of stands, rather than to prove a model is correct. The ability to predict levels in yield can be tested by comparison with sample-plot data or stands with known plot-histories. Confidence in the inferences drawn from the model's behaviour can be obtained from extensive running of the model using designed experiments.

INTRODUCTION

The past few years have seen a proliferation of computer simulation models used to predict some aspects of the growth and yield of trees and stands.[1] Once a model has been constructed and before the model or the results of the model can be used in practical forest management, the question of validity must be answered. This paper attempts to express the ideas formed over the past few years and used in validating a stand growth model for *Pinus radiata* D. Don in Kaingaroa State Forest (Elliott and Goulding, 1976). Some of the ideas were taken from Goulding (1972), who constructed a distance-independent individual-tree model for the growth of *Pseudotsuga menziesii* (Mirb.) Franco (Douglas fir) in British Columbia, Canada. Both of these models were aimed at predicting growth and yield for use in long-term timber management plans, but the concepts of validation used apply equally to those growth models with other objectives. The article should also be of use to the forest manager faced with the decision as to the applicability of the latest yield table or growth model.

Validation is but one step in the process of building a simulation model. Despite apparent differences, much of the

*Forest Research Institute, Private Bag, Rotorua.

[1]For a collection of articles, see Fries (1974). Newnham (1964) produced the first individual-tree computer simulation model; more recent examples are those by Mitchell (1975) and Daniels and Burkhart (1975).

philosophy behind building any model is similar, whether it be an operations research model of a logging and trucking system (*e.g.*, Bonita, 1972) or an ecological model of dynamic succession (*e.g.*, Bledsoe and Van Dyne, 1971). There are many texts on simulation; one of the older and better known is that by Naylor *et al.* (1966), from which Fig. 1 has been adapted. This shows validation as a distinct step in a multistage sequential process, but in reality many of the steps are interchangeable and will merge and overlap. In particular, finishing one model (especially one predicting the growth and yield of trees) does not mean the end of the project; rather, with changing goals, increasing computer power, and more data available, the objectives can be redefined and more components included.

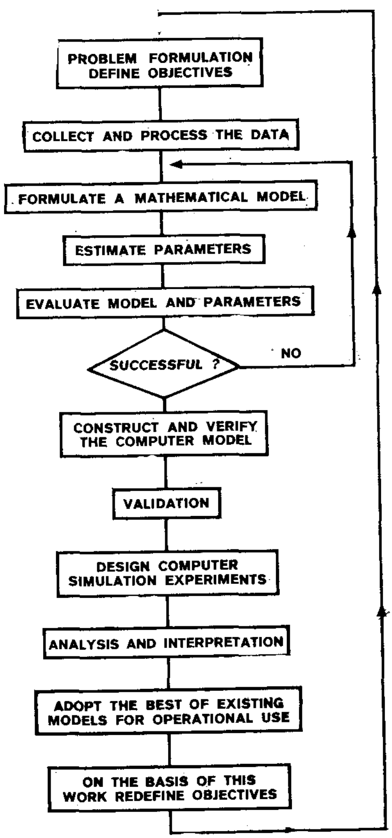

FIG. 1: *Stages in the design of a simulation model.*

This cyclic process depends to a large extent on the success of the validation and model analysis stages. Because of this, Fig. 1 provides a useful structure on which to base a description of a new growth model or yield prediction method, and any such description must include some discussion on all the stages, especially Validation, and Analysis and Interpretation.

DEFINITION OF VALIDATION

Figure 1 shows three stages which are generally considered part of validating a model. These are:

(1) Evaluate model and parameters;

(2) Verify the computer model;

(3) Validation.

These stages are equivalent to the approach proposed by Naylor and Finger (1967). First, construct a set of hypotheses and hence components of the model, using all available information including statistical estimation from data; secondly, attempt to verify any assumptions of the model by independent testing; thirdly, test the whole model's ability to predict the behaviour of the real system.

The stage "Evaluate model and parameters" occurs during the construction of the model. It refers to the process of independent experimental testing of the hypotheses and assumptions of the individual components of the model. The form of the equations representing the components, the way the model is structured, and the estimates of the parameters must be checked to ensure (a) that the components conform with the knowledge available from the literature, (b) that the parameters have been estimated using correct statistical techniques, and (c) that the estimation errors conform with any assumptions embodied by the techniques used on the basic data. The data must be checked to ensure that there are no latent or hidden variables in the analysis. It is at this stage that any deficiencies in the structure of the model should be found and corrected.

The stage "Verify the computer model" refers to the process of ensuring that the logic and arithmetic operations of the computer program are correct. In the final analysis, there is only one way to ensure that the program code has been verified, and that is to take one or more examples of input and output and work through each segment of the model by hand.

The "Validation" stage is a formal, independent process concerned with evaluating the model as a whole. Before this stage is entered there should be a degree of *a priori* confidence in the model. This is not enough, and the model as a whole needs to be tested because the effects of interactions between the components of the model will be as important as the effects of the components themselves. To quote Forrester (1960):

> The defence of a model rests primarily on the individual defence of each detail and policy, all confirmed when the total behaviour of the model system shows the performance characteristics associated with the real system.

It is this process of testing the model as a whole which is described in the remainder of the paper.

Van Horn (1969) defined validation as: "The process of building an acceptable level of confidence that an inference about a simulated process is a correct or valid inference about the actual process."

Note that the stress of the definition is on the inferences drawn from using the model rather than on the "correctness" or otherwise of the model itself. The validation procedure is therefore not a process of proving that the model is correct and can be used mechanically, but rather that the conclusions obtained and decisions made as a result of using the model are defensible and that the model's predictions are well grounded. Even (perhaps especially) a predictive yield table or model will have its figures modified by local experience or "recoverable yield percentages", and it will be the conclusions drawn from the predictive process that will require validating rather than the direct results.

Both Forrester (1960) and Van Horn (1969), although dealing primarily with simulation of the business system, give much of the philosophy of model construction, validation, and analysis that is directly applicable to growth models. They emphasise that the validity of a model cannot be divorced from the objectives for which the model was constructed. Caswell (1976) suggested that the design process of modelling is a search for agreement between the properties of the model and a set of demands placed on it by the designer. There are large differences between models designed primarily for predictive purposes and models designed to gain a better understanding of the system, and the validation procedure must differ accordingly. Growth models for forest management tend to be of the former category; in fact, they will often be a black-box in a much larger forest-management system (which itself will require validating if the system is a simulation model). Invariably comparisons will be made between regimes in terms of yield, but, provided changes to the growth model's input result in valid changes to the output, why these changes occur in terms of fundamental biological principles will often be answered only outside the direct involvement of the model. This is because growth and yield models for management tend to work at a coarse level of resolution of a biological system. The variables of interest are only parts of the biological unit, *e.g.*, stem volume inside bark or the cross-sectional area of the stem at 1.4 m above ground. The fundamental biological processes of growth and competition (*e.g.*, ion-exchange between competing roots and the soil, or light absorption by needles) are not represented in the model, nor often are variables specifically incorporating root, branch, or needle growth. Management growth models are unlikely, therefore, to explain why a stand of trees grows the way it does, except in the broadest, most empirical terms.

It is very unlikely that a model can be proved to be valid. Most of the tests suggested attempt to disprove the model. Because the model is an artificial system with a domain of applicability, at some level of resolution of behaviour or with some particular enquiry a model will fail. The validation procedure suggested for growth models is therefore a sequence of tests in order, increasing in severity until the model fails. If the model's domain encompasses the region of interest defined by the objectives, and if decisions made on the basis of inferences drawn from the analysis of the model improve forest management, then the growth model can be said to be valid. Although some statistical tests are suggested, these are limited in scope and some emphasis must be placed on qualitative value judgements.

OBJECTIVES

Growth models for use in long-term management planning tend to have as their objectives primarily the prediction of levels of yield, and secondarily the understanding of to what extent changes in treatment affect yield and average size. The growth model for *P. radiata* at Kaingaroa had such typical objectives, that is:

(1) To be able to predict the level of total volume yield, basal area, top height, and stems per hectare at any age after the initial establishment period over the site qualities encountered;

(2) To be constructed in a form suitable for use in a forest management system; it must be easy to use, speedy in execution time, and use average or total stand variables for input and control;

(3) As far as possible to predict differences in the effects of different regimes over all parts of the forest.

The model is a stand model, using average or per hectare values for input and output. Various equations predict the annual growth and mortality as a function of standing values, and these predictions are then added to the standing values to obtain a new set at age +1. The model predicts the effects of thinning but it is limited in that it excludes the direct effects of pruning, spatial distribution of trees, and fertiliser application except as indirect influences on site index, on stand density, or on the average condition of the data used to construct the model.

The aims of the validation procedure used were to answer two questions:

(1) How good is the model at predicting levels of growing stock?

(2) How close is the model's behaviour to reality when stands are influenced by thinning and initial spacing?

It is important to distinguish between the two questions. Figure 2 illustrates the growth patterns of two growth models simulating the development of a permanent sample plot. If the comparison was based on a rotation age of 30 years, model 2 would be judged the better, but the shape of the growth patterns in model 1 more nearly matches that of the permanent sample plot, despite the initial abnormal mortality. Inferences drawn from model 2 about the age of maximum mean annual increment are likely to be wrong. However, in practice a comparison with just one plot is meaningless owing to the very high variability that can be encountered. particularly when the plot is sampling a stand rather than laid out in a carefully designed experiment. Conclusions can be drawn only from an adequate-sized sample.

Testing a model's predictive ability requires comparative data of good quality. Some of the results of this testing go towards testing the inferences drawn from the model's behaviour, but a thorough testing involves running a model through an extensive experimental design and comparing the predictions with known experimental results, theory, and practice. This is discussed later.

TESTS OF A MODEL'S PREDICTIVE ABILITY

These tests can make use of statistical analysis to a large degree. Known starting points of different stands represented by sample plot data are input to the model, growth is simu-

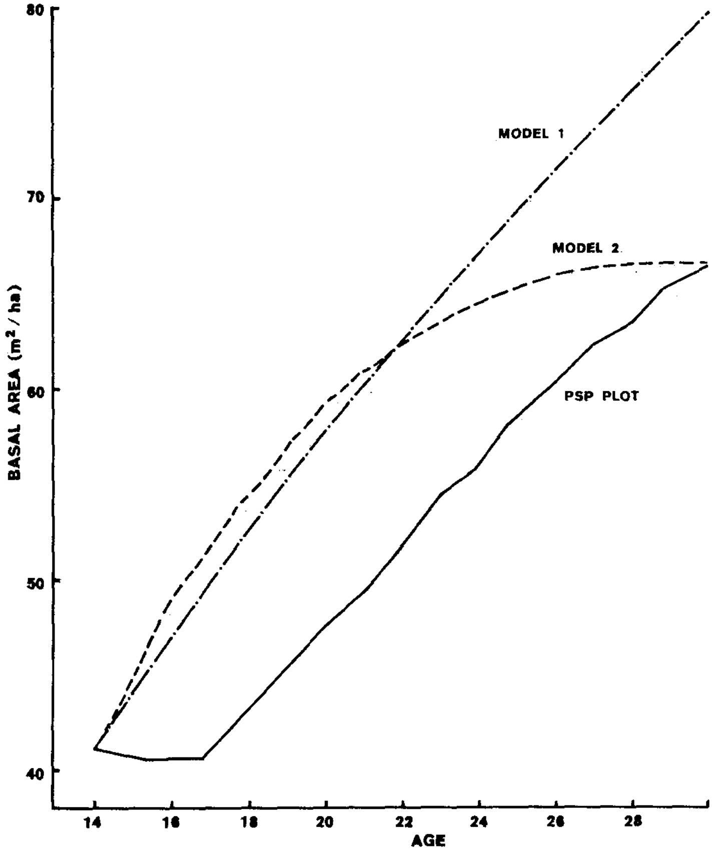

FIG. 2: *Illustration of the difference between prediction and behaviour.*

lated over the lifetime described by data (including treatments), and the simulated results are compared with the actual values. Both permanent sample plot data and data from detailed forest inventories can be used.

With permanent sample plots, several measurements over long periods of time are required. With *P. radiata* rotations of 20 to 45 years, a 15- to 20-year span was considered the minimum desirable. Plots measured over shorter periods have considerable variability because of measurement and sampling error and, importantly, do not allow growth patterns to be compared (see Fig. 2).

McEwen (1976) indicated that volumes calculated from permanent sample plot data were not accurate, with errors of up to 6% at older ages. Contrary to popular belief, there was no consistent bias in consecutive measurements and the periodic volume increment was poorly measured with errors as high as +71%. The effect of this imprecision in the measurements of growth can be avoided only if enough sample data are used. It implies that tests of the precision of a model are likely to prove misleading.

Many of the data used for testing a model tend to have been used in estimating the parameters of the model. Quite obviously it would be better if the test data were completely independent from the estimation data, and in a model consisting of completely independent regressions, test data should not have been used to estimate parameters. Similarly, where a model has been "calibrated" by comparing simulated results with real growth trends and deriving adjustment factors used within the model (*e.g.*, Mitchell, 1975) data for testing *must* be totally independent of those used in the calibration procedure, and a formal validation procedure after calibration is mandatory. However, many simulation growth models are feedback models; that is, the change over one time interval is predicted and added to the stand values at the beginning of the interval to obtain the stand values at the beginning the next interval. The individual regressions may fit the data reasonably well for predictions of changes over one time interval, but the model as a whole may be in error owing to the effects of interaction between errors in individual components and compounding of errors over time. If the feedback is positive, a 5% bias in one component (unlikely to be detected as significant during regression analysis, for example) will result in a final bias of about 100% after only 15 time intervals. With the current trend in New Zealand Forest Service regimes of heavy early thinning to final stocking, a simulation model will be predicting little or no mortality and may well be operating in a positive feedback situation with regard to basal area growth. When long-term plots are used for testing, the data will be used in an entirely different way from those used for parameter estimation, and compounding of errors can be detected. This is less true when there is only a short period between the first and last measurements of a plot.

One drawback of using the same data occurs when they come from one or two major sources or data collection systems. Any data system errors will not be discovered during testing; it will be assumed that the data truly represent the real world, and this could well be an invalid assumption, especially when only research plots are used. Setting aside a random selection of plots from the data base to use for testing does not overcome this problem; if the data coverage is complete with adequate replication, a random selection will be very similar to the data base and the additional benefits in testing will be marginal. However, where some of the plots are obviously special cases and are directly related to the type of stands for which the model is to be used (*e.g.*, plots from other organisations or from different data base systems), it may be worth setting aside these particular plots for use in the validation procedure.

An interesting idea is contained in the Jack-knifing (Miller, 1974) and Cross-validation techniques (Stone, 1974); it involves estimating the components of the model by partitioning the data into sets, each of which has one observation omitted. These techniques could prove useful with smaller incomplete sets of data where all the data could be used to estimate the parameters, and the dangers caused by the model being dependent on one or two observations at the extremes of the data set could be avoided. The system of components as a whole making up the model would still require validation, but the data would be used quite differently in testing from the way they were used for estimation.

Therefore, although an independent set of data for the validation procedure is preferable, when a model is of the feedback type and is constructed of many individually estimated and interacting components (and has not been "calibrated"), the validation of the model as a whole can be carried out on data that were used in estimating the parameters of the independent components. At the very least the model should compare well with the original data, and this is not otherwise necessarily true for these types of models.

The stand development of each of 12 to 15 permanent plots with reliable measurements and distributed over the model's domain can be compared by initialising the model with the same values as the first measurement of the plot and simulating the growth and treatment over the lifetime of the plot. Fewer plots make it difficult to determine significant differences between actual and predicted values because the confidence interval of the mean estimated error becomes too large. High variability in the errors may obscure significant differences in all the following tests:

(a) Graph actual and predicted values against time for each plot for each of the major variables of interest. This is best done by hand to allow time for judgement and deliberation.

(b) Perform paired t tests on errors in final predictions of net growth over the whole period of measurement for a group of plots with similar lengths of time between the first and last measurements. If the average error over all plots is significantly different from zero, check whether the bias is large by practical standards. If the bias cannot be found to be significantly different from zero, check whether the variability of the errors is so high that it is impossible to precisely determine what is the average error. The use of the χ^2 test as suggested by Freese (1960) is not recommended in this situation; it is a test of precision rather than accuracy. The variability of the errors is due to errors in determining actual growth, to sampling errors inherent in the use of a plot as a sampling unit of a stand, as well as to the lack of precision in the model's estimates from one plot to another. Moreover, the test is highly susceptible to non-normality in the data in contrast to the more robust t test (Box, 1953). Where a model predicts tree size distributions, these can be tested using the Kolmogorov-Smirnov test against the actual size distribution.

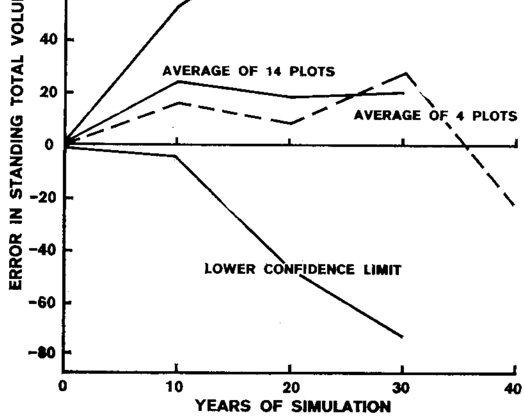

FIG. 3: *Results of paired* t *tests over 10, 20, and 30 years of simulation (from Goulding, 1972).*

Similar tests can be made for errors in growth over fixed intervals of time to ensure that there is no trend in the average error with the length of simulation. In Fig. 3 the mean error in standing total volume (equal to the error in net growth over the time period as these plots were unthinned) is plotted against length of simulation for 14 plots and also for a subset of 4 plots measured over 40 years. The average growth of all the plots was 374 m³/ha over the 30-year period. The graph shows that, on the average, errors did not accumulate over time. The increasing width of the confidence limits for the mean error indicates an increase in the variability of the errors of individual plots and shows the difficulty of simulating the development of an individual sample plot, especially (as in this example) because of the unpredictability of tree mortality.

(c) Perform simple linear regression of actual growth as a function of predicted growth. The regression should be significant, the slope coefficient not significantly different from 1, and the intercept not significantly different from zero. The plots in this trial must come from a wide range of conditions and have different amounts of growth over the period. If they effectively belong to one population, then the different values for growth over the interval of measurements may be random observations about a common mean and the regression will not be significant, despite the model predicting the average growth of the population very well.

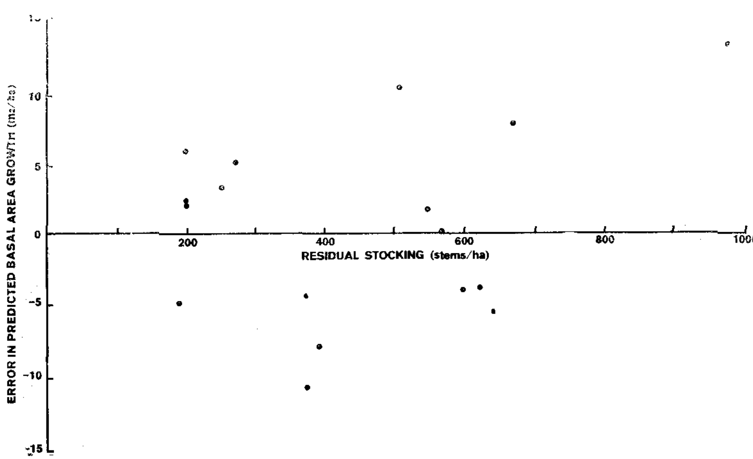

FIG. 4: *Errors in basal area growth predicted by the Kaingaroa growth model v. residual stocking after last treatment.*

(d) Perform regression analysis of errors in net growth prediction as a function of stand values—for example, site index, residual stocking after last treatment, or thinning intensity. Figure 4 illustrates the errors in net basal area growth (including any thinnings) predictions compared with the stocking after final treatment for plots measured over 10 to 20 years with an average net growth of 36 m^3/ha, as predicted by the Kaingaroa growth model. If the model is biased, the relationship between the bias and the stand value may be more complex than a simple linear relationship, and multiple regression may be necessary to obtain a significant relationship. If such a regression were significant it would indicate that the model was biased for various treatments or subsets of the data.

The last test (d) can be performed on plots with a shorter interval between first and last measurements than the other tests if it is important to widen the coverage of the test data by increasing the number of test plots available.

The advantage of tests of this sort ((a) to (d)) is that they allow precise comparison of the model with measured trends. However, there are disadvantages in addition to the usual lack of data coverage. Management growth models will be used to predict yields where the stand conditions are not well known, sometimes based only on prescriptions rather than measurements and often starting at the age of establishment. Testing the model against permanent plots may not test the "stand generation" capabilities of the model and the interactions of errors in that component with errors in the growth routines. Moreover, plots tend to be well maintained, established in fully stocked parts of the stand, away from stand edges or gaps, and with treatments tightly controlled. All these factors can tend to produce overestimates of yield (see Bruce, 1977). If a model is to be used for management planning, especially at the tactical level, then the model should be tested against actual stands managed for production.

A similar exercise to that used on permanent plots (tests (a) to (d)) can be carried out to predict the standing values of some reasonably mature stands (with *P. radiata*, say over 18 years old), although there is little point in reproducing the life history of regimes that are very different from current ones. The growth model program will require control cards containing data on treatment, and so the past history of each stand in the exercise must be known reasonably well—*i.e.*, numbers surviving after establishment and details of any treatments and thinnings necessary (in the Kaingaroa growth model this includes the age of thinning and the residual stocking—the numbers and volume removed need not be known, though obviously actual and predicted thinning yields could not then be compared). The standing crop values at the oldest convenient age must be determined precisely (this is more important than in an ordinary inventory). Each stand can then be simulated from the starting point and the prescription until the age of last measurement, and the predicted final values compared with assessment values. If the predicted value falls within the confidence limits of the assessment of the standing crop, then there is no justification for stating that the error is different from zero; hence the need for precise inventory.

In such a trial there is usually no opportunity to check the growth curve over time. If the prescribed treatment differs from what was actually carried out, if there was a change in site index over time, or if the stand is not uniform with respect to site and treatment, then it is unlikely that the growth model will predict final values at all well. This implies that stands selected for such a trial may have to be smaller in area than would normally be thought of for management. Any such problems with the tests may indicate need for improvement in the actual management—that is, the operational control of treatments or the delineation of stands.

TESTS OF THE MODEL'S BEHAVIOUR

One can never prove that two "machines" are identical just by comparing input-output transformations, no matter how large a sample is used (Van Horn, 1969). As the objectives of growth models for management tend to go beyond mere prediction of levels and require some information on differences between treatments, analysis of the model's behaviour must be carried out. This analysis must demonstrate what inferences about the yield of the species under study can be drawn from the model when the stands are subjected to differing treatments. Any model will have such inferences which go far beyond those originally conceived when the individual components were designed and the parameters estimated from data.

For example, from Fig. 2 the relationship between the mean annual increment of net basal area and age differs markedly between models 1 and 2. This relationship is derived from the model as a whole and could not have been predicted from one of the components alone. When either of the models or their results are incorporated in other work, such as planning the long-term cut of a forest estate or optimising the economic value of a given regime, these properties may have a substantial influence on the final results. It is therefore very important first to discover what are the inferences about growth behaviour incorporated in a model, and then to validate these inferences. In this aspect, the comments of Forrester (1960) are directly applicable in that qualitative judgement may have to be relied upon to a very large extent.

A growth model can be viewed as a subject on which detailed experiments may be carried out. These experiments must be well planned; trials of random ideas or minor variations of the latest regime are unlikely to be enlightening. When a model has random elements, the experimental design and analysis become critical; otherwise the user will become swamped by a large number of growth predictions resulting in confusing and conflicting opinions. Even with deterministic models the use of standard statistical designs is desirable. In particular, response surface designs, usually second order, appear appropriate (see Mead and Pike, 1975; Myers, 1971). The variable of interest predicted by the model, *e.g.*, total volume yield, is expressed as a function of independent variables depicting treatment. Goulding (1972) used a second-order central composite design to examine the effect on maximum mean annual increment of gross volume per hectare of stems per hectare at age 20, site index, and thinning intensity.

There exists the problem of multiple response where no single variable in particular is of interest. For example, total volume production at a given age is meaningless alone. Value judgements about the growth and yield of regimes are based on several parameters. The tree size distribution, the maximum mean annual increment of volume, and the merchantable yield/age relationship are the major response variables, with tree shape, numbers of stems, and basal area also being of interest. Predicted values of these parameters will be taken in combination according to the objectives of the study and the type of crop being grown. The varying degrees of error in the predictions of the parameter values must also be taken in combination.

For *P. radiata* in New Zealand several experiments are possible for determining the response of the above parameters to:

(a) Initial stems per hectare, unthinned, and site index.

(b) Initial stems per hectare, site index, and residual stocking in a regime with one or two early thinnings to waste to achieve final stocking.

(c) Thinning intensity or level of main-crop growing stock (perhaps expressed as relative spacing (Beekhuis, 1966) or percentage of maximum mean annual volume increment removed (Johnston *et al.*, 1967)) and site index for a regime of multiple thinnings from a given initial stocking.

(d) Level of volume removed and site index for a given regime consisting of just one or two commercial thinnings.

Note the inclusion of site index or some other aspect of site quality in all the above regimes. Small differences in the results of treatments are unlikely to be significant. It should be obvious that many other experiments are possible and the objectives of the model must be borne in mind. If the experiments cover a wide enough range the model's predictions will become unreliable and areas outside the domain of the model's applicability will become apparent. Figure 5 illustrates part of the results of an experiment with the Kaingaroa growth model, similar to (b) above. Only one site index is shown, for an initial stocking of 1500 stems/ha.

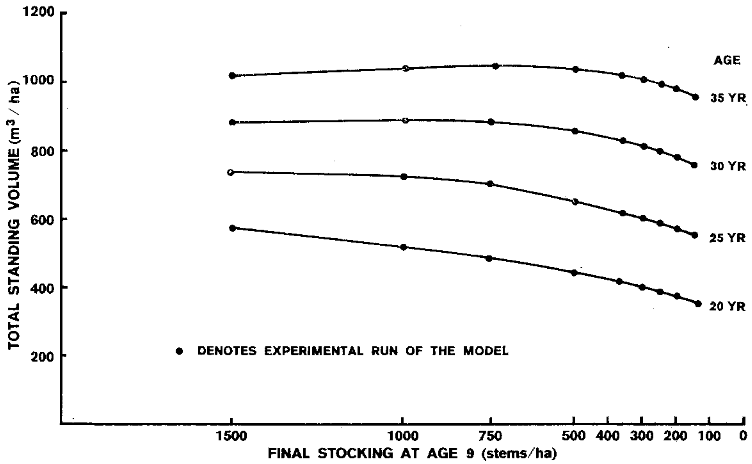

FIG. 5: *Response curves of total standing volume after thinning to final stocking at age 9 (site index 30 m).*

The results of the experiments can be summarised and compared with known empirical data or existing theory, the latter being unfortunately voluminous. Outside expertise may be called upon, and in fact may be necessary owing to the obvious bias of the model builder. Van Horn (1969) suggested a "Turing" test in which output from the model's experiments was compared with actual data by independent experts and any flaws or artefacts commented upon. For growth models the independent experts would be local foresters. In view of the interaction of spacing, site index, and treatment (see Assman, 1970) and the fact that individual experimental trials are often located on only one site, considerable confusion exists in the literature as to the response of the crop to treatment. If the model differs from theory it could be the theory that is wrong.

CONCLUSIONS

Following the steps of construction of a model with a high *a priori* validity, verifying that the logic and arithmetic of the program are correct, and after validation and analysis, the experimenter should be able to state whether a given trend in response to differing treatments predicted by the model is an artefact of the model or is a trend exhibited by the real system. The combination of age, site qualities, and treatments where the model appears to predict reliable results (here termed the model's domain of applicability) should be established and areas of weakness highlighted. It is hoped that the above ideas prove useful with other growth models and that future descriptions of new simulation models include extensive sections on validation and analysis.

REFERENCES

Assman, E., 1970. *The Principles of Forest Yield Study.* Pergamon Press, Oxford. 506 pp.

Beekhuis, J., 1966. Prediction of yield and increment in *Pinus radiata* stands in New Zealand. *N.Z. For. Serv., For. Res. Inst. Tech. Pap. No. 49.* 40 pp.

Bledsoe, L. J.; Van Dyne, G. M., 1971. A compartment model simulation of secondary succession. In Patten, B. C. (ed.), *Systems Analysis and Simulation in Ecology. I*: 488-511. Academic Press, New York.

Bonita, M. L., 1972. *A Simulation Model for Planning and Control of Forest Harvesting Operations.* Ph.D. thesis, Univ. of British Columbia.

Box, G. E. P., 1953. Non-normality and tests on variances. *Biometrika, 40*: 318-35.

Bruce, D., 1977. Yield differences between research plots and managed forests. *J. For., 75* (1): 14-7.

Caswell, H., 1976. The validation problem. In Patten, B. C. (ed.), *Systems Analysis and Simulation in Ecology. IV*: 313-25.

Daniels, R. F.; Burkhart, H. E., 1975. Simulation of individual tree growth and stand development in managed loblolly pine plantations. *Div. For. & Wildlife Resources V.P.I. S.U.,* FWS-5-75. 69 pp.

Elliott, D. A.; Goulding, C., 1976. The Kaingaroa growth model for radiata pine and its implications for maximum volume production. *N.Z. Jl For. Sci., 6* (2): 187.

Forrester, J. W., 1960. *Industrial Dynamics.* MIT Press, Cambridge. 464 pp.

Freese, F., 1960. Testing accuracy. *For. Sci., 6*(2): 136-45.

Fries, J., 1974. *Growth Models for Tree and Stand Simulation. Proceedings.* Research Notes 30, Royal College of Forestry, Stockholm. 379 pp.

Goulding, C. J., 1972. *Simulation Techniques for a Stochastic Model of the Growth of Douglas fir.* Ph.D. thesis, Univ. of British Columbia. 233 pp.

Johnston, D. R.; Grayson, A. J.; Bradley, R. T., 1967. *Forest Planning.* Faber, London. 541 pp.

McEwen, A. D., 1976. *An Examination of the Accuracy of the Data Collected from the Yield Research Plots of the Forest Research Institute.* Ph.D. thesis, Univ. of Canterbury. 452 pp.

Mead, R.; Pike, D. J., 1975. A review of response surface methodology from a biometric viewpoint. *Biometrics, 31*: 803-51.

Miller, R. G., 1974. The Jack-knife—A review. *Biometrika, 61*(1): 1-16.

Mitchell, K. J., 1975. Dynamics and simulated yield of Douglas fir. *For. Sci. Monograph 17.* 39 pp.

Myers, R. H., 1971. *Response Surface Methodology.* Allyn and Bacon, Boston. 246 pp.

Naylor, T. H.; Finger, J. M., 1967. Verification of computer simulation models. *Man. Sci., 14*(2): 92-101.

Naylor, T. H.; Balintfy, J. L.; Burdick, D. S.; Chu. K., 1966. *Computer Simulation Techniques.* Wiley, New York. 352 pp.

Newnham, R. M., 1964. *The Development of a Stand Model for Douglas fir.* Ph.D. thesis, Univ. of British Columbia. 201 pp.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. B., 36*(2): 111-47.

Van Horn, R., 1969. Validation. In Naylor, T. H. (ed.), *The Design of Computer Simulation Experiments*: 232-5. Duke Univ. Press, Durhan N.C.