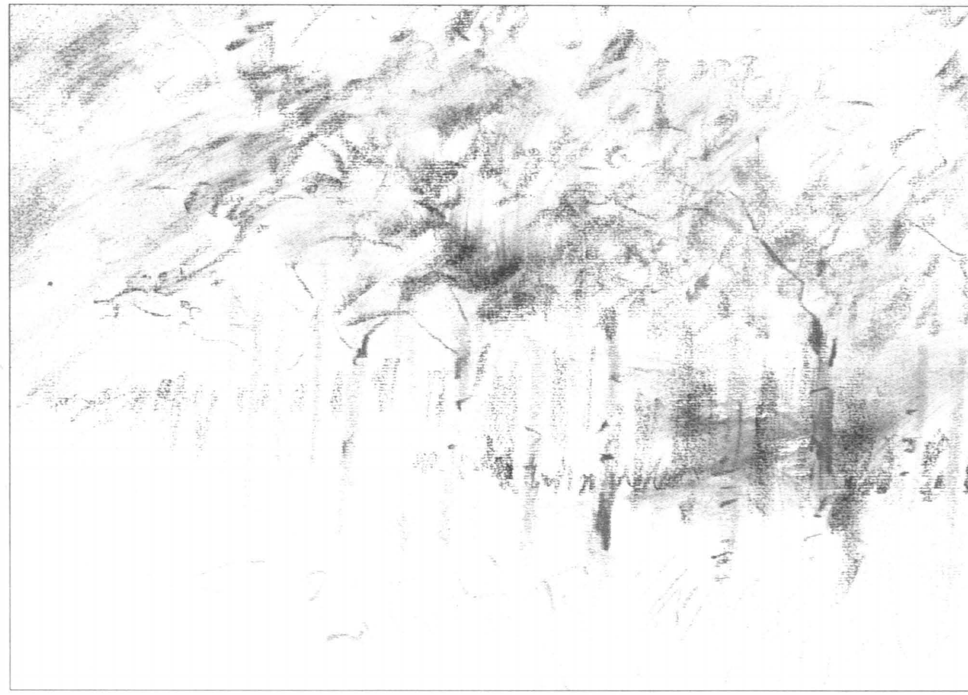
ACTA FORESTALIA FENNICA



Mauno Pesonen, Arto Kettunen and Petri Räsänen

Non-Industrial Private Forest Landowners' Choices of Timber Management Strategies: Genetic Algorithm in Predicting Potential Cut

250 · 1995

ACTA FORESTALIA FENNICA

Publishers The Finnish Society of Forest Science
The Finnish Forest Research Institute

Editors Editor-in-chief Eeva Korpilahti
Production editors Tommi Salonen, Seppo Oja

Editorial Office Unioninkatu 40 A, FIN-00170 Helsinki, Finland
Phone +358 0 857 051, Fax +358 0 625 308, E-mail silva.fennica@metla.fi, WWW Home Page http://www.metla.fi/publish/acta/

Managing Board Erkki Annila (The Finnish Forest Research Institute), Jari Kuuluvainen (The Finnish Forest Research Institute), Esko Mikkonen (University of Helsinki), Lauri Valsta (The Finnish Forest Research Institute), Harri Vasander (University of Helsinki), and Seppo Vehkamäki (University of Helsinki)

Editorial Board Per Angelstam (Grimsö Wildlife Research Station, Sweden)
Julius Boutelje (Swedish University of Agricultural Sciences, Sweden)
Finn H. Brække (Norwegian Forest Research Institute, Norway)
J. Douglas Brodie (Oregon State University, USA)
Raymond L. Czaplewski (USDA Forest Service, USA)
David Ford (University of Washington, USA)
George Gertner (University of Illinois, USA)
Martin Hubbes (University of Toronto, Canada)
William F. Hyde (Virginia Polytechnic Institute and State University, USA)
Jochen Kleinschmit (Lower Saxony Forest Research Institute, Germany)
Michael Köhl (Swiss Federal Institute for Forest, Snow and Landscape Research, Switzerland)
Noel Lust (University of Gent, Belgium)
Bo Långström (Swedish University of Agricultural Sciences, Sweden)
William J. Mattson (USDA Forest Service, USA)
Robert Mendelsohn (Yale University, USA)
Hugh G. Miller (University of Aberdeen, United Kingdom)
John Pastor (University of Minnesota, USA)
John Sessions (Oregon State University, USA)
Jadwiga Sienkiewicz (Environment Protection Institute, Poland)
Richard Stephan (Federal Research Centre for Forestry and Forest Products, Germany)
Elon S. Verry (USDA Forest Service, USA)
S.E. Vompersky (Russian Academy of Sciences, Russia)
A. Graham D. Whyte (University of Canterbury, New Zealand)
Claire G. Williams (Texas A&M University, USA)

Aim and Scope Acta Forestalia Fennica publishes monographs in forest science. These may be original research articles or comprehensive reviews aiming at a synthesis of knowledge in a particular topic. The series covers all aspects of forest science, both basic and applied subjects. Manuscripts are subject to peer review.

ACTA FORESTALIA FENNICA

250 · 1995

Mauno Pesonen, Arto Kettunen and Petri Räsänen

Non-Industrial Private Forest Landowners' Choices of Timber Management Strategies: Genetic Algorithm in Predicting Potential Cut

The Finnish Society of Forest Science — The Finnish Forest Research Institute

Pesonen, M., Kettunen, A. & Räsänen, P. 1995. Non-industrial private forest landowners' choices of timber management strategies: Genetic algorithm in predicting potential cut. Acta Forestalia Fennica 250. 28 p.

The factors affecting the non-industrial, private forest landowners' (hereafter referred to using the acronym NIPF) strategic decisions in management planning are studied. A genetic algorithm is used to induce a set of rules predicting potential cut of the landowners' choices of preferred timber management strategies. The rules are based on variables describing the characteristics of the landowners and their forest holdings. The predictive ability of a genetic algorithm is compared to linear regression analysis using identical data sets. The data are cross-validated seven times applying both genetic algorithm and regression analyses in order to examine the data-sensitivity and robustness of the generated models.

The optimal rule set derived from genetic algorithm analyses included the following variables: mean initial volume, landowner's positive price expectations for the next eight years, landowner being classified as farmer, and preference for the recreational use of forest property. When tested with previously unseen test data, the optimal rule set resulted in a relative root mean square error of 0.40.

In the regression analyses, the optimal regression equation consisted of the following variables: mean initial volume, proportion of forestry income, intention to cut extensively in future, and positive price expectations for the next two years. The R^2 of the optimal regression equation was 0.34 and the relative root mean square error obtained from the test data was 0.38.

In both models, mean initial volume and positive stumpage price expectations were entered as significant predictors of potential cut of preferred timber management strategy. When tested with the complete data set of 201 observations, both the optimal rule set and the optimal regression model achieved the same level of accuracy.

Keywords genetic algorithms, strategic decision making, timber management strategies.
Authors' address Finnish Forest Research Institute, PL 18, 01301 Vantaa, Finland.
Telefax +358-0-857 05 809, **E-mail** mauno.pesonen@metla.fi
Accepted December 1, 1995

ISBN 951-40-1491-X
ISSN 0001-5636
Tampere 1995, Tammer-Paino Oy

Contents

Preface

This study is part of the project 'Optimisation of Regional Cutting Budgets' conducted at The Finnish Forest Research Institute. The main objective of the project is to develop a new calculation system for determining regional cutting budgets. Special thanks are due to the following project partners for their help and advice during the research process: The Finnish Forest Industries Federation, The Central Union of Agricultural Producers and Forest Owners, The Forest Centre Tapio, The Regional Forestry Board of Pohjois-Savo, The University of Joensuu and The National Board of Taxation.

The authors would also like to thank the following persons for their helpful comments on the manuscript: Dr. J.N.R. Jeffers, Dr. J. Kuuluvainen, Dr. L. Lönnstedt, Dr. E. Mikkonen, Dr. T. Pukkala, Mr O. Salminen, Dr. D. Schmoldt and Dr. L. Valsta. Thanks are also extended to Mr E. Pekkinen, for checking the English language. The responsibility for any remaining errors lies with the authors.

1 Introduction

1.1 Strategic Decisions in Management of Non-Industrial Private Forests

The behaviour of non-industrial private forest (NIPF) landowners has been studied widely in many industrialised countries where forests in private ownership amount to significant proportions. In these countries, forest industry receives a great amount of timber from NIPF lands, and there is also a growing demand for recreational and other non-timber uses of non-industrial private forests. Many landowner studies have focused on relating the landowners' characteristics to specific aspects of past and future behaviour, e.g. harvesting (Kuuluvainen and Salo 1991), forestry investments (Romm et al. 1987), management intentions (Greene and Blatner 1986), and objectives and motivations of NIPF landowners (Kurtz and Lewis 1981, Bliss and Martin 1989).

It has been generally accepted that, by knowing the factors affecting the behaviour of landowners, it is possible to guide public measures to activate NIPF management (e.g. Gramann et al. 1985). It is also known that NIPF landowners form a diverse group with a variety of objectives and intentions. Moreover, most NIPF landowners have long-time perspectives and strategic views concerning management (Lönnsted 1989). However, few studies have been made on the strategic decisions made in NIPF management (Lönnstedt and Törnqvist 1990, Hansson et al. 1990, Pukkala and Kangas 1993) and the factors that affect these decisions have been, more or less, neglected.

Understanding the strategic decisions of NIPF landowners is important for several reasons. First, it helps in the planning of public measures to promote timber management on NIPF lands. Second, predictions concerning the supply of timber from private forests for the future development of forest industry can be based on knowledge of these decisions (Lönnstedt and Roos 1993). Third, a strategic view is needed in the development of day-to-day NIPF management planning.

At the strategic level of decision making, NIPF landowners lack information about the decision alternatives and their consequences (Kangas et al. 1992). So far, NIPF management planning in Finland has been basically tactical. Forestry plans are usually based on the presentation of a single cutting budget with no real decision alternatives. Thus, the various goals of Finland's over 400 000 NIPF landowners are almost totally ignored in State-funded forest management planning. This being the case, it is important to bring landowners into the planning process by offering them alternatives for the strategic use of their forest property.

1.2 Machine Learning

Modelling the strategic decision making of NIPF landowners, like any other attempt to model human behaviour, is a complex and multidimensional task. New approaches to retrieving knowledge from complex problem domains are offered by *machine learning* methods. Many machine learning techniques are capable of dealing with quantitative as well as qualitative variables with linear and non-linear dependencies. They are often non-parametric in character, noise-tolerant, and non-sensitive to fixed hypotheses. Moreover, the results obtained from machine learning analyses, particularly in the form of production rules, can be directly used to construct rule-based expert systems (Guan and Gertner 1991).

Genetic algorithms, the machine learning approach addressed in this study, have been suggested for various applications in forest science, including the maintenance and management of forestry-related, knowledge-based systems (Foster 1993), modelling of natural processes and construction of rule-based expert systems (Guan and Gertner 1991), substituting for and complementing the traditional tools of statistical analysis (Liepins et al. 1990), and for inducing rules from complex forestry databases (Jeffers 1991). The genetic algorithm used in this study (BEAGLE) has been applied to discrimination of *Populus* clones by measurement of poplar leaves (Jeffers 1991) and to automated rule induction from forest regeneration database (Saarenmaa 1992).

In this study, a rule-learning genetic algorithm is applied to analysing NIPF landowners' strategic decisions. This kind of an approach has not been reported in any earlier studies of landowner behaviour. Neither have genetic algorithms, as yet, been applied in the same sense.

1.3 Aim of the Study

The aim of this study is to predict non-industrial, private forest landowners' potential cut using a genetic algorithm and linear regression. Furthermore, the factors affecting the landowners' strategic decisions in NIPF management planning are studied. A genetic algorithm is used to induce a set of production rules predicting NIPF landowners' preferred timber management strategies. The rules are based on the landowners' objectives and characteristics of the landowners and their forest holdings. The dependent variable is the preferred timber management strategy, potential cut, described as *average annual removals of the first five-year planning period* (*m^3/ha/a*). The performance of genetic-based machine learning is compared to more traditional statistical analysis. The theoretical framework of the study is presented in Fig. 1.

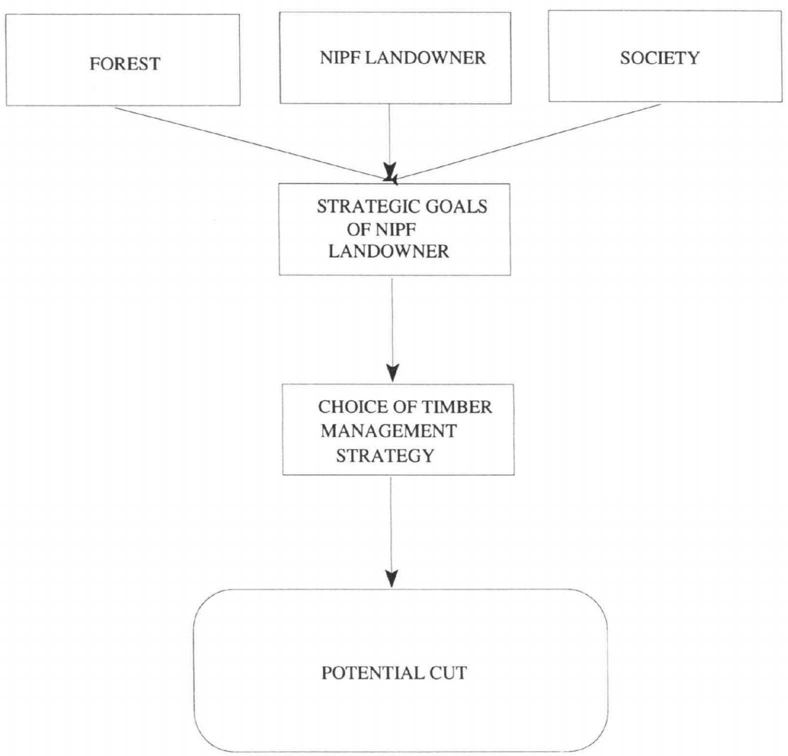


Figure 1. Frame of reference applied in this study.

2 Genetic Algorithms

2.1 Introduction to Genetic Algorithms

Genetic algorithms (GA) can be defined as inductive, problem-solving methods applying the principles of biological evolution and natural selection (Alander 1992). The basic idea in genetic algorithms is that of developing a mathematically simple, yet robust, basic structure and make it into an effective search method appropriate for solving various types of complex problems (Goldberg 1989). Genetic algorithms have been applied to a variety of problem domains, including integer and combinatorial optimisation, heuristic search, rule induction and biological simulation (Davis 1991).

Genetic algorithms differ from traditional optimisation and search procedures in four ways (Goldberg 1989):

1 GAs work based on encoding of the parameter set, not the parameters themselves.
2 GAs conduct searches within a population of points, not a single point.
3 GAs use payoff (objective function) information, not derivatives or other auxiliary knowledge.
4 GAs use probabilistic transition rules, not deterministic rules.

A standard genetic algorithm usually operates on a population of bit strings. In principle, the operation of a genetic algorithm includes copying of bit strings and mutual changes between string parts. The only assumption in applying genetic algorithms is that the alternative solutions to the problem at hand can be represented as character strings of definite length. The bit-string encoding of alternative solutions is merely to facilitate the operation of genetic algorithms – the actual information included in bit-strings can be, for example, classification or production rules (Goldberg 1989).

Genetic algorithms produce variation in the populations of candidate solutions through genetic operators. The basic genetic operators include *reproduction, recombination* (i.e. crossover) and *mutation* (Goldberg 1989). Reproduction creates new information structures, starting from the initial population, so that the surviving descendants relate to the above average candidate solutions. In recombination, bit-strings exchange a random part of their information structures. Mutation creates new individuals by randomly altering the value of one or more bits with pre-specified probability. With a population of classification rules, the above operations are basically the same except that the information structures under manipulation consist of parts of rules (Forsyth 1989). An example of recombination is presented below (Goldberg 1989).

Let A_1 and A_2 represent bit-strings of length $l = 5$

$A_1 = 0\ 1\ 1\ 0\ 1$
$A_2 = 1\ 1\ 0\ 1\ 0$

Assuming that, in choosing the crossover site, we obtain a random number $k = 2$ (indicated by the separator symbol |), the resulting crossover leads to new strings A'_1 and A'_2 where the specified parts of the information structures have been exchanged:

$$A_1 = 0\ 1 \mid 1\ 0\ 1 \qquad \Rightarrow \qquad A'_1 = 0\ 1\ 0\ 1\ 0$$
$$A_2 = 1\ 1 \mid 0\ 1\ 0 \qquad\qquad\qquad A'_2 = 1\ 1\ 1\ 0\ 1$$

A fitness function is needed to make the selection between the alternative problem solutions. The best individuals of the population of alternative solutions are cross-bred and the most compatible solutions are selected in terms of the fitness function. The procedure is iterated until the desired outcome has been achieved (Alander 1992).

Adopted from Davis (1991), a top-level description of a rule-manipulating genetic algorithm in action can be outlined as follows:

1 Initialise a population of rules at random.
2 Evaluate each rule in the population.
3 Create new rules by mating current rules; apply genetic operators as the parent rules mate.
4 Delete inferior members of the population to make room for the new rules.
5 Evaluate the new rules and insert them into the new population.
6 If time is up, or a desired outcome is reached, stop and return the best rules; if not, go to 3.

2.2 Schema Theorem

The basic functions of genetic algorithms are primarily based on an essential paradigm of the theory of GAs called *Schema Theorem* (Goldberg 1989). Schema Theorem may be described by considering the effect of reproduction, crossover (recombination), and mutation on the growth or decay of schemata from generation to generation. Prior to any mathematical formulation of the functions of a genetic algorithm, some important concepts have to be defined.

A bit-string is a binary-coded set of characters of finite length, e.g. the seven-bit string A = 0011010 can be represented as A = $a_1,a_2,a_3,a_4,a_5,a_6,a_7$. In order to describe the ideas of the schemata, a *, which stands for a wild-card symbol, (i.e., a * can be either 0 or 1), is added to form a character set {0, 1, *} .

A population of individual strings is defined A_j, $j = 1, 2, \ldots, n$, and it is contained in the population $A(t)$ at time t (boldface denotes a population).

A schema (Holland 1975) is a similarity template describing a subset of strings with similarities at certain string positions. For example, a schema H = *11*0** is represented by string A = 0111000 because of the matching positions at certain string positions.

Order of a schema H, denoted by $o(H)$, is the number of fixed positions (the number of 1's and 0's) in the template, e.g. $o(*11*0**) = 3$.

Defining length of a schema H, denoted by $\delta(H)$, is the distance between the first and the last specific string position, e.g. $\delta(*11*0**) = 3$, because the distance between the first (2) and the last (5) fixed position is $(5 - 2) = 3$.

In analysing the function of a genetic algorithm, effects of genetic operators on manipulation of the schemata must be considered first. As a result of reproduction on a particular schema, an ever increasing number of samples are given to the more fit strings. However, reproduction alone introduces no new points in the search space. Crossover leaves a schema non-manipulated unless it cuts the schema, but it may disrupt the schema in doing so. Mutation at normal, low rates does not disrupt a particular schema very frequently. Thus, it can be concluded that the more fit, short-defining-length schemata (*building blocks)* survive from generation to generation by having exponentially increasing probabilities to survive and reproduce (Goldberg 1989).

During reproduction, a string is copied according to its fitness. After picking a non-overlapping population of size n, with replacement from population $A(t)$, it is expected to have $m(H,t + 1)$ representatives of the schema H in the population at time $t + 1$. Thus, the reproductive schema growth equation (1) may be written as follows:

$$m(H,t+1) = m(H,t)\frac{f(H)}{\bar{f}} \quad (1)$$

where
$m(H,t + 1)$ = number of representatives of schema H at time $t + 1$
$m(H,t)$ = initial number of representatives of schema H at time t
$f(H)$ = average fitness of strings representing schema H at time t
$\bar{f}$ = average fitness of the entire population of strings

Schemata with fitness values above the population average will receive an increasing number of trials in the next generation, while schemata with fitness values below the population average will receive a decreasing number of trials. This procedure is carried out with every schema H in population A in parallel (Goldberg 1989).

The mathematical form of growth (or decay) of schemata can be reduced to the general form of *compound interest equation* by assuming that a particular schema H remains above-average by an amount cf, where c stands for a constant. Under this assumption, the schema difference equation (2) may be written as follows (starting at $t = 0$ and assuming a stationary value of c):

$$m(H,t) = m(H,0) \cdot (1 + c)^t \quad (2)$$

Crossover is a structured, yet randomised, information exchange between strings. Crossover creates new structures with a minimum of disruption to the allocation strategy dictated by reproduction. This results in exponentially increasing (or decreasing) proportions of schemata in the population.

A lower boundary to crossover survival probability p_s can be calculated for any schema. Because a schema survives when the crossover site falls outside the defining length, the survival probability under simple crossover is $p_s = 1 - \delta(H) / (l - 1)$. If crossover itself is performed by random choice, e.g. with probability p_c for a particular mating, the survival probability may be given by expression (3).

$$p_s \geq 1 - p_c \cdot \frac{\delta(H)}{l-1} \quad (3)$$

where
$(l-1)$ = number of possible crossover sites in a schema of length l

which reduces to the earlier expression when $p_c = 1.0$.

The combined effect of reproduction and crossover can be defined by calculating the number of a particular schema H expected in the next generation. Assuming independence of the reproduction and crossover operations, the estimate (4) is obtained:

$$m(H,t+1) \geq m(H,t) \cdot \frac{f(H)}{\bar{f}}\left[1 - p_c \cdot \delta(H)/(l-1)\right] \quad (4)$$

The combined effect of crossover and reproduction is obtained by multiplying the expected number of schemata for reproduction alone by the survival probability under crossover p_s. Schema H grows or decays depending on the multiplication factor. With both crossover and reproduction, this factor depends on whether the schema is above or below the population average, and whether the schema is of relatively short or long in terms of defining length (Goldberg 1989).

Mutation is the random alteration of a single bit position with probability p_m. In order that a schema H would survive, all the specified positions must themselves survive. Therefore, since a single allele survives with probability $(1 - p_m)$, and since each of the mutations is statistically independent, a particular schema survives when each of the $o(H)$ fixed positions within the schema survives. The probability of a surviving mutation is obtained by multiplying the survival probability $(1 - p_m)$ by itself $o(H)$ times, $(1 - p_m)^{o(H)}$. For small values of p_m ($p_m << 1$), the schema survival probability may be approximated by the expression $1 - o(H) \cdot p_m$. Thus, ignoring small cross-product terms, the combined effect of reproduction, crossover, and mutation can be given by expression (5) (Goldberg 1989):

$$m(H,t+1) \geq m(H,t) \cdot \frac{f(H)}{\bar{f}}\left[1 - p_c \cdot \delta(H)/(l-1) - o(H)p_m\right] \quad (5)$$

Expression (5) can be explained so that the number of representatives of schema H at time t+1 is equal to, or greater than the product of reproductive schema growth (1), survival probability in crossover (3), and probability of a surviving mutation (with small terms ignored).

2.3 BEAGLE

The applied genetic, rule-learning system called BEAGLE (**B**iologic **E**volutionary **A**lgorithm **G**enerating **L**ogical **E**xpressions) displays the essential ideas of the genetic algorithm in the form of evolutionary rule induction. The operation of BEAGLE is based on the presentation of alternative problem solutions as classification rules. The genetic operators are analogous to the general presentation of a genetic algorithm, except that the population under genetic manipulation consists of rules and parts of rules (Forsyth 1981). The principle of the evolutionary learning is to save the best rules and generate new structures by recombining and mutating the rule parts that survive. In that way, the process gradually creates rules that predict the target better and better.

The BEAGLE system consists of six main

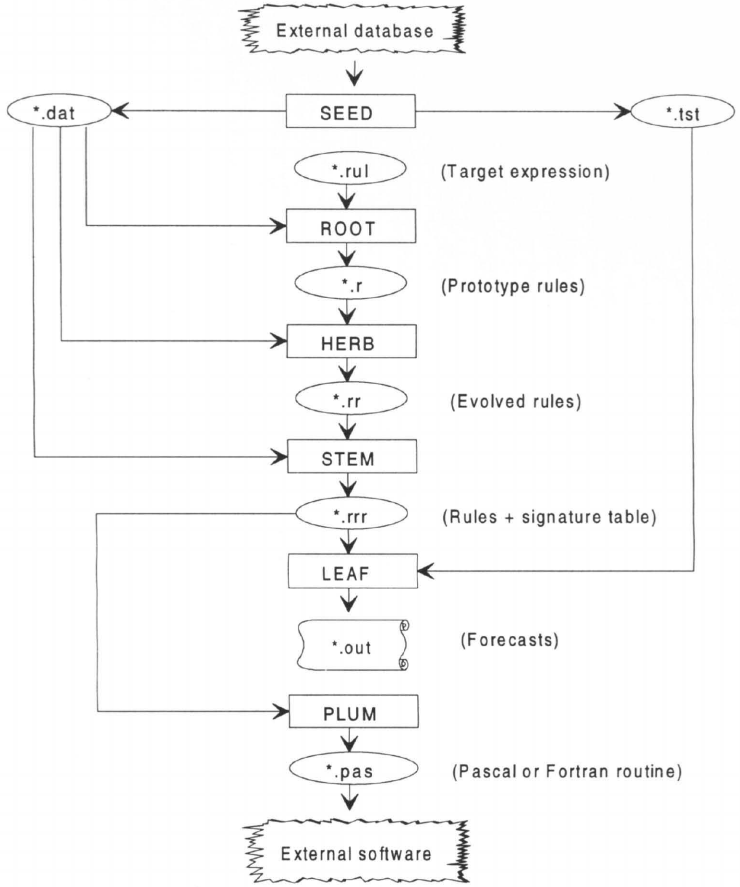


Figure 2. Main functions, output files and components of the applied genetic algorithm (Forsyth 1989).

modules performing the following tasks (Fig. 2): importing of external data files and randomly splitting data into subsets for learning and testing (SEED), creating the initial rule set at random and testing the user-suggested rules (ROOT), generating of new rules by applying genetic operators (HERB), testing and ranking the generated rules (STEM), and applying rules to the test data (LEAF). Furthermore, there is a subroutine (PLUM) which can be used to convert rules into other computer languages.

The rule-language BEAGLE operates on consists of variables from the initial data and numeric constants linked by logical, comparison and arithmetic operators (Forsyth 1989). The applied operators are:

LOGIC		ARITHMETIC	
!	Logical negation (NOT)	+	Addition
&	Logical conjunction (AND)	–	Subtraction
|	Logical disjunction (OR)	*	Multiplication

COMPARISON

=	Equality (EQ)
<>	Inequality (NE)
>	Superiority (GT)
<=	Non-superiority (LE)
>=	Non-inferiority (GE)

An example of the rule language and the scoring of rules:

((MEANVOL > 122.2730) >
(TAXREAL > (MEANVOL > 146.5312)))

		Target expression	
		true	false
Rule	true	80	18
	false	33	71
	score	57.08	

According to the rule language, the following interpretation of the example rule is obtained: the rule is true IF *mean initial volume* is between *122.3* and *146.5 m³/ha* AND *choice of the forest taxation basis* is site productivity OR *mean initial volume* is greater than *146.5 m³/ha*. After the rule, there is a score table which can be intepreted as follows:

80	true positives (rule true – target true)
18	false positives (rule true – target false)
33	false negatives (rule false – target true)
71	true negatives (rule false – target false)
57.08	rule-quality score (depends on predicting power and length of rule)

The BEAGLE decides upon survival of the rules according to the rule-quality score, which can be seen as a fitness value of an individual rule. The rule-quality score is obtained by calculating the proportion of variance accounted for the actual and predicted target values (Table 2) and by scaling the value to the range 0–100. Furthermore, a penalty of $0.25*L$, where L is the number of variables, constants and operators in the rule, is deducted. Thus, the scoring slightly favors shorter rules. The rule-quality score can be described as the percentage of the maximum departure from expectation due to pure chance. As a rule of thumb, a score of 50 is good and any score over 60 excellent (Forsyth 1981).

When the target expression deals with numeric variables, the truth values of individual rule-target combinations indicate whether the rule predicts the target value to be over or under the actual mean target value of the training data. In the example rule, the first of the above rule scores implies that 80 cases out of 202 fell in the category where both the predicted and the actual target values were above the mean target value of the training data. Respectively, in 18 cases the target value predicted by the rule was above the mean, but the actual target value fell under the mean. In statistics, the corresponding figures would form a contingency table.

In the following, all rules are presented as verbally interpreted expressions in order to facilitate the review of the results. Furthermore, it should be noted that conclusions concerning the prediction accuracy of an individual rule are not valid; rules must always be seen as complete sets denoting a certain target value (see chapter 4.1).

3 Material

3.1 Description of Data

The data were collected from within the jurisdiction of the Regional Forestry Board of Pohjois-Savo. The basic information on the forest holdings consisted of forestry plans made according to the TASO planning system for NIPFs (Ranta 1991). The descriptive information on NIPF landowners, their forestry goals, and characteristics of their forest holdings were collected by means of a two-phase mail inquiry (Pesonen 1994). The two-phase survey was necessary because the landowners had to be asked in advance for their permission to use the data from their forestry plans. In this study, the final sample included 201 NIPF landowners with forest holdings in Pohjois-Savo, comprising at least 5 ha of forest land and an up-to-date forestry plan.

In the first phase, the landowners were asked some questions about the characteristics of their ownership, economic situation and educational background. In addition, the questionnaire enquired about the landowners' preferences for alternative uses for their forest property.

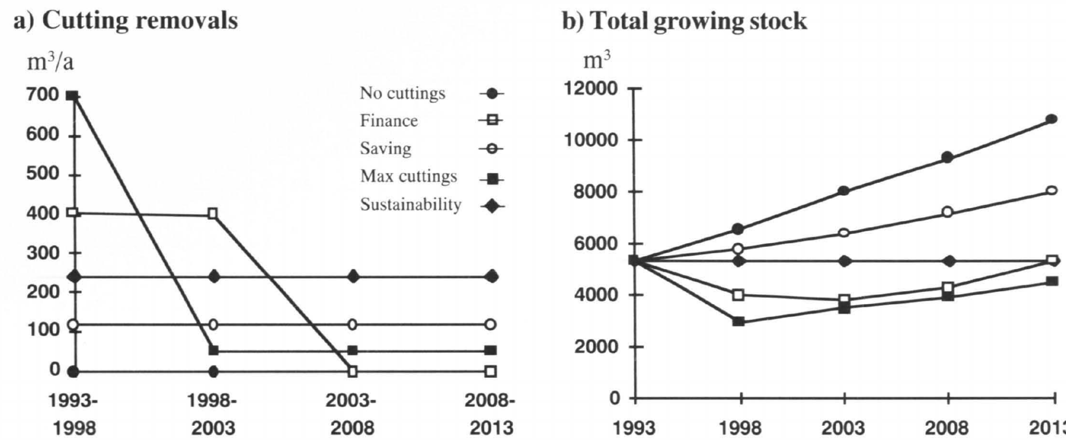
In the second phase, NIPF landowners were asked to prioritise a group of alternative timber management strategies for their forest holdings, computed and presented individually for each landowner (Fig 3). Furthermore, both questionnaires included some questions about the landowners' conceptions of the choice between two alternative forest taxation bases for the next 13-year transition period[1].

3.2 Choice of Timber Management Strategy

Each NIPF landowner was provided with five alternative timber management strategies covering a period of 20 years. The strategies were computed using the MELA system, an LP-based system developed in Finland for long-term timber management planning (Kilkki and Siitonen 1976, Siitonen 1983). The MELA system is based on the tree-level simulation of feasible management schedules for forest stands and the simultaneous selection of the production programme for the forestry unit (Siitonen 1993). The strategies were described for each landowner with the objective and constraint variables derived from the MELA parameters (Fig. 3). The total planning period was 20 years, divided into four 5-year intervals. In the calculations, the forest-holding-level development of several forest characteristics was described and illustrated for the landowners (Pesonen 1994).

In principle, the main differences between the strategies can be described in terms of intensity and the recurrence of removals. The objective variable used in the optimisations was maximisation of the income from timber sales for the first planning period (the constraints for each strategy are presented below). With potential cut of the first five-year planning period, *m³/ha/a*, calculated from the initial data, the applied strategies were as follows:

[1] The Finnish forest taxation system was changed in 1993 when the former site-productivity system of taxation was replaced by a system based on realized income. In the spring of 1994, landowners had to choose between the two taxation systems to be applied during the coming 13-year transition period. With the transition period over, all landowners will be taxed according to their realized income from timber sales. By choosing site-productivity based taxation for the transition period, landowners with an abundance of timber ripe for cutting are in a position to realize the accumulated increment which has already been taxed once. The choice between the two alternatives is affected mainly by potential cut during the transition period, the estimated value of annual increment under site-productivity taxation, and the individual landowner's marginal tax rate (Pesonen et al. 1995).



Figures 3a and **3b.** Alternative timber management strategies described in terms of development of removals (3a) and total growing stock (3b) during the planning period (an example of calculations made for each NIPF landowner, representing a sample case of the forest holdings). *Characteristics of the holding: forest area = 50.8 ha, initial growing stock = 5332 m³, mean initial volume = 105 m³/ha and average of the realized commercial cuttings during 1988–1992 = 347 m³/a.*

1 NO CUTTINGS — *Total abstinence from cuttings* (0.00 *m³/ha/a*)
- removals set to zero

2 SAVING — *Utilisation of approx. half of the sustained allowable cut* (2.46 *m³/ha/a*)
- removals set to half of the removals of SUSTAINABILITY

3 SUSTAINABILITY — *Practising forestry on a sustained yield basis* (4.91 *m³/ha/a*)
- even flow of removals over time
- even flow of stumpage earnings over time
- even amount of clear-cut areas over time
- volume of sawtimber at the end of period equal to, or greater, than in the beginning
- market value of growing stock at the end of period at least the same as in the beginning

4 FINANCE — *Utilisation of most of the allowable cut during first 10-year period* (8.78 *m³/ha/a*)
- even flow of removals during the first and the last two planning periods
- market value of the growing stock at the end of planning period at least the same as in the beginning

5 MAX CUTTINGS — *Immediate utilisation of the total allowable cut* (17.87 *m³/ha/a*)
- even flow of removals during the last three planning periods

The NIPF landowners were asked to prioritise the timber management strategies according to their personal goals and preferences for forest use. In prioritisation, the strategies were compared pairwise and preferences were derived by means of the Analytic Hierarchy Process (Saaty 1980).

The Analytic Hierarchy Process (AHP) is a mathematical method for analysing complex decision problems with multiple criteria. The method is based on the hierarchical presentation of the decision problem as interrelated decision elements (Fig. 4). On each level of the hierarchy, pairwise comparisons are made between the decision elements. As a result, the weights are computed to represent the decision maker's preferences concerning each decision element. The decision alternatives on the lowest level of the hierarchy are

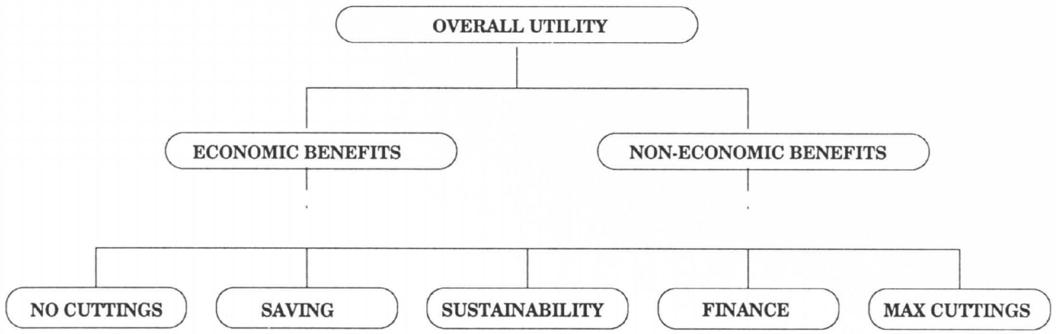


Figure 4. Decision hierarchy in selecting the preferred timber management strategy.

ranked according to the relative weights of the corresponding decision elements (Saaty 1980). Recent applications of the AHP to forest management planning include those by Mendoza and Sprouse (1989), Kangas (1992), Pukkala and Kangas (1993), and Pesonen (1994).

When applied to the present problem (Fig. 4), the overall benefit from forest property was placed at the highest hierarchy level. The NIPF landowners were first asked to compare the benefits from the economic and non-economic uses of their forest holdings. Second, pairwise comparisons were made between the management strategies, considering economic and non-economic benefits separately. The AHP process resulted in the relative priorities for each strategy being scaled 0–1. For each landowner, the strategy with the highest global priority (i.e. one that maximises the overall benefit) thus represented the preferred alternative (Pesonen 1994).

3.3 Problem Definition

A total of twenty-three variables were used to predict the NIPF landowners' choices of timber management strategies (Table 1). The main criterion for the selection of independent variables was significance in the context of landowner behaviour, as demonstrated in earlier studies (e.g. Järveläinen 1988, Karppinen & Hänninen 1990, Kuuluvainen & Salo 1991).

The dependent variable defined for the analyses was derived from the landowners' choices of timber management strategies and it was described as *potential cut of the first five-year planning period of the preferred strategy ($m^3/ha/a$)*. Thus, the dependent variable represented potential cut of the timber management strategy chosen by the NIPF landowner. Potential cut of the first planning period was used because it was considered to best reflect the differences between the strategies as regards the intensity and recurrence of cuttings. However, despite the differences between potential cut, the overall removals during the 20-year planning period came very close to each other in the cases of the strategies SUSTAINABILITY, FINANCE, and MAX CUTTINGS.

The data were cross-validated in both the genetic algorithm and regression analyses in order to examine the data-sensitivity and robustness of the generated models. A total of seven cross-validated data sets were produced with a random split of 70% (i.e. 141 observations) in the data sets for learning and modelling, and 30% (i.e. 60 observations) in the data sets for testing. The same randomly formed data sets were used in both methods. Thus, from the model point of view, the test data used in the genetic algorithm and regression analyses consisted of previously unseen cases.

A logarithmic transformation was required for the dependent variable in order to linearise the effects of independent variables. Prior to testing, the dependent variable was back-transformed and the bias caused by logarithmic transformation was rectified.

The model defined for the analyses was of the form

Table 1. Summary of the initial variables for the genetic algorithm and linear regression analyses.

Variable	Measurement	Mean	Min-Max	SD
RESPERM	Binary, permanent residence on farm	0.63	0.1	0.48
RESNOT	Binary, no residence on farm	0.26	0.1	0.44
FARMER	Binary, farmer	0.58	0.1	0.49
EINCOME	Proportion of gross earned income of household's gross total income, %	24.4	0–100	31.8
AINCOME	Proportion of gross agricultural income, %	42.3	0–100	32.6
FINCOME	Proportion of gross forestry income, %	17.6	0–80	16.6
FOLAND	Area of forest land, ha	69.0	8.9–609	67.5
FALAND	Area of farm land	10.0	0–49.6	11.3
EXCUT	Binary, intention to cut extensively in future	0.30	0.1	0.46
INCUT	Binary, intention to cut intensively in future	0.11	0.1	0.31
TIMBERG	Binary, landowner's view of timber production possibilities of his forest holding is 'good'	0.43	0.1	0.50
TIMBERP	Binary, view of timber production possibilities is 'poor'	0.12	0.1	0.33
ECONOMIC	Binary, preference of economic benefits of forest property	0.70	0.1	0.46
AVECUT	Average commercial cuttings of last 5-year period, $m^3/ha/a$	2.34	0–10.3	2.27
MEANVOL	Mean initial volume, m^3/ha	117.6	37.1–231.7	38.3
AHPCUT[1]	Preference of regular income from timber sales	0.27	0.02–0.70	0.16
AHPREC[1]	Preference of recreational use of forest property	0.14	0.01–0.52	0.08
TAXPROD	Binary, choice of site productivity based forest taxation	0.22	0.1	0.42
TAXREAL	Binary, choice of realized income based forest taxation	0.40	0.1	0.49
PRICE2HI	Binary, positive stumpage price expectations for next two years	0.10	0.1	0.30
PRICE2LO	Binary, negative stumpage price expectations for next two years	0.30	0.1	0.46
PRICE8HI	Binary, positive stumpage price expectations for next eight years	0.61	0.1	0.49
PRICE8LO	Binary, negative stumpage price expectations for next eight years	0.11	0.1	0.31
STRATEGY[2]	Dependent var., harvest rate of preferred timber management strategy, $m^3/ha/a$	5.44	0.65–18.0	2.70

1) Preferences of alternative uses of forest property (scaled to 0–1) expressed in the first inquiry and calculated with AHP-method

2) The NIPF landowners that chose the 'NO CUTTINGS' strategy were excluded from the final analyses, see discussion section

$$\log(y) = a + bx_1 + cx_2 + \ldots + lx_m \qquad (6)$$

where

$\log(y)$ = dependent variable, logarithm of the average removals of first five-year planning period

$a,b,c,\ldots,l$ = constant and regression coefficients

$x_1,x_2,\ldots,x_m$ = independent variables

and the formula for rectifying the bias of estimated values of dependent variable caused by the transformation was

$$y = \exp(a + bx_1 + cx_2 + \ldots + lx_m + \frac{\delta^2}{2}) \qquad (7)$$

where

y = dependent variable, back-transformed

δ^2 = variance of residuals

The internal operations of the genetic algorithm required that the target values cross the zero line. Therefore, the average value of the logarithmic dependent variable was subtracted from each observation. Thus, the task of the genetic algorithm was to generate rules predicting the average removals of the preferred timber management strategy of each landowner.

Seven cross-validated data sets were used in the genetic algorithm analysis. Starting from randomly created initial populations of rules, with ten rules per population, the genetic algorithm performed the evolutionary learning process and resulted in seven independent sets of four rules. Compared to the initial sets of rule created at random, the fitness of the rules improved considerably during the evolution process and rules also became more concise and easier to interpret. A total of 200 learning generations was obtained for each data set.

After obtaining the improved sets of rules from the genetic learning procedure, the sets of rules were evaluated in terms of their prediction accuracy. One rule was then excluded from each set of rules on the basis of not improving the overall predictive ability within the context of the remaining three rules (Forsyth 1981).

4 Results

4.1 Rules Induced by Genetic Algorithm

A total of seven sets of rules, with three rules in each set, was obtained after the execution of the genetic learning process (Appendix). The resulting final sets of rules were evaluated according to the truth-value combinations of the respective rules (Table 2). The truth values (0, 1) indicate whether an individual rule is considered true (1) or false (0). Thus, the combination "100" indicates that rule number 1 is true and rules 2 and 3 are false.

On being interpreted, the rules must always be viewed as complete sets denoting a certain truth-value combination and a value of the target expression of that particular set of rules. For example, the first set of rules with the truth value combination "000" (meaning all three rules are false), predicts the average annual removals of 2.52 $m^3/ha/a$ for the landowners that fall into the "000"

Table 2. Summary of estimated / observed target values of the rule truth value combinations derived from the seven cross-validated data sets. All values are in $m^3/ha/a$, describing the average removals of the first five-year planning period. The observed target values are given in parentheses.

Truth values / Rule set	1	2	3	4	5	6	7	Mean	TMS[1]
000	2.52 (3.01)	3.41 (3.02)	2.68 (4.71)	3.50 (3.84)	2.85 (4.20)	2.99 (3.67)	3.75 (3.69)	*3.10* *(3.73)*	***2.46*** **(SAVING)**
001	3.22 (4.71)	4.40 (3.69)	3.85 (3.62)	4.67 (4.38)	3.96 (4.21)	4.31 (4.85)	3.94 (4.18)	*4.05* *(4.23)*	
010	4.00 (3.93)	2.97 (4.55)	4.34 (4.69)	2.89 (5.20)	4.35 (4.20)	4.49 (4.56)	4.12 (4.39)	*3.88* *(4.50)*	
100	4.81 (5.56)	5.98 (4.44)	4.75 (6.69)	5.08 (6.31)	5.24 (4.42)	5.52 (5.50)	4.48 (6.47)	*5.12* *(5.63)*	***4.91*** **(SUST.)**
011	5.88 (5.31)	5.23 (4.30)	4.16 (6.38)	5.09 (4.86)	5.53 (4.44)	5.72 (4.29)	5.43 (5.86)	*5.29* *(5.06)*	
110	5.93 (6.61)	5.98 (5.91)	6.23 (6.18)	7.08 (7.16)	5.90 (6.53)	6.45 (7.46)	6.22 (6.26)	*6.26* *(6.59)*	
101	6.59 (6.16)	4.97 (5.57)	5.63 (6.79)	4.76 (6.12)	6.76 (6.44)	6.77 (4.88)	6.18 (5.90)	*5.95* *(5.98)*	
111	6.27 (6.27)	6.74 (6.31)	6.72 (5.24)	6.69 (6.30)	6.24 (5.99)	6.20 (6.75)	6.86 (6.55)	*6.53* *(6.20)*	***8.78*** **(FINAN.)**
Average	**4.90** (5.20)	**4.96** (4.72)	**4.80** (5.54)	**4.97** (5.52)	**5.10** (5.05)	**5.31** (5.25)	**5.12** (5.41)	**5.02** (5.24)	

[1] Refers to the timber management strategy that best corresponds the particular truth value combination.

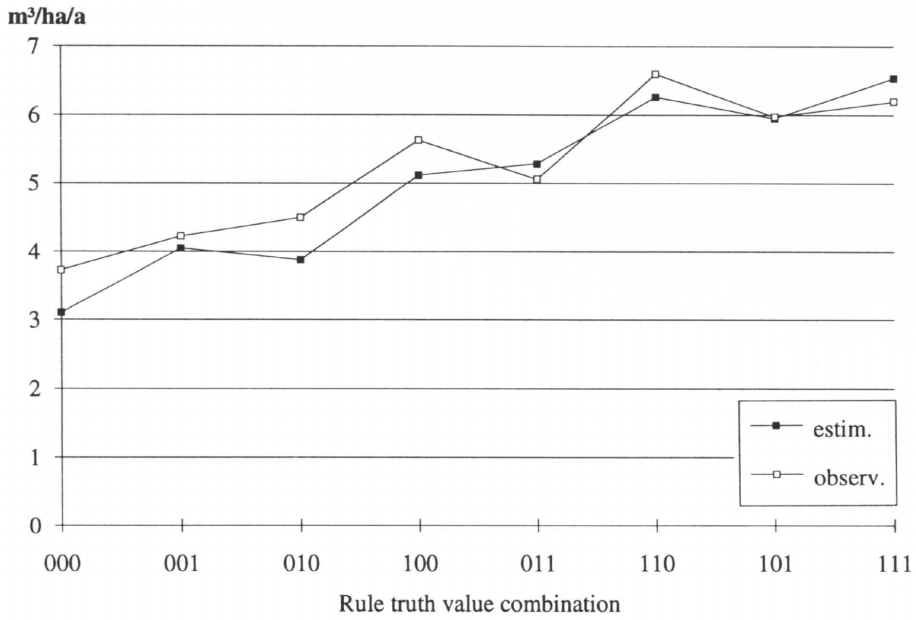


Figure 5. Average of estimated and observed target values in cross-validated data sets according to the rule truth value combination. (Observed values are presented as group averages of individual sets of rules)

category (Table 2).

The complete matrix of eight different truth value combinations times seven sets of rules can be viewed in the same fashion, together with the estimated and observed values of target expression (Table 2).

Finally, the generated rules were tested with the previously unseen test data. Thus, each of the seven sets of rules were tested with the respective sets of test data. The predictive abilities of the rules were examined by charting the estimated target values against the group average of the actual target values in each rule-combination category (Fig. 5). In order to obtain a comparable measure for the fitness of the rules, the relative root mean square errors (RMSE) and correlation coefficients were calculated for the predicted and observed target values of test data sets in the same way as in the regression analyses. Furthermore, the appearance of the independent variables in each set of rules was summarised (Table 3).

The cross-validated test data sets consisted of 60 cases each, of which, according to the measures provided by the genetic algorithm, an average of 41 (i.e. 68% of the sample, Table 3) were predicted correctly. The average target value for the category 'All rules true' (denoted by "111") was 6.53 $m^3/ha/a$ and for 'All rules false' (denoted by "000") it was 3.10 $m^3/ha/a$. The average of the relative root mean square errors (rRMSE) of the test data sets was 0.42 and the Pearson correlation coefficient between the actual and estimated target values was 0.35 (Table 3).

The seven final sets of rules had to be applied as such because they originated from separate data sets; i.e. mixing the rules from different sets would have led to misinterpretation of the actual predictive abilities. Thus, the performance of the rule sets must be compared quantitatively to find out the optimal set of rules. In comparison, a balance should be reached between the complexity of a set of rules vs. the fitness (predictive power) and robustness of rules. Therefore, the sets of rules were examined by considering the parameters as a number of variables in the rule set, relative RMSE, and the correlation coefficient between the predicted and observed target values in each set (Fig. 6). Judged by RMSEs, the rule sets per-

Table 3. Appearance of independent variables in rule sets derived from seven cross-validated data sets and summary of test results. (Appearance of a variable in resp. rule set is denoted by '1').

Variable / Rule set	1	2	3	4	5	6	7	Total
MEANVOL	1	1	1	1	1	1	1	7
RESPERM			1					1
RESNOT						1		1
FARMER	1		1	1	1			4
EINCOME		1		1	1			3
AINCOME				1			1	2
FINCOME		1	1	1	1			4
EXCUT					1			1
INCUT							1	1
TIMBERP		1			1	1		3
ECONOMIC			1					1
AHPCUT						1		1
AHPREC	1						1	2
TAXREAL		1	1			1		3
PRICE2LO			1	1				2
PRICE8HI	1		1					2
Test results								
'FITNESS'	70.00	60.42	69.77	69.77	65.91	72.73	67.92	68.07
RMSE	0.395	0.408	0.486	0.401	0.386	0.429	0.415	0.417
Pearson correl.	0.412	0.396	0.138	0.372	0.371	0.335	0.436	0.352
coeff. between estimated and observed dep. var.								

formed in a stable manner in the prediction stage, although the number of variables varied considerably between sets of rules. The correlation coefficients varied more than did the relative RMSEs since, compared to RMSE, the Pearson correlation coefficient is more sensitive to variance in the dependent variable to be tested.

4.2 Regression Analysis

Stepwise linear regression analyses were carried out to compare the performance of the genetic learning procedure to more conventional data analysis. Stepwise analyses were applied because the method of selecting the independent variables from the group of initial variables is somewhat analogous to that of the genetic algorithm. In regression analysis, the same, previously formed data sets were used so that the model was built using the cross-validated training data sets of the genetic process and tested using similar sets of the test data. The initial variables were also analogous to those of the genetic learning process, including the logarithmic transformation of the dependent variable.

In the regression analyses, all the independent variables of significance were entered into the models after four steps (forward stepping with alpha-to-enter of 0.15). The regression equations obtained from the analyses included seven different combinations of a total of nine independent variables with F-Ratios varying from 16.7 to 28.8 (Table 4). The average of the adjusted squared multiple R of the regression models was 0.37. When tested with the previously unseen data, the average for the relative RMSE was 0.41 and the correlation coefficient between estimated and the actual values of the dependent variable was 0.48 (Table 4).

As with genetic algorithm analyses, the seven independent regression equations were compared to determine the optimal model with respect to robustness, fitness and predictive ability. Thus,

Table 4. Regression models and test results of the cross-validated data sets. (T-values are given in parenthesis and significance is denoted by asterisk, so that * = 0.05, ** = 0.01 and *** = 0.001 -significance level).

Variable/Model	1	2	3	4	5	6	7
CONSTANT	0.577*** (5.98)	0.726*** (5.98)	0.639*** (5.59)	0.607*** (5.34)	0.897*** (6.35)	0.633*** (4.87)	0.706*** (6.48)
MEANVOL	0.007*** (7.94)	0.007*** (7.47)	0.007*** (8.61)	0.007*** (8.09)	0.008*** (8.19)	0.008*** (8.01)	0.006*** (7.52)
EINCOME	–	–	–	–	–0.002 (–1.57)	–	–
FINCOME	0.005* (2.35)	–	0.005* (2.37)	0.009*** (3.29)	–	0.004 (1.46)	0.004* (2.30)
FOLAND	–	–	–	–0.001* (–1.96)	–	–	–
EXCUT	–	–0.132 (–1.65)	–0.141 (–1.83)	–	–0.231** (–2.96)	–0.129 (–1.58)	–
INCUT	0.247* (2.22)	0.227 (1.85)	–	0.392*** (3.31)	–	–	0.241** (2.52)
AHPREC	–	–	–	–	–1.012* (–2.32)	–	–
PRICE2HI	0.379** (3.21)	0.279* (2.34)	–	0.374** (2.99)	0.311** (2.53)	0.289* (2.07)	0.338** (3.03)
PRICE8LO	–	–	–	–	–		0.195 (1.92)
Test results							
R^2	0.373	0.329	0.373	0.384	0.408	0.338	0.359
F-Ratio	21.86	18.13	28.76	18.46	20.32	18.90	16.70
P(tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RMSE	0.395	0.378	0.426	0.486	0.409	0.380	0.417
Pearson correl. coeff. between estimated and observed dependent var.	0.447	0.555	0.487	0.375	0.488	0.556	0.448

the regression equations were examined by charting the R^2 and the number of variables in the different models together with the RMSEs calculated from the test data sets (Fig. 7). Similarly to genetic algorithm analyses, the regression models were stable in terms of R^2 and RMSE, although there was substantial variance in the number of variables included in the models.

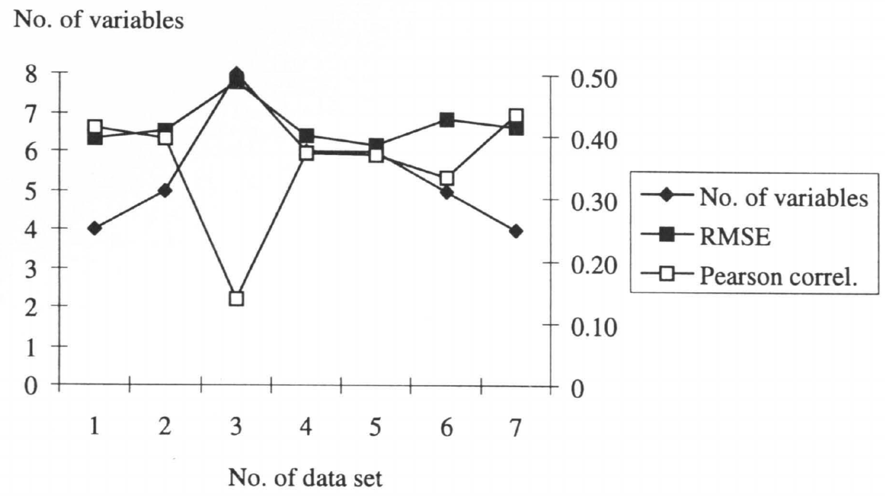


Figure 6. Number of variables, RMSE, and Pearson correlation coefficient between observed and the estimated target values in the seven cross-validated test data sets.

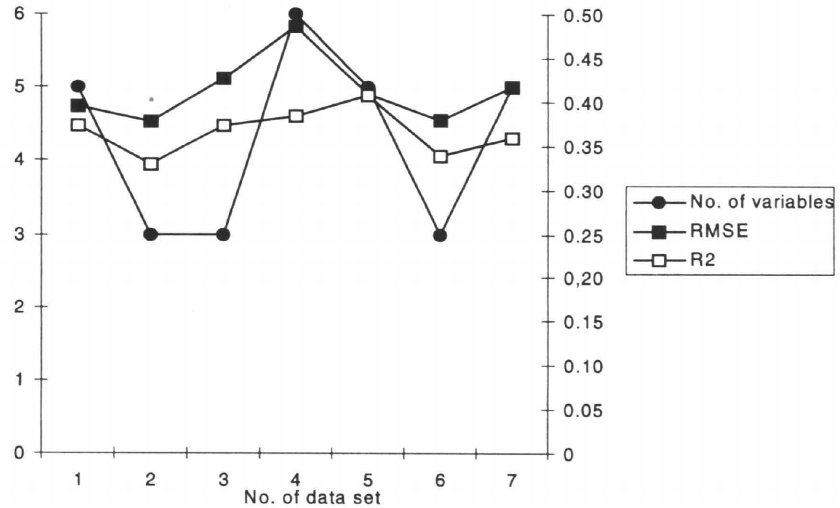


Figure 7. Number of variables, R^2, and relative RMSE according to the regression models derived from the seven cross-validated data sets.

5 Discussion

5.1 Comparison of Methods

One of the aims of this study was to compare the two modelling approaches. The use of identical data sets and the principle of selecting the best predictive variables according to a specific strategy facilitated the comparison. The results of the predictive power and the accuracy of the two methods were compared in two phases: 1) by averaging, over the data sets, the RMSEs and the correlation coefficients between the predicted and observed values of the planned removals, and 2) by selecting the optimal set of rules and the regression equation, and then comparing their performance.

In the genetic algorithm approach, the highest correlation and the lowest RMSE were obtained from rule sets 1 and 7 with the least number of variables used in the rules. Furthermore, there seemed to be obvious problems with rule set 3 with its many variables and weak predictive power. Of the leading sets, rule set number1 was selected on the basis of its slightly better RMSE (0.395 vs. 0.415) which is indicative of higher prediction accuracy (Table 3).

In comparing the regression analyses, equations 5 and 6 can both be introduced as candidates for the optimal model. On the one hand, equation 5 has more variables, but gives a better fitness than does equation 6. On the other hand, equation 6 returns the lowest RMSE and shows the highest correlation with the least number of variables (Table 4). On the basis of predictive ability and simplicity of the model, equation number 6 was chosen to be the optimal model for the purposes of this study.

The predictive abilities of the optimal set of rules and regression equation were summarised and the average figures of cross-validated data sets were presented (Table 5). Furthermore, both the genetic algorithm and regression analyses were recalculated using the optimal models and the complete data set of 201 observations. In the genetic algorithm analysis, recalculation resulted in a fitness of 76.1% and an RMSE value of 0.38, together with a Pearson correlation coefficient of 0.55. Respectively, the parameters derived from recalculation with the optimal regression model were: R^2 value of 0.34, RMSE value of 0.38, and Pearson correlation coefficient of 0.55. Thus, when tested with the complete data set, the optimal set of rules (1) and the optimal regression equation (6) performed equally well.

Table 5. Summary of predictive abilities of the optimal rule set (number one) and regression equation (number six). Averages of resp. parameters in cross-validated data sets are given in parenthesis.

	RMSE	Corr. coeff.
Rule set one from genetic algorithm	0.395	0.412
(Average of the seven sets)	(0.417)	(0.352)
Regression equation no. six	0.380	0.556
(Average of the seven models)	(0.413)	(0.479)

In the genetic algorithm analysis, the variables used as predictors of potential cut of NIPF landowners included the following variables: mean initial volume (MEANVOL), landowner's positive price expectations for the next eight years (PRICE8HI), landowner being classified as a farmer (FARMER), and preference for the recreational use of one's forest property (AHPREC).

When formulated into production rules, the effect of the variables on potential cut may be interpreted as follows:

Potential cut increases when

1) the mean initial volume increases,
2) the landowner is classified as farmer,
3) the landowner anticipates rising stumpage prices during the next eight years, and
4) the landowner's preference for recreational use of his forest property decreases.

In the regression analysis, an analogous interpretation of the regression equation results in the following conclusions (with variables mean initial volume (MEANVOL), proportion of forestry income (FINCOME), intention to cut extensively in future (EXCUT), and positive price expectations for the next two years (PRICE2HI)):

Potential cut increases when

1) the mean initial volume increases,
2) the proportion of forestry income increases,
3) the landowner intends to cut intensively in the future, and
4) the landowner anticipates rising stumpage prices during the next two years.

For both methods, mean initial volume was the most significant predictor of potential cut of the preferred timber management strategy. Furthermore, the parameters common to both methods included positive stumpage price expectations. This is consistent with, for example, the results obtained by Kuuluvainen (1989); analogous variables were discovered to be significant predictors of landowners' realized annual timber sales.

5.2 Reliability of Results

The central role of mean initial volume as a predictor in both analyses is partly due to the factors related to the original definition of the timber management strategies. On the one hand, by strongly referring to the actual potential cut of each forest holding, the mean initial volume "sets the level" around which potential cut varies. On the other hand, the fact that the landowners were consistent in their choices of timber management strategies as regards the actual timber production possibilities of their forest holdings, implies the landowners' rationale of decision making – at least in regard to the benefits to be gained from timber production. In more a general context, this has also been stated by Romm et al. (1987) and Hyberg (1987), among others.

The proportions of the alternative sources of total income for households indicate the decision makers' economic dependency on agriculture and forestry. The analyses showed that NIPF landowners, who obtained a considerable proportion of their gross annual income from forestry, and particularly from agriculture, were willing to cut intensively in the future. Furthermore, the subdivision of landowners into farmers and non-farmers influenced significantly the future potential cut. In the case of Finland, this may be interpreted as indicating that there is still a link between traditional farming and forestry, with forestry representing a supplementary source of income (e.g. Järveläinen 1988, Hyttinen 1992).

The average level of potential cut was increased by leaving the strategy group 'NO CUTTINGS' out of the analysis. Preliminary inspections of the data showed that landowners who had chosen the strategy of 'NO CUTTINGS' formed a heterogeneous group, whose behaviour could not be modelled along with the others. The strategy group 'NO CUTTINGS' could have been included in the analyses by using a tobit/logit formulation of the regression model as was done by Kuuluvainen & Salo (1991), Gramman et al. (1985), and Royer (1987). In that case, the genetic approach would have required two separate models: one for identifying the 'NO CUTTINGS' group from among the others, and the other for predicting potential cut of the rest of the landowners. This is due to the dualistic character of the genetic algorithm; numeric models must be analysed separate from logistic models.

When generalising the results of this study, it must be noted that the results suffer from the bias caused by the auto-selection of the data during the study phases. First, the initial population was formed on the basis of the landowner having a forestry plan for the property; it is possible that owners of forest holdings with forestry plans are more oriented towards timber production than owners of 'non-planned' forest holdings. Second, the two-phase survey could also have further selected the active, timber-production-oriented landowners. Third, the final data were small in number, consisting of only 201 landowners, and the strategy group that favoured the non-timber uses of forest (NO CUTTINGS) had to be excluded from the analyses.

5.3 Applicability of Genetic Algorithm

Methodologically, the genetic algorithm proved applicable in predicting the planned removals of preferred timber management strategies. However, with respect to the comparison parameters used in this study, the performance of the regression analysis proved to be slightly better. This may be due to the linear response surface of the independent variables. Despite small differences between the predictive abilities of the applied methods, the results from the cross-validated data sets imply the robustness and reliability of both methods in the sense of data-dependency. Moreover, when tested with the complete data set of 201 observations, both the optimal rule set and the optimal regression model reached the same level of accuracy.

The fact that the number of variables as well as selection of variables varied substantially between the BEAGLE runs is related to the small number of cases in the final analyses and random effect of the cross-validation. The analyses showed that besides the mean initial volume (MEANVOL) being the most powerful predictor, there were several independent variables with approximately equal explanatory power. Thus, the number of cases being small and the random division of them into test and learning data may have affected the variables that were picked up during each run. The same random character of cross-validated data sets had an influence on the regression analyses as well.

The genetic algorithm could have been more effective in bringing out the possible effects of the qualitative variables and of the non-linear dependency. It is also important to note that the genetic algorithm applied to this study has been proved to be better in dealing with logical problems, i.e. target expressions which obtain only the values 'true' or 'false' (PC / BEAGLE 1987). In that sense, reformulation of the problem into a logical one and the inclusion of logistic regression analysis as a reference method would be an interesting topic for future research.

Although the genetic algorithm did not exceed the predictive ability of regression analysis, its major implications are based on the applicability of the generated set of rules. The results from the genetic algorithm analysis may be interpreted as suggesting the grouping of the landowners, as homogeneously as possible, according to planned removals. The rules classifying NIPF landowners into these strategy groups could be used to precompute the decision alternatives for the different landowner categories. The preferred strategy could then be defined by specifying the individual goal structure of each NIPF landowner. This procedure could then be used as a part of an expert system in NIPF forest management planning as suggested by Pesonen and Kettunen (1993).

5.4 Conclusions

According to results obtained from both the genetic algorithm and the regression analyses, mean initial volume and stumpage price expectations were the most significant predictors of potential cut of NIPF landowners with properties in eastern Finland. Thus, it may be concluded that landowners are more or less aware of the possibilities of their forest holdings for timber production, and that landowners with plenty of timber are also willing to make use of their possibilities to cut. Furthermore, in making their choices on strategic management, the landowners follow, and also react to, anticipated changes in timber prices.

The most powerful application of the results of this study is based on the use of the preferred timber management strategies in calculating the regional allowable cut. The strategies can be generalised over a particular forestry area. By taking into notice the owners of non-planned forest holdings and other landowners beyond the scope of this study, the potential allowable cut represented by NIPF land can be derived from the timber management strategies (Pesonen 1994). This allowable cut, based on the landowners' objectives, can be taken into consideration in the planning of future investments by forest industry, for example. Moreover, the predictions of potential cut connected to the preferred strategies can be used to calculate potential cut for landowner groups, whose characteristics are known in advance.

Despite the fact that the principles of genetic algorithms were developed more than 25 years ago, very few applications have been reported in the field of forest science so far. However, genetic algorithms, along with other machine-learning methods, have great potential in solving complex problems in forestry-related domains. When adopting these methods, studies are needed in which different machine-learning approaches are compared in the solving of a variety of problems. An extension of this study would be to include a comparison of the neural network- and inductive-learning paradigms as supplementary prediction tools.

According to the results obtained by Pesonen (1994), when he used the same data as in this study, it was shown that the presentation of timber production possibilities for the landowners had a positive effect on their orientation towards timber production; 23% of the landowners were prepared to increase their cuttings after having seen the alternative timber management strategies. An important conclusion, and one that concerns this study as well, is that by increasing the availability of information on the strategic alternatives in forest management, and landowners' awareness of their timber production potential, it is possible to activate landowners into practising more intensive management on their forest holdings. This is crucial for the development of forest management planning on NIPF lands.

References

Alander, J. T. 1992. Optimal GA. In: Alander, J. T. (ed.) Geneettiset algoritmit – Genetic algorithms. Helsinki Univ. of Technology. Technical Report TKO-C53. Espoo. p. 164–179.

Bliss, J. C. & Martin, A.J. 1989. Identifying NIPF management motivations with qualitative methods. Forest Science 35: 601–622.

Davis, L. 1991. A Genetic algorithms tutorial. In: Davis, L. (ed.) Handbook of genetic algorithms. Van Nostrand Reinhold. N.Y. p. 1–99.

Forsyth, R. S. 1981. BEAGLE – A Darwinian approach to pattern recognition. Kybernetes 10(3): 159–166.

— 1989. The evolution of intelligence. In: Forsyth, R. S. (ed.) Machine learning: Principles and techniques. Chapman and Hall. London. p. 65–82.

Foster, M. A. 1993. Approaches and tools to ease the maintenance of knowledge-based systems. AI Applications. 7(2,3): 49–53.

Goldberg, D. E. 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley Publishing Company, Ann Arbor, MI. 412 p.

Gramann, J. H., Marty, T. & Kurtz, W. B. 1985. A logistic analysis of the effects of beliefs and past experiences on management plans for nonindustrial private forests. Journal of Environmental Management 20: 347–356.

Greene, J. L. & Blatner, K. A. 1986. Identifying woodland owner characteristics associated with timber management. Forest Science 32(1): 135–146.

Guan, B. T. & Gertner, G. 1991. Machine learning and its possible roles in forest science. AI Applications. 5(2): 27–36.

Hansson, K., Lönnstedt, L. & Svensson, J. 1990. Decision support system for nonindustrial private forest owners. Swedish University of Agricultural Sciences. Department of Forest-Industry-Market Studies. Report 12.

Holland, J. H. 1975. Adaptation in natural and artificial systems. Ann Arbor. The University of Michigan Press. MI. 183 p.

Hyberg, B. T. 1987. Multiattribute decision theory and forest management: A discussion and application. Forest Science 33(4): 835–845.

Hyttinen, P. 1992. Toimintojen optimaalisuus runsasmetsäisissä maatilayrityksissä. Summary: Optimality of activities on farms with large woodlots. University of Joensuu. Publications in Sciences 25.

Järveläinen, V-P. 1988. Hakkuumahdollisuuksien käyttöön vaikuttavat tilakohtaiset tekijät maan länsi- ja itäosissa. Summary: Factors affecting the use of the allowable cut in western and eastern parts of Finland. Folia Forestalia 707.

Jeffers, J. N. R. 1991. Rule induction methods in forestry research. AI Applications. 5(2): 37–44.

Kangas, J. 1992. Metsikön uudistamisketjun valinta – monitavoitteiseen hyötyteoriaan perustuva päätösanalyysimalli. Summary: Choosing the regeneration chain in a forest stand: A decision analysis model based on multi-attribute utility theory. University of Joensuu. Publications in Sciences 24.

— , Päivinen, R. & Varjo, J. 1992. Integrated forest planning. University of Joensuu. Faculty of Forestry. Research Notes 2.

Karppinen, H. & Hänninen, H. 1990. Yksityistilojen hakkuumahdollisuuksien käyttö Etelä-Suomessa. Summary: Actual and allowable cut in nonindustrial private woodlots in southern Finland. Folia Forestalia 747.

Kilkki, P. & Siitonen, M. 1976. Principles of a forest information system. Proc. XVI IUFRO World Congress, Division IV. p. 154–163.

Kurtz, W. B. & Lewis, B. 1981. Decision-making framework for nonindustrial private forest owners: An application in the Missouri Ozarks. Journal of Forestry 5: 285–288.

Kuuluvainen, J. 1989. Nonindustrial private timber supply and credit rationing. Swedish University of Agricultural Sciences Department of Forest Economics. Report 85.

— & Salo, J. 1991. Timber supply and life cycle harvest of nonindustrial private forest owners: An empirical analysis of the Finnish case. Forest Science 37(4): 1011–1029.

Liepins, G., Goeltz, R. & Rush, R. 1990. Machine learning techniques for natural resource data analysis. AI Applications. 4(3): 9–18.

Lönnstedt, L. 1989. Goals and cutting decisions of private small forest owners. Scandinavian Journal of Forest Research 4: 259–265.

— & Törnqvist, T. 1990. The owner, the estate and the external influences: Nonindustrial private forest owners' management decisions. Swedish University of Agricultural Sciences. Department of Forest-Industry-Market Studies. Report 14.

— & Roos, A. 1993. Cutting levels for nonindustrial forestry: An estimate based on the state of forests and the goals of forest owners. Swedish University of Agricultural Sciences. Department of Forest-Industry-Market Studies. Report 26.

Mendoza, G. & Sprouse, W. 1989. Forest planning and decision making under fuzzy environments: An overview and illustration. Forest Science 35(2): 481–502.

PC / BEAGLE. 1987. User Guide. Pathway Research Ltd. Nottingham, U.K. 55 p.

Pesonen, M. 1995. Nonindustrial private forest landowners' choices of timber management strategies and potential allowable cut: Case of Pohjois-Savo. Acta Forestalia Fennica 247. 31 p.

— & Kettunen, A. 1993. An expert system for forest management planning and optimization of regional cutting budgets. AI Applications 7(2,3): 35–38.

— , Räsänen, P. & Kettunen, A. 1995. Modelling nonindustrial private forest landowners' strategic decision making by using logistic regression and neural networks: Case of predicting the choise of forest taxation basis. Silva Fennica 29(2): 171–186.

Pukkala, T. & Kangas, J. 1993. A heuristic optimization method for forest planning and decision making. Scandinavian Journal of Forest Research 8: 560–570.

Ranta, R. 1991. Metsätalouden suunnittelulaskelmat (in Finnish). In: Tapion taskukirja. Kustannus Oy Metsälehti. p. 334–337.

Romm, J., Tuazon, R. & Washburn, C. 1987. Relating forestry investment to the characteristics of nonindustrial private forestland owners in northern California. Forest Science 33(1): 197–209.

Royer, J. P. 1987. Determinants of reforestation behavior among southern landowners. Forest Science 33(3): 654–667.

Saarenmaa, L. 1992. Induktiivinen oppiminen metsänviljelyn tietokannan tulkinnassa. Summary: Inductive learning in knowledge acquisition from the forest regeneration database. University of Helsinki, Department of Forest Ecology. Publications 3.

Saaty, T. L. 1980 The Analytic Hierarchy Process. McGraw-Hill, N.Y. 287 p.

Siitonen, M. 1983. A long term forestry planning system based on data from the Finnish national forest inventory. Proc. IUFRO subject group 4.02 meeting in Finland. University of Helsinki, Department of Forest Mensuration and Management. Research Notes 17: 195–207.

— 1993. Experiences in the use of forest management planning models. Silva Fennica 27(2): 167–178.

Total of 40 references

Appendix. Verbal intepretation of target expression and rule sets induced from seven cross-validated data sets.

Target expression: Logarithm of the removals of preferred strategy subtracted by average value of removals within respective data set.

DATA SET 1.

Rule 1: The rule is true IF *mean initial volume* is greater than 112.5 m^3/ha.

Rule 2: The rule is true IF *price expectations of next eight years* are positive (rise in prices) AND *landowner is classified as farmer*.

Rule 3: The rule is true IF *preference of recreational use of forest property* is smaller than 0.12

DATA SET 2.

Rule 1: The rule is true IF *mean initial volume* is greater than 105.2 m^3/ha.

Rule 2: The rule is true IF *landowner's view of the timber production possibilities of his forest holding* is 'moderate' or 'good' AND *share of gross earned income* of household's total annual income is less than 20 % AND his *residence on the farm* is 'permanent' or 'temporary'.

Rule 3: The rule is true IF *landowner has at least some forestry income* OR *choice of forest taxation basis* is site productivity taxation.

DATA SET 3.

Rule 1: The rule is true IF *mean initial volume* is greater than 105.1 m^3/ha.

Rule 2: The rule is true IF *price expectations of next eight years* are positive (rise in prices) AND *landowner is classified as farmer*.

Rule 3: The rule is true IF *landowner prefers noneconomic uses of forest* AND his *price expectations of next two years are negative* (fall in prices) OR *landowner prefers economic uses of forest AND* his *price expectations of next two years* are not negative (stable prices or rise in prices).

DATA SET 4.

Rule 1: The rule is true IF *mean initial volume* is greater than 117.3 m^3/ha.

Rule 2: The rule is true IF *landowner has at least some agricultural income* AND *he does not have gross earned income at all*.

Rule 3: The rule is true IF *landowner has at least some forestry income* AND his *price expectations of next two years* are not negative (stable prices or rise in prices) AND *landowner is classified as farmer*.

DATA SET 5.

Rule 1: The rule is true IF *mean initial volume* is greater than 113.7 m^3/ha.

Rule 2: The rule is true IF *landowner has at least some forestry income* AND *he is classified as farmer AND landowner's view of the timber production possibilities of his forest holding* is 'moderate' or 'good'.

Rule 3: The rule is true IF *landowner's intention to cut* is 'sustainable cuttings' or 'heavy cuttings' OR *share of gross earned income* of household's total annual income is less than 25 %.

DATA SET 6.

Rule 1: The rule is true IF *mean initial volume* is greater than 117.3 m^3/ha.

Rule 2: The rule is true IF *landowner's preference of regular income from timber sales* is greater than 0.24

Rule 3: The rule is true IF *landowner's choice of forest taxation basis* is site productivity taxation AND his *residence on the farm* is 'permanent' or 'temporary' AND *landowner's view of the timber production possibilities of his forest holding* is 'moderate' or 'good'.

DATA SET 7.

Rule 1: The rule is true IF *mean initial volume* is greater than 119.9 m^3/ha.

Rule 2: The rule is true IF *landowner has at least some agricultural income* AND *he does not have gross earned income at all*.

Rule 3: The rule is true IF *landowner's intention to cut* is 'heavy cuttings' AND his *preference of recreational use of forest property* is 0.00

Submission of Manuscripts Manuscripts should be submitted in triplicate to Acta Forestalia Fennica, Unioninkatu 40 A, FIN-00170 Helsinki, Finland. Detailed instructions to authors have last been printed in Acta Forestalia Fennica 244, are available from the editorial office, and can be found on Acta's WWW pages at http://www.metla.fi/publish/acta/

Publication Schedule Acta Forestalia Fennica is published intermittently, 3 to 6 numbers per year.

Subscriptions and Exchange Subscriptions and orders for back issues should be addressed to Academic Bookstore, Subscription Services, P.O. Box 23, FIN-00371 Helsinki, Finland, Phone +358 0 121 4242, Fax +358 0 121 4450. Subscription price for 1995 is 70 FIM per issue. Exchange inquiries should be addressed to the Finnish Society of Forest Science, Unioninkatu 40 B, FIN-00170 Helsinki, Finland, Phone +358 0 658 707, Fax +358 0 191 7619, E-mail sms@helsinki.fi

Statement of Publishers Acta Forestalia Fennica has been published since 1913 by the Finnish Society of Forest Science. In 1989 Acta Forestalia Fennica was merged with Communicationes Instituti Forestalis Fenniae, started in 1919 by the Finnish Forest Research Institute. In the merger, the Society and Forest Research Institute became co-publishers of Acta Forestalia Fennica. The Finnish Society of Forest Science is a nonprofit organization founded in 1909 to promote forest research. The Finnish Forest Research Institute, founded in 1917, is a research organization financed by the Ministry of Agriculture and Forestry.

Abstracting Articles in Acta Forestalia Fennica are abstracted and indexed in Agrindex, Biological Abstracts, Current Advances in Ecological Sciences, Current Advances in Plant Sciences, Ecological Abstracts, Forest Products Journal, Forestry Abstracts, International Bibliography of Periodical Literature, Life Sciences Collection.

ACTA FORESTALIA FENNICA

The Finnish Society of Forest Science
The Finnish Forest Research Institute

ISBN 951-40-1491-X
ISSN 0001-5636

ACTA FORESTALIA FENNICA 250 · 1995

Cover: Jarmo Koivunen