

Metrics for Distribution Similarity Applied to the Bucking to Demand Procedure

Jukka Malinen
Teijo Palander
University of Joensuu
Joensuu, Finland

ABSTRACT

In the computerized bucking to demand procedure bucking is done according to a given price list and demand matrix, which defines the demands for different log length-diameter class proportions. To achieve as good a log length-diameter distribution as possible, the computer compares demand and actual output to appropriately direct bucking. A comparison has been made with a variable called distribution level, which, however, is unable to distinguish between error that is close to the optimum log length-diameter class proportion and error that is further away. In addition, the distribution level does not distinguish between log length-diameter classes, even though error in one class can be far more undesirable than in another.

In this study, bucking to demand using the distribution level was compared to bucking to value and bucking to demand using the penalty segmented distribution level, squared distribution level, chi-square formula and flexible penalty segmented distribution level. The bucking outcome employing these various techniques was achieved by using a bucking simulator and artificially generated stand and stem data.

The results show that the best bucking outcomes were produced by methods with a squared error term, i.e. the squared distribution level, chi-squared formula and flexible penalty segmented distribution level. In addition, it was possible to direct error toward preferred log length-diameter classes without substantial loss in overall goodness of fit

Keywords: *Cut-to-length method, bucking to demand, distribution level, log distribution.*

INTRODUCTION

Total revenue from harvested stems and stands depends greatly on how stems are cut up. Since the cut-to-

The authors are, respectively, Senior Assistant of Forest Mensuration Science, and Senior Lecturer of Forest Technology, Faculty of Forestry, University of Joensuu.

length method uses bucking decisions made during the logging process itself, operators who use harvesters play a critical role in such a man-machine system. Furthermore, the operators' role has increased so much that some decision support has had to be developed to control bucking. Information technology, such as microcomputers, mobile phones and wireless data communication systems, has helped operators to make the right decisions. In addition to information technology, good decision support for wood procurement requires properly used decision-making models, methods and computer software. [13]

Modern harvesters are able to measure the stem while processing and to predict the taper function for the rest of the stem. Using these measurements, taper predictions and the price list given for various log quality and length-diameter combinations the harvester's computer is capable of calculating optimal bucking for the stem. This procedure, called *bucking to value*, has been widely used in the Nordic countries.

The bucking to value procedure described above has not, however, satisfied the needs of sawmill managers. In customer-oriented sawing, the raw material also has to meet the wishes of other interest groups for log length-diameter demand. In practice, it has been quite difficult to determine log values so that the log distribution demanded can be obtained. The *bucking to demand* (or *bucking to order*) procedure has been developed as a first step to solving this problem. [2]

In the bucking to demand procedure, the desired log length-diameter distribution is given to the computer program as a list, in addition to the price list. In practice, there are two different approaches for adapting the actual output of log length-diameter distribution according to demand: the adaptive price lists approach and the near-optimum approach [18]. In the former, the optimization is done according to the log price values, but the computer itself changes the values within given limits to obtain the log length-diameter distribution required. Changes are made according to the differences between the output and demand matrices. In the near-optimum approach, the computer calculates a group of possible bucking solutions for the stem within the given tolerance from the maximum revenue, and selects the best solution from this group according to the demand.

One of the biggest problems in the bucking to demand procedure is how to evaluate the similarity of the demand and the actual output matrices. The most used similarity indicator variable is *distribution level* (or *target assortment percentage* – TAP), which has been used for over a decade [2, 15, 16, 11], but the formula was not published until Lukkarinen and Vuorenperä. [6]. It is similar to ge-

netic distance [3] but illustrates similarity as a percentage. Identical distributions receive a value of 100 percent and completely different matrices receive a value of zero percent but values usually fall between 40 and 90 percent [5] depending on the bucking optimization procedure and the structure of the log length-diameter matrices.

The distribution level has been widely adopted in practice, both in bucking optimization and in wood procurement planning. In addition to comparisons of demand and actual output while harvesting, the distribution level has been used in wood procurement studies to determine the accuracy of pre-information, and in wood procurement planning to compare stands to be cut according to different bucking control parameters [e.g. 19, 17, 20, 9, 7].

Although the distribution level is simple and easy to use, it has some drawbacks. All log length-diameter classes are weighted equally; in practice, however, some differences from the optimum are far less desired than others. In addition to this, change in a log length-diameter class proportion from the optimum decreases the distribution level by the same amount, regardless of the previous direction and distance from the optimum log length-diameter class proportion.

To allow sawmill managers to specify preferred log length-diameter classes, Weijo [20] used coefficients to weight log length-diameter classes individually. With different weights it was possible to specify some classes over others, but this technique still suffered from equal reduction of distribution level regardless of the distance from the optimum. Kirkkala *et al.* [4] proposed modification of the distribution level, called a penalty segmented distribution level, to overcome this problem. This technique enabled determination of log length-diameter classes or areas in which deviation from the optimum solution did not affect the distribution level. However, outside the defined tolerance, this technique suffers from the same problems as the original distribution level.

For the bucking to demand procedure, a parameter depicting similarity of the demand and actual output matrices should penalize dissimilarity differentially depending on user preferences for length-diameter class, direction of deviation and distance from optimum proportion. It should be possible to define user preferences for each length-diameter class and for both surplus and shortage separately.

This study seeks to improve the bucking to demand procedure by investigating various distribution similarity parameters to compare two matrices, such as the demand matrix and the actual outcome matrix. The influence of distribution similarity parameters on bucking outcome

was determined with the bucking simulator, which makes it possible to buck the same stems repeatedly and compare outcomes from the same stem population. The procedures considered were bucking to value and bucking to demand using six different techniques: distribution level, penalty segmented distribution level, squared distribution level, chi-square formula and flexible penalty segmented distribution level.

STUDY MATERIAL AND METHODS

Study Data

The study data were generated by a stand generator [8] developed from that of Oinas and Sikanen [12]. The purpose of the stand generator is to produce artificial stands and trees for the purposes of wood procurement studies at the University of Joensuu. Stands for a given wood procurement area are created by producing location, area and mean stand variables from models based on observations in North Karelia. According to mean stand variables, the stand generator produces stand and stock tables, which can be bucked by the bucking simulator.

The stem data generated consisted of 7 spruce-dominated stands, which were clear-cut. The average stand age was 111 years and the total volume of spruce was 1535m³. The average dominant height was 24.5 metres, median diameter 32.3cm and median height 24.1 metres. The diameter distribution of the stem data covered diameters from 12 to 57 (Figure 1).

Bucking Simulations

The bucking to value procedure was included in order to provide a yardstick for experiments on bucking to demand. In the bucking to value procedure, optimization was done according to prices given for the different log length-diameter combinations. Optimization was based on dynamic programming [1]. In the bucking to demand procedures, the stem population was bucked by the bucking simulator, using the near-optimum approach, which achieves a high distribution level sooner than does the use of the adaptive price list approach [11]. In the near-optimum approach, the bucking simulator calculated a group of possible bucking solutions for the stem within the given tolerance from the maximum revenue by using bucking to value approach and dynamic programming, and selected the best solution from this group according to the demand.

The constant value for the pulpwood in the price list was 24.45 US dollars/m³ and the constant value for saw



Figure 1. The diameter (cm) distribution of the study data

timber was 49.15 US dollars/m³. These values were actual prices in North Karelia in week 40 of the year 2002 [10]. The demand matrix used (Table 1) was modified from that introduced by Lukkarinen [5].

Bucking to demand using distribution level is presently used in modern harvesters. It is based on the buck-

ing to value procedure, but the log length-diameter distribution is directed to demand by adaptive price lists or a near-optimum approach. In this process, the computer uses distribution level as an indicator variable to depict the similarity between the demand and the actual output. The distribution level (DL) was calculated using the following formula:

Table 1. Proportions of demand for diameter classes used in the study. Length classes (dm) are in columns and diameter classes are in rows (cm).

	34	37	40	43	46	49	52	55	58	61	Total
15	0	0	0.5	0.15	0.15	0.1	0.05	0.05	0	0	100%
17.7	0	0	0.1	0.16	0.16	0.16	0.16	0.16	0.05	0.05	100%
19.0	0.1	0.1	0	0.1	0.1	0.1	0.22	0.14	0.09	0.05	100%
21.1	0.1	0.1	0	0.1	0.1	0.1	0.22	0.14	0.09	0.05	100%
24.2	0.0	0	0	0.1	0.2	0.15	0.23	0.17	0.1	0.05	100%
27.9	0.0	0	0	0.1	0.2	0.15	0.23	0.17	0.1	0.05	100%
32.5	0.0	0	0	0.1	0.2	0.15	0.23	0.17	0.1	0.05	100%

$$DL = 100 * \left(1 - \frac{\sum_{i=1}^k |D_{di} - D_{oi}|}{2}\right)$$

in which

k = number of log length-diameter classes.

D_{di} = proportion of demand for log length-diameter class i .

D_{oi} = proportion of outcome of log length-diameter class i .

Kirkkala et al. [4] suggested a modification of the DL for bucking to demand called the penalty segmented distribution level (PSDL). The basis is the same as in the DL, but this variant offers the saw-mill manager or person who is responsible for making the bucking control parameter file called the apt-file the opportunity to provide some log length-diameter classes tolerance (Figure 1), where shortage or/and surplus does not affect the distribution level. The PSDL can be formulated as follows:

$$PSDL = 100 * \left(1 - \frac{\sum_{i=1}^k \max[0, |D_{di} - D_{oi}| - T_i \cdot D_{di}]}{2}\right)$$

in which

T_i = tolerance (proportion of the demand) for the outcome shortage or surplus in log length-diameter class i .

While there is a need for differential penalties in distribution level regarding the previous shortage or surplus, it is possible to use the squared distribution level (SDL) to tolerate deviation near the optimal outcome and to penalize when the previous deviation is greater (Figure 2). The formula for the SDL is:

$$SDL = 100 * \left(1 - \frac{\sum_{i=1}^k (D_{di} - D_{oi})^2}{2}\right)$$

One of the most widely used goodness-of-fit tests in evaluating one distribution against another is the Pearson chi-square test, which has been considered a good all-round test for discrete data [14]. The chi-square (χ^2) formula can be used in the bucking to demand procedure by minimizing:

$$\chi^2 = \sum_{i=1}^k \frac{(D_{di} - D_{oi})^2}{D_{di}}$$

By combining the properties of the PSDL and SDL, it is possible to obtain an indicator variable which is able to weight every log length-diameter class shortage and surplus individually, and take the previous deviation from the optimum into account (Figure 3). The formula for the flexible penalty segmented distribution level (FPSDL) is:

$$FPSDL = 100 * \left(1 - \sum_{i=1}^k \frac{(D_{di} - D_{oi})^2}{2F_i}\right)$$

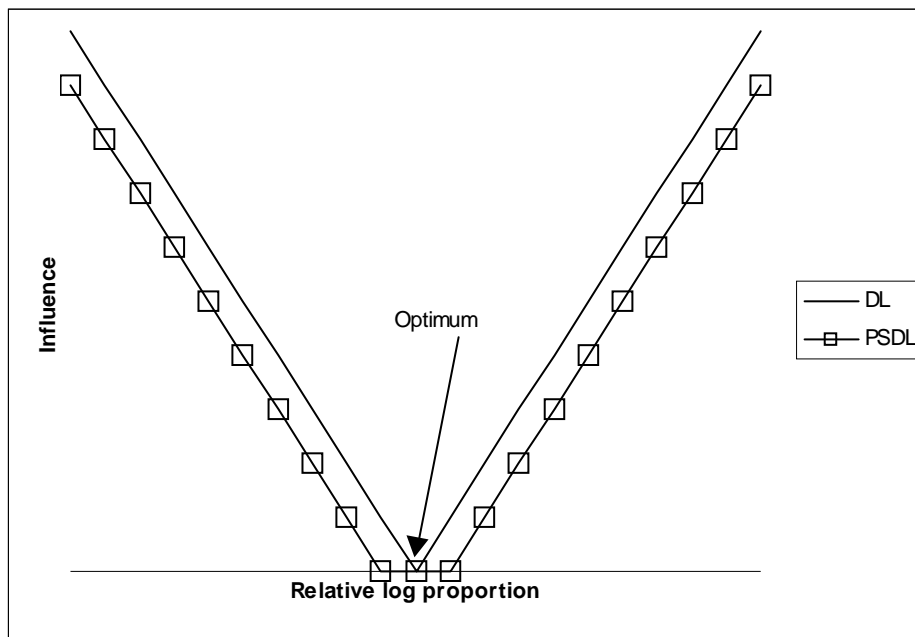


Figure 2. Influence on distribution similarity variables as a function of relative log proportion in one log length-diameter class for the distribution level (DL) and penalty segmented distribution level (PSDL). In the penalty segmented distribution level the tolerance is defined as equal in both shortage and surplus.

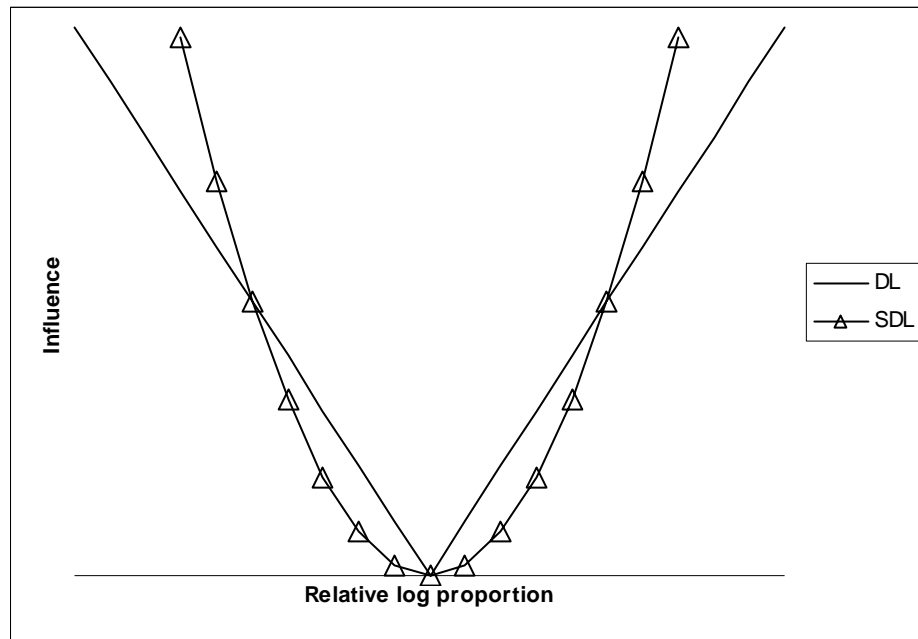


Figure 3. Influence on distribution similarity variables as a function of relative log proportion in one log length-diameter class for the distribution level (DL) and the squared distribution level (SDL).

in which

F_i = flexible parameter.

The FPSDL with a flexible parameter value of 1 equals the SDL. This could be used as a basic value, which can be increased or decreased for the desired log length-diameter classes according to the user preferences. By increasing the flexible parameter the user can give more tolerance to particular length-diameter class and by decreasing the tolerance is reduced.

The PSDL and the FPSDL were evaluated by modifying the 52 dm length class. Tolerances in the PSDL were 5% and 20%, and the flexible parameters in the flexible penalty segmented distribution level were 0.5 and 2.

Evaluation of the Bucking Simulation Outcomes

The effect of the various similarity indicator variables: DL, PSDL, SDL, chi-squared formula and FPSDL on the bucking to demand outcome was examined by using distribution level and squared distribution level in an evaluation of overall goodness. In addition to overall goodness, the performance of the PSDL and the FPSDL was evaluated by examining the overall distribution and the specific distribution of the 52 dm length class, the length class in which tolerances and flexibility parameters were defined.

Since there are considerable weaknesses in the DL in depicting differences between demand and actual output of log length-diameter distribution, the DL should be

viewed critically when it is used to evaluate the performance of the similarity indicator variables in the bucking to demand procedure. However, it was included as an evaluation variable since it is the variable most often used to depict similarity of the demand and actual output matrices in practice. The other evaluation variable, SDL, differs from the DL by taking into account error variance. The SDL weights average goodness of estimates but uses square errors to punish large variation.

RESULTS

As expected, the DL and the SDL for the bucking to value procedure was considerably lower in the evaluation (Table 2) than the values for the bucking to demand procedures. This clearly shows the better usability of the bucking to demand procedure compared to bucking to value. As a bucking technique, the SDL outperformed the DL in the bucking to demand procedure. This was certainly expected in regard to evaluation by SDL, but even evaluations with the DL indicated the effectiveness of the bucking to demand procedure with the SDL.

Bucking with the chi-squared formula performed better than the DL, especially in evaluating with the SDL, but did not produce as good a result as the bucking to demand with the SDL. The usability of the bucking to demand procedure with the chi-square formula depends on user preferences. Should an increase in the proportion of logs in a log length-diameter class decrease the influence of the deviation?

Table 2. Distribution levels (DL) and squared distribution levels (SDL) as evaluation variables for the bucking to value procedure and the bucking to demand procedure with the distribution level (DL), the squared distribution level (SDL), the chi-squared formula (χ^2), the penalty segmented distribution level (PSDL) and the flexible penalty segmented distribution level (FPSDL).

Bucking procedure	Similarity indicator in the bucking simulation	Evaluation variables	
		DL (%)	SDL (%)
Bucking to value		58.65	77.20
Bucking to demand	DL	87.19	96.30
	SDL	87.74	99.33
	χ^2	87.42	99.25
	PSDL (tolerance 5 %)	86.91	97.04
	PSDL (tolerance 20 %)	86.10	96.95
	FPSDL (flex 0.5)	87.66	99.29
	FPSDL (flex 2)	87.76	99.31

The performance of the PSDL and the FPSDL should be evaluated, both by analyzing the effect of the tolerance and the flexible parameter on log length-diameter classes where it is defined and as the effect on the overall bucking outcome. The log demands and outcomes of 52 dm length class (Table 3) show that there is a tendency to have a surplus in the 15.0 cm and 17.7 cm diameter classes, and in others a shortage. By adding a tolerance of 5% and 20% to the 52 dm length allowed the optimization to tolerate more surplus in the two smallest diameter classes and at the same time to tolerate shortage in other classes. In the FPSDL, the flexible parameter 0.5 weighted the op-

timization process to tolerate deviation in the 52 dm length class less, while the flexible parameter 2 weighted the optimization process to tolerate more deviation in this length class.

The PSDL gave more freedom to the 52 dm length, using a tolerance of 20% (Table 3). On the contrary, in this length class the 5% tolerance produced a better DL. While evaluating by the DL, the overall goodness dropped with both tolerances, while for both tolerances the SDL for the bucking to demand procedure with the PSDL was better than the SDL for the bucking to demand procedure with the DL.

Table 3. Log demands (%), outcomes (% from diameter class overall outcome) and distribution levels as an evaluation variable for the 52 dm length class (DL_{52dm}) for the bucking to demand procedure with the distribution level (DL), the squared distribution level (SDL), the chi-squared formula (χ^2), the penalty segmented distribution level (PSDL) and the flexible penalty segmented distribution level (FPSDL) for the 52 dm log length.

Diameter class (cm)	Demand	DL	χ^2	SDL	PSDL (T=5)	PSDL (T=20)	FPSDL (F=0.5)	FPSDL (F=2)
15.0	5.0	5.0	8.5	9.7	5.2	6.2	8.1	11.4
17.7	16.0	16.3	17.0	17.0	15.6	17.6	16.0	17.0
19.0	22.0	10.3	13.6	16.1	11.4	11.4	18.1	13.2
21.1	22.0	8.3	14.0	16.1	8.4	9.3	18.4	13.5
24.2	23.0	16.6	16.2	18.3	15.3	15.1	20.0	16.3
27.9	23.0	19.9	16.9	18.9	20.9	19.2	20.4	16.9
32.5	23.0	15.4	16.4	18.8	15.9	13.2	20.4	16.6
	DL_{52dm}	78.6	79.8	84.8	79.1	76.2	90.6	78.1

The bucking technique FPSDL produced a better distribution (Table 3) for the 52 dm length class when using a tightening flexible parameter (0.5) and gave more freedom when using the loosening flexible parameter (2). The overall goodness remained quite stable with both flexible parameters, demonstrating that the tightening flexible parameter pushed error from the 52 dm length class to other classes, while the loosening flexible parameter adopted more error in the 52 dm length class.

DISCUSSION

The aim of this study was to improve the bucking to demand procedure by presenting and testing various distribution similarity parameters to compare two matrices, such as the demand matrix and the actual outcome matrix. The efficiency of the similarity parameters was determined by the bucking simulator, which was based on the bucking to demand procedure with the near-optimum approach, and with artificial stand data generated by a stand generator.

The results of this study suggest that bucking with the DL can be replaced by the distribution similarity parameter that gives more flexibility in formulating user preferences for log length-diameter distribution. In addition, since techniques with squared error term performed better than other techniques, the use of the DL as an indicator parameter to depict similarity of demand and actual outcome should be reconsidered.

Bucking to demand with the SDL produced the best results with both evaluation variables. If the user has no preference for one log length-diameter class importance over another, the SDL seems to be the most suitable parameter among those discussed in this study. However, if the user is willing to direct harmfulness for shortages and surpluses for the log length-diameter classes, the FPSDL can be used as well. The findings show that the use of different flexibility parameters in the FPSDL does not affect overall performance, but directs the error in the desired log length-diameter classes. The same conclusion can be made considering the usability of the bucking to demand procedure with the chi-square formula. If the user preferences are that an increase in the proportion of logs in a log length-diameter class should decrease the influence of the deviation, the chi-square formula might be used as well.

In practice, the bucking control parameters are given with two matrices, the price matrix and the demand matrix. The SDL would not change practice, but with either the PSDL or with the FPSDL two matrices are insufficient. In those situations, the bucking control parameters should

be formulated using four matrices: price, demand, shortage tolerance or flexibility and surplus tolerance or flexibility. This technique could easily be added to bucking control programs. For example, bucking to demand with the FPSDL would have basic flexible parameter 1 in all the log length-diameter classes for both shortage and surplus. If needed, values for desired log length-diameter classes can be decreased or increased, independently in both directions.

Although the SDL and its flexible modification were the best techniques, this does not mean that these techniques provide optimal solutions. However, it has now been shown that there are simple ways to reduce the problems of bucking to demand with the DL. The DL has a strong foothold in practice; but its disadvantages are many and it is hoped that this and other studies will offer new techniques for cut-to-length harvesting.

ACKNOWLEDGEMENTS

This study was conducted at the Faculty of Forestry, University of Joensuu. The study was part of the projects "Pre-information of timber stock" and "Timber flow model in wood procurement planning" funded by the Finnish Foundation for Research on Natural Resources and "Group decision support in wood procurement", funded by the Academy of Finland. The authors would like to thank Mr Lauri Mehtätalo for valuable comments on the manuscript.

AUTHOR CONTACT

Jukka Malinen can be reached by e-mail at --
jukka.malinen@joensuu.fi

REFERENCES

- [1] Bellman, R. E. 1954. The theory of dynamic programming. *Bull. Am. Math. Soc.* 60: 503-516.
- [2] Bergstrand, K-G. 1990. *Fördelningsapatering - ett sätt att tillgodose sågverksönskemål. [Bucking to demand, the way to fulfil sawmill objectives]. Skogforsk Resultat Nr 11/1990.*
- [3] Gregorius, H. R. 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Mathematical Biosciences* 41: 253-271.

- [4] Kirkkala, A., L. Sikanen, P. Harstela, T. Ruha and T. Tarnanen. 2000. Sakkosegmentoitu tavoitejakauma apteeraustuloksen arvioinnissa ja jakauma-asteen laskennassa. [Penalty segmented demand distribution for evaluating bucking outcome and calculating distribution level]. *Metsätieteen aikakauskirja* 1/2000: 59-61.
- [5] Lukkarinen, E. 1999. Jakauma-apteeraus täyttää sahan toiveet. [Bucking to demand fulfills sawmill objectives]. *Ponsse uutiset* 2/99: 7.
- [6] Lukkarinen, E. and T. Vuorenpää. 1997. Hakkuukoneen apteerauksen ohjausta konttorista käsin. [Directing harvester bucking from the office]. *Koneyrittäjä* 1/1997: 12-13.
- [7] Malinen, J. 2003. Locally adaptable non-parametric methods for estimating stand characteristics for wood procurement planning. *Silva Fennica* 37(1): 109-120.
- [8] Malinen, J. and T. Palander. 2003. Impact of pre-information from timber stock and back-hauling on timber flow. Manuscript.
- [9] Malinen, J., M. Maltamo and P. Harstela. 2001. Application of most similar neighbour inference for estimating marked stand characteristics using harvester and inventory generated stem databases. *International Journal of Forest Engineering*. 12(2): 33-41.
- [10] Metinfo. 2002. Forest information services. Research-based, forest-related information services and expert systems. <http://www.metla.fi/metinfo/index-en.htm> (as accessed on 4.2.2003).
- [11] Möller, J. and I. von Essen. 1997. Fördelningsaptinger en fungerande metod även på smp trakter och vid liten tillåten värdeavvikelse. [Dimension-apportionment merchandising-a useful method even for small sites with minor value deviation]. *Skogforsk Resultat* Nr 14/1997.
- [12] Oinas, S and L. Sikanen. 2000. Discrete event simulation model for purchasing of marked stands, timber harvesting and transportation. *Forestry* 73(3): 283-301.
- [13] Palander, T. 1998. Tactical models of wood-procurement teams for geographically decentralized group decision-making. D.Sc. (Agr. and For.) thesis. University of Joensuu, Faculty of Forestry. Academic Dissertation. ISBN 951-708-705-5.
- [14] Reynolds, M.R.Jr., T.E. Burk and W-C. Huang. 1988. Goodness-of-fit tests and model selection procedures for diameter distribution models. *For. Sci.*34(2):373-399.
- [15] Sondel, J. 1993. Ändamålsanpassad aptering. [Bucking to order]. *Skogforsk Resultat* Nr 18/1993.
- [16] Sondel, J. and I. von Essen. 1995. Apteringsdatortest 1995 – studier av sex apteringssystem. [Merchandising-computer trials 1995 – Studies on six merchandising systems]. *Skogforsk Redogörelse* Nr 4/1996.
- [17] Tommola, M., M. Tynkkynen, J. Lemmetty, P. Harstela and L. Sikanen. 1999. Estimating the characteristics of a marked stand using k-nearest-neighbor regression. *Journal of Forest Engineering* 10(2): 75-81.
- [18] Virtanen, J-P. 1998. Jakauma-apteeraus. [Bucking to demand]. Department of Mathematical Sciences, University of Tampere, Report C 59.
- [19] Väättäinen, K. 1998. Hakkuukoneen tuottaman puutavaralajikertymän ja tukkijakauman ennustaminen SpruceOpti –ohjelmalla. [Prediction of harvester produced timber assortments and log distribution with the SpruceOpti program]. Msc thesis. University of Joensuu, Faculty of Forestry.
- [20] Weijo, A. 2000. Runkopankki Metsähallituksen puunhankinnanohjauksen apuvälineenä. [Stem database as an instrument for wood procurement control in Metsähallitus]. Msc thesis. University of Joensuu, Faculty of Forestry.