

Empirical Prediction Models for *Vaccinium myrtillus* and *V. vitis-idaea* Berry Yields in North Karelia, Finland

Marjut Ihalainen, Kauko Salo and Timo Pukkala

Ihalainen, M., Salo, K. & Pukkala, T. 2003. Empirical prediction models for *Vaccinium myrtillus* and *V. vitis-idaea* berry yields in North Karelia, Finland. *Silva Fennica* 37(1): 95–108.

Forest berries and the outdoor experiences related to berry collection are important goods and services provided by Finnish forests. Consequently, there is a need for models which facilitate the prediction of the impacts of alternative forest management options on berry yields. Very few such models are available. In particular, empirical models are lacking. Models used in forest management should express the effect of variables altered in forest management such as stand density and mean tree size. This study developed empirical models for bilberry and cowberry yields in North Karelia. The data consisted of 362 measurements of 40 m² sample plots. The plots were located in clusters. The same plot was measured over 1 to 4 years. Besides berry yield some site and growing stock characteristics of each plot were measured. A random parameter model was used to express the berry yield as a function of site fertility, growing stock characteristics, and random parameters. The random part of the models accounted for the effect of plot, measurement year, and cluster. The fixed predictors of the model for bilberry were stand age and forest site type. Stand basal area, mean tree diameter and forest site type were used to predict cowberry yields. The most significant random parameter was the plot factor. The fixed model part explained only a few per cent of the variation in berry yields. The signs of regression coefficients were logical and the model predictions correlated rather well with the predictions of earlier models.

Key words berry yield index, mixed models, *Vaccinium myrtillus*, *Vaccinium vitis-idaea*

Authors' address Ihalainen & Pukkala: University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, Finland; Salo: The Finnish Forest Research Institute, Joensuu Research Centre, P.O. Box 68, FIN-80101 Joensuu, Finland

E-mail marjut.ihalainen@joensuu.fi; kauko.salo@metla.fi; timo.pukkala@joensuu.fi

Received 12 February 2002 **Accepted** 21 October 2002

1 Introduction

Objectives other than those based solely on wood production carry increasing weight in forestry decision-making. Not only the public but also private forest landowners, both industrial and non-industrial, value the multiple-use aspects of forests (Kangas 1998). According to a survey made by Kangas and Niemeläinen (1996), such aspects as the vitality of forests, scenic beauty, biodiversity, and berries and mushrooms are regarded as the most important values of forests in Finland. Picking forest berries, for example, provides many kinds of utilities for people. It is a very popular outdoor recreation activity; 87% of Finns collect berries and mushrooms occasionally or frequently (Kangas and Niemeläinen 1996). In addition, berries are widely collected both for household consumption and trade. The total value of the harvested wild berry crop was estimated at EUR 115 million in 1997 and EUR 95 million in 1998. These values include both picking for home use and commercial picking (Saastamoinen et al. 2000).

To integrate non-wood forest products and services in numerical forest planning, production functions applicable to planning calculations are needed. However, very few such functions are available. In particular, there is a lack of functions which are based on empirical studies (Kangas 1998). Four production functions have been prepared in Finland for bilberry (*Vaccinium myrtillus*) and cowberry (*Vaccinium vitis-idaea*), which are the most common forest berries in Finland (Pukkala 1988, Muhonen 1995, Ihalainen and Pukkala 2001, Ihalainen et al. 2002), but only the models produced by Pukkala (1988) are based on empirical but scanty measurements. One reason for this situation is that it is very laborious to collect enough modelling data for empirical models. Expert modelling, which was used in three of the studies mentioned above and which is not resource demanding, can only relieve the acute need for prediction models and produce temporary models for forest planning cheaply and quickly (Kangas and Mononen 1997).

When one develops empirical berry yield models for forest planning purposes, site and growing stock characteristics are the most rea-

sonable predictors. This is because the site and stand characteristics are known in forest planning and forest management greatly affects the stand characteristics. Berry yields also depend on the state of ground vegetation and weather conditions. Variables like this, however, are seldom useful in forest planning because they are usually unknown.

To construct models that reliably describe the effect of trees and the site on the berry yield, one needs to gather large quantities of empirical data for many years (Belonogova and Kuchko 1979). There are many factors for the high temporal variation in berry yields, such as frost and variations in precipitation, temperature and pollination success (e.g. Nousiainen 1983, Solantie 1983, Wallenius 1999). In addition, the berry crop and weather conditions from the previous year (e.g. Laine 1988, Laakso et al. 1990) and the harshness of the winter (e.g. Raatikainen and Vänninen 1988) have an effect on the berry production of the subsequent growing season. Berry yields also vary spatially within a stand since wild berries are distributed in a clustered manner. Therefore, if the aim of the study is to create a berry yield prediction model for a certain geographical region, a comprehensive sample plot network is needed.

The aim of this study was, on the basis of field measurements, to formulate empirical prediction models for bilberry and cowberry yields. Berry yields were measured in four consecutive years. By means of the mixed model technique it was possible to take into account the annual variation of berry yields in the models. As a by-product, the statistical method produced annual berry yield indices for the study years. The models of the present study were prepared for forest planning purposes and they can be utilised in regions which belong to the middle boreal coniferous zone in Finland.

2 Material and Methods

2.1 Study Area

The study material consisted of the bilberry and cowberry yields of 362 berry sample plots which were inventoried during 1981–1984 and of the site

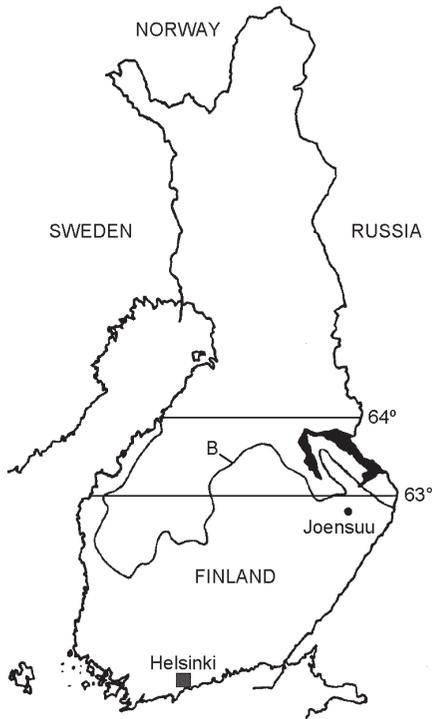


Fig. 1. Location of the study area (black). B = border between southern and middle boreal vegetation zones (Salo 1993).

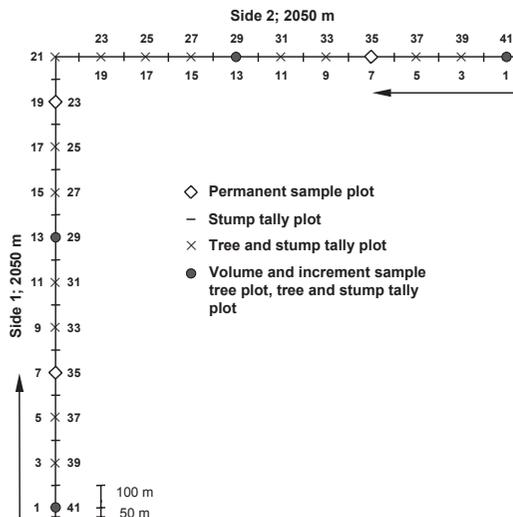


Fig. 2. Layout of a cluster used in the Finnish National Forest Inventory (Salo 1993). Plot numbering depends on walking direction (arrows).

and growing stock characteristics corresponding to the sample plots as measured in 1980. When the berry yield was assessed on more occasions than in a single year the same sample plot also occurs more than once in the data.

The plots are situated in the transitional area between the southern and middle boreal coniferous zones (Ahti et al. 1968, Kalliola 1973), most of it belonging to the middle boreal coniferous zone (Fig. 1). The study area is located within the Nurmes and Lieksa districts and is owned by the Forest and Park Service of Finland. The area has a mean elevation of 185 m a.s.l., and an effective temperature sum (> 5 °C) during the growing season of about 1000 d.d. (Sevola 1983).

According to the 7th Finnish National Forest Inventory (7NFI), a network of permanent sample plots was established in the study area in 1980 (Salo 1993). The established sample plot network was denser than the normal cluster density in the Finnish National Forest Inventory: the clusters were located at 4 km intervals instead of the normal 8 km. Three permanent sample plots were established in each cluster (Fig. 2). Four sub-plots of 10 m² were delineated at the corners of each permanent sample plot (Fig. 3). These four rectangular sub-plots formed one 40 m² berry sample plot. The berry sample plots were used as observations in the modelling.

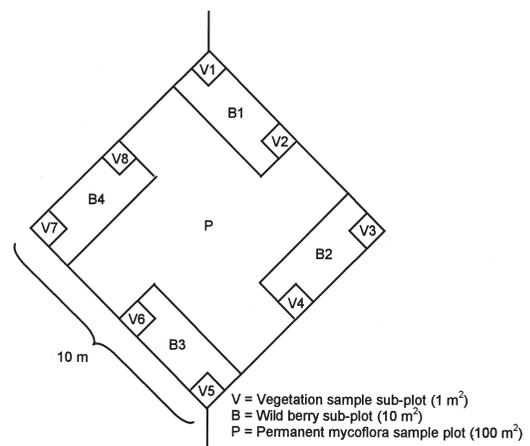


Fig. 3. Layout of the permanent sample plot (100 m²) and the location of four wild berry sub-plots which formed one berry sample plot (40 m²) (Salo 1993). Permanent mycoflora sample plot in this figure is equal to permanent sample plot in Fig. 2.

2.2 Measurements

When the permanent sample plots were established in 1980, the NFI crew recorded the site and stand characteristics according to the field instructions of the 7th NFI (Valtakunnan metsien... 1977). On average, the berries were inventoried three times during the growing seasons of 1981 and 1982, and twice in 1983. In 1984 the berry sample plots were inventoried once. On each 10 m² sub-plot, the berry species, the dominant height of the berry vegetation, the number of unripe berries and the number of ripe berries were recorded. The three latter variables were determined separately for different berry species. The berries were not picked when the berry yield was inventoried and special care was taken to avoid trampling the berry vegetation.

In the years when berries were inventoried several times during the growing season the intervals between the different inventory times were long. In the case of bilberry, for example, the second inventory in 1981 was frequently too late, i.e. a considerable proportion of the ripe berries had dropped off, birds had fed on them or people had picked the berries. In the case of cowberry, instead, the second inventory was very often too early with respect to the development of ripe berries.

For these reasons, the biomass of bilberries and cowberries for each sample plot was determined by means of two different methods and the higher biomass was used. Firstly, the number of ripe berries was multiplied by the mean weight of one ripe berry (0.31 g for bilberry and 0.23 g for cowberry). The mean weight was calculated from 30 ripe berries picked and weighed from each sample plot. Secondly, the number of unripe berries was multiplied by the mean weight of one ripe berry and this value was multiplied by 0.8. The coefficient 0.8 is the proportion of those unripe berries which will develop into ripe ones. The multiplier was calculated on the basis of Salo's data collected in 2000 (unpublished material) from permanent experimental plots established in different areas of Finland in order to study the yields of several economically important wild berries (Salo 1999).

In the present study all sample plots which were located on mineral soil sites were included in the

study material. This means that those berry sample plots which did not have any bilberry or cowberry vegetation were also included in the material.

2.3 Characteristics of the Sample Plots

Most sample plots belonged to medium or rather poor soil fertility classes (VMT, DeMT, EVT) (Lehto and Leikola 1987). 13% of the sample plots represented poor mineral soil sites (ECT) whereas rich soil sites (GOMT) were very rare (Table 1).

The permanent sample plots from a single inventory cluster usually belonged to different stands, since they were located at a considerable distance from each other (Fig. 2). The growing stock characteristics were estimated for the stand within which the berry plot was located. None of the berry plots crossed a stand boundary. The following stand characteristics were available in the data set used in this study: mean diameter, stand basal area, stand age and dominant tree species (Table 2).

Most stands making up the study material represented commercially managed forests. All of the stages of stand development were represented and the whole range of variation in stand density was covered (Table 2). Some plots were in virgin forests. Most sample plots (81%) were situated in pine-dominated stands and 15% were spruce-dominated stands. The remaining plots were treeless.

2.4 Statistical Analysis

The distributions of bilberry and cowberry yields were skewed so that the proportions of zero and small values were emphasised in the data. In the case of bilberry, 29% of the observations were zeros. For cowberry, the corresponding figure was 39%. In order to linearize the relationships and convert the residuals to resemble normal distributions, several transformations of the response (y) were attempted. Logarithms were found to be the best transformation. To avoid taking logarithms of zeros, one was added to the berry yields. Thus, the predicted variable in the modelling was $\ln(y + 1)$.

Table 1. Number of sample plots representing various forest site types in the middle boreal vegetation zone (Pohjanmaa-Kainuu) in 1981–1984. The site fertility description corresponding to each forest site type is given below each type.

Year	GOMT rich	VMT, DeMT medium	EVT rather poor	ECT poor	Total
1981	3	19	15	6	43
1982	1	24	40	9	74
1983	1	47	70	16	134
1984	–	37	59	15	111
Total	5	127	184	46	362

GOMT *Geranium-Oxalis-Myrtillus* Type

VMT *Vaccinium-Myrtillus* Type

DeMT *Deschampsia-Myrtillus* Type

EVT *Empetrum-Vaccinium* Type

ECT *Empetrum-Calluna* Type

Table 2. Mean and range of bilberry and cowberry yields and some of the stand characteristics in the study material.

Characteristic	Minimum	Mean	Maximum	Unit
Bilberry yield	0	13	130	kg/ha
Cowberry yield	0	3	95	kg/ha
Mean diameter	0	14	40	cm
Stand basal area	0	14	46	m ² /ha
Stand age	0	72	205	year

The bilberry and cowberry observations were correlated both spatially and temporally. The berry yield observations on the same cluster are examples of spatially correlated observations. It is also likely that in a certain year the berry yields are more similar to each other than would be the average situation. Further, the berry yields measured on the same sample plot in several years are correlated observations. For these reasons the mixed model technique, which is commonly applied to hierarchical and complex data sets (Lappi 1986), was employed. The MIXED procedure of SAS software (SAS Institute Inc. 1992) was used for model fitting.

The mixed model contains two parts: a fixed part and a random part (see e.g. Penner et al. 1995). In the present study the fixed part consisted of the site and stand characteristics. Both the site properties and the dominant tree species were taken into account by creating various dummy variables. In addition, several transformations (e.g. the square root of the mean diameter) and

interactions (e.g. site dummy variable \times mean diameter) of variables measured on an interval or ratio scale were used as additional potential predictors. The parameters of the fixed part were estimated using the Generalized Least Squares (GLS) technique. The potential variables of the random part included the year, cluster and sample plot as well as some of the interactions of these variables (e.g. year \times cluster).

Several combinations of potential predictors were tried. A certain predictor was accepted for the model if it was statistically significant (the significance level used was 0.05). The random part was formulated by using the same rule: if a random effect was found to be statistically significant it was included in the random part. However, the random year effect was imposed on the model. This made it possible to calculate the berry yield indices (I_t) for the study years (cf. Henttonen 1990, Miina 2000). The following formula was used for this calculation:

$$I_t = 100e_t + 100 \quad (1)$$

where e_t is an estimate of the random year effect ($t = 1981, \dots, 1984$). The expectation value of I_t is 100, which depicts the average level of the berry yields in 1981–1984. The estimates of the year effects (e_t) multiplied by one hundred can be interpreted as percentage deviations from the average level.

The logarithmic predictions for bilberry and cowberry yields are usually calculated using only the fixed part of the mixed model. In response to the logarithmic predictor of berry yields a correction term needs to be applied when the prediction is transformed into the arithmetic scale. A multiplicative correction factor ($\exp(\hat{\sigma}^2 / 2)$, where $\hat{\sigma}^2$ is the sum of variance estimates of the random effects) suggested by Baskerville (1972) was tried but it resulted in biased back-transformed predictions. In consequence, a ratio estimator for bias correction in the logarithmic regression was applied for Eqs. 6 and 8 (Snowdon 1991). The proportional bias in logarithmic regression was estimated from the ratio of the mean berry yield added by one ($\bar{y} + 1$) and the mean of the back-transformed predicted values from the regression $\overline{\exp\{\ln(\hat{y} + 1)\}}$. Hence, the ratio estimator was

$c = \overline{(y+1)} / \exp\{\ln(\hat{y}+1)\}$. The logarithmic predictions for berry yields are transformed into absolute berry yields as follows:

$$\hat{y} = \exp\{\ln(\hat{y}+1)\} \times c - 1 \tag{2}$$

where \hat{y} is berry yield in kilograms per hectare.

The following statistics were calculated for the models created in this study:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-1}} \tag{4}$$

$$\text{bias} = \frac{\sum (y_i - \hat{y}_i)}{n} \tag{5}$$

where R^2 is the degree of determination of the model, RMSE is the root mean square error and the bias is an absolute bias; n is the number of observations, \bar{y} is the mean berry yield (calculated from the measured values; see Table 2) and y_i and \hat{y}_i are the measured and predicted berry yields, respectively.

3 Results

3.1 Prediction Model for Bilberry Yield

The mixed model for predicting bilberry yields is as follows (standard errors of the regression coefficients are given in parentheses):

$$\ln(y_{tk} + 1) = 0.0830 + 0.0103 t_g + 0.9904 D_1 + 0.4997 D_2 + e_t + e_k + \varepsilon_{tk} \tag{6}$$

(0.2733) (0.0016) (0.2686) (0.2565)

where

y_{tk} = bilberry yield in sample plot k in year t (kg/ha)

t_g = mean age of trees (year)

D_1 = site dummy: $D_1 = 1$, if the forest site type is medium (VMT, DeMT) (Table 1), and $D_1 = 0$ otherwise

D_2 = site dummy: $D_2 = 1$, if the forest site type is rather poor (EVT) (Table 1), and $D_2 = 0$ otherwise

e_t = random effect of year t (between-year variation)

e_k = random effect of sample plot k (between-

Table 3. Variance estimates for the random parts of prediction models 6 and 8.

Variable	Estimate	Standard error	p-value
Model 6			
Year	0.0546	0.0449	0.2241
Sample plot	0.9264	0.1367	0.0001
Random error	0.6171	0.0652	0.0001
Total	1.5981		
Model 8			
Year	0.0807	0.0624	0.1963
Year × cluster	0.0766	0.0367	0.0368
Sample plot	0.4159	0.0668	0.0001
Random error	0.2515	0.0391	0.0001
Total	0.8247		

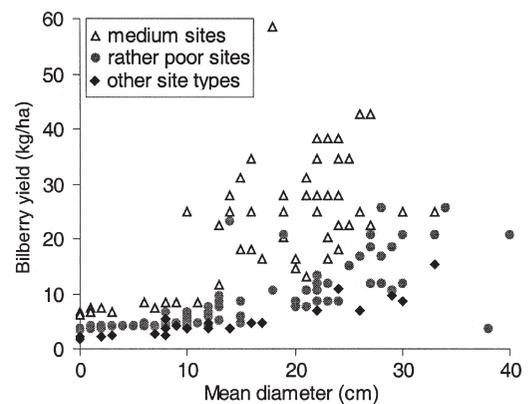


Fig. 4. Predicted bilberry yields for the study stands as a function of the mean diameter of trees. The predictions were calculated by using Model 7. Medium, rather poor and other (rich and poor) site types are marked with different symbols (see Table 1).

sample plot variation)

ε_{tk} = random error (within-sample plot and within-year variation)

According to Model 6, the most abundant bilberry yields may be found in a mature stand which is situated on a mineral soil site of medium fertility (*Vaccinium-Myrtillus*, *Deschampsia-Myrtillus* types). Rather poor soil sites (*Empetrum-Vaccinium* type) also produce good bilberry crops. These facts can also be concluded from Fig. 4,

where the mean diameter of trees correlates positively with the bilberry yield and, in addition, the slope of the relationship is highly dependent on the forest site type.

The fixed model part explains 6% of the total variance in the back-transformed bilberry yield ($R^2 = 0.06$). The RMSE for the back-transformed berry yield is 22.06 kg/ha.

The random part of Model 6 indicates that bilberry yields varied randomly between the different sample plots (Table 3). A major part of the residual variation (58%) was caused by the random sample plot effect. The within-sample plot and within-year variation accounted for almost 40% of the variation and the proportion explained by the year effect was merely a few percentage points.

When the prediction model for bilberry yield is applied in practice the following form of Eq. 6 is used:

$$y_b = \exp(0.0830 + 0.0103 t_g + 0.9904 D_1 + 0.4997 D_2) \times 2.4507 - 1 \quad (7)$$

where y_b refers to the bilberry yield (kg/ha). The multiplier 2.4507 in Eq. 7 is a ratio estimator for bias correction.

3.2 Prediction Model for Cowberry Yield

The model devised for predicting cowberry yields consists of the following equation (standard errors of the regression coefficients are given in parentheses):

$$\ln(y_{ijk} + 1) = 1.0560 + 0.0005 D_3 d_g^2 - 0.1196 \sqrt{G} + e_t + e_{ij} + e_k + \varepsilon_{ijk} \quad (8)$$

(0.1818) (0.0002) (0.0308)

where

y_{ijk} = cowberry yield in sample plot k in cluster j in year t (kg/ha)

D_3 = site dummy: $D_3 = 1$, if the forest site type is rather poor (EVT) or poor (ECT) (Table 1), and $D_3 = 0$ otherwise

d_g = mean diameter of trees (cm)

G = stand basal area (m^2/ha)

e_{ij} = random effect of cluster j in year t (year \times cluster interaction)

ε_{ijk} = random error (within-sample plot, within-year, within-cluster variation)

others as in Eq. 6

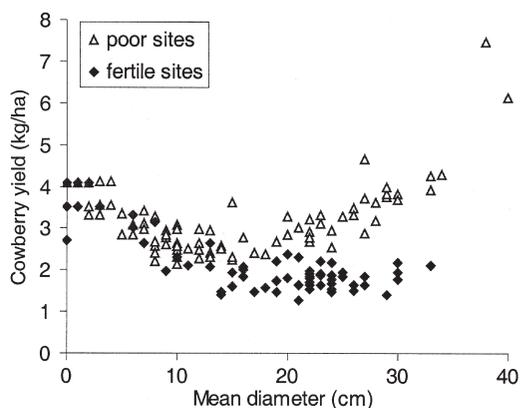


Fig. 5. Predicted cowberry yields for the study stands as a function of the mean diameter of trees. The predictions were calculated by using the transformed version (see Formula 2) of Model 8. Fertile (i.e. medium and rich) sites are marked with different symbols than poor (i.e. rather poor and poor) sites (see Table 1).

Eq. 8 and also Fig. 5 indicate that forests of the *Empetrum-Vaccinium* type or poorer produce the highest cowberry yields. On poor sites the stage of stand development affects yields so that the best yields can be found either in recently clear-felled open areas and in young seedling and sapling stands or in mature stands (Fig. 5). On mineral soil sites of medium and rich fertility cowberry production seems to decrease when the forest grows older (Fig. 5). Stands with low

density of trees are the most suitable for cowberry collection.

The fixed part of Model 8 explains 3% of the total variance in a back-transformed berry yield ($R^2 = 0.03$). The RMSE for a back-transformed berry yield is 7.87 kg/ha.

All random components except the year effect were statistically significant (Table 3). This suggests that in a given year the level of cowberry yields varied from cluster to cluster. In addition, variation occurred in cowberry crops between different sample plots. Almost one third of the

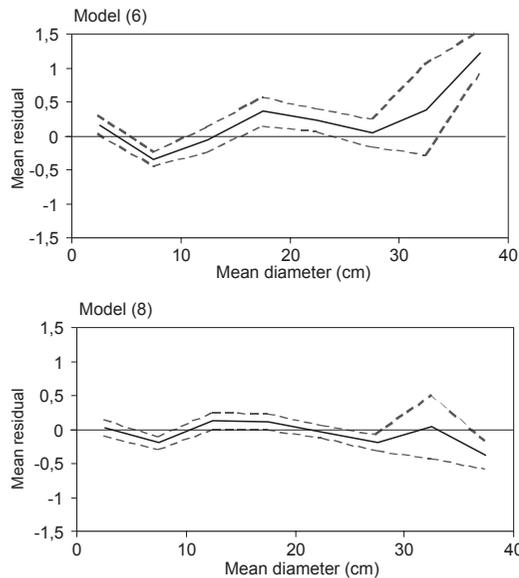


Fig. 6. Mean residuals (in log-scale) of the fixed part of Models 6 and 8 as a function of the mean diameter of the trees (dashed lines indicate the standard error of the mean).

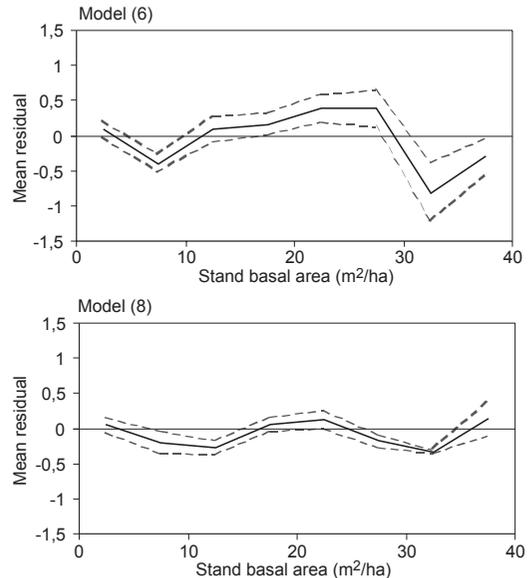


Fig. 7. Mean residuals (in log-scale) of the fixed part of Models 6 and 8 as a function of stand basal area (dashed lines indicate the standard error of the mean).

variation consisted of random error. The variance in the random sample plot effect accounted for half of the total variance in the random effects. The proportions of year \times cluster and year were 9% and 10%, respectively.

The logarithmic predictions for cowberry yields calculated by means of Model 8 are back-transformed into absolute berry yields using Eq. 2. In this case, a ratio estimator for bias correction is 1.7713.

3.3 Model Validation

The mean residuals of the fixed part of Models 6 and 8 were generally close to zero and did not depend on the mean diameter of trees (Fig. 6) or the stand basal area (Fig. 7). The largest absolute values for mean residuals, especially in the case of Eq. 6, were found in the largest diameter and stand basal area classes ($d_g > 30$ cm, $G > 30$ m²/ha). This was assumed to be due to the low number of observations. The variation in residuals was larger for bilberry than for cowberry.

The average bias of Models 6 and 8 on the absolute scale of yield was equal to zero as a result of the ratio estimator used for bias correction. Hence, on average the models produced unbiased predictions for berry yields in the modelling data.

The validity of the models was also tested by comparing the model predictions with previous models, using the data from the study made by Ihalainen and Pukkala (2001) (495 stands on mineral soil sites). In the case of bilberry the predicted yields computed using Eq. 7 from the present study correlated significantly with the predictions calculated by using the models of Pukkala (1988), Muhonen (1995), Ihalainen and Pukkala (2001) and Ihalainen et al. (2002); the correlations were 0.410, 0.308, 0.796 and 0.707, respectively (Fig. 8). Typical of the bilberry yield prediction model in the present study is a very clear effect of site fertility class on the berry yields (Fig. 8).

In the case of cowberry the predicted yields (transformed into kg/ha) calculated by using Eq. 8 correlated significantly but not very strongly with the models established by Muhonen (1995),

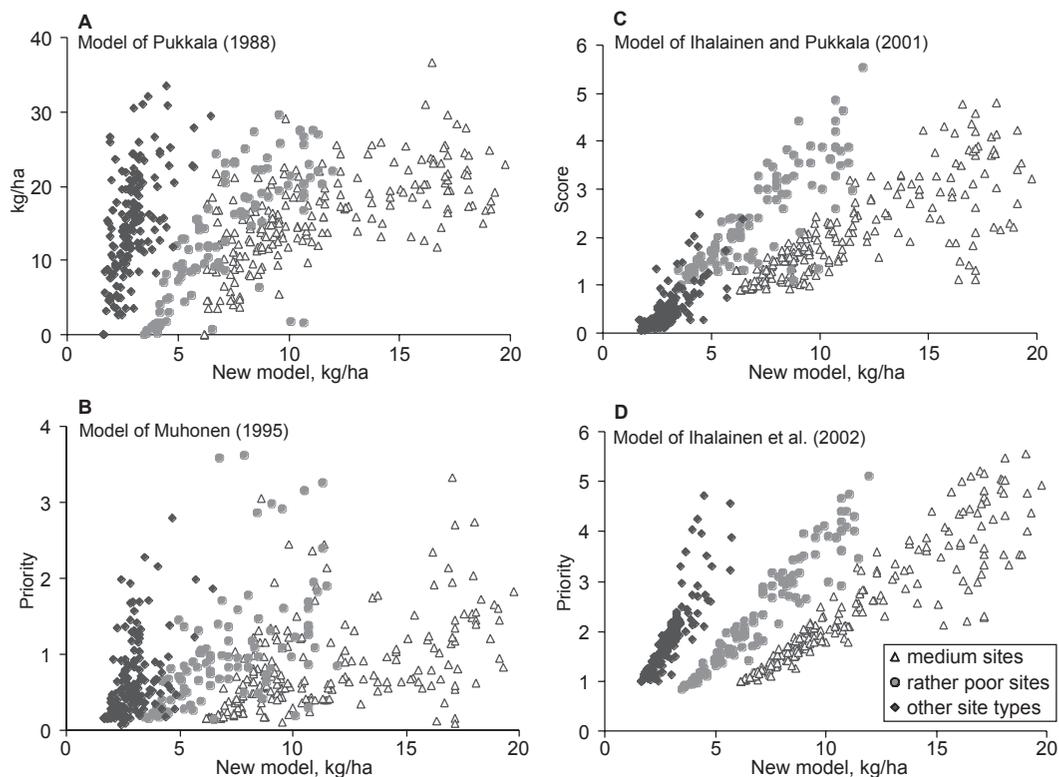


Fig. 8. Correlation between the predicted bilberry yields calculated by Eq. 7 of this study and the models of A) Pukkala (1988), B) Muhonen (1995), C) Ihalainen and Pukkala (2001) and D) Ihalainen et al. (2002). Medium, rather poor and other site types are marked with different symbols (see Table 1).

Ihalainen and Pukkala (2001) and Ihalainen et al. (2002); the correlations were 0.407, 0.406 and 0.210, respectively (Figs. 9 B–D). The correlation with the model made by Pukkala (1988) was poor, 0.090 (Fig. 9 A). It has also been found earlier (Ihalainen and Pukkala 2001, Ihalainen et al. 2002) that cowberry yields predicted by using the model set up by Pukkala (1988) do not correlate very strongly with the other models.

3.4 Berry Yield Indices

Fig. 10 indicates that the bilberry yield was highest in 1983 in comparison with the yields for the other study years. 1982, in contrast, was a poor bilberry year. Bilberry production in 1981 was somewhat higher and in 1984 somewhat lower than the average level of the study period.

The most abundant cowberry crop was found in 1981 (Fig. 10). It seems that during the four-year study period the cowberry yields decreased year by year. In 1981–1983 berry production was nevertheless above the average level for the study period, but in 1984 the cowberry yield was very poor.

4 Discussion

4.1 Analysis of the Methods

In the present study prediction models for bilberry and cowberry yields were prepared using data sets which were based on empirical measurements of berry yields and site and growing stock characteristics. The berry yields were measured in four

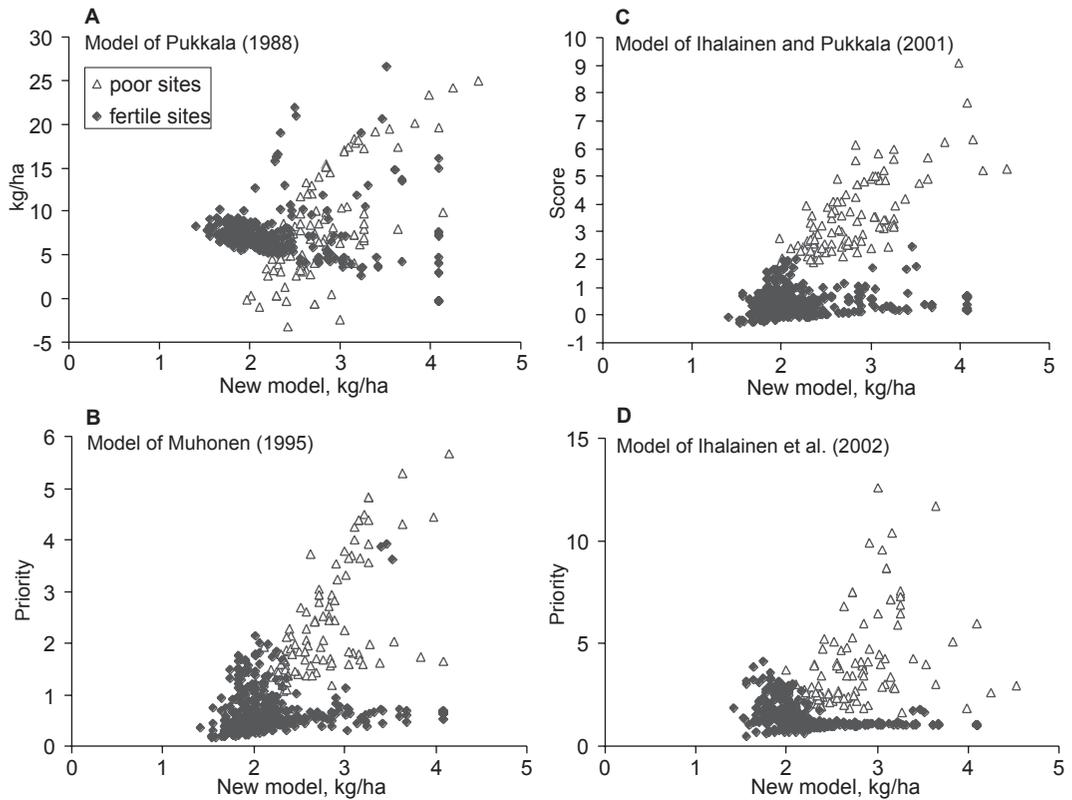


Fig. 9. Correlation between the predicted cowberry yields calculated by using the back-transformed version (see Formula 2) of Eq. 8 in this study and the models produced by A) Pukkala (1988), B) Muhonen (1995), C) Ihalainen and Pukkala (2001) and D) Ihalainen et al. (2002). Fertile (i.e. medium or more fertile) sites are marked with different symbols than poor (i.e. rather poor or poorer) sites (see Table 1).

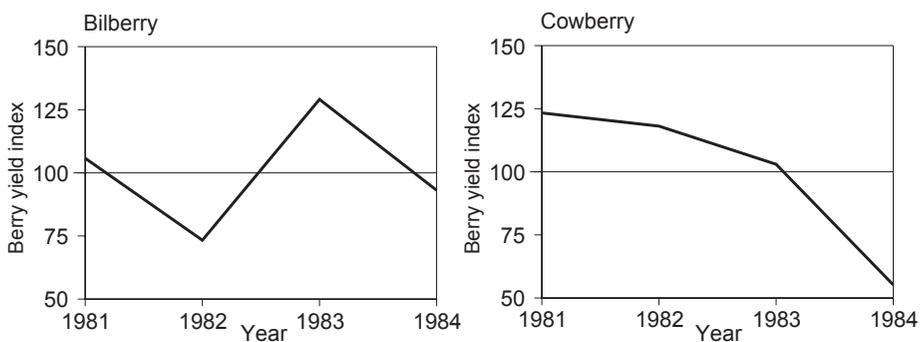


Fig. 10. Bilberry and cowberry yield indices for the study years 1981–1984. The indices are produced by Models 6 and 8. The index value of 100 is the average level of berry production during the study period.

consecutive years in sample plots which were located in clusters. The models were constructed as mixed linear models with several random effects. Until now, the annual variation in berry yields has not been taken into account when preparing prediction models for bilberry and cowberry yields. The use of random components provided estimates of the magnitude of different sources of variation. Penner et al. (1995) stated that, particularly when observations are nested within a hierarchical arrangement, with each level containing its own sources of variation, the analysis of the variance components can improve the model.

The modelling data contained some features and limitations which caused problems and may also affect the model predictions. The scarcity of explanatory variables was one problem: only a few growing stock characteristics could be used in the modelling. In addition, none of these variables was measured separately for each tree species and for each canopy layer, which would have been advantageous for the development of prediction models for forest planning purposes. However, the variables most essential to the description of forest structure were available for this study.

Further, a problematic feature of the study material was the abundance of zero observations in both the bilberry and the cowberry yield data. Almost one third of the berry sample plots had no bilberries, and the corresponding figure for cowberry was about 40%. Because bilberry and cowberry occur in a patchlike manner (e.g. Sepponen and Viitala 1982, Laakso et al. 1990) and the berry sample plots were quite small it is not surprising that a great number of sample plots had no berry vegetation or had only sterile vegetation (Salo 1995).

In this study, an alternative approach of modelling would have been to estimate the probability of occurrence (presence/absence) of berries by logistic functions and then in the second step estimate the abundance (given presence of berries). However, when the logistic model is used in applications, a threshold probability is needed to divide the predictions into presence and absence cases, which complicates the usage of the model.

In the light of the experience gained in this study further studies should pay special atten-

tion to the inventory arrangements. It would be reasonable to time berry yield inventories to coincide only with periods when the berries are ripe, since it is quite difficult to estimate the berry yields reliably on the basis of unripe berries. Further, in order to increase the accuracy of prediction models several inventories of berry yields should be made during the period when bilberries or cowberries are worth picking. This, however, requires that berries within the sample plots should be picked and not merely counted so that the same berries are not included in several assessments. An inventory method like this would require considerable resources but it would also increase both the accuracy and the reliability of berry yield predictions.

One advantage of the statistical method used in this study, namely the mixed model technique, was that it was possible to estimate the berry yield indices for the study years. This may be the first occasion on which these indices have been applied in berry yield studies. This approach has been used earlier for computing growth indices for different tree species (e.g. Henttonen 1990, Hökkä 1997, Miina 2000). Even though in this study the berry yield indices were similar to the annual average berry yields (Salo 1982, 1983), this may not necessarily hold true in other cases in which the data have a hierarchical structure. If the indices and average values do not correspond with each other, one should rely on the indices rather than the mean values. This is because the possible imbalance in the data can be taken into account when the indices are computed. In the present study, for example, the imbalance was caused by the fact that different quantities of berry sample plots were inventoried in each study year. In addition, different sample plots were visited in different years. Hence, if in a case like the present one the annual average berry yields are calculated on the basis of data sets collected in different years and then compared with each other, the comparison may be misleading. The use of the mixed model technique will eliminate this problem when the year effect is included in the random part of the model (see e.g. Henttonen 1990).

4.2 Analysis of the Results

The degree of determination (R^2) which was calculated for the fixed part of the models using back-transformed units was low, ranging from 0.03 to 0.06. Most probably this resulted from the problematic features of the modelling data, as discussed above. A major part of the variation in the berry yields was accounted for by the random plot factor. The plot factor most probably explained any variation caused by the patchy occurrence of berry vegetation and also the effect of those stand and site variables which were not used as predictors. However, the models were very significant and no large biases were detected in the residuals of the fixed model part when plotted against the mean diameter of the trees and the stand basal area.

According to the prediction model for bilberry yield, the best crops may be found in a mature stand, whereas openings and seedling and sapling stands are not very suitable for bilberry gathering. This result is supported by many earlier studies (e.g. Zvorykina 1970, Eriksson et al. 1979, Jaakkola 1983, Raatikainen et al. 1984). Further, forests of medium fertility seem to produce the highest bilberry yields. Berry production on rather poor mineral soil sites is also good. Raatikainen et al. (1984), Belonogova (1988) and Kuchko (1988) have earlier stated that forests of the *Myrtillus* and *Vaccinium* types are the most advantageous with respect to high bilberry yields in the southern boreal vegetation zone. Thus, it seems that forests of medium and rather poor fertility have the best bilberry production both in the southern and middle boreal vegetation zones.

Like many earlier studies (e.g. Raatikainen 1978, Eriksson et al. 1979, Jäppinen et al. 1986) the prediction model in this study also indicates that cowberry produces good crops in stands with low density of trees. This is quite obvious since cowberry is a photophilous plant which does not thrive in closed stands (Belonogova 1993). In addition, the results of the present study suggest that the best cowberry yields can be found on poor mineral soil sites (*Empetrum-Vaccinium* type or poorer). On poor sites, gaps and young seedling and sapling stands on the one hand and mature stands on the other seem to be most suitable for cowberry collection. These findings are similar to

those of many previous studies (e.g. Raatikainen et al. 1984, Ihalainen and Pukkala 2001, Ihalainen et al. 2002).

It is very likely that the models for this study produce underestimates for the berry yields. There are at least two potential explanations for this. Firstly, the berry yields during the study period were for the most part low in the province of North Karelia, within which the study area is located (Jäppinen et al. 1986, Wallenius 1999). Fortunately, the models can be easily calibrated to correspond to a typical berry year. Secondly, the models predict the mean yield, which is much lower than the yield from sites where berries are usually collected. Eriksson et al. (1979) found that in Sweden 70–80% of all forested land had either no or only a rare occurrence of bilberry blossoms and ripe bilberries during the four years under observation. Thus, only 20–30% of the forest produced bilberries either moderately or abundantly. In the case of cowberry, low fertility occurred on about 80–90% of the forest land (Eriksson et al. 1979). For these two reasons it is justified to assume that the berry yields from areas known to be good berry sites are like to be very much greater than the predictions computed by the models in the present study.

The models established in the present study were developed for forest planning purposes. Since the prediction models for bilberry and cowberry yields produce absolute berry yield predictions (kg/ha), they can also be utilised when estimating the berry production of a particular geographical region which is located within the middle boreal coniferous zone in Finland. Thus, the secondary use of the models is that they provide new possibilities for estimating the regional resources and biological supply of the major wild berries in Finland. However, revised models based on a larger set of empirical measurements would permit more reliable estimates to be made.

Acknowledgements

The authors wish to thank Jari Miina, D.Sc. (For.), for his guidance in the statistical analysis and also for his valuable comments on the manuscript. This study was funded by the Finnish Society of

Forest Science, the University of Joensuu, and the Graduate School in Forest Sciences.

References

- Ahti, T., Hämet-Ahti, L. & Jalas, J. 1968. Vegetation zones and their sections in northwestern Europe. *Annales Botanici Fennici* 5: 169–211.
- Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research* 2: 49–53.
- Belonogova, T.V. 1988. Yield forecasting and optimization of berry harvesting in the forests of Southern Karelia, USSR. *Acta Botanica Fennica* 136: 19–21.
- 1993. Changes in bilberry and cowberry yields under the influence of forestry measures. *Aquilo Series Botanica* 31: 17–20.
- & Kuchko, A.A. 1979. Plodonosenie cherniki v bereznejakah karelii. *Rastitelnye Resursy* 15: 415–421.
- Eriksson, L., Ingelög, T. & Kardell, L. 1979. Blåbär, lingon, hallon. Förekomst och bärproduktion i Sverige 1974–1977. Summary: Bilberry, lingonberry, raspberry. Occurrence and production in Sweden 1974–1977. Sveriges lantbruksuniversitet, Avdelning för landskapsvård, Rapport 16. 124 p.
- Henttonen, H. 1990. Kuusen rinnankorkeusläpimitan kasvun vaihtelu Etelä-Suomessa. Summary: Variation in the diameter growth of Norway spruce in Southern Finland. *Helsingin yliopiston metsänarvioimistieteen laitoksen tiedonantoja* 25: 1–88.
- Hökkä, H. 1997. Models for predicting growth and yield in drained peatland stands in Finland. *Metsäntutkimuslaitoksen tiedonantoja* 651 [Finnish Forest Research Institute, Research Papers]. 45 + 53 p.
- Ihalainen, M. & Pukkala, T. 2001. Modelling cowberry (*Vaccinium vitis-idaea*) and bilberry (*Vaccinium myrtillus*) yields from mineral soils and peatlands on the basis of visual field estimates. *Silva Fennica* 35(3): 329–340.
- , Alho, J., Kolehmainen, O. & Pukkala, T. 2002. Expert models for bilberry and cowberry yields in Finnish forests. *Forest Ecology and Management* 157: 15–22.
- Jaakkola, I. 1983. Rovaniemen maalaiskunnan marjasatoinventointi. *Metsäntutkimuslaitoksen tiedonantoja* 105: 137–143.
- Jäppinen, J.-P., Hotanen, J.-P. & Salo, K. 1986. Marja- ja sienisadot ja niiden suhde metsikkötunnuksiin mustikka- ja puolukkatyyppin kankailla Ilomantsissa vuosina 1982–1984. Summary: Yields of wild berries and larger fungi and their relationship to stand characteristics on MT and VT-type mineral soil sites in Ilomantsi, eastern Finland, 1982–1984. *Folia Forestalia* 670. 25 p.
- Kalliola, R. 1973. Suomen kasvimaantiede. Werner Söderström Oy, Porvoo. 308 p.
- Kangas, J. 1998. Non-wood forest goods and benefits and the choice of forest management practices. In: Lund, H.G., Pajari, B. & Korhonen, M. (eds.). Sustainable development of non-wood goods and benefits from boreal and cold temperate forests. *EFI Proceedings* 23: 203–210.
- & Mononen, A. 1997. Ekologiseen asiantuntemukseen perustuvan numeerisen mallin tuottaminen metsäalueen biodiversiteetin arviointiin. *Metsätieteen aikakauskirja* 2/1997: 225–238.
- & Niemeläinen, P. 1996. Opinion of forest owners and the public on forests and their use in Finland. *Scandinavian Journal of Forest Research* 11: 269–280.
- Kuchko, A.A. 1988. Bilberry and cowberry yields and the factors controlling them in the forests of Karelia, USSR. *Acta Botanica Fennica* 136: 23–25.
- Laakso, S., Lehtinen, P. & Raatikainen, M. 1990. Kuusen ja männyn vaikutus mustikan esiintymiseen, kukintaan ja marjasatoon. Summary: The influence of spruce and pine on the cover values, flowering and berry yields of the bilberry (*Vaccinium myrtillus* L.). *Memoranda Societatis pro Fauna et Flora Fennica* 66: 47–53.
- Laine, K. 1988. Long-term variations in plant quality and quantity in relation to cyclic microtine rodents at Kilpisjärvi, Finnish Lapland. *Acta Universitatis Ouluensis, Ser. A* 198. 33 + 54 p.
- Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. *Communicationes Instituti Forestalis Fenniae* 134. 69 p.
- Lehto, J. & Leikola, M. 1987. Käytännön metsätyypit. Kirjayhtymä, Helsinki. 96 p.
- Miina, J. 2000. Dependence of tree-ring, earlywood and latewood indices of Scots pine and Norway spruce on climatic factors in eastern Finland. *Ecological Modelling* 132: 259–273.
- Muhonen, T. 1995. Mustikka- ja puolukkasatojen ennustaminen kasvupaikka- ja puustotunnusten

- avulla – asiantuntemukseen perustuva lähestymistapa. Master's thesis of forest planning and economics. University of Joensuu, Faculty of Forestry. 37 p.
- Nousiainen, H. 1983. Eräiden *Vaccinium*-lajien pölytysbiologiasta, kukinnasta ja marjonnasta. Metsäntutkimuslaitoksen tiedonantoja 90: 66–86.
- Penner, M., Penttilä, T. & Hökkä, H. 1995. A method for using random parameters in analyzing permanent sample plots. *Silva Fennica* 29(4): 287–296.
- Pukkala, T. 1988. Monikäytön suunnitteluohjelmisto Monsu: ohjelmiston toiminta ja käyttö. University of Joensuu, Joensuu. 40 p.
- Raatikainen, M. 1978. Puolukan sato, poiminta ja markkinointi Pihtiputaan kunnassa. Summary: The berry yield, picking, and marketing of *Vaccinium vitis-idaea* L. in the commune of Pihtipudas. *Silva Fennica* 12(2): 126–139.
- & Vänninen, I. 1988. The effects of the 1984–1985 cold winter on the bilberry and lingonberry yield in Finland. *Acta Botanica Fennica* 136: 43–47.
- , Rossi, E., Huovinen, J., Koskela, M-L., Niemelä, M. & Raatikainen, T. 1984. Metsä- ja suomarjasadot Väli-Suomessa. Summary: The yields of the edible wild berries in central Finland. *Silva Fennica* 18(3): 199–219.
- Saastamoinen, O., Kangas, K. & Aho, H. 2000. The picking of wild berries in Finland in 1997 and 1998. *Scandinavian Journal of Forest Research* 15: 645–650.
- Salo, K. 1982. Metsänhoitotoimenpiteet ja marja- ja sienisadot. Pohjois-Karjalan Luonto, Pohjois-Karjalan Luonnonsuojelupiiri ry:n vuosijulkaisu. 12. vuosikirja. Pohjois-Karjalan Kirjapaino Oy, Joensuu. p. 14–18.
- 1983. Marja- ja sienisatojen seuranta VMI-tutkimuksen osana Pohjois-Karjalassa. Metsäntutkimuslaitoksen tiedonantoja 90: 122–134.
- 1993. The composition and structure of macrofungus communities in boreal upland type forests and peatlands in North Karelia, Finland. *Karstenia* 33: 61–99.
- 1995. Non-timber forest products and their utilization. In: Hytönen, M. (ed.). Multiple-use forestry in the Nordic countries. The Finnish Forest Research Institute. p. 117–155. ISBN 951-40-1421-9.
- 1999. Principles and design of a prognosis system for an annual forecast of non-wood forest products. In: Niskanen, A. & Demidova, N. (eds.). Research Approaches to Support Non-Wood Forest Products Sector Development: Case of Arkhangelsk Region, Russia. *EFI Proceedings* 29. p. 35–44.
- SAS Institute Inc. 1992. SAS technical report P-229, SAS/STAT Software: Changes and enhancements, release 6.07. SAS Institute Inc., Cary, NC. 620 p.
- Sepponen, P. & Viitala, L. 1982. Metsämarjasadon arvioinnin menetelmäongelmia. Summary: Methods of assessing the forest berry crop. *Luonnon Tutkija* 86: 175–180.
- Sevola, Y. 1983. Metsähallinnon Nurmeksen hoitoalueen voimaperäinen puunkasvatus: seuranta-järjestelmä ja tuloksia. Summary: Intensive timber growing in a state forest district: monitoring system and results. *Folia Forestalia* 574. 83 p.
- Snowdon, P. 1991. A ratio estimator for bias correction in logarithmic regression. *Canadian Journal of Forest Research* 21: 720–724.
- Solantie, R. 1983. Suomen luonnonmarjat, ilmasto ja sää. Metsäntutkimuslaitoksen tiedonantoja 90: 38–49.
- Valtakunnan metsien inventoinnin kenttätyön ohjeet, yleinen osa. 1977. Metsäntutkimuslaitos, metsänarvioimisen tutkimusosasto. Helsinki. 59 p.
- Wallenius, T.H. 1999. Yield variations of some common wild berries in Finland in 1956–1996. *Annales Botanici Fennici* 36: 299–314.
- Zvorykina, K.V. 1970. Plodonosenie cherniki v lesu i na vyrubkah. *Rastitelnye Resursy* 6: 550–555.

Total of 43 references