SILVA FENNICA

Silva Fennica 46(2) research articles
www.metla.fi/silvafennica · ISSN 0037-5330
The Finnish Society of Forest Science · The Finnish Forest Research Institute

Using Cost-Plus-Loss Analysis to Define Optimal Forest Inventory Interval and Forest Inventory Accuracy

Antti Mäkinen, Annika Kangas and Mikko Nurmi

Mäkinen, A., Kangas, A. & Nurmi, M. 2012. Using cost-plus-loss analysis to define optimal forest inventory interval and forest inventory accuracy. Silva Fennica 46(2): 211–226.

In recent years, optimal inventory accuracy has been analyzed with a cost-plus-loss methodology, where the total costs of inventory include both the measurement costs and the losses from the decisions based on the collected information. Losses occur, when the inaccuracies in the data lead to sub-optimal decisions. In almost all cases, it has been assumed that the accuracy of the once collected data remains the same throughout the planning period, and the period has been from 10 up to 100 years. In reality, the quality of the data deteriorates in time, due to errors in the predicted growth. In this study, we carried out a cost-plus-loss analysis accounting for the errors in (stand-level) growth predictions of basal area and dominant height. In addition, we included the inventory errors of these two variables with several different levels of accuracy, and costs of inventory with several different assumptions of cost structure. Using the methodology presented in this study, we could calculate the optimal inventory interval (life-span of data) minimizing the total costs of inventory and losses through the 30-year planning period. When the inventory costs only to a small extent depended on the accuracy, the optimal inventory period was 5 years and optimal accuracy RMSE 0%. When the costs more and more heavily depended on the accuracy, the optimal interval turned out to be either 10 or 15 years, and the optimal accuracy reduced from RMSE 0% to RMSE 20%. By increasing the accuracy of the growth models, it was possible to reduce the inventory accuracy or lengthen the interval, i.e. obtain clear savings in inventory costs.

Keywords value of information, prediction error, inventory error
Addresses *Mäkinen*, Simosol Oy, Rautatietori 4, FI-11130 Riihimäki, Finland; *Kangas* and *Nurmi*, University of Helsinki, Department of Forest Sciences, Helsinki, Finland
E-mail antti.makinen@simosol.fi
Received 8 December 2010 **Revised** 20 January 2012 **Accepted** 20 January 2012
Available at http://www.metla.fi/silvafennica/full/sf46/sf462211.pdf

1 Introduction

Typical decisions for a forest owner include such decisions as to when a given stand is treated (if at all), and how the stand is treated. All the decisions include uncertainty, concerning which state of nature will occur (Hirsleifer and Riley 1979). For example, the realized amount of timber in the future may vary due to 1) the uncertainty in initial estimates of standing stock obtained in a forest inventory, 2) errors in the predicted development (growth) of the stand or 3) events such as fire or insect attacks. The decision maker has then two options: either to make the optimal choice with the current information or to reduce the uncertainty by collecting more information.

Value of information (VOI) in decision making (ex ante) can be defined as the difference between the expected value of a given decision with and without the information (e.g. Hirshleifer and Riley 1979, Lawrence 1999, Birchler and Bütler 2007, Kangas 2010). Thus, the value of information stems from the possibility of making better decisions if new, more accurate information is available. In forestry, practical analyses about the value of information are rare (see, however Knoke 2002, Amacher et al. 2005, Duvemo 2009, Kangas et al. 2010). The decision-related aspects of data acquisition have been studied using cost-plus-loss analysis, however. In cost-plus-loss analysis, the expected losses due to sub-optimal decisions caused by inaccurate data are added to the total costs of the forest inventory (e.g. Hamilton 1978, Burkhart et al. 1978, Ståhl et al. 1994).

In the studies carried out so far, it has been assumed that decision maker is maximizing the net present value (NPV) of the forest area (e.g. Eid 2000, Holmström et al. 2003, Eid et al. 2004, Duvemo et al. 2007, Borders et al. 2008, Islam et al. 2009, Mäkinen et al. 2010). The decisions have been about scheduling of thinnings and clearcuts (Holmström et al. 2003, Eid et al. 2004). The losses are then due to carrying out the treatment out too late or too early, with respect to the predicted development. Most of the studies have concentrated on a set of basic forest inventory variables, obtainable from a specified inventory method. An exception is the work of Eid (2000) which considered pre-defined accuracy levels rather than specified methods. In all cases it has been assumed that the data quality is constant throughout the planning period (from 10 up to 100 years), although this assumption is known not to be correct (Duvemo and Lämås 2006). In practice, the inventory interval has typically been 10–15 years (Koivuniemi and Korhonen 2006).

Predictions concerning the future development are made with growth models. The errors of initial data and growth predictions propagate in time, meaning that the longer the prediction period, the lower the quality of the predictions at the end of the period (Gertner and Dzialowy 1984, Gertner 1987, Mowrer 1991, Kangas 1997). It should be the more profitable to invest on accurate data, the longer the period in which this data can be used, i.e. the more decisions can be made with the collected data (e.g. Karnon 2002). As the errors propagate through the system, the expected losses are assumed to increase and at some time the expected losses increase to a level where collecting new data is more profitable than using the old data. This defines the optimal inventory interval, i.e. the life span of the initial data. There may also be interactions between the prediction errors and the initial accuracy, meaning that an initial data set with given accuracy is more valuable when the growth models are more accurate (c.f. Ståhl et al. 1994). Holopainen et al. (2010) concluded that the errors in growth models were the most important source of error in the expected NPVs, but the combined effect of growth prediction errors and initial errors was less than the sum of the individual effects. It indicates that the different (random) errors to some extent may compensate each other.

The first attempt to analyze the effect of growth errors on the life-span of inventory data was carried out by Ståhl et al. (1994), but it only included a very crude growth model and presented a couple of fairly theoretical examples. A following attempt (Pietilä et al. 2010) included a real stand-level growth and yield simulator, SIMO, and error models that were based on observed errors in predictions. The results show that the expected losses from updating the data with growth models can be as high or in long term even higher than the expected losses from the errors in the initial data (at worst 900 €/ha/60 year period for the studied area). However, their study did not include the errors in the initial forest inventory, the inventory costs nor the interaction between the errors in the initial data and growth predictions.

The aims of this study are: 1) to present a method for estimating the optimal inventory interval with given assumptions on growth prediction errors and inventory errors, 2) to test the method by evaluating the expected losses due to growth projection and forest inventory errors in a simple harvest scheduling (or forest planning) problem, 3) to estimate the optimal inventory interval (i.e. life-span of data) for various assumed inventory cost structures and 4) to examine how adjusting the growth model to each stand would affect the life-span. In the last two tasks, several different assumptions of the relation of inventory costs and accuracy are used. Assumptions are needed as the errors are simulated and not related to any real inventory method.

2 Materials and Methods

2.1 Materials

As the input data for the forest planning computations we used a sample of 99 forest stands from a stand-wise field inventory database collected from central Finland. This type of data has been, and still is, the most common input data in Finnish forest planning. In this study, this data was assumed to be error-free. The data was used to get a distribution of stand properties that represents a typical forest estate.

The average size of a single forest stand was ~2 ha. Each stand record contained information about stand location, site class, soil characteristics and a number of aggregate attributes describing standing stock properties such as basal area (G), mean diameter (basal area median diameter D_{gM}), mean stand age (T), mean height (height of the mean tree H_{gM}) and dominant height (H_{dom}). As a large proportion of the stands contained multiple tree species, the aggregate attributes were stratified by tree species so that each stratum in a stand had its own standing stock properties. For a more detailed description about the data, see Pietilä et al. (2010).

2.2 Methods

2.2.1 Forest Planning Computations

The forest planning computations, i.e. growth simulation and harvest scheduling, in this study were done using SIMO (SIMulation and Optimization) software (Rasinmäki et al. 2009). It consists of modifiable simulation and optimization modules with built-in capability for Monte Carlo analysis.

The planning period for which the simulations and optimization were done was 30 years, consisting of six five-year time steps. A 30-year period was selected for this study based on the earlier work by Pietilä et al. (2010). Based on that study, 30-year period most likely includes the optimal inventory interval. A forest planning, or harvest scheduling problem, was solved by first generating multiple alternative harvest schedules for each stand with the simulator module and then using the optimizer module for finding a schedule that maximizes the net present value of the whole forest without any forest estate-level constraints. The interest rate used for calculating the present values was 3%.

The growth projections in non-seedling stands were based on stand-level growth models of G and H_{dom} by Vuokila and Väliaho (1980) for Scots pine (*Pinus sylvestris*) and Norway spruce (*Picea abies*); growth model of G by Mielikäinen (1985) for silver birch (*Betula pendula*) and white birch (*Betula pubescens*); and growth models of H_{dom} by Oikarinen (1983) for silver birch and by Saramäki (1977) for white birch. The response variable in all of the growth models was the increment percentage I_x of attribute x during a five-year simulation time step. The explanatory variables in Scots pine and Norway spruce G growth models were G, H_{dom}, T, and site class, and in H_{dom} growth models the variables were T, H_{dom} and site class. For birches and other deciduous species, the explanatory variables were G, T and site class for G growth models and H_{dom}, T and site class for H_{dom} growth models. Other stand-level forest attributes were updated on each time step using various models with G and H_{dom} as explanatory variables (for details, see Pietilä et al. 2010).

The growth of Scots pine seedling stands was predicted using models of Huuskonen and Miina

(2006). In Norway spruce and birch seedling stands the growth was predicted using simple models, which predict the age T for the stand to reach 1.3 meter mean height, using tree species and site class as explanatory variables (Hynynen et al. 2002). This approach was used as no other model for the development of seedling stand growth has been published.

For simulating harvests (e.g. thinnings and regeneration harvests), tree lists were generated for each stand on each time step using theoretical diameter distribution models by Kilkki et al. (1989), Siipilehto (1999) and Kangas and Maltamo (2003). The distribution models predicted diameter distributions with one-centimeter classes using stand-level forest attributes as explanatory variables. The bucking of each stem in the tree list was optimized for maximum value by constructing stem profiles with a taper curve model (Laasasenaho 1982), applying a bucking algorithm and using different prices for sawlogs and pulpwood.

The optimizer module was used for selecting the optimal harvest schedule for each stand from the set of simulated alternatives. We used a simple heuristic HERO algorithm (Kangas and Pukkala 1998) for the optimizations and the objective function at single stand-level was to maximise stand's *NPV* defined as

$$\mathrm{NPV} = \sum_{t=1}^{30}\left[\frac{C_t}{(1+r)^t}\right] + \frac{P_{30} + P_{L,30}}{(1+r)^{30}}$$

where C_t is cash flow, i.e. net income or cost, at year t, r is the rate of interest, and P_{30} and $P_{L,30}$ are net present value of standing stock and land for the stand at the end of the 30-year planning period, respectively. The P_{30} and $P_{L,30}$ values were predicted using models by Pukkala (2005) that use standing stock characteristics, site fertility, stand location and timber assortment prices as explanatory variables. As there were no estate-level constraints, the optimization algorithm could easily find the optimal solution just by iterating through each schedule in the limited set of simulated schedules per stand and selecting the one yielding the highest objective function value.

The stochasticity in the forest planning computations was accounted for by adding random variation into the growth predictions and by simulating a forest inventory procedure with random forest inventory errors.

2.2.2 Simulation of Growth Projection Errors

Growth prediction errors in non-seedling stands were generated by introducing a random error component $\varepsilon_{t,x}$ to the predicted increments of attributes G and H_{dom}. The growth prediction error was composed of between-stand variation u_x and within-stand (i.e. periodical) variation $e_{t,x}$ so that the total error of attribute x at time t was

$$\varepsilon_{t,x} = u_x + e_{t,x}$$

The stand effect u can be interpreted as the average growth level of the stand over time, when compared to the expected curve, and the periodical variation as the variation around this average growth level between subsequent periods. Normally distributed random error component u_x was generated once at t_0 for each stand and normally distributed component $e_{t,x}$ was generated at the beginning of each simulation time step. The correlation between the random error components of variables G and H_{dom} and between the strata in each stand was accounted for when the values of u_x were generated (see Pietilä et al. 2010). The variance of the error term $\varepsilon_{t,x}$ was taken from the observed growth prediction errors in Finnish conditions (Haara and Leskinen 2009) and divided into components u_x and $e_{t,x}$ by applying the results of Kangas (1999) and using proportions 0.365 and 0.635 for u_x and $e_{t,x}$, respectively. This means that the correlation of total errors between the periods is assumed to be constant, i.e. var(u)/(var(e) + var(u)) = 0.635 (for details see Pietilä et al. 2010). Thus, the value of attribute ($\hat{x}$), affected by prediction error, at time $t+5$ was given by

$$\hat{x}_{t+5} = \hat{x}_t\left[\left(1 + \frac{I_{t,x}}{100}\right) + \varepsilon_{t,x}\right]$$

where $I_{t,x}$ is the predicted increment percentage of attribute x and $\varepsilon_{t,x}$ was the prediction error at time t.

Projecting forest growth using the stand-level growth models and random error components described above yielded in quite similar variances in the growth predictions as observed by Haara and Leskinen (2009, see Pietilä et al. 2010 for details). Haara and Leskinen (2009), however, only studied two periods, and the development of the error variance in longer time periods is very poorly known.

In Norway spruce and birch seedling stands, the variation in the growth predictions was accounted for by adding a normally distributed random error to the initial stand age (T) of reaching the 1.3 meter mean height. The error ε_T was also divided into between-stands and within stand error components. Thus, the initial erroneous age $\hat{T}$ for a stand was given by

$$\hat{T} = T(1 + \varepsilon_T)$$

where T is the true age of the stand. In Scots pine seedling stands, the errors were simulated in the same way as in advanced stands (see Pietilä et al. 2010).

In addition to the growth simulations affected by the growth prediction errors, the forest growth was simulated also without the error components. This was done in order to simulate forest inventory, in which the attribute values affected by the growth prediction errors ($\hat{x}$) were replaced with the true attribute values (x), unaffected by the growth prediction or forest inventory errors.

We estimated the effect of adjusting the models to each stand, by examining what would happen if the local stand-level bias could be removed. In effect, it means that the u-term in the error model was assumed to be zero, and only the within-stand (between-period) effects $\varepsilon_{t,x}$ were included in the analysis. It means that we assumed that the growth predictions could be improved with additional measurements (i.e. measurement of the stand effect, or the mean growth level in the stand). Similar improvements could, of course, be obtainable by estimating new improved growth models.

2.2.3 Simulation of Forest Inventory Errors

The random growth prediction errors were responsible for a certain proportion of the total variance in the forest attribute values at any given time t. The rest of this variance was due to forest inventory errors.

Forest inventory errors were taken into account by simulating a forest inventory every nth year, where n is the forest inventory interval. In forest inventory, the values of $\hat{G}$ and $\hat{H}_{dom}$ were substituted with the true values of G and H_{dom}. After that, random forest inventory error components δ_G and δ_{Hdom} were added to the values of G and H_{dom}. Forest inventory errors δ_G and δ_{Hdom} were generated from normal distribution $N(0, \sigma^2)$ with no bias and standard deviation σ getting values 0%, 5%, 10%, 15%, 20% and 25%. The possible correlations between δ_G and δ_{Hdom} were not regarded for the sake of simplicity. As normal distribution is not limited, generating random errors to the forest variables can in principle lead to both negative and extremely large variable values. However, as the errors were relative, and the highest standard deviation we used for the forest inventory errors was 25 (%), the probability of getting random relative errors outside limits [–100%..100%] was only ~0.00006. Because of this there was no need to limit the random errors. The same applies for the random growth projection errors.

The inventory interval n had values 5, 10, 15, 20 and 30 years for every inventory error level σ, so that the total number of (n, σ) combinations was $6 \times 5 = 30$. The simulation of forest inventory was a four-step procedure:

1) Substitute the current values of attributes $\hat{G}$ and $\hat{H}_{dom}$, affected by cumulated growth prediction errors and the previous forest inventory errors, with the true values G and H_{dom} unaffected by the errors (but affected by all the decisions, suboptimal or optimal, carried out within the period).
2) Generate normally distributed random forest inventory errors δ_G and δ_{Hdom}.
3) Add the forest inventory errors to the values of G and H_{dom} so that

$$\hat{G} = G\left(1 + \frac{\delta_G}{100}\right) \text{ and } \hat{H}_{dom} = H_{dom}\left(1 + \frac{\delta_{H_{dom}}}{100}\right)$$

4) Update other forest stand-level attributes using the new $\hat{G}$ and $\hat{H}_{dom}$ values.

The top-level simulation logic is depicted in flow chart format in Fig. 1. The flow chart contains

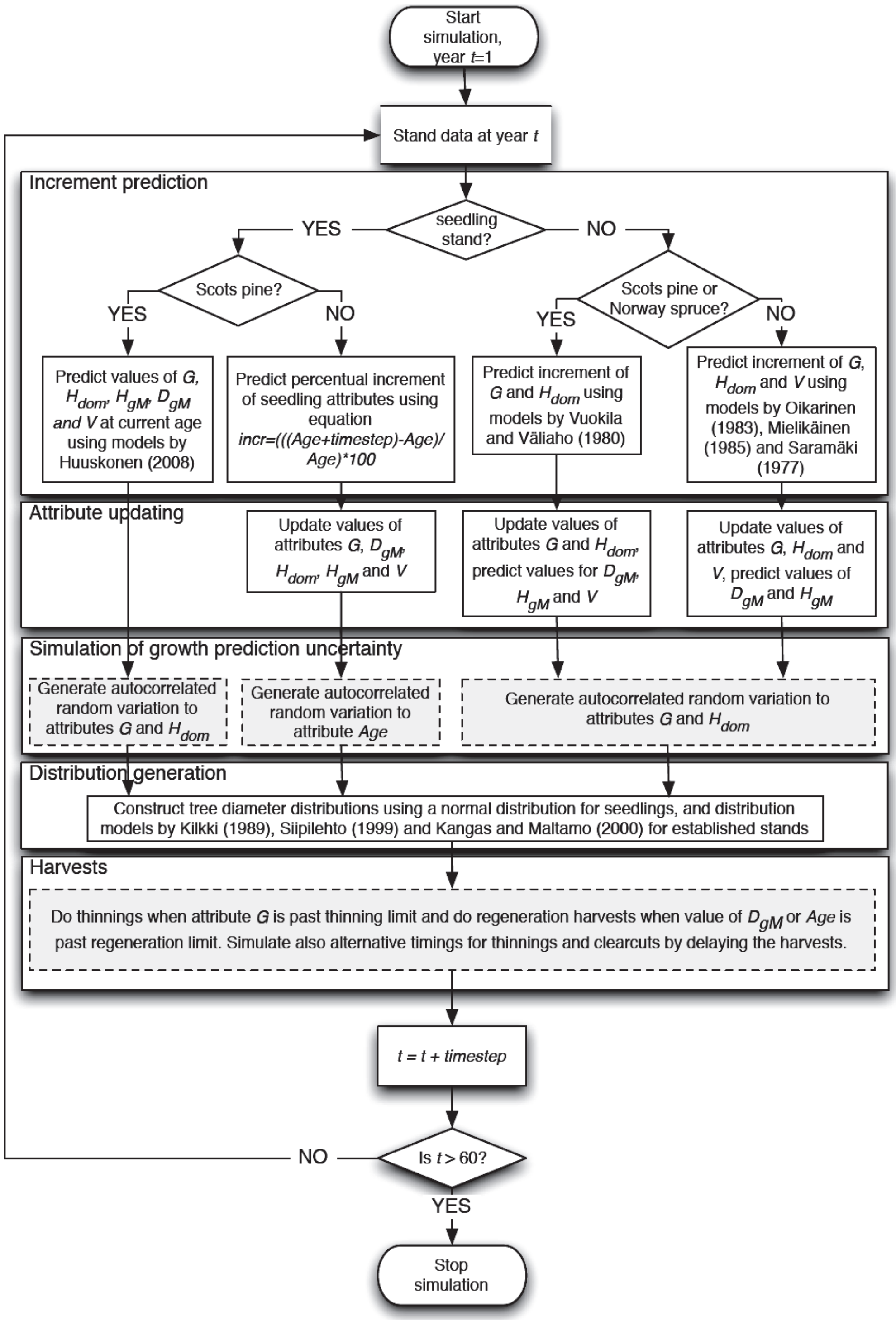

Fig. 1. Simulation logic flowchart (adapted from Pietilä et al. 2010).

both the deterministic components, such as the growth prediction and harvest simulations, and the stochastic components for simulating forest inventory errors and growth prediction errors.

2.2.4 Simulating Inoptimality Losses with Monte Carlo Method

The inoptimality losses due to the forest inventory and growth prediction errors were studied using Monte Carlo simulation approach with 100 realisations for each stand and (n,σ) pair combination. The number of realisations was considered to be sufficient as the inoptimality loss estimates seemed to be very stable using 100 realisations. The stability was verified by repeating the analysis with the same inventory error and interval levels for 20 times. The relative standard deviation of the average losses between the repeats was under 1%. More realisations would have stabilised the estimates even more, but also the computational time would have grown notably.

Optimal harvest schedule s for any stand with given input data vector $\mathbf{X}$ can be computed applying a simulation and optimisation model f_s giving $s = f_s(\mathbf{X})$. In case of stochastic forest inventory and growth prediction errors with given inventory interval n and inventory standard error level σ, the model gives $\hat{s} = f_s(\hat{\mathbf{X}}, n, \sigma)$. Due to the stochasticity involved in the growth predictions and inventory errors, each realisation of $\hat{s}$ generated with the model may be different.

A net present value v for any stand with given s can be computed with model $g(\mathbf{X},s)$. Model g uses a similar simulation module as model f_S, but with predefined set of harvests and no optimisation, as alternative harvest schedules are not generated.

In order to determine the inoptimality losses due to the growth prediction and inventory errors, we needed to determine a true NPV value v_k (unaffected by errors) for each stand k (k = 99). First we defined the optimal schedule s_k by

$$s_k = f_s(\mathbf{X}_k)$$

after which we could get the true NPV value v_k by

$$v_k = g(\mathbf{X}_k, s_k)$$

where $\mathbf{X}_k$ is an input attribute vector for stand k.

Next, the inoptimality losses L_{ijk} were determined for every (n,σ) combination i ($i = 1,\ldots,30$), realisation j ($j = 1,\ldots,100$), and stand k ($k = 1,\ldots,99$) by

$$\hat{s}_{ijk} = f_s(\hat{\mathbf{X}}_{jk}, n_i, \sigma_i)$$

$$\hat{v}_{ijk} = g(\mathbf{X}_{jk}, \hat{s}_{ijk})$$

$$L_{ijk} = \hat{v}_{ijk} - v_k$$

where $\mathbf{X}_k$ is input attribute vector for stand k and $\hat{\mathbf{X}}_{jk}$ is the jth input attribute vector with simulated inventory errors for stand k.

2.2.5 Cost Assumptions in Cost-Plus-Loss Calculations

We did not want to restrict the "inventory method" to any specific method used, but wanted to include all possible ways to produce stand-level information for planning, e.g. traditional, partly visual stand inventory, sampling and modern remote sensing based (ALS, aerial photos or satellite images) methods. Thus, we simulated the error levels rather than used the published errors from any method. As the inventory errors were simulated, and did not "mimic" any known inventory method it was not possible to include any real costs for the analysis, and it would have been difficult to make a detailed model for the costs. The calculations were thus carried out with several different assumptions on the cost structure. The costs were assumed to depend on three parameters: the number of inventories carried out in 30 year period (m), the level of costs in the least accurate inventory (cl), and the dependency of additional costs on the accuracy of the inventory (p).

The number of inventories carried out m was calculated from the inventory interval: if the interval n is 5 years, then number of inventories is 30/5 = 6. Thus, the used values of m were 6, 3 (n = 10), 2 (n = 15), 1.5 (n = 20) and 1 (n = 30). Each inventory was assumed to cost the same amount of money, but inventory costs occurring in the future were discounted in the same way as the incomes, using an interest rate of 3%.

The level of the costs in the inventory with

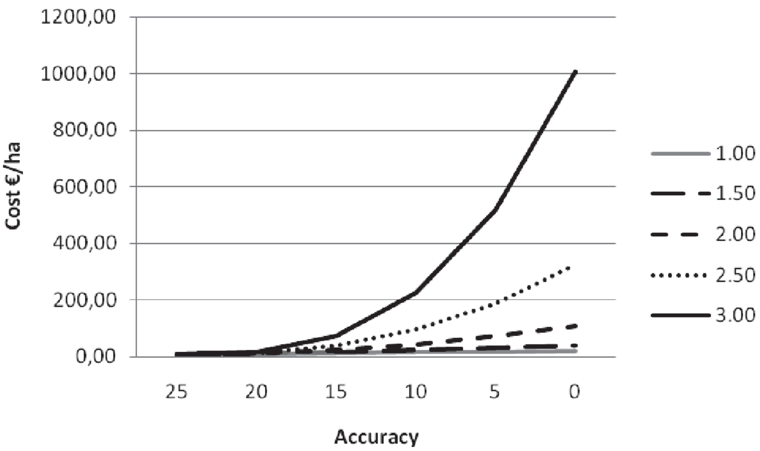

Fig. 2. The costs of one inventory, as a function of accuracy level, with the different power parameters p and level parameter 8 €/ha.

lowest accuracy, *cl*, was defined based on published costs. Uuttera et al. (2002) estimated that the traditional Finnish field inventory with an Root Mean Square Error (RMSE) 20% (depending on the variable in question) had a cost 7.9 €/ha, and Uuttera et al. (2006) gave a cost of 9.2 €/ha. Eid et al. (2004) gave cost 5.53 €/ha for a visual interpretation of aerial photos in Norway with accuracy ~25% and 11.39 €/ha for a laser scanning inventory with accuracy ~10%. Based on these, we used as a level of cost for the lowest accuracy inventory method 4–16 €/ha.

Holmström et al. (2003) gave a cost of 50 SEK/5 ha (about 1 €/ha) for an inventory based on stand registers (with accuracy ~25%), and 2080 SEK/5 ha (about 46 €/ha) for the most accurate inventory with 10 sample plots in a stand (~5%) in Sweden. Thus, the costs were assumed to increase 40-fold when the accuracy increased. Based on these, we used different power parameters (powers p 1, 1.5, 2, 2.5, and 3) giving the increase from lowest to highest accuracy from 2.25-fold ($p=1$) to 126-fold ($p=3$).

In this study, we used a formula

$$c = m(cl + 2^p)$$

to calculate the costs of the inventory per hectare. The resulting inventory costs for level $cl=8$ and different power parameters p are shown in Fig. 2. The highest accuracy here was an inventory with zero RMSE, which would be possible only if all trees were measured, without any error, in the field.

3 Results

The total average losses per hectare during the 30-year period are shown in Tables 1 and 2. Quite obviously, the losses were smallest with shortest inventory interval (meaning smallest possible growth prediction errors) and perfectly accurate inventory with relative RMSE 0%. The difference between the 0% RMSE and 5% RMSE was, however, quite small. What is most notable from these figures is that when the losses with the 5-year interval increase from 188 to 591 (3.14-fold) as the inventory accuracy decreases, they increase only 1.65-fold when the inventory period is 30 years (Fig. 3). In other words, with the least accurate inventories the losses do not much increase as inventory interval increases, but with the most accurate inventories the trend is very clear.

When the cost models were included, the level was assumed medium ($cl=8$) and the dependency of cost on accuracy was linear ($p=1$), the estimated optimal inventory (cost-plus-loss 265.1 €/ha) was

Table 1. The average losses, in €/ha with different inventory periods (n) and accuracy levels (σ).

σ	n				
	5	10	15	20	30
0	–188.06	–250.27	–293.58	–374.22	–419.97
5	–208.68	–271.99	–300.03	–391.76	–466.04
10	–307.00	–339.10	–358.63	–427.79	–482.92
15	–399.74	–442.84	–461.65	–497.41	–559.66
20	–509.14	–531.66	–588.60	–594.21	–625.22
25	–591.25	–638.53	–653.70	–694.32	–695.25

Table 2. The average relative losses. in percentage, with different inventory periods (n) and accuracy levels (σ).

σ	n				
	5	10	15	20	30
0	–2.66	–3.31	–3.95	–4.77	–4.72
5	–2.80	–3.69	–3.86	–4.99	–5.61
10	–3.98	–4.40	–4.60	–5.44	–5.91
15	–5.18	–5.55	–5.70	–6.38	–6.84
20	–6.50	–6.71	–7.28	–7.45	–7.50
25	–7.17	–8.05	–7.83	–8.48	–7.86

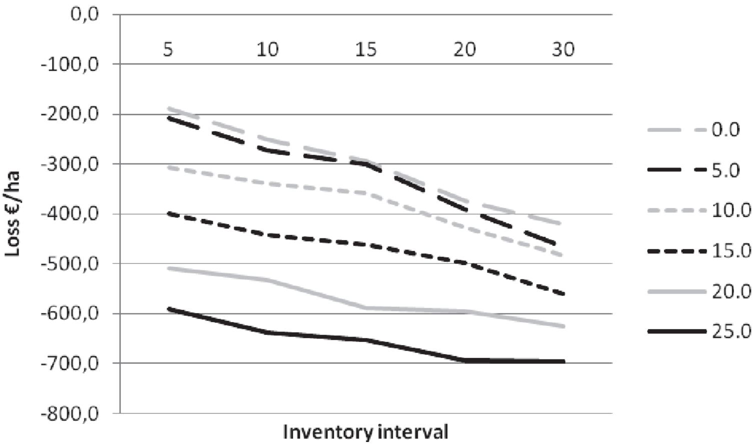

Fig. 3. The losses due to inventory and growth prediction errors at different inventory intervals.

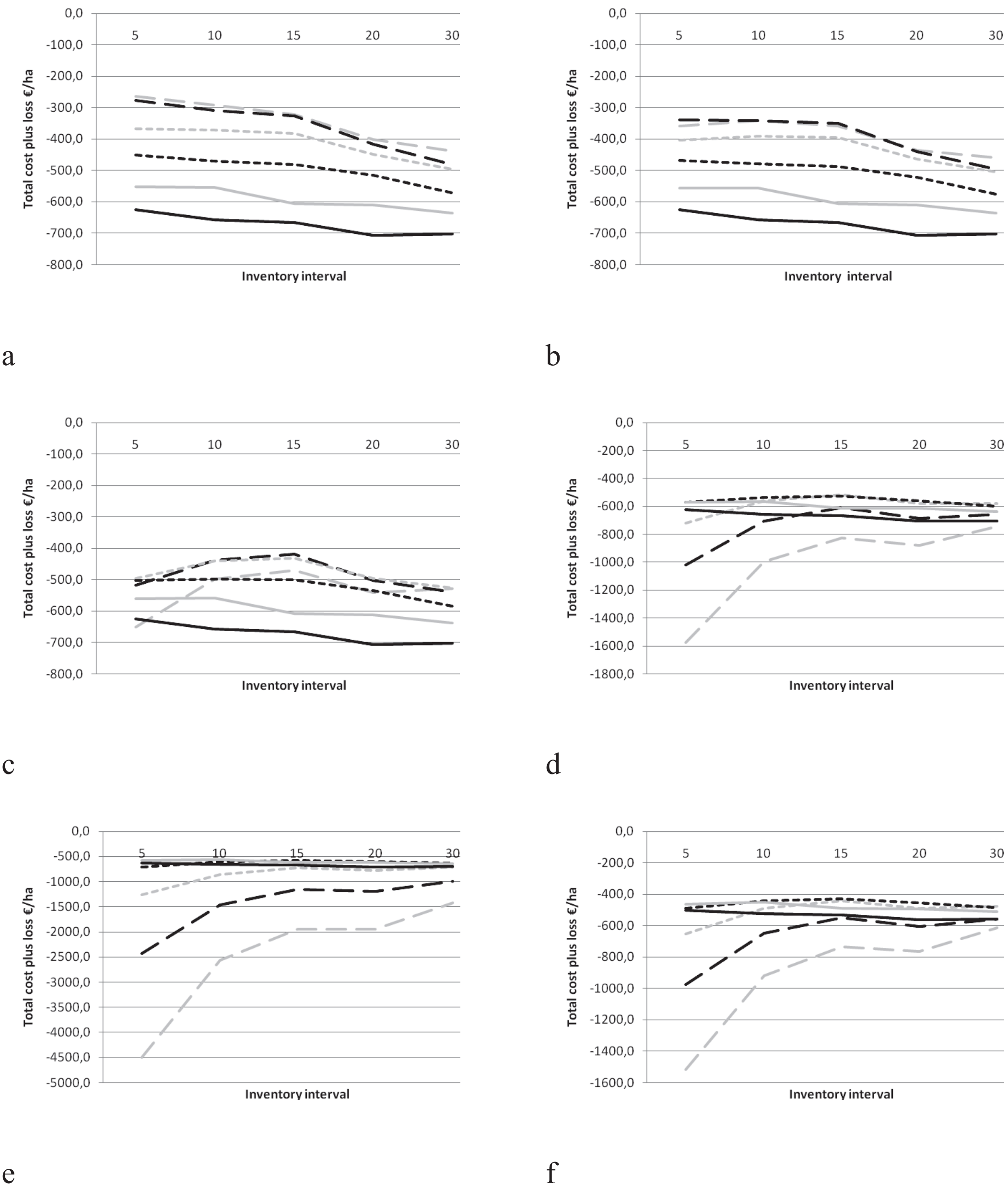

Fig. 4. The total costs and losses with 5 different power parameters (p) in the cost model, with $cl=8$. The legends are as in Fig. 3, solid black line RMSE=25%, solid grey line 20%, dashed black line 15%, dashed grey line 10%, long dashed black line 5% and long dashed grey line 0%.

Table 3. The average absolute total costs plus losses, in €/ha with different inventory periods (n) and accuracy levels (σ), with cost parameters cl = 8, and p = 2.5. The optimal accuracy and inventory period is given in bold.

σ			*n*		
	5	10	15	20	30
0	–1575.7	–995.3	–825.9	–878.0	–744.2
5	–1017.7	–706.3	–610.4	–685.4	–655.1
10	–718.6	–560.1	**–516.5**	–577.2	–579.1
15	–570.9	–534.7	–527.3	–559.6	–599.7
20	–567.6	–563.0	–611.0	–615.4	–638.9
25	–625.5	–656.9	–666.8	–706.7	–703.3

inventory with RMSE 0% and inventory period of 5 years (Fig. 4a). When the dependency of cost and accuracy was "steeper", i.e. power p was 1.5, the estimated optimal inventory interval was still 5 years and the estimated optimal accuracy 5% (b). When the power increased to 2, the estimated optimal accuracy level remained at 5%, but the estimated optimal inventory interval changed to 15 years (c). With yet steeper dependency, p = 2.5, the estimated optimal accuracy was 10% and the estimated optimal inventory interval remained at 15 years (d, see also Table 3). With the steepest dependency, p = 3 (e), the estimated optimal interval dropped to 10 years and the estimated optimal accuracy to 20%. Thus, with the highest costs it was more profitable to reduce the interval than to improve the accuracy (σ). Thus, when increasing the accuracy of inventory is more and more expensive, the optimal accuracy reduces from 0% to 20%, and the optimal inventory interval increases from 5 years to 15 years. With our assumptions, longer intervals are never optimal, however, neither is the worst accuracy level, 25%.

Changing the cost level (*cl*) from 4 to 16 had almost no effect. The only changes were observed with power parameter p = 1.5. Then, when the *cl* was 12 or 16, the RMSE 0% and 10-year interval were estimated to be optimal. At lower cost levels, the estimated optimal accuracy level was 5% and the optimal inventory interval alternated between 5 and 10 years.

When the results were calculated with an adjusted growth model (i.e. *u* was set to 0), the results were calculated only with 10-year inventory interval and inventory RMSE 0%. In this case, the total losses were reduced by 30.8%, i.e. from about 250 to about 175 €/ha. If the similar reduction was assumed to all losses, and the cost level was assumed to be 8, the results were exactly the same as in the original case of growth model accuracy in all but two cases. In the case of power parameter $p = 1.5$, the improved growth model increased the life span from 5 to 10 years, but the optimal RMSE reduced from 5% to 0%. This meant 40.1 €/ha/30 years (30.6%) savings in actual inventory costs. In the case of $p = 2.5$, the optimal inventory interval remained at 15 years but the optimal RMSE level increased from 10% to 15%, giving a saving of 92.4 €/ha/30 y (58.4%) in inventory costs (Fig. 4f, see also Table 4). Thus, in some cases the improved growth model either allowed for a longer inventory interval or less accurate data. Even then, with all the tested cost structures, the estimated optimal RMSE was never higher than 20% and the estimated optimal inventory interval never longer than 15 years.

Naturally, there was certain amount of variation in the losses and also in the estimated optimal inventory intervals. The relative standard deviations of the losses for each (n,σ) combination are found in Table 5. The relative SDs vary from 2.65% to 9.23% so that the relative SDs are smaller when inventory interval is short and data is accurate and larger when interval is longer and data more inaccurate. For the estimated optimal inventory interval and accuracy combinations, the relative SDs varied from 4.51% and 6.15%.

Table 4. The average absolute total costs plus losses, in €/ha with improved growth predictions, different inventory periods (n) and accuracy levels (σ), with cost parameters cl = 8, and p = 2.5. The optimal accuracy and inventory period is given in bold.

σ			*n*		
	5	10	15	20	30
0	–1517.8	–918.2	–735.5	–762.7	–614.8
5	–974.2	–649.7	–548.0	–604.0	–558.1
10	–654.8	–489.6	–441.9	–488.2	–478.7
15	–487.8	–442.6	**–431.3**	–456.1	–483.2
20	–461.7	–452.5	–488.6	–491.8	–508.8
25	–502.5	–524.1	–530.9	–562.3	–558.6

Table 5. The relative standard deviations (SD) of the losses, in percentage with different inventory periods (n) and accuracy levels (σ).

σ			*n*		
	5	10	15	20	30
0	2.65	3.42	4.00	4.85	7.04
5	2.99	3.61	4.26	5.06	7.01
10	2.99	4.51	5.06	5.27	6.67
15	5.10	5.49	6.15	6.15	7.99
20	6.27	6.40	7.19	7.09	8.56
25	7.09	7.48	8.20	8.06	9.23

4 Discussion

In this paper we have continued the analysis of the effect of growth prediction errors (Pietilä et al. 2010) on the expected losses in a more general case. We simplified the assumptions from the previous study so that we did not assume an autocorrelation in the within-stand error component, but they were assumed to be mutually independent. Here we also assumed the stand effect to be constant rather than depend on stand age. It meant that the final autocorrelation within each stand was assumed as fixed, 0.635. This meant smaller total autocorrelation especially in the younger stands (see Pietilä et al. 2010).

In this case, all the simulated scenarios proved to be feasible. In Pietilä et al. (2010) 10 simulated scenarios (out of 54900) had to be rejected as the simulated values were in the end so unrealistic that SIMO was unable to produce a result for them. It means that the assumptions used here are probably more realistic, although no new evidence on the true autocorrelation exists. It seems, however, that the autocorrelation parameter is very important to know, if accurate estimates of losses due to growth prediction errors are required. It can be seen from the estimated losses: the estimated losses in the longest inventory intervals (30 years) used in this study were 308 €/ha, while the similar losses in the earlier study were 768 €/ha (Pietilä et al. 2010). Partly this difference is due to the fact that in the earlier study the losses were calculated for 60 years, and here only for 30 years: more years means more decisions and, in turn, more decisions mean more expected losses. However, the difference is too large to be explained purely by the time scale (earlier 12.8 €/ha/y, now 10.27 €/ha/y), especially as all the incomes and costs are discounted. Thus, the remaining difference is due to the changed assumptions, i.e. the autocorrelation.

In this study, we also added the effects of inventory errors into the analysis. The losses increased with increasing RMSE of inventory, and shorter inventory interval elevated the rate of the increase in losses. With least accurate inventories the inventory interval did not affect much. It means that inventory errors and growth prediction errors ”interact”, i.e. the different errors may somewhat cancel each other out. It also means that increasing the accuracy of inventory does not seem to be as useful when the inventory period is long, than when the inventory period is short.

We also included cost-plus-loss analyses with different assumptions of the cost structure. The results were remarkably stable: 25% RMSE, which is close to the traditional field inventory used in Finland until recently, was not optimal in any of the studied cases, and 20–30 year inventory intervals were neither ever optimal. Thus, it seems that with any reasonable costs and accuracy levels (i.e. with RMSE > 0, and p > 1), the estimated optimal interval is either 10 or 15 years with the current growth models. The results thus seemed very robust to the assumptions made. To define if the optimal interval is 10 or 15 years, the additional cost of more accurate and precise data is crucial.

Even improving the growth models so that the losses could be reduced by 30% did not change the estimated optimal interval. Again, the estimate on optimal inventory interval was either 10 or 15 years. Improving the growth model made it possible in some cases either to lengthen the life span of data or to reduce the accuracy of the initial data. This would mean direct cost savings, in the example studied here as much as 30–58%. However, here we assumed that improving the growth model does not introduce additional costs, while in reality adjusting the growth models for the stand effect would mean measurements of past growth. This would increase the inventory costs. In addition, we assumed that the stand effect could be estimated without error. Thus, this study gives the value of perfect information about stand effect, and real savings from sample information would be smaller. This remains to be studied in the future.

In this study we assumed also that the improvement of growth models would be similar with each inventory interval and each inventory accuracy level. It is possible that adjusting the growth models for local bias would reduce the losses less in shorter inventory interval than in long ones, which could possibly make the long intervals more advantageous in some conditions. If the stand effects were larger in mature stands than in young stands, it might have an effect on the life span. This needs to be studied in further analyses, however. Likewise, the effect of the autocorrelation coefficient needs to be analysed in future studies. For instance, assuming the stand effect constant in time may also overestimate the value of adjusting the model, as the stand effects (average stand growth) may be less than perfectly correlated in time (see e.g. Holm 1980).

In this analysis, the forest inventory error was introduced to basal area and height measurements, and these variables were assumed to have the same accuracy level. If the effects were analysed separately for all variables, it would be possible to find out if one of the variables would be more important than the other, i.e. if the efforts should be concentrated on a certain variable. For instance, in Kangas et al. (2011), it seems evident that a similar error in basal area is much more dangerous than in height. However, in real forest inventory method, the height is typically assessed much more accurately than basal area, as there is much within-stand variation in basal area. Moreover, laser scanning based forest inventory directly measures the height, but basal area is model based.

In order to implement the complex simulation model used in this study, we had to make a number of assumptions. For instance, we assumed the diameter distribution models, height, volume and taper models perfect, as well as the stand productive value models. As these error sources are similar to all inventory accuracy levels and inventory intervals, they do not affect the optimal inventory intervals (unless there are very complex interactions between the models). They could affect to the estimated losses, however, which are most probably underestimates. We believe that these assumptions were justified in order to concentrate on the addressed sources of uncertainty.

Although this study seems to give a definite recommendation of 10–15 years inventory period, there is still a lot of work that needs to be done. The effects of biases in the models and inventory estimates, for instance, may have a more pronounced effect on the losses than random errors. Had we considered a specific inventory method, these complications could have been taken into account in inventory errors (see Mäkinen et al. 2010). With respect to the growth prediction errors, the effect of possible biases needs to be studied in future. As the bias in growth predictions might vary according to the site quality, species, stand age and region, this is not a simple task.

Due to the uncertainties discussed above, the estimates of the losses are also uncertain. The uncertainties can be measured by the standard deviations (Table 5) of the losses in the 100 realization to the extent the sources of uncertainties were included in the analysis. The estimated losses may be quite close to each other near the estimated optimal inventory accuracy/inventory interval combinations (Tables 3 and 4). Thus, the loss function is rather flat around the optimum. When this is combined to the effect of uncertainty in the estimated losses, it can be concluded that small deviations in the inventory interval or inventory accuracy are not detrimental.

5 Conclusions

In this study, we proposed a cost-plus-loss –based method for estimating the optimal inventory interval that takes into account the given forest inventory errors as well the errors in forest growth projections. This method is not restricted to any single forest inventory method, but can be applied to any combination of forest inventory methods for which the accuracy and the true costs are known. The methodology can be utilised in calculating how the errors in different variables affect to the optimal inventory interval, so that the inventory efforts can be concentrated optimally. We showed that there is an important link between the quality of forest growth and yield models and inventory accuracy that should be acknowledged, both when estimating growth models and in planning forest inventory. The results also imply that measuring past growth from the stands and using that for improving the growth predictions may be a very efficient way to improve the quality of forest planning databases.

References

Amacher, G.S., Malik, A.S. & Haight, R.G. 2005. Nonindustrial private landowners, fires, and the wildland-urban interface. Forest Policy and Economics 7: 796–805.

Birchler, U. & Bütler, M. 2007. Information economics. Routledge advanced texts in economics and finance. London. 462 p.

Borders, B.E., Harrison, W.M., Clutter, M.L., Shiver, B.D. & Souter, R.A. 2008. The value of timber inventory information for management planning. Canadian Journal of Forest Research 38: 2287–2294.

Burkhart, H.E., Stuck, R,D., Leuschner, W.A. & Reynolds, M.A. 1978. Allocating inventory resources for multiple-use planning. Canadian Journal of Forest Research 8: 100–110.

Duvemo, K. 2009. The influence of data uncertainty on planning and decision processes in forest management. Ph.D. thesis. Department of Forest Resource Management, Faculty of Forest Sciences, SLU, Umeå.

— & Lämås, T. 2006. The influence of forest data quality on planning processes in forestry. Scandinavian Journal of Forest Research 21: 327–339.

— , Barth, A. & Wallerman, J. 2007. Evaluating sample point imputation techniques as input in forest management planning. Canadian Journal of Forest Research 37: 2069–2079.

Eid, T. 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fennica 34(2): 89–100.

— Gobakken, T. & Naesset, E. 2004. Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scandinavian Journal of Forest Research 19: 512–523.

Gertner, G.Z. 1987. Approximating precision in simulation projections: an efficient alternative to Monte Carlo methods. Forest Science 33(1): 230–239.

— & Dzialowy, P.J. 1984. Effects of measurement errors on an individual tree-based growth projection system. Canadian Journal of Forest Research 14(3): 311–316.

Haara, A. & Leskinen, P. 2009. The assessment of the uncertainty of updated stand-level inventory data. Silva Fennica 43(1): 87–112.

Hamilton, D.A. 1978. Specifying precision in natural resource inventories. In: Integrated inventories of renewable resources: proceedings of the workshop. USDA Forest Service, General Technical Report RM-55: 276–281.

Hirshleifer, J. & Riley, J.G. 1979. The analytics of uncertainty and information – an expository survey. Journal of Economic Literature Vol. XVII. p. 1375–1421.

Holm, S. 1980. Att utnyttja historisk tillväxt i tillväxtprognoser. Hugin rapport 19. p. 88. (In Swedish).

Holmström, H., Kallur, H. & Ståhl, G. 2003. Cost-plus-loss analyses of forest inventory strategies based on kNN-assigned reference sample plot data. Silva Fennica 37(3): 381–298.

Holopainen, M., Mäkinen, A., Rasinmäki, J., Hyytiäinen, K., Bayazidi, S., & Pietilä, I. 2010. Comparison of various sources of uncertainty in stand-level net present value estimates. Forest Policy and Economics 12: 377–386.

Huuskonen, S. & Miina, J. 2006. Stand-level growth models for young Scots pine stands in Finland. Forest Ecology and Management 241: 49–61.

Hynynen, J. et.al. 2002. Models for predicting stand development in MELA system. Metsäntutkimuslaitoksen tiedonantoja 835.

Islam, N., Kurttila, M., Mehtätalo, L. & Haara, A. 2009. analyzing the effects of inventory errors on holding-level forest plans: the case of measurement error in the basal area of the dominated tree species. Silva Fennica 43(1): 71–85.

Kangas, A. 1997. On the prediction bias and variance in long-term growth projections. Forest Ecology and Management 96: 207–216.

— 1999. Methods for assessing uncertainty in growth and yield predictions. Canadian Journal of Forest Research 29: 1357–1364.

— 2010. Value of forest information. European Journal of Forest Research 129:863–874.

— & Maltamo, M. 2003. Calibrating predicted diameter distribution with additional information in growth and yield predictions. Canadian Journal of Forest Research 33: 430–434.

— & Pukkala, T. 1998. HERO heuristic optimization for forest management planning: applications and experiences. In: Saarikko, J. (ed.). Resource Technology ’98 Nordic. International Symposium on Advanced Technology Environmental and Natural Resources, Rovaniemi, Finland.

— , Lyhykäinen, H. & Mäkinen, H. 2010. Value of quality information in timber bidding. Canadian Journal of Forest Research 40:1781–1790.

— , Mehtätalo, L., Mäkinen, A. & Vanhatalo, K. 2011. Sensitivity of harvest decisions to errors in stand characteristics. Silva Fennica 45: 693–709.

Karnon, J. 2002. Planning of efficient allocation of research funds: an adapted application of a non-parametric Bayesian value of information analysis. Health Policy 61:329–347.

Kilkki, P., Maltamo, M., Mykkänen, R. & Päivinen, R. 1989. Use of the Weibull function in estimating the basal-area diameter distribution. Silva Fennica 23: 311–318.

Knoke, T. 2002. Value of complete information on red heartwood formation in beech (Fagus sylvatica). Silva Fennica 36: 841–851.

Koivuniemi, J. & Korhonen, K. 2006. Inventory by compartments. Chapter 16. In: Kangas, A. & Maltamo, M. (eds.). Forest inventory, methods and applications. Managing Forest Ecosystems Vol. 10. Springer, Dordrecht. p. 267–278.

Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. Communicationes Instituti Forestalis Fenniae 108. 74 p.

Lawrence, D.B. 1999. The economic value of information. Springer. 393 p.

Mäkinen, A., Kangas, A. & Mehtätalo, L. 2010. Correlations, distributions, and trends in forest inventory errors and their effects on forest planning. Canadian Journal of Forest Research 40(7): 1386–1396.

Mielikäinen, K. 1985. Koivusekoituksen vaikutus kuusikon rakenteeseen ja kehitykseen. Communicationes Instituti Forestalis Fenniae 133. 79 p. (In Finnish).

Mowrer, H.T. 1991. Estimating components of propagated variance in growth simulation model projections. Canadian Journal of Forest Research 21(3): 379–386.

Oikarinen, M. 1983. Etelä-Suomen viljeltyjen rauduskoivikoiden kasvumallit. Communicationes Instituti Forestalis Fenniae 113. (In Finnish).

Pietilä, I., Kangas, A., Mäkinen, A. & Mehtätalo, L. 2010. Influence of growth prediction errors on the expected losses from forest decisions. Silva Fen-

nica 44(5): 829–843.
Pukkala, T. 2005. Metsikön tuottoarvon ennustemallit kivennäismaan männiköille, kuusikoille ja rauduskoivikoille. Metsätieteen aikakauskirja 3: 311–322. (In Finnish).
Rasinmäki, J., Mäkinen, A. & Kalliovirta, J. 2009. SIMO: An adaptable simulation framework for multiscale forest resource data. Computers and Electronics in Agriculture 66: 76–84.
Saramäki, J. 1977. Ojitettujen turvemaiden hieskoivikoiden kehitys Kainuussa ja Pohjanmaalla. Metsäntutkimuslaitoksen julkaisuja 91(2). (In Finnish).
Siipilehto, J. 1999. Improving the accuracy of predicted basal-area diameter distribution in advanced stands by determining stem number. Silva Fennica 33: 281–301.
Ståhl, G. 1994. Optimizing the utility of forest inventory activities. Ph.D. thesis, Swedish University of Agricultural Sciences, Department of Biometry and Forest Management, Umeå.
Uuttera, J., Hiltunen, J., Rissanen, P., Anttila, P. & Hyvönen, P. 2002. Uudet kuvioittaisen arvioinnin menetelmät – arvio soveltuvuudesta yksityismaiden metsäsuunnitteluun. Metsätieteen aikakauskirja 2002(3): 523–531. (In Finnish).
— , Anttila, P., Suvanto, A. & Maltamo, M. 2006. Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. Metsätieteen aikakauskirja 2006(4): 507–519. (In Finnish).
Vuokila, Y. & Väliaho, H. 1980. Viljeltyjen havumetsiköiden kasvumalllit. Metsäntutkimuslaitoksen julkaisuja 99(2). (In Finnish).

Total of 46 references