Value of forest information

Annika Susanna Kangas

Abstract Traditionally, the quality of forest inventory data has been measured with root mean square error of interesting variables. In many occasions, however, even this information has not been available for all variables of interest, as the main focus has always been on the accuracy of timber volume estimates. In recent years, the quality of data as a basis for decision-making has also been considered, using so-called cost-plus-loss analysis. In such analysis, the optimal data acquisition method is defined to be the one which minimizes the total costs of inventory, i.e., the direct inventory costs and the losses due to suboptimal decisions based on incorrect data. It would, however, be possible to go even further, and estimate the value of certain information, or even information concerning a certain forest variable/parameter in decision-making. This is possible when utilizing Bayesian decision theory. It would enable inventory researchers to concentrate on important issues, and managers to invest optimally on data acquisition process. This review presents the research carried out in valuation of information in different fields and discusses the possibilities to use it in forestry applications.

Keywords Bayesian decision analysis · Cost-plus-loss · Monte Carlo

A. S. Kangas (✉)
Department of Forest Resource Management,
University of Helsinki, P.O. BOX 27, 00014 Helsinki, Finland
e-mail: annika.kangas@helsinki.fi

Defining the quality of forest information

Forest decisions are carried out by different people in many different situations and levels. Forest owner decides whether to harvest his/her stands or to sell the forest holding altogether. Forest companies decide how much and what kind of timber they want to buy, what should be the harvesting order for the stands and how the trees should be bucked to produce maximum value as forest products. At higher level, industry makes decisions concerning future investments. State makes decisions concerning the forest policy, for instance concerning taxes and subsidies or concerning nature conservation areas. All these decisions require good quality information.

According to Nohr (2001), quality of information can be defined as the sum of all requirements that information is expected to have, in order to fulfill specific information needs. The quality of information can be described with several criteria, for instance accuracy, reliability, relevancy, timeliness, completeness and presentation (Rolph and Bartram 1994). Relevancy in this context means that information decreases uncertainty in some specific decisions, and reliability refers to the reliability of the data source or data acquisition method in general (Kätsch 2006).

Information can have value from two different sources: intrinsic value and value in the decision-making (Birchler and Bütler 2007). Value of information (VOI) in decision-making can be defined as the difference between the project value with information and the project value without the information (e.g., Hirshleifer and Riley 1979; Birchler and Bütler 2007). This implies that there must be at least two possible options for the decision-maker to select from, and also uncertainty as to the consequences of these options. If there is no uncertainty, acquiring information cannot add value, which also makes information worthless

from decision-making point of view. Subtracting the cost of acquiring the information gives the net worth of information.

Economics of information, based on Bayesian statistics, has in recent years been an important subject of study in many research fields. Traditionally, it has important applications in insurance and advertising (e.g., Hirshleifer and Riley 1979). Currently, information economics is also used to evaluate diagnostic tests and research in the case of selecting best health care interventions (e.g., Karnon 2002; Ades et al. 2004; Claxton and Sculpher 2006; Chalabi et al. 2007), to evaluate supply chain management practices (e.g., Ketzenberg et al. 2007), to evaluate data obtained from remote sensing spacecrafts versus the costs of launching such systems (Macauley 2006), and to evaluate environmental data acquisition and research (e.g., Dakins et al. 1996; Kim et al. 2003).

In forestry, the VOI has not yet been systematically analyzed, although there are a couple of studies for very specific applications (Knoke 2002; Amacher et al. 2005). Value of forestry decisions may also be very hard to estimate, as besides economic issues forest decisions have also social and environmental effects. Forestry decisions may also affect in very far-away future. Because of such complications, Barth et al. (2006) concluded that it was only possible to evaluate data in qualitative terms. They gave five criteria for evaluating data for national and regional levels:

1. degree of detail
2. accuracy of variables
3. consistency between variables
4. spatial completeness
5. spatial consistency of errors

The first of these criteria, degree of detail, refers to the number of different forest variables concerning which information is collected. Consistency means that the variables at same point have logical relationships. Spatial completeness means that information is available from every point in the forest area.

Yet, even these requirements are not as straightforward as they might seem at first glance. For instance, spatial completeness can always be obtained using interpolation methods such as kriging or with imputation methods if suitable carrier data exist (e.g., Temesgen et al. 2003; Duvemo et al. 2007). Moreover, the degrees of detail can be increased by predicting the missing variables. For instance, diameter distribution can be predicted based on visual stand assessment or laser scanning data (e.g., Maltamo 1997; Gobakken and Næsset 2005). Irrespective of the problems at the task, it would be extremely important to know the value of data in decision-making in order to collect relevant and good quality information.

This review presents the traditional approach to data quality, and approach based on the cost-plus-loss analyses. Then, the concepts and methods for analyzing VOI using Bayesian decision theory are presented, the issues affecting VOI as well as the methods suitable for multi-criteria decision-making situations. Finally, some conclusions and directions for future research are presented.

Selecting the data acquisition method in forestry

In forestry, decisions concerning the acquisition of information have mainly been based on the accuracy of the data and/or the costs of acquiring the data. Optimal data acquisition has usually been understood as the sampling design giving minimum variance for certain estimates with a given budget (e.g., Scott and Köhl 1993; Ståhl 1994a). Usually, the accuracy of mean volume is used as the most important measure concerning the data quality. The problem with this approach is that in practice we do not know which variables have highest value in decision-making, and therefore should have best accuracy.

Another option has been to minimize the costs, given accuracy requirements for certain estimates. The problem with this approach is that we do not have definite accuracy requirement based on decision-making. Thus, we have ended up using ad hoc quality requirements. Yet, another option has been to minimize the weighted sum of the inventory costs and accuracy of certain estimates (e.g., Päivinen 1987). The problem of weighting the costs and accuracy of different variables remains, however.

In practice, it may be that optimal data acquisition is not even strived for. It may be that the method is selected based on tradition, in order to maintain comparability with previously obtained results. Inventory may also be specifically designed for monitoring change. Thus, it aids decisions concerning managing the change, besides aiding decision-making in a certain period. Then, selecting an optimal method is even more problematic than in one inventory. In complicated sampling designs, optimization may only be carried out in small sub-problems, for instance for selecting number of plots given the sampling method, or for selecting number of sub-plots in a case of cluster sampling.

In sampling-based estimation, it is possible to obtain analytical estimates for the accuracy of data, based on the sampling design and the number of sample plots. This covers field samples and estimates obtained using remote sensing material in ratio or regression estimation. If the sampling-based approach does not apply, the accuracy estimates are obtained with empirical tests. This concerns, e.g., the kNN estimates used in many remote sensing-based inventories (e.g., Tomppo 2006a, b), and the same holds

also for traditional (partly visual) field assessments (e.g., Haara and Korhonen 2004).

Typically, the empirical tests just include a couple of methods and a small test area (e.g., Hyyppä et al. 2000). It would simply be too expensive to arrange a thorough test. The relative accuracy of the methods tested typically varies from case to case because of differences in the areas tested, and details of the methods. This can make selecting the optimal method quite complicated. For instance, the accuracy of remote sensing inventories may vary as a function of field data used as reference plots in kNN approach (e.g., Duvemo et al. 2007), due to technical details such as number of pulses per square meter in laser scanning or due to the estimated models used in area-based laser scanning approach (e.g., Uuttera et al. 2006).

Overall, we still have a fairly good understanding of the relative accuracy of all relevant methods for any inventory problem, at least with respect to total stand volume. Forest managers in Finland have for several decades used partly visual standwise forest inventory in their decision-making. In this traditional method, the root mean square error (RMSE) of total volume in Finland has been 24.8% (Haara and Korhonen 2004). Forest managers in practice have been satisfied with this level and are willing to accept remote sensing methods that have at least the same quality (e.g., Uuttera et al. 2002). In general, the accuracy of remote sensing-based methods has been considered quite poor, as these methods have larger RMSE than traditional inventory (Hyyppä et al. 2000). On the other hand, managers are very interested in the area-based applications of laser scanning, since RMSE from around 10 to 20%, i.e., lower than in traditional methods that has been reported in tests (e.g., Næsset 2002, 2004; Uuttera et al. 2006). The satisfaction is, however, more based on tradition and practical experience than true knowledge of the VOI.

It is also not easy to estimate the real costs of the information, especially if the methods to be compared differ very much. In field inventory based on a fixed number of sample plots with fixed measurement schedule, the costs per plot are easy enough to calculate. In traditional field inventory, where the number of observations is not fixed but depends on the surveyor, the task needs to be carried out with empirical tests or based on measured area versus the observed working costs. The costs of remote sensing material are also fairly easy to calculate, although the costs per hectare depend on the area covered. The difficult part, however, is that these methods require an indefinite amount of field observations in addition to remote sensing material, on which the costs (and accuracy) depend on.

The most difficult part in all cases is the cost of work carried out in office, calculating the results. In comparing samples of different size, the difference in the office work

might not be important, but in comparing laser scanning and plot inventory it may be. Some estimates for the total inventory costs have, however, been calculated. In Finland, the estimated cost for traditional inventory was 7.9 €/ha (Uuttera et al. 2002), and in Norway the cost of area-based laser inventory was estimated to be 11.39 €/ha, and method based on photo-interpretation 5.53 €/ha (Eid et al. 2004).

Cost-plus-loss analysis

The traditional approach based on the mean square errors of the estimates does not produce any information regarding the usefulness of the measured information for decision-making purposes. This aspect has been studied using cost-plus-loss analysis, in which the expected losses due to suboptimal decisions caused by inaccurate data are added to the total costs of the forest inventory (Hamilton 1978; Burkhart et al. 1978). This aspect has in recent years been intensively studied in forestry context (e.g., Ståhl 1994a; Eid 2000; Holmström et al. 2003; Eid et al. 2004; Juntunen 2006; Holopainen and Talvitie 2006; Barth et al. 2006; Duvemo et al. 2007).

The hardest part of cost-plus-loss analysis is to define the losses. In the studies carried out so far, it has been assumed that the decision-maker is maximizing the net present value (NPV) of the forest area, and the losses have been defined in terms of NPV (e.g., Holmström et al. 2003; Eid et al. 2004; Juntunen 2006; Holopainen and Talvitie 2006). Also the decisions considered have been fairly similar. For instance, in Holmström et al. (2003), the decisions analyzed were the scheduling of thinnings and clearcuts, in Eid et al. (2004) the timing of final harvests at stand level.

Cost-plus-loss analysis can be carried out either with analytical methods or with simulation. Analytical method is possible, if the losses can be calculated as a function of accuracy. This function could be, for instance, a quadratic function of the accuracy (e.g., Cochran 1977; Ståhl et al. 1994). Typically, in forest studies, the analysis has been carried out using simulation. In simulation, there are also two possible approaches. It is possible to either use real forest data and observed errors or simulated data (or real data with simulated errors).

In the first case, there is only one observed error per each method in each stand (e.g., Eid et al. 2004). The validity of the results depends on the representativeness of the data set used, but on the other hand the errors in each stand and each variable are realistic for that method, and only the future consequences need to be simulated. This approach is especially suitable for comparing existing methods.

With simulated data, on the other hand, it is possible to simulate several realizations of possible errors in each

stand and for each variable (e.g., Eid 2000). Then, it is possible to also analyze the effect of errors in different variables. This method is more general and does not depend on the availability of data. On the other hand, accurate information on the probability distributions of the errors is needed in this approach. Typically, normal distributions have been assumed, but measurement errors could well have asymmetrical distributions and the skewness can vary depending on the true value (e.g., Canavan and Hann 2004). Incorrect correlations among the different errors may also cause incorrect conclusions on the VOI. Positive correlation in errors of, say mean height and diameter, increases error variances in all variables predicted based on these two variables, such as future growth estimates. In such case, the losses will be underestimated if independent errors are assumed.

The losses in each stand i can be calculated from the difference between the NPV from the optimal decision and the NPV from the suboptimal decision based on erroneous data (Eid et al. 2004):

$$NPV_{loss\,i} = NPV_{opt\,i} - NPV_{err\,i} \qquad (1)$$

In this case, it is assumed that there is only one observation per each method from each stand, as is the case with real data. If there are several observations from each stand as in simulation case, the loss in the stand is calculated as their mean. The expected loss in the test area can then be calculated as mean loss in the stands as

$$NPV_{loss} = \frac{1}{n}\sum_{i=1}^{n} NPV_{loss\,i} \qquad (2)$$

The NPV of suboptimal decision is based on true data, not on the estimated NPV with erroneous data (Fig. 1). The anticipated optimal NPV value using erroneous data may be even higher than the true optimum, but that is not loss, it is regret.

Results of Eid (2000) based on simulated data show that different variables have different value in decision-making (Table 1; Fig. 2). Based on his studies, it would be best to concentrate on estimating site quality and age, while most data acquisition techniques concentrate on basal area and mean height. According to Eid (2000), the effect of errors in age and height is highest in stands that are close to maturity, i.e., age in which the final felling would be optimally carried out (Fig. 2). The errors in site quality have, on the other hand, largest effect in young stands. Overall, the effect was the largest when the final felling period was close and the uncertainty was large (Table 2). Thus, besides the accuracy in itself, also other issues affect to the expected losses.

Cost-plus-loss has been used in real data sets to compare different data acquisition methods. Holmström et al. (2003)
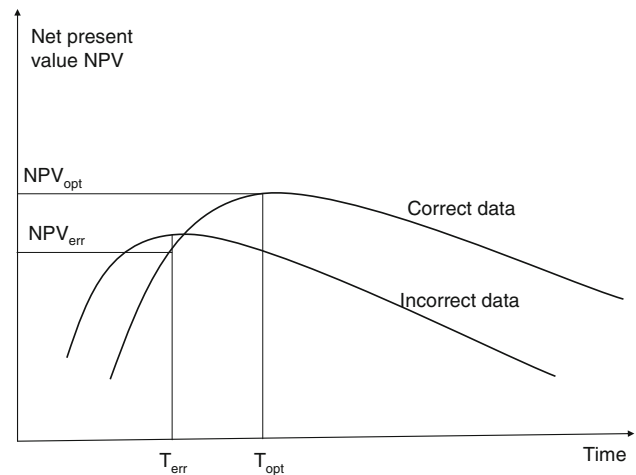

Fig. 1 The principle of cost-plus-loss analysis

Table 1 The expected losses (NOK/ha) in final felling decisions due to random errors in different forest variables (Eid 2000)

Error level (%)	Basal area	Mean height	Site quality	Age	Simultaneous
10	0	28	131	105	243
15	1	63	210	240	499
20	3	147	277	497	931

compared imputation approaches using either traditional field data or aerial photographs as carrier data, and field inventory with five to ten plots per stand with two data sets with simulated errors. In their case, the optimal data acquisition methods depended on the assumed stand size (because both the costs of field inventory and the losses depended on size), on the stand composition in the test area and also on interest rate. If large proportion of the stands were overmature, accurate information was not useful, but if large part of them were not mature but close to it, accurate data were useful.

In another study, Eid et al. (2004) compared photo-interpretation and laser scanning, and found laser scanning clearly more profitable than photo-interpretation in two different test areas with real data. This result may, however, partly depend on the composition of the forest area used in the test. They also calculated losses from the whole rotation period, while Holmström et al (2003) only included the losses from the first 10 years.

The studies carried out so far are inevitably simplifications of the true situation. Either the collected information is used infinitely (Eid et al. 2004; Holopainen and Talvitie 2006) or true data are assumed for later decisions (Holmström et al. 2003). The effect of errors in growth and yield models is not accounted for, i.e., the fact that the data deteriorates in time is ignored (see also Duvemo and

Fig. 2 The average losses due to errors in basal area, mean height, site quality and age as a function of relative maturity (Eid 2000)
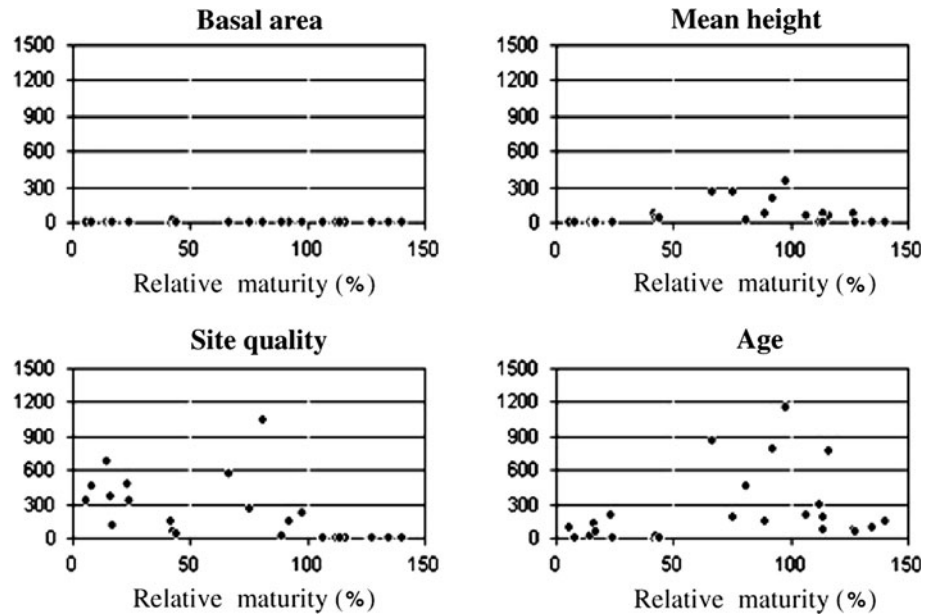

Table 2 The proportions of the optimal period of final harvest due to 15% level of random variation in all variables, seven example stands (Eid 2000)

| Stand no.	Reference data set		Generated data sets											Expected loss (NOK/ha)
	Age	Period of final harvest	1	2	3	4	5	6	7	8	9	10	>10	
1	3	10										61	39	410
2	8	10								34	48	18		631
3	13	10								15	39	39	7	740
4	23	8							4	74	21	1		180
5	54	2	34	38	19	9								1,342
6	64	1	90	6	3	1								990
7	77	1	99		1									292

Lämås 2006). Therefore, the (not discounted) losses should be, on average, the larger the farther in the future the decisions occur.

The deterioration of data due to errors in growth and yield models can be analyzed theoretically using Monte Carlo simulation (e.g., Kangas 1997), or by comparing data predicted with a specific forest simulator to observed data (e.g., Kangas 1999; Haara 2005; Mäkinen et al. 2008; Välimäki and Kangas 2009; Fig. 3). Problem with growth is, of course, that true growth is only known when it has happened. So, studies accounting for accuracy of growth simulators need to be based either on expensive permanent data or on simulation studies. So far simulations have only been carried out for very small data sets and small simulator (e.g., Mowrer and Frayer 1986; Kangas 1997). Building a Monte Carlo simulator for this kind of task is not easy, as there might be several hundreds of models in a large simulator, and their relations are largely unknown.

The forest simulators may, however, also introduce other complications. For instance, in Finnish cost-plus-loss studies (Holopainen and Talvitie 2006; Juntunen 2006), the expected losses are on much higher level than on the other studies mentioned (Eid 2000; Holmström et al. 2003; Eid et al. 2004). In Holopainen and Talvitie (2006), the mean loss varied from 375 to 1,014 €/ha with 3% interest rate, and in Juntunen (2006) from 64 to 130 €/ha with 4% interest rate, while in other studies the mean losses varied from 7 to 51 €/ha. In the case of Holopainen and Talvitie (2006), large part of this is probably due to the higher error level in that study (e.g., relative error of mean diameter varied from 15 to 23%), and the fact that the errors were assumed to be systematic over- or underestimates.

However, it seems evident that the characteristics of the growth and yield simulator also have a large effect on the results. For instance, the growth models used in Norway in GAYA-JLP simulator are not sensitive to the basal area

Fig. 3 The development of the prediction errors of mean height, relative to the predicted mean height, in time (Välimäki and Kangas 2009)
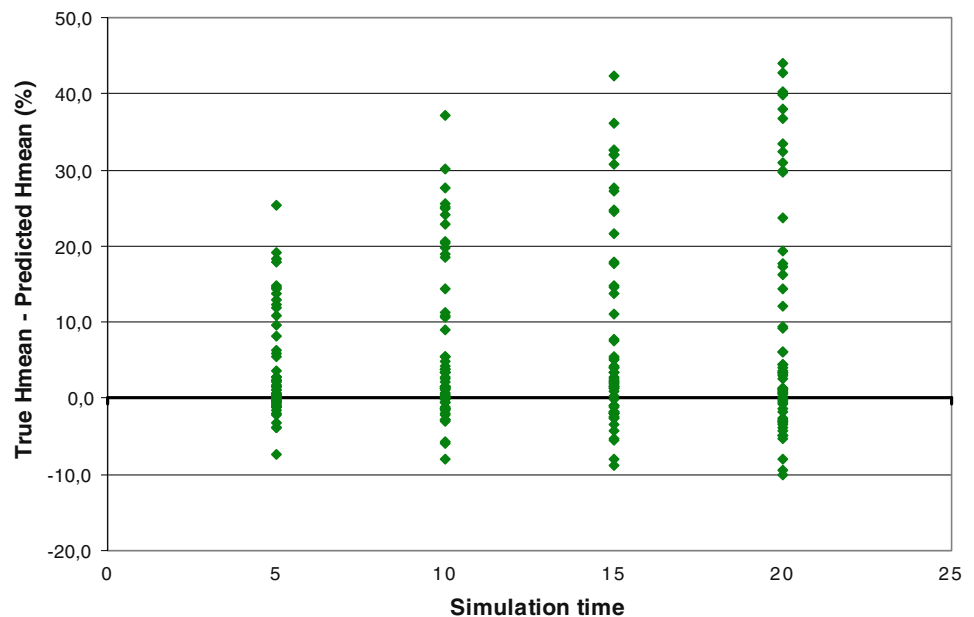
(i.e., basal area is not among the independent variables in the growth predictions), and thus the VOI of basal area seems negligible (Eid 2000). The growth models used in MOTTI simulator are formulated so that the effect of thinnings, i.e., varying basal area, could be accounted for (Hynynen et al. 2002). Thus, these models used by Holopainen and Talvitie (2006) may be more sensitive to errors in basal area. Juntunen (2006), on the other hand, used older MELA models, which can be assumed to be less sensitive to basal area than the new models, as the models underestimated thinning effects in tests (Ojansuu et al. 1991). It means that the expected losses will vary also according to the models used, and this could be utilized also in model building (Eid 2003).

Another problem is that the prediction errors in growth and yield models are affected by the errors in inventory data. For instance, it may be that if basal area is underestimated in inventory, the predicted growth will be overestimated and vice versa. This behavior, of course, also depends on the growth simulator, and the effect of competition in the model. In such cases, it may be very difficult to find out how the quality of data in fact deteriorates. It may even be that no deterioration appears to happen or the relative RMSE of basal area, for instance, in fact decreases in time. An example of this is in Fig. 4, where relative errors increased in the stands with accurate original data quality, but actually decreased in the stands where the original data were of poor quality.

Accounting for the errors of simulators means that it should be the more profitable to invest on accurate data, the longer the period in which this data can be used (e.g., Karnon 2002). It also means that if growth and yield

models are inaccurate, it would be more profitable to make field measurements often, even if they are not very accurate (e.g., Ståhl et al. 1994). Neither the useful life-span of forest data is, however, known, nor its dependency on the original quality of the data. So far, the only attempt to analyze the life-span of data was carried out by Ståhl et al. (1994).

Value of information

The VOI can be calculated based on Bayesian decision theory (e.g., Raiffa and Schlaiffer 1967; Hirshleifer and Riley 1979). In Bayesian decision theory, it is assumed that there is a set of possible decisions $d \in D$. The utility $U(d,s)$ of each decision d depends on which state of nature $s \in S$ is actually going to happen. Thus, there is uncertainty concerning the consequences of the decisions. There exists some prior information concerning the probability of the possible events, $p(s)$. Then, we can maximize the expected value of the decision using the information available as (e.g., Ades et al. 2004):

$$\max_{d \in D} E_s(U), \quad E_s(U) = \sum_{s \in S} U(s,d)p(s). \qquad (3)$$

This is generally not equal to the decision made ignoring uncertainty (e.g., Ståhl 1994a), which implies that it is possible to experience losses simply due to ignoring the uncertainty (Kim et al. 2003). If the uncertainty concerning the future states can be described with a continuous probability distribution, the situation is described as (see also Ståhl 1994a, p. 8):
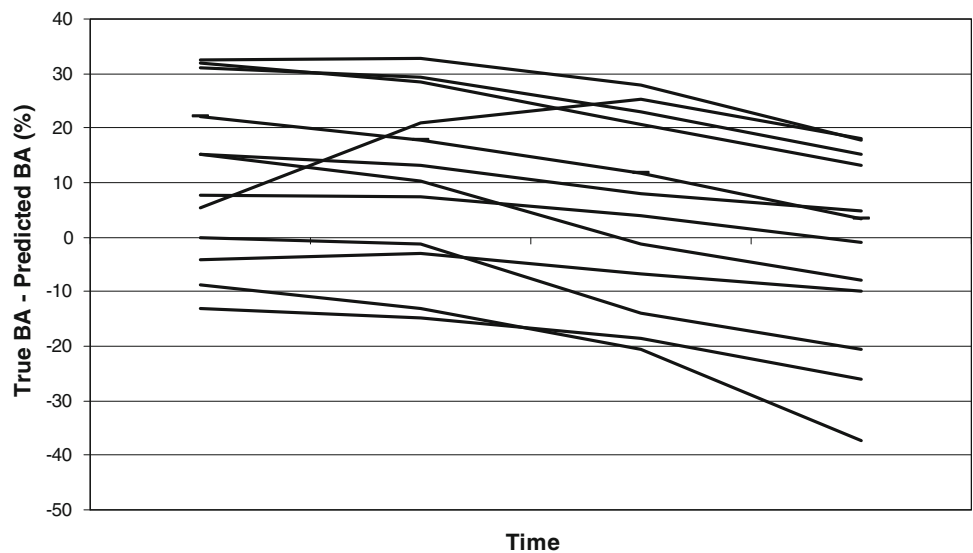
$$\text{Max}_{d \in D} E_s(U), \quad E_s(U) = \int\limits_{s \in S} U(s,d)f(s)ds. \tag{4}$$

In Bayesian sense, the maximum uncertainty would be a case where all the possible states of nature (in discrete case) are assumed equally possible. If it is possible to get some information concerning the states of nature, the uncertainty can be reduced. The new information (message $m \in M$) is used to update the prior probability $p(s)$ to a posterior probability $p(s|m)$ as (Hirshleifer and Riley 1979; Birchler and Bütler 2007):

$$p(s|m) = \frac{p(m|s)p(s)}{p(m)} \tag{5}$$

where $p(s|m)$ describes the conditional probability of observing state of nature s, after observing a message m, $p(m|s)$ is the conditional probability of observing message m, if the state of nature is s and $p(m)$ describes the probabilities of the messages m in general. This, in turn, can be found by averaging the conditional distribution $p(m|s)$ over all possible states s (or in continuous case integrating over the distribution of s):

$$p(m) = \sum_{s \in S} p(m|s)p(s) \tag{6}$$

Given a certain message, the decision-maker needs to maximize the expected value of the decision using the posterior distribution as (e.g., Ades et al. 2004):

$$\text{Max}_{d \in D} E_s(U|m), \quad E_s(U|m) = \sum_{s \in S} U(s,d)p(s|m). \tag{7}$$

Since the actual message is not known beforehand, the results need to be averaged over the distribution of the possible messages as (e.g., Ades et al. 2004)

$$E_m \text{Max}_{d \in D} E_s(U|m), \tag{8}$$

which also implies that it is not meaningful to calculate the value of a certain message, but rather the source of messages (Birchler and Bütler 2007). Then, the difference between the expected utilities (e.g., Ades et al. 2004)

$$E_m \text{Max}_{d \in D} E_s(U|m) - \text{Max}_{d \in D} E_s(U), \tag{9}$$

describes the VOI about the states of nature. If the utility is expressed in terms of money, the VOI is also expressed in money.

In forestry, the message would be an estimate derived from a sample from the forest. Besides states of nature, the uncertainty can also be due to some parameters θ of a decision model (e.g., Ståhl 1994a; Ades et al. 2004). For instance, this could mean parameters explaining the future development, such as mortality rate of seedlings. The uncertainty considered can even be structural uncertainty concerning the shape of these decision models (Claxton and Sculpher 2006, p. 1064). It should be noted that excluding relevant decision alternatives might also have a large impact on VOI (Claxton and Sculpher 2006, p. 1063).

Example

This situation can be illustrated with a forest owner pondering whether to harvest immediately or after 10 years. The interest rate is 3%, and the land value is, for the sake of simplicity, ignored. The possible states of nature are that the stand is still growing well (value growth 4% per year) or that it is growing not so well (1.5%). The forest owner does not have a clue as to which of these options would be true, and so the prior probability of both events is 0.5. The
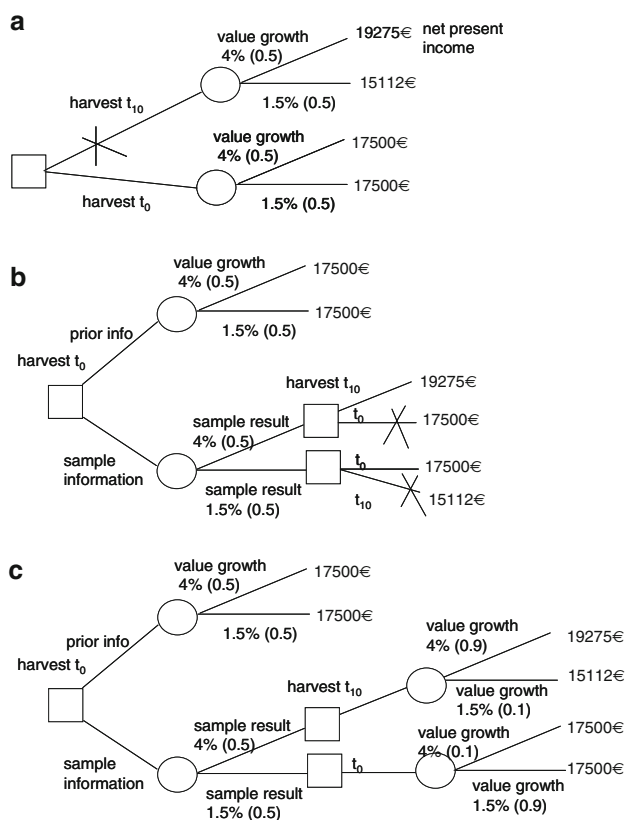
Fig. 5 The decision tree with **a** pure prior information **b** with perfect sample information and **c** with imperfect sample information

problem can be described with a decision tree (Fig. 5a). In this case, the expected utility of harvesting later is $0.5 \times 19{,}275 + 0.5 \times 15{,}112 = 17{,}193$, and that of harvesting now is $0.5 \times 17{,}500 + 0.5 \times 17{,}500 = 17{,}500$, and so the optimal decision based on prior information would be to harvest immediately.

Of course, the owner could take a sample of the stand before deciding. In case the sample is perfect, the optimal decision for the owner would be to harvest later, given the message is good growth, and harvest now if the message is not so good growth. As the prior probabilities of these events were equal, the probability of observing the messages is also equal, 0.5 (Fig. 5b). In this case, the expected value of the decision (Eq. 7) is $0.5 \times 19{,}275 + 0.5 \times 17{,}500 = 18{,}387$, giving as the VOI $18{,}387 - 17{,}500 = 887$. After the owner knows the message, for instance, that the growth is not so good after all, he realizes that he could have made the same decision based on pure prior information. Thus, the VOI is the value of the sample, not a certain message.

Usually, the information is not perfect, however. If the sample information is correct only with probability 0.9 (denoted with quality of data q) irrespective of true growth, it means that $p(m = 4\%|s = 4\%) = p(m = 1.5\%|s = 1.5\%) = 0.9$, and respectively, that $p(m = 4\%|s = 1.5\%)$

$= p(m = 1.5\%|s = 4\%) = 0.1$. In this case, the probability of getting the message of good growth, $p(m = 4\%)$, is the probability of growth being 4% and observing it correctly + probability of growth being 1.5% and observing 4% incorrectly $= pq + (1 - p)(1 - q) = 0.5 \times 0.9 + (1 - 0.5)(1 - 0.9) = 0.5(0 \times 9 + 0.1) = 0.5$, and likewise for the message of not so good growth. Then, posterior probability of truly 4% growth, conditional of obtaining sample value $4\% = p(s = 4\%|m = 4\%) = (0.5 \cdot 0.9)/0.5 = 0.9$, and conditional to sample value 1.5%, 0.1. In this case, the expected value of the decision, $E(U) = 0.5 \cdot (0.9 \times 19{,}275 + 0.1 \times 15{,}112) + 0.5(0.1 \times 17{,}500 + 0.9 \times 17{,}500) = 18{,}179$ (Fig. 5c), giving as VOI $18{,}179 - 17{,}500 = 679$, i.e., much lower than in the case of perfect information. The value of the information does not, however, only depend on the quality of information, but also on the quality of prior information. If the forest owner was pretty sure that the growth would in fact be not so good (with prior probability 0.8), then similar quality sample would only be worth 127.

In case the information acquired is perfect, the estimated VOI is called expected value of perfect information (EVPI), and in case of sample, expected value of sample information (EVSI), or expected value of imperfect information (EVII) (e.g., Karnon 2002; Kim et al. 2003). It should be noted, however, that perfect information can never be obtained. There is always inherent variation in nature which cannot be reduced by collecting new information (e.g., Ståhl 1994a). Thus, EVPI serves as an upper bound on what could be a cost of a data acquisition policy (e.g., Kim et al. 2003).

If the VOI as a function of information quality, $v(q)$, as well as the cost for acquiring such data, $c(q)$ can be obtained, the optimal policy in data collection policy is to maximize the net VOI (Cochran 1977; Ståhl 1994a, p. 12; Birchler and Bütler 2007)

$$\underset{q}{\text{Max}}\Delta v, \quad \Delta v = v(q) - c(q) \qquad (10)$$

This approach leads to the same data acquisition solution as the cost-plus-loss analysis.

The biggest problem of the cost-plus-loss analyses made so far is that they do not account for the prior information (see also Duvemo and Lämås 2006). Such information is, however, always available in forestry. It could be earlier national forest inventories, traditional field data updated with growth models or even the planting year of the stand and growth and yield table. Therefore, while the data acquisition policy obtained with these two methods is similar, Bayesian value analysis may reveal cases where acquiring new data is simply not profitable (see also Ståhl 1994a).

The losses from cost-plus-loss analysis can, however, be interpreted as EVPI with the given information. So, the

difference between the expected loss (or EVPI) based on the prior uncertainty and the expected loss with the updated information results the value of updated, but imperfect information (Dakins et al. 1996; Karnon 2002, p. 335)

$$EVII = EVPI_{prior} - EVPI_{updated} \qquad (11)$$

Based on that, the value of laser scanning in Eid et al. (2004), given that photo-interpretation data already exist, would be 37.5 €/ha for Våler and 33 €/ha for Krødsherad. Thus, even though prior information with the traditional method were assumed to be available, investing in laser scanning data would be profitable (net benefits from 21.6 to 26.1 €/ha).

Ståhl et al. (1994) used the Bayesian approach for selecting the optimal inventory method in the case of multitemporal inventories. This situation is, in principle, equal to the cases presented here, where the possible different inventory results represent the messages of good and not so good growth in the example. Like in the example, the messages were assumed discrete, although the distributions in general were assumed continuous. However, Ståhl et al. (1994) did not calculate the value of the data, but the expected value of the decisions, given different inventory data. This approach could, in future studies, also be used in defining the useful time-span of certain quality forest information.

Like in cost-plus-loss analysis, the VOI can be calculated either with analytical methods or with simulation. The work of Ståhl et al. (1994) suffered from the fact that analytically the Bayesian problem is only solvable with very simple distributions, e.g., in the case where all distributions are normal. Such a case can also be easily simulated (see, e.g., Ades et al. 2004), but for more complicated distributional assumptions even simulation approach is demanding. Currently, however, there are efficient numerical tools such as Markov Chain Monte Carlo (e.g., Gelfand and Smith 1990; Green and Murdoch 1998; Rosenthal 2007) and easy-to-use programs such as BUGS (http://www.mrc-bsu.cam.ac.uk/bugs/). These would enable more realistic applications of Bayesian analysis.

What affects the value of information?

Value of information in the example above depended on the quality of the information, i.e., accuracy of the data, and the quality of prior information. VOI is not, however, simple as that, but also depends on many other issues. In fact, when Ketzenberg et al. (2007) made a meta-analysis of all studies analyzing VOI in supply chain management, they realized that apparently similar data in some situations had large value, and in some cases small value. In their analysis they were able to detect five general issues

affecting the VOI (Fig. 6), namely level of uncertainty (or quality of data), marginal information (or quality of prior information), marginal uses of information, sensitivity to uncertainty and responsiveness to uncertainty.

Sensitivity to uncertainty has also been noticed in forestry studies. Ståhl (1994a), Eid (2000) and Holmström et al. (2003) all noted that if the next treatment was evident, there would be no need for new information, and therefore, the VOI would be low. For example, assuming the only goal of decision-maker is NPV, the next treatment in old stands would in any case be immediate final harvest, no matter what is the volume or basal area.

The marginal uses of information mean that information is used for several purposes, which of course increases its value. In forestry it would mean that in addition to thinking from the forest owner perspective, i.e., the VOI in the timing of harvests decisions, the VOI for the buyer of the timber should also be analyzed. Knoke (2002), for instance, studied how the information concerning the development of quality of beech could improve the value of management decisions. Quality is obviously much more important to the buyer than to seller, unless good quality wood has higher price. Ståhl (1994a, p. 41) concludes that same data is probably more valuable to industrial forest owner than to private forest owner. It may be so, since they can make decisions concerning the bucking, harvesting and logistics based on the same data. Therefore, also the VOI for the other parties should be analyzed in the future, if the true VOI is required. The question remains, however, who should pay for the information in such situations?

The responsiveness to the data means that information is not valuable, if decisions are not made based on the information. It means that if spatial considerations, such as clustering harvests, introduce greater possible losses to the decision-maker than the timing of harvests, the forestry data is not as valuable as would seem (see also Duvemo and Lämås 2006). The same goes with possible even-flow
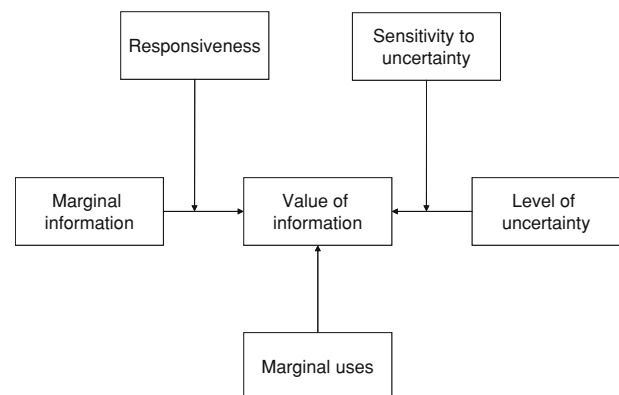
Fig. 6 The aspects affecting the value of information (Ketzenberg et al. 2007)

considerations, or restrictions due to capacity for harvesting and transporting the timber.

Also the fluctuations in price may weaken the responsiveness of the forestry decisions to forestry information. Ståhl (1994b) studied the effect of uncertain timber prices on the optimal data policy. He concluded that with fluctuating prices the optimal data acquisition policy would be to measure less than in the case of fixed prices. It means that the VOI is lower, if the owners do not act based on that but based on some other information. In Ståhl (1994b) the difference was, however, very small, but this may be due to the fact that only part of the timber value was assumed to fluctuate in time.

Multi-criteria decision analysis

When there are several different objectives in the decision-making, and the objectives cannot be measured in monetary terms the problem of VOI is much more complicated (see also Barth et al. 2006). It is not easy to define losses due to, for instance, under- or overestimated sustainable flow of timber from a forest area. Such losses could include losses of industry due to increased/decreased import of logs (Barth and Ståhl 2007), but also due to increased unemployment, decreased taxes to the state, decreased biodiversity and so on.

In principle, the VOI in a multi-criteria case is calculated in exactly the same way as when maximizing the NPV of stands. Two different approaches for the analysis can be used. In the first case, all the benefits are given monetary values. Amacher et al. (2005), for instance, assessed the value of fire risk information, in cases where the forest owner valued both the timber and non-timber benefits of the forest. In medicine, the quality adjusted life years of patients are given a monetary value (e.g., Karnon 2002). In this case, the analysis is carried out in principle exactly as above. The monetary values could be based on, for instance, to stated-preference analyses such as willingness to pay or accept analyses or choice modeling (see, e.g., Bennett and Blamey 2001; Bateman et al. 2002).

The second approach is to use utility functions for the benefits (e.g., Kangas et al. 2008). It means that the net utility of information, defined based on multiple criteria, is searched for. Kim et al. (2003), for instance, analyzed the value of several research projects in Lake Erie, based on a hierarchical utility model including social, ecological and economic aspects. The ten lower-level criteria included the annual sport and commercial harvest of several fish species, PCB concentration in one species and different productivity measures.

In this approach, the VOI is defined using utility as a measure instead of money, which can be hard to interpret.

However, it is always possible to express the value in terms of some of the criteria. Kim et al. (2003) proposed an approach where the utility of two different alternatives are compared. The amount of one criterion variable in the least preferred alternative is then increased (ceteris paribus) until the expected utility of the alternatives is equal, giving the utility difference in terms of that criterion. This method is commonly used in decision analysis, namely in Even swaps method (see, e.g., Hammond et al. 1998a, b; Kangas et al. 2008). If one of the criteria used is a monetary criterion, say net income, it is therefore possible to express the VOI also in terms of money.

Interpreting the results from a multi-criteria analysis is, however, not necessarily easy. Kim et al. (2003) used a panel of six different persons, who had to give both the prior probabilities for the parameters, and the relative values for the ten criteria. In terms of the sport walleye harvest (one of the criteria), ton/ha, value of the information in one research proposal varied from 0 to 22,570. The value was sensitive to both the prior probabilities and weights, but especially so to the weights of the criteria. For instance, giving equal weight to all criteria produced 0 value for all research projects with all prior probabilities.

Another application of VOI analysis in multi-criteria case has been published by Azondékon and Martel (1999), but based on an outranking method instead of utility function. Outranking methods, however, do not produce ratio-scale utility values, but just (complete or incomplete) rank order for the different options. Therefore, there is no analytical method yet to calculate the VOI based on these results.

Conclusions

Selecting the data acquisition methods has in forestry been based on past experiences and the stated accuracy of the used methods. The decisions as to which data is good enough to use for management decisions have been mainly based on the accuracy of stand volume (e.g., Uuttera et al. 2002). Since managers are used to having certain quality data, they might be happy with the accuracy, even if they had to collect additional data in order to make the actual decisions. Such additional costs are also typically ignored when selecting cost-efficient data acquisition method.

Likewise, the decisions as to what constitutes too expensive a method are only based on the costs of the traditional methods. The costs of suboptimal decisions are never realized, and thus not accounted for. This may also be the reason why accurate forest data are generally judged to be too expensive; the VOI is not seen or it is underestimated, as the losses due to poor-quality data remain unknown to decision-makers (Kätsch 2006).

The relevancy of collected forest data has never been questioned. We collect certain kinds of data without thinking if it is really needed in decision-making. However, it may be that we do not even know what decisions are made based on collected data, let alone what information is really needed or used for these decisions. Thus, it would be very important to examine how the collected data are actually used. It may be that quite different information is relevant to different decisions makers. It might also mean that we could obtain the information needed for decision-making with much smaller costs than currently, if we were to concentrate on the most relevant information. Or that we could improve our decisions although the data acquisitions costs remain the same.

Moreover, it would be very useful to know the life-span of the forest information, both as it is and updated with growth simulators. Currently, also the timing of the inventories is based on tradition and experiences of forest managers rather than true knowledge.

It would be high time, therefore, to change from developing new methodology for inventory to analyzing what are the real needs of varying decision-makers in forestry and develop methods that best serve their needs. We need to know, what forest characteristics are most needed for decision-making and how these characteristics can be assessed as cost-efficiently as possible.

References

Ades AE, Lu G, Claxton K (2004) Expected value of sample information calculations in medical decision modeling. Med Decis Making 24(2):207–227

Amacher GS, Malik AS, Haight RG (2005) Nonindustrial private landowners, fires, and the wildland-urban interface. For Policy Econ 7:796–805

Azondékon SH, Martel JM (1999) "Value" of additional information in multicriterion analysis under uncertainty. Eur J Oper Res 117:45–62

Barth A, Ståhl G (2007) Determining sampling size in a national forest inventory by cost-plus-loss analysis. In: Barth A (ed) Spatially comprehensive data for forestry scenario analysis—consequences of errors and methods to enhance usability. Doctoral thesis No 2007:101. SLU, Faculty of Forest sciences

Barth A, Lind T, Peterson H, Ståhl G (2006) A framework for evaluating data acquisition strategies for analyses of sustainable forestry at national level. Scand J For Res 21:94–105

Bateman I, Carson R, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E, Pearce D, Sugden R, Swanson J (2002) Economic evaluation with stated preferences. Edward Elgar, Cheltenham

Bennett J, Blamey R (2001) The choice modeling approach to environmental valuation. Edward Elgar, Cheltenham

Birchler U, Bütler M (2007) Information economics. Routledge advanced texts in economics and finance. Routledge, London, 462 pp

Burkhart HE, Stuck RD, Leuschner WA, Reynolds MA (1978) Allocating inventory resources for multiple-use planning. Can J For Res 8:100–110

Canavan SJ, Hann DW (2004) The two-stage method for measurement error characterization. For Sci 50:743–756

Chalabi Z, Epstein D, McKenna C, Claxton K (2007) Uncertainty and value of information when allocating resources within and between healthcare programmes. Eur J Oper Res 119:530–539

Claxton KP, Sculpher M (2006) Using value of information analysis to prioritise health research. Some lessons from recent UK experience. Pharmacoeconomics 24(11):1055–1068

Cochran WG (1977) Sampling techniques, 3rd edn. Wiley, New York

Dakins ME, Toll JE, Small MJ, Brand KP (1996) Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. Risk Anal 16:67–79

Duvemo K, Lämås T (2006) The influence of forest data quality on planning processes in forestry. Scand J For Res 21:327–339

Duvemo K, Barth A, Wallerman J (2007) Evaluating sample point imputation techniques as input in forest management planning. Can J For Res 37:2069–2079

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

Eid T (2003) Model validation by means of cost-plus-loss analyses. In: Amaro A, Reed D, Soares P (eds) Modelling forest systems. CABI Publishing, Wallingford, pp 295–305

Eid T, Gobakken T, Næsset E (2004) Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scand J For Res 19:512–523

Gelfand AE, Smith AFM (1990) Sampling based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409

Gobakken T, Næsset E (2005) Weibull and percentile models for lidar-based estimation of basal area distribution. Scand J For Res 20:490–502

Green PJ, Murdoch DJ (1998) Exact sampling for Bayesian inference: towards general purpose algorithms. In: Bernardo JM et al (eds) Bayesian statistics, vol 6. Oxford University Press, Oxford, pp 301–321

Haara A (2005) The assessment of uncertainty of updated stand-level inventory data. The uncertainty of forest management planning data in Finnish non-industrial private forestry. Dissertationes Forestales 8, University of Joensuu

Haara A, Korhonen K (2004) Kuvioittaisen arvioinnin luotettavuus. Metsätieteen aikakauskirja 2004:489–508

Hamilton DA (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

Hammond JS, Keeney RL, Raiffa H (1998a) Even swaps: a rational method for making trade-offs. Harv Bus Rev 1–11

Hammond JS, Keeney RL, Raiffa H (1998b) Smart Choices: a practical guide to making better decisions. Harvard Business School Press, Boston

Hirshleifer J, Riley JG (1979) The analytics of uncertainty and information—an expository survey. J Econ Lit 17:1375–1421

Holmström H, Kallur H, Ståhl G (2003) Cost-plus-loss analyses of forest inventory strategies based on kNN-assigned reference sample plot data. Silva Fenn 37:381–398

Holopainen M, Talvitie M (2006) Effects of data acquisition accuracy on timing of stand harvests and expected net present value. Silva Fenn 40:531–543

Hynynen J, Ojansuu R, Hökkä H, Siipilehto J, Salminen H, Haapala P (2002) Models for predicting stand development in MELA System. Metsäntutkimuslaitoksen tiedonantoja, vol 835, 116 pp

Hyyppä J, Hyyppä H, Inkinen M, Engdahl M, Linko S, Yi-Hong Z (2000) Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. For Ecol Manage 128:109–120

Juntunen R (2006) Puustotiedon laadun vaikutus metsänkäsittelyn optimoinnin tuloksiin—UPM Metsän laserkeilausaineiston ja kuviotiedon vertailu. Pro Gradu. Helsingin yliopisto 76 s

Kangas A (1997) On the prediction bias and variance of long-term growth predictions. For Ecol Manage 96:207–216

Kangas A (1999) Methods for assessing the uncertainty of growth and yield predictions. Can J For Res 292:1357–1364

Kangas A, Kangas J, Kurttila M (2008) Decision support for forest management. Managing forest ecosystems, vol 16. Springer, Berlin

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

Kätsch C (2006) Precision forestry and information—information management a forgotten task? In: Ackerman PA, Längin DW, Antonides MC (eds) Precision forestry in plantations, semi-natural and natural forests. Proceedings of the international precision forestry symposium. Stellenbosch University, South Africa, pp 175–186

Ketzenberg ME, Rosenzweig ED, Marucheck AE, Metters RD (2007) A framework for the value of information in inventory replenishment. Eur J Oper Res 182:1230–1250

Kim JB, Hobbs BF, Koonce JF (2003) Multicriteria Bayesian analysis of lower trophic level uncertainties and value of research in Lake Erie. Hum Ecol Risk Assess 9:1023–1057

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

Macauley MK (2006) The value of information: measuring the contribution of space derived earth science data to resource management. Space Policy 22:274–282

Mäkinen A, Kangas A, Välimäki E, Kalliovirta J, Rasinmäki J (2008) Comparison of treewise and standwise forest simulators with the aid of quantile regression. For Ecol Manage 255:2709–2717

Maltamo M (1997) Comparing basal area diameter distributions estimated by tree species and for the entire growing stock in a mixed stand. Silva Fenn 31:53–65

Mowrer HT, Frayer WE (1986) Variance propagation in growth and yield projections. Can J For Res 16:1196–1200

Næsset E (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens Environ 80:88–99

Næsset E (2004) Accuracy of forest inventory using airborne laser-scanning: evaluating the first Nordic full-scale operational project. Scand J For Res 19:554–557

Nohr H (2001) Management der Informationsqualität. Working papers knowledge management Nr. 3. Fachhochschule Stuttgart, Studiengang Informationwirtschaft, 22 pp

Ojansuu R, Hynynen J, Koivunen J, Luoma P (1991) Luonnonprosessit metsälaskelmassa (MELA)—Metsä 2000versio. Metsäntutkimuslaitoksen tiedonantoja, vol 385, 59 pp

Päivinen R (1987) Metsän inventoinnin suunnittelumalli. Summary: a planning model for forest inventory. University of Joensuu, Publications in Sciences No. 11, 179 pp

Raiffa H, Schlaiffer R (1967) Applied statistical decision theory. Wiley, New York

Rolph P, Bartram P (1994) The information agenda: harnessing relevant information in a changing business environment. Management Books 2000, London

Rosenthal JS (2007) AMCMC: an R interface for adaptive MCMC. Short communication. Comput Stat Data Anal 51:5467–5470

Scott CT, Köhl M (1993) A method for comparing sampling design alternatives for extensive inventories. Mitteilungen der Eidgenössischen Forschunganstallt für Wald, Schnee und Landschaft. Band 68, Heft 1

Ståhl G (1994a) Optimizing the utility of forest inventory activities. Swedish University of Agricultural Sciences, Department of Biometry and Forest Management, Report 27

Ståhl G (1994b) Optimal forest inventory intensities under deterministic and stochastic stumpage value assumptions. Scand J For Res 9:405–412

Ståhl G, Carlson D, Bondesson L (1994) A method to determine optimal stand data acquisition policies. For Sci 40:630–649

Temesgen H, LeMay VM, Froese KL, Marshall PL (2003) Imputing tree-lists from aerial attributes for complex stands of southeastern British Columbia. For Ecol Manage 177:277–285

Tomppo E (2006a) The Finnish national forest inventory. Chapter 11, pp 179–194. In: Kangas A, Maltamo M (eds) Forest inventory, methods and applications. Managing forest ecosystems vol 10. Springer, Dordrecht, 362 pp

Tomppo E (2006b) The Finnish multi-source national inventory—small area estimation and map production. Chapter 12, pp 195–224. In: Kangas A, Maltamo M (eds) Forest inventory, methods and applications. Managing forest ecosystems vol 10. Springer, Dordrecht, 362 pp

Uuttera J, Hiltunen J, Rissanen P, Anttila P, Hyvönen P (2002) Uudet kuvioittaisen arvioinnin menetelmät—Arvio soveltuvuudesta yksityismaiden metsäsuunnitteluun. Metsätieteen Aikakauskirja 3/2002:523–531

Uuttera J, Anttila P, Suvanto A, Maltamo M (2006) Yksityismetsien metsävaratiedon keruuseen soveltuvilla kaukokartoitusmenetelmillä estimoitujen puustotunnusten luotettavuus. Metsätieteen Aikakauskirja 4/2006:507–519

Välimäki E (2006) Kasvumallien toiminnan validointi ylitiheissä metsiköissä. Pro gradu -työ. 62 s

Välimäki E, Kangas A (2009) Kasvumallien toiminnan validointi ylitiheissä metsiköissä. Metsätieteen Aikakauskirja (to appear)