HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

The role of growth predictions in the value of information analysis

Annika Kangas

Antti Mäkinen, Ilona Pietilä, Lauri Mehtätalo &

Mikko Nurmi

4.5.2011

What defines a good decision?

- Good decisions are based on analysing the consequences of the possible actions
- The quality of decisions depends on the quality of the information concerning these consequences
- Quality of information concerning
  - the current status of forests → forest inventory
  - the future development → growth and yield predictions
- Typically forest inventory and growth models are considered separately, but they are connected to each others

The three roles of growth models

1. The validity of growth models used in predictions defines the validity of value-of-information calculations

2. The reliability of growth models defines the life-span of collected inventory data

3. Improving the growth models can save money from forest inventory

Value of information (VOI) analysis definition

- VOI =
  the expected monetary value of the decision with the new information
  - the expected monetary value of a decision without the new information
  (Hirshleifer & Riley 1979, Birchler & Bütler 2007, Kangas 2010)

- Subtracting the cost of data acquisition gives the net worth of data

VOI analysis

- VOI cal be calculated for for a given inventory method (e.g. plot sampling)
- for a given variable (e.g. basal area)
- and even for a given measurement (e.g. height sample tree measurement)

Based on Bayesian decision theory

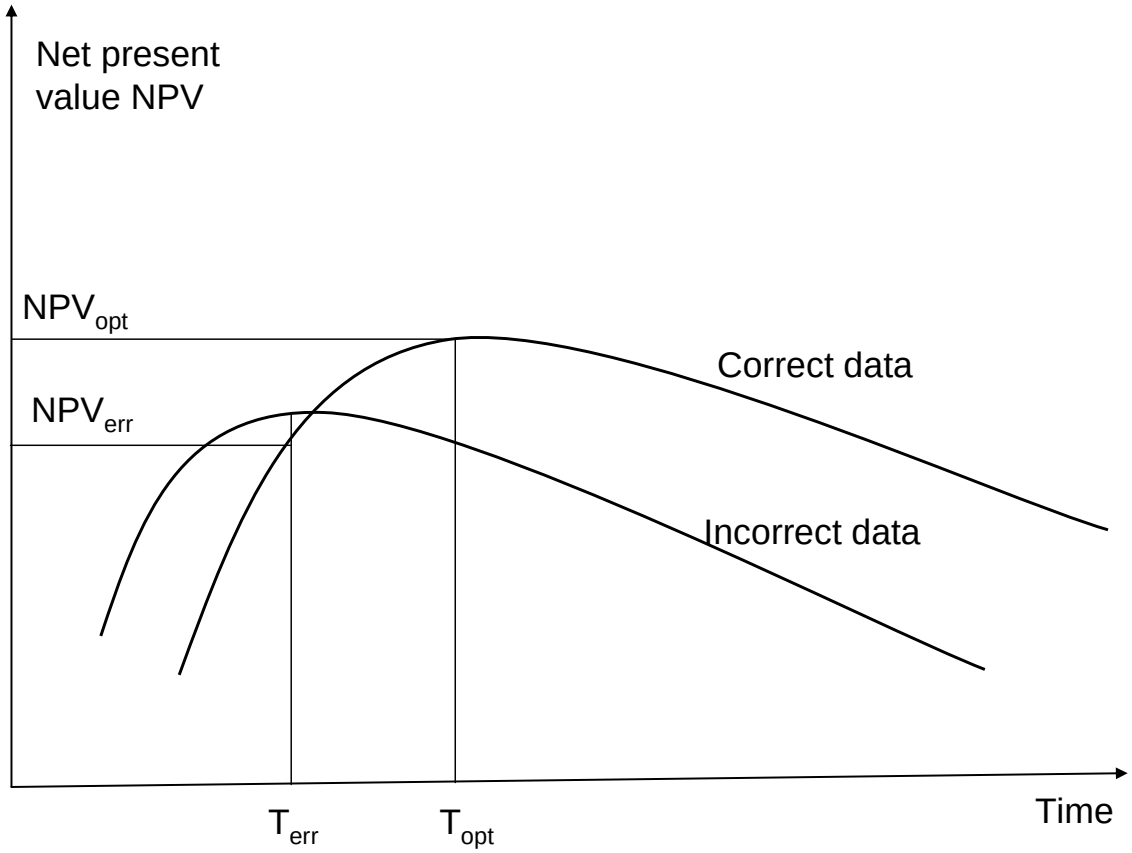
$$VOI(S) = E_y \max_a E_{x|y}\pi(x,a) - \max_a E_x\pi(x,a)$$

- *S* denotes the information source
- *x* denotes the possible states of nature concerning the uncertain variable
- *y* denotes the possible messages from an information source
- $\max_a E_x\pi(x,a)$ the expected payoff of the optimal decision *a* without the additional information
- $\max_a E_{x|y}\pi(x,a)$ the expected payoff of the optimal decision *a* given message *y* and state *x*
- E_y averaged over all possible messages *y*

In a more simple way

- Calculate the losses in net present value (NPV) due to uncertain information
  - Loss = $NPV_{optimal\ decision}$ - $NPV_{inoptimal\ decision}$
- Average loss can be interpreted as expected value of perfect information (EVPI)
  - result of cost-plus-loss analysis
- Expected value of imperfect information (EVII) can be calculated from differences in losses
  - EVII = $Loss_{Prior}$ - $Loss_{Updated}$

Losses in NPV

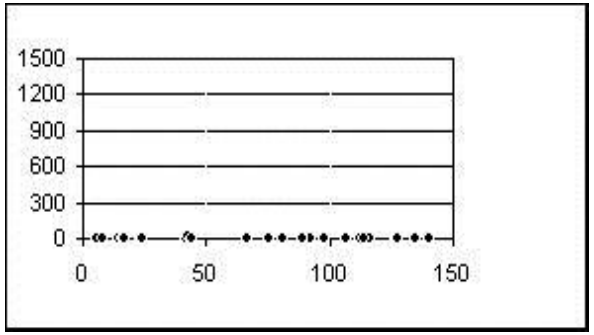


1. How does growth model validity affect VOI?

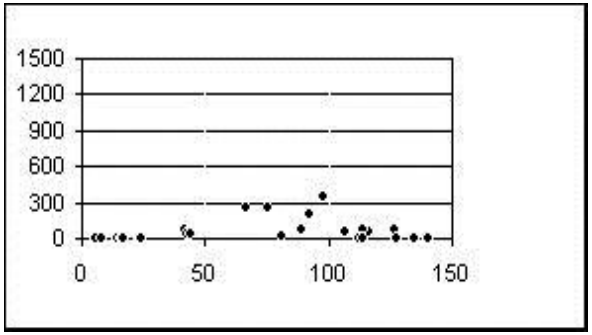
- Making optimal harvest scheduling decisions is based on the assumption that future forest growth under different treatment options can be correctly predicted
  - VOI analysis for any given dataset varies due to the selected growth and yield model
  - The more sensitive the model is to any variable in the growth and yield model, the higher the VOI
  - Only the independent variables of the forest growth and yield simulator (consisting of a group of growth and yield models) can have nonzero value of VOI
  - If growth model is biased with respect to one independent variable, estimate of VOI of this variable will also be biased

Example from Eid 2000

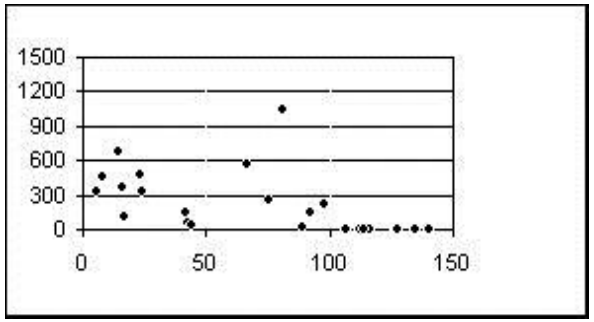
Expected losses (NOK/ha) due to errors in different variables, as a function of relative maturity



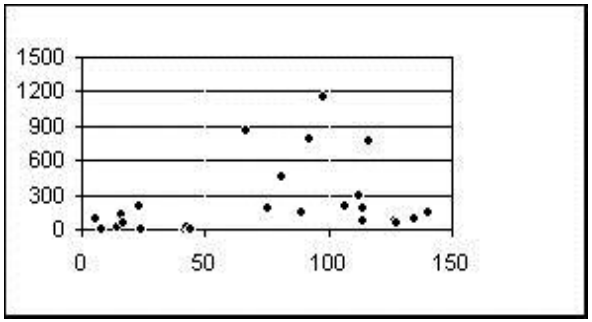
Basal area



Mean height



Site index



Age

Example from Eid 2000

- Basal area has zero EVPI, as basal area is not an independent variable in the used growth models
- The model is sensitive to age and site index, and they have a high EVPI

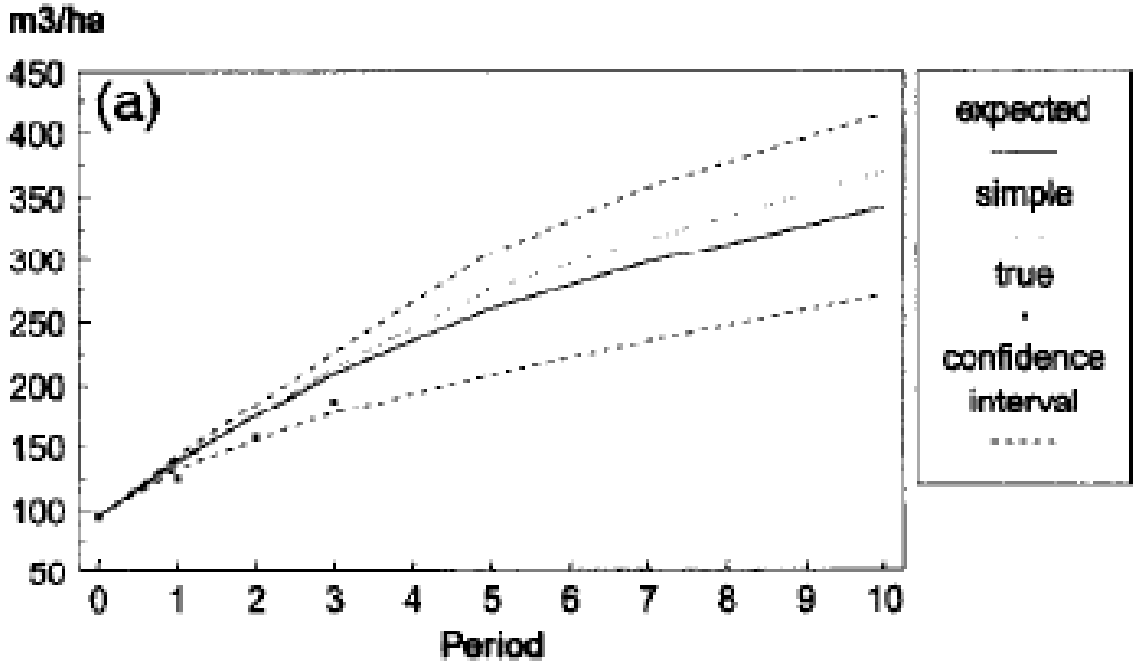
- Yet, in reality basal area affects the growth of the forests through competition of trees

2. How does the growth model affect the life-span of data?

- The errors of initial data and growth predictions propagate in time
  - The longer the prediction period, the lower the quality of the data at the end of the period (Gertner & Dzialowy 1984, Gertner 1987, Mowrer 1991, Kangas 1997)
- The expected losses increase with increasing uncertainty
  - 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, the life span of the data
  - Depends also on data collecting costs

Example from Kangas 1997

- The development of confidence interval of volume prediction of one plot in 10 5-year periods



How can the life-span be studied?

- The pre-requisite for analyzing the VOI of accurate growth predictions is that we can simulate the propagation of errors in time
- In the example, we assume a stand-level growth and yield simulator (SIMO)
  - The increment percentage of basal area and dominant height is predicted
  - using stand age, basal area and dominant height as independent variables
  - volumes, timber assortments and NPVs are calculated with models based on these basic variables

Error model

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

- The error variance is based on observed prediction errors of mean height and basal area (Haara & Leskinen 2009)
- The error of variable x at time t is $\varepsilon_{t,x}$
- It is divided
  - to stand effect u (between-stand variation)
  - and residual variation e (within-stand or between-period variation)

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

Autocorrelation

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

- The stand effect $u_{t,x}$ is assumed fixed in time (perfect autocorrelation in time)
- The between-period effect $e_{t,x}$ is assumed either
  - autocorrelated with AR(1) process (Pietilä et al. 2010)
  - uncorrelated in time (Mäkinen et al. 2011)
- Total autocorrelation $corr(\varepsilon_{t,x}, \varepsilon_{t+5,x})$ was assumed to
  - vary from 0.41 to 0.78 (Pietilä et al. 2010)
  - be fixed 0.65 (Mäkinen et al. 2011)

The losses due to uncertainty of growth predictions (Pietilä et al. 2010)

length of inventory interval (years)	inoptimality loss (euro ha^{-1})				inoptimality loss (%)			
	average	max value	min value	standard deviation	average	max value	min value	standard deviation
5	229.6	1287.7	0	205.1	3.3	19.9	0	3.5
10	341.3	3581.2	0	481.9	4.5	28.5	0	5.4
15	392.8	3316.3	0	493.8	5.7	40.9	0	7.4
20	685	3931.3	0	659.6	9.1	35.2	0	6.9
30	768.3	4132.8	0	717.1	10.4	40.9	0	8.4
60	859.6	3358.2	0	712.5	11.6	36.3	0	8.3

Losses due to growth prediction errors

Including the inventory errors

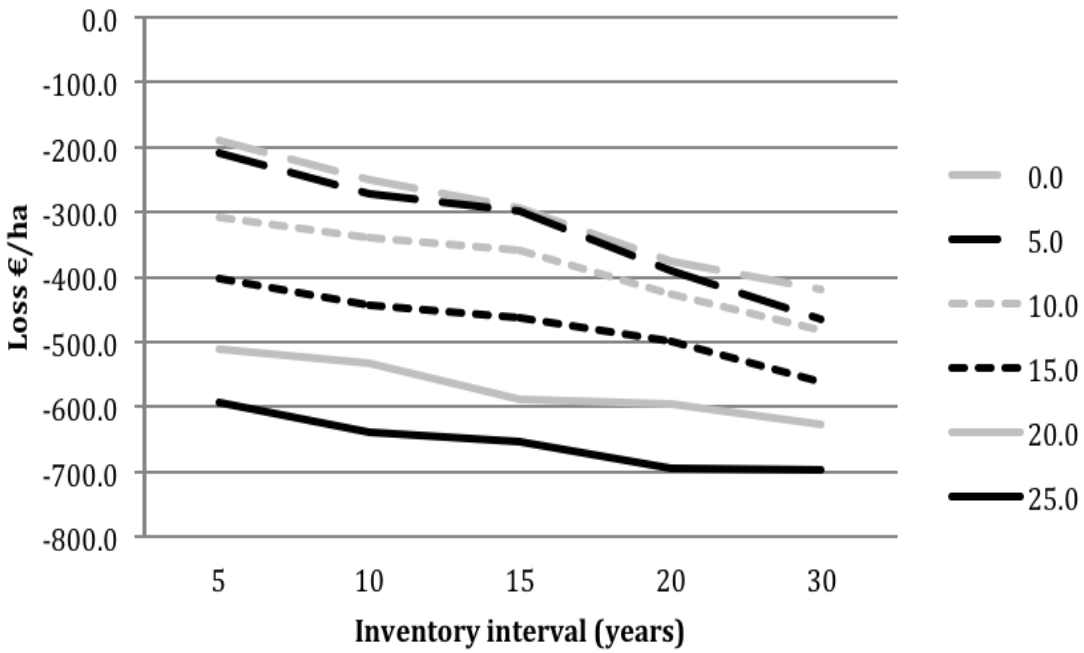
- If we want to define the life-span of the data, we need to include both the inventory errors and the inventory costs

- The SD of inventory errors for both dominant height and basal area was assumed to be either 0%, 5%, 10%, 15%, 20% or 25%

- The costs were described with a simple model
  - number of inventories needed per the 30-year period
  - level parameter giving the general cost level
  - parameter defining the dependency of costs on accuracy of the inventory

Total losses (Mäkinen et al. 2011)

- The total losses of inventory errors and growth prediction errors with different inventory accuracy
- The increase of losses is greatest for the most accurate inventories



Total cost plus loss (Mäkinen et al. 2011)

- Costs of inventory + losses due to inoptimal decisions define the optimal inventory accuracy (10%) and the optimal life-span (15 years) with one given cost assumption

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

3. How can we save money by improving growth models?

- Since the growth prediction errors define the optimal life-span of collected data, we can either
  - lengthen this life-span (interval between inventories) or
  - reduce the accuracy of the original inventory
  - or both

- We can improve the growth predictions by
  - removing (reducing) the stand-level bias u → adjusting the models for a given stand with sample tree measurements of past growth
  - or by estimating new models with additional independent variables

Example (Mäkinen et al. 2011)

- When the stand-effect u was assumed to be zero in the analysis, the total losses in the case of 10-year inventory interval were reduced from 250€/ha to 175€/ha
  - the value of the (perfect) information on the stand-effect is 75€/ha
  - and thus it seems that measuring sample trees for the past growth should be profitable

Example (Mäkinen et al. 2011)

- When a similar (about 30%) reduction of losses was assumed for all inventory intervals, the optimal inventory interval was 15 years and the optimal inventory accuracy was 15%
  - the savings in inventory costs were 92€/ha/30years

	n				
σ	**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

Conclusions

- It would be important to develop the growth models and planning the inventories at the same time
  - in order to select the variables that best explain the growth as independent variables, not just the ones that are traditionally measured
- It should be profitable to include measurement of past growth into the group of measured variables
  - but it depends largely on the true autocorrelation of the stand effects in time
  - and the measurement accuracy of the stand effect
  - but both aspects should be studied in the future

Conclusions

- In the studies carried out so far, we have made many simplifying assumptions
  - assumptions concerning the distributions and autocorrelations of errors
  - assumption that the growth prediction error is homogeneous in different age groups and site classes
  - assumption that the inventory accuracy is the same for all independent variables…

- So, information on VOI on different situations is very uncertain and needs additional studies

References

- Pietilä, I., Kangas, A., Mäkinen, A., & Mehtätalo, L. 2010. Influence of growth prediction errors on the expected losses from forest decisions. Silva Fennica 44: 829–843.
- Mäkinen, A., & Kangas, A. & Nurmi, M. 2011. Using cost-plus-loss analysis to define optimal forest inventory interval and forest inventory accuracy. Submitted to Silva Fennica.
- Kangas, A. 2010. Value of forest information. European Journal of Forest Research 129:863-874.
- Eid, T. 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fennica 34(2): 89–100.