Modelling the probabaility of incorrect harvest decisions due to errors in stand characteristics

Lauri Mehtätalo, Annika Kangas, and Kalle Vanhatalo

[1] University of Helsinki

April 17, 2009
Forest, Wildlife and Wood Sciences for Society Development
Prague, Czech Republic

Outline of the presentation

Forest planning in Finland

- Forest area is divided to compartmets

Forest planning in Finland

- Forest area is divided to compartmets
- An inventory is carried for each compartment, where some important characteristics are assessed by tree species:
  - Basal area (G),
  - Basal area median diameter (DGM),
  - mean height (H),
  - site fertility class
  - Stand age

Forest planning in Finland

- Forest area is divided to compartmets
- An inventory is carried for each compartment, where some important characteristics are assessed by tree species:
  - Basal area (G),
  - Basal area median diameter (DGM),
  - mean height (H),
  - site fertility class
  - Stand age
- "Descriptor trees" of the stand are generated
  - By first predicting the diameter distribution and H-D curve, and
  - then systematically sampling trees from the distribution

Forest planning in Finland

- Forest area is divided to compartmets
- An inventory is carried for each compartment, where some important characteristics are assessed by tree species:
  - Basal area (G),
  - Basal area median diameter (DGM),
  - mean height (H),
  - site fertility class
  - Stand age
- "Descriptor trees" of the stand are generated
  - By first predicting the diameter distribution and H-D curve, and
  - then systematically sampling trees from the distribution
- Alternative treatment chains for the planning predioid are simulated for each compartment by utilizing
  - A tree-level growth simulator,
  - Rules determining maturity for different treatments (mainly thinnings and final cuts),
  - Simulated harvests

Forest planning in Finland

- Forest area is divided to compartmets
- An inventory is carried for each compartment, where some important characteristics are assessed by tree species:
  - Basal area (G),
  - Basal area median diameter (DGM),
  - mean height (H),
  - site fertility class
  - Stand age
- "Descriptor trees" of the stand are generated
  - By first predicting the diameter distribution and H-D curve, and
  - then systematically sampling trees from the distribution
- Alternative treatment chains for the planning predioid are simulated for each compartment by utilizing
  - A tree-level growth simulator,
  - Rules determining maturity for different treatments (mainly thinnings and final cuts),
  - Simulated harvests
- The forest plan is the combination of these harvest schedules that maximizes the utility to the forest owner, or at least satisfies her.

The effect of measurement errors on the forest plan

- The stand characteristics include errors, which range from 10 to 30%.

The effect of measurement errors on the forest plan

- The stand characteristics include errors, which range from 10 to 30%.
- The errors lead to inoptimality losses if the erors change the suggested schedule

The effect of measurement errors on the forest plan

- The stand characteristics include errors, which range from 10 to 30%.
- The errors lead to inoptimality losses if the erors change the suggested schedule
- We could ask: How much we can afford for more accurate data?
  - Answers could be searched for by using the cost-plus-loss approach or the value of information (VOI) approach

The effect of measurement errors on the forest plan

- The stand characteristics include errors, which range from 10 to 30%.
- The errors lead to inoptimality losses if the erors change the suggested schedule
- We could ask: How much we can afford for more accurate data?
  - Answers could be searched for by using the cost-plus-loss approach or the value of information (VOI) approach
- Practitioners could formulate the question as:
  How much errors can be allowed in stand charcteristics if the aim is to end up to a schedule that does not essentially differ from the correct schedule?

The aim of this study

- To study the probabaility of incorrect decision due to varying amounts of errors and stand properties.

The aim of this study

- To study the probabaility of incorrect decision due to varying amounts of errors and stand properties.
- Criteria for correct decision:
  - Correct type of treatment (i.e., no treatment, first thinning, later thinning or final cut), and
  - the timing of a scheduled final felling differs at most 3 years from the correct timing or
  - the timing of a scheduled thinning differs at most 2 years from the correct timing.

  for the next ten year period

The aim of this study

- To study the probabaility of incorrect decision due to varying amounts of errors and stand properties.
- Criteria for correct decision:
  - Correct type of treatment (i.e., no treatment, first thinning, later thinning or final cut), and
  - the timing of a scheduled final felling differs at most 3 years from the correct timing or
  - the timing of a scheduled thinning differs at most 2 years from the correct timing.

  for the next ten year period
- The proportion of correctly scheduled thinnings was computed in different classes of
  - DGM, G, site type, and the type of next correct treatment
  - Realized errors in DGM, G, and H (eD, eG, and eH) (NOT the error variances or standard deviations)

The aim of this study

- To study the probabaility of incorrect decision due to varying amounts of errors and stand properties.
- Criteria for correct decision:
  - Correct type of treatment (i.e., no treatment, first thinning, later thinning or final cut), and
  - the timing of a scheduled final felling differs at most 3 years from the correct timing or
  - the timing of a scheduled thinning differs at most 2 years from the correct timing.

  for the next ten year perioid
- The proportion of correctly scheduled thinnings was computed in different classes of
  - DGM, G, site type, and the type of next correct treatment
  - Realized errors in DGM, G, and H (eD, eG, and eH) (NOT the error variances or standard deviations)
- The probability of correct decision was modeled on stand characteristics an the realized errors.

Data

- 157 spruce-dominated stands from Southern Finland

	G, m^2/ha	D, cm	Site type
min	5	6	2
median	22	19	3
max	39	36	4

Data

- 157 spruce-dominated stands from Southern Finland

	G, m^2/ha	D, cm	Site type
min	5	6	2
median	22	19	3
max	39	36	4

- Defining the correct schedule
  - The ground-measured stand characteristics were regarded as correct
  - The schedule suggested by the General guidelines of Tapio was regarded as correct
  - Note: the effect of errors in formulating the descriptor trees from stand characteristics is omitted.

Data

- 157 spruce-dominated stands from Southern Finland

	G, m^2/ha	D, cm	Site type
min	5	6	2
median	22	19	3
max	39	36	4

- Defining the correct schedule
  - The ground-measured stand characteristics were regarded as correct
  - The schedule suggested by the General guidelines of Tapio was regarded as correct
  - Note: the effect of errors in formulating the descriptor trees from stand characteristics is omitted.
- Simulating the erroneous schedules
  - Errors up to 30% were simulated to D, H and G separately and jointly
  - A total of 183054 realizations were used in this study

Data

- 157 spruce-dominated stands from Southern Finland

	G, m^2/ha	D, cm	Site type
min	5	6	2
median	22	19	3
max	39	36	4

- Defining the correct schedule
  - The ground-measured stand characteristics were regarded as correct
  - The schedule suggested by the General guidelines of Tapio was regarded as correct
  - Note: the effect of errors in formulating the descriptor trees from stand characteristics is omitted.
- Simulating the erroneous schedules
  - Errors up to 30% were simulated to D, H and G separately and jointly
  - A total of 183054 realizations were used in this study
- Schedules were simulated using the SIMO software developed at the University of Helsinki.

Anticipated effects of errors

Anticipated effects of errors

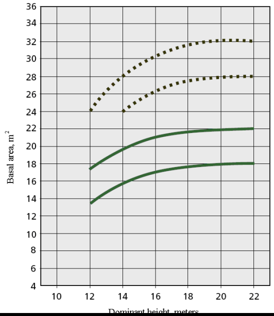
- Final felling
  - The maturity for final felling is defined based primarily on mean diameter (limit 26-28 cm) and secondarily on stand age (limit 70).
  - Only errors on mean diameter should have an effect
  - Overestimation for stands above 26cm should have no effect.

Anticipated effects of errors

- Final felling
  - The maturity for final felling is defined based primarily on mean diameter (limit 26-28 cm) and secondarily on stand age (limit 70).
  - Only errors on mean diameter should have an effect
  - Overestimation for stands above 26cm should have no effect.
- First thinning
  - Carried out when dominant height is between 12-16 meters
  - The number of stems is reduced to 900-1000 stems per ha

Anticipated effects of errors

- Final felling
  - The maturity for final felling is defined based primarily on mean diameter (limit 26-28 cm) and secondarily on stand age (limit 70).
  - Only errors on mean diameter should have an effect
  - Overestimation for stands above 26cm should have no effect.
- First thinning
  - Carried out when dominant height is between 12-16 meters
  - The number of stems is reduced to 900-1000 stems per ha
- Later thinning
  - Based on the thinning model



The logistic model for incorrect decisions

Let random variable OK_{ki} specify whether schedule i for stand k was correct ($OK_{ki} = 1$) or incorrect ($OK_{ki} = 0$). Assume that

$$OK_{ki} \sim Bernoulli(\pi)\,,$$

where

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{1ki} + \ldots + \beta_p x_{pki}\,.$$

The predictors $x_{1ki}, \ldots, x_{pki}$ are include the stand characteristics, relative errors and their interactions.

The logistic model for incorrect decisions

Let random variable OK_{ki} specify whether schedule i for stand k was correct ($OK_{ki} = 1$) or incorrect ($OK_{ki} = 0$). Assume that

$$OK_{ki} \sim Bernoulli(\pi)\,,$$

where

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{1ki} + \ldots + \beta_p x_{pki}\,.$$

The predictors $x_{1ki}, \ldots, x_{pki}$ are include the stand characteristics, relative errors and their interactions.

- Positive and negative errors are anticipated to have different effects – separate terms for positive and negative errors are needed.

The logistic model for incorrect decisions

Let random variable OK_{ki} specify whether schedule i for stand k was correct ($OK_{ki} = 1$) or incorrect ($OK_{ki} = 0$). Assume that

$$OK_{ki} \sim Bernoulli(\pi)\,,$$

where

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{1ki} + \ldots + \beta_p x_{pki}.$$

The predictors $x_{1ki}, \ldots, x_{pki}$ are include the stand characteristics, relative errors and their interactions.

- Positive and negative errors are anticipated to have different effects – separate terms for positive and negative errors are needed.
- The trends are curvilinear – several transformations are needed

The logistic model for incorrect decisions

Let random variable OK_{ki} specify whether schedule i for stand k was correct ($OK_{ki} = 1$) or incorrect ($OK_{ki} = 0$). Assume that

$$OK_{ki} \sim Bernoulli(\pi)\,,$$

where

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_{1ki} + \ldots + \beta_p x_{pki}.$$

The predictors $x_{1ki}, \ldots, x_{pki}$ are include the stand characteristics, relative errors and their interactions.

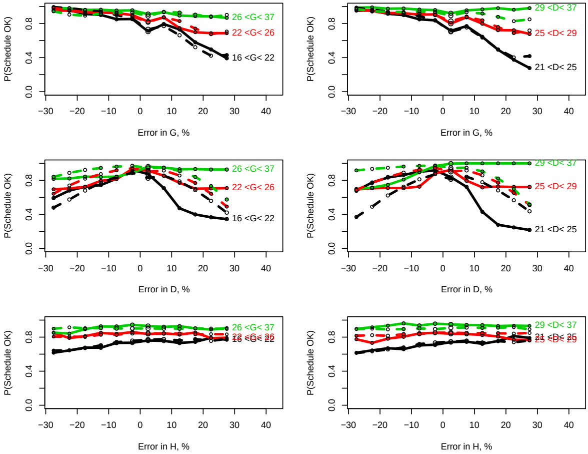
- Positive and negative errors are anticipated to have different effects – separate terms for positive and negative errors are needed.
- The trends are curvilinear – several transformations are needed
- The effect of errors depends on stand properties, e.g., overestimation of mean diameter in a stand with $D = 26cm$ has different effect than in a stand with $D = 20cm$ – interactions between errors and stand characteristics are needed.

The set of predictors in our tentative models

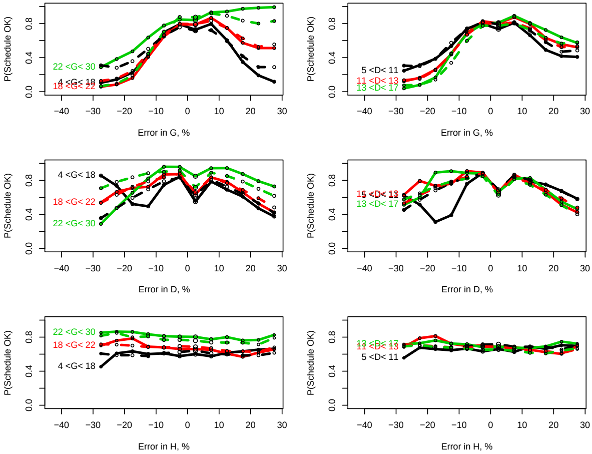
G	$eGneg^2$	D*eGneg	later thinning*eGpos
D	$eGneg^3$	D*eGpos	first thinning*eGneg
eST	eDneg	eGpos*eDpos	later thinning*eGneg
ST3	eDpos	eGpos*eDneg	first thinning*eDpos
ST4	eHneg	eGneg*eDpos	later thinning*eDpos
first thinning	eHpos	eGneg*eDneg	first thinning:*eDneg
later thinning	D*eDneg	ST2*eG	later thinning*eDneg
eGpos	D*eDpos	ST3*eG	G*first thinning
$eGpos^2$	G*eHneg	clearcut*eH	G*later thinning
$eGpos^3$	G*eGpos	first thinning*eH	D*first thinning
eGneg	G*eGneg	first thinning*eGpos	D*later thinning

- We are not yet satisfied with the model

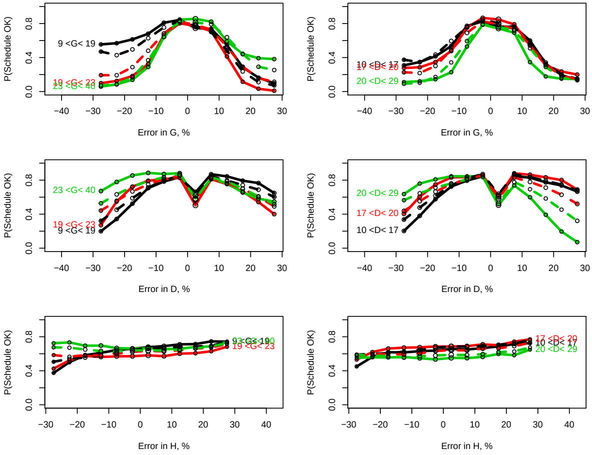
True (solid) and modelled (dashed) proportion in final felling stands



True (solid) and modelled (dashed) proportion in first thinning stands



True (solid) and modelled (dashed) proportion in later thinning stands



Discussion

- The models could be used to compute the error level which gives the specified probability for correct schedule as

$$1 - \alpha = \widehat{\pi} = \frac{\exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \ldots + \widehat{\beta}_p x_p\right)}{1 + \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \ldots + \widehat{\beta}_p x_p\right)}$$

Discussion

- The models could be used to compute the error level which gives the specified probability for correct schedule as

$$1-\alpha=\widehat{\pi}=\frac{\exp\left(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\ldots+\widehat{\beta}_p x_p\right)}{1+\exp\left(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\ldots+\widehat{\beta}_p x_p\right)}$$

- Could also be used in finding stands where good-quality data is most important

Discussion

- The models could be used to compute the error level which gives the specified probability for correct schedule as

$$1 - \alpha = \widehat{\pi} = \frac{\exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \ldots + \widehat{\beta}_p x_p\right)}{1 + \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \ldots + \widehat{\beta}_p x_p\right)}$$

- Could also be used in finding stands where good-quality data is most important
- Does not provide information about how severely the schedule is incorrect (i.e., no prize for the failure to find correct schedule)

Discussion

- The models could be used to compute the error level which gives the specified probability for correct schedule as

$$1-\alpha=\widehat{\pi}=\frac{\exp\left(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\ldots+\widehat{\beta}_p x_p\right)}{1+\exp\left(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\ldots+\widehat{\beta}_p x_p\right)}$$

- Could also be used in finding stands where good-quality data is most important
- Does not provide information about how severely the schedule is incorrect (i.e., no prize for the failure to find correct schedule)
- Finding flexible enough functions for our models was found problematic. Final models might be based on spline regression.

THANK YOU!