Estimating individual tree growth with non-parametric methods

Susanna Sironen

Xqlyhuvlw\ ri Mrhqvxx
Idfxow\ ri Iruhvwu\

JOENSUUN YLIOPISTO

Introduction to non-parametric methods

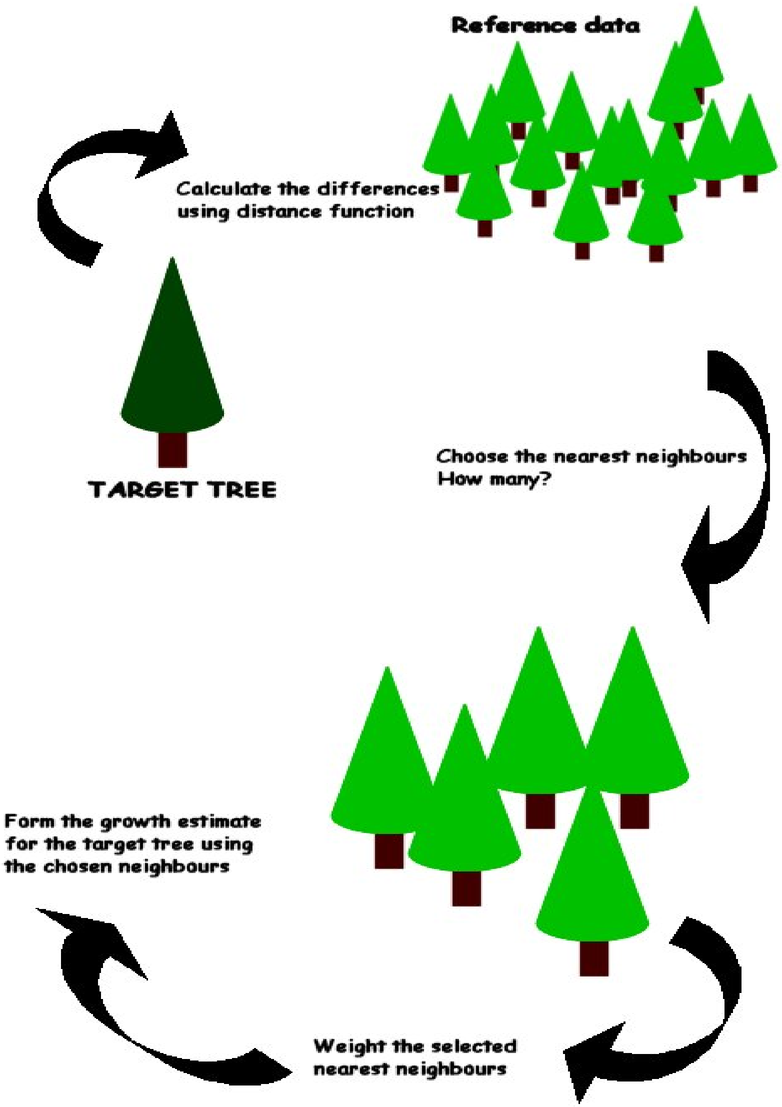
Growth is predicted as a weighted average of the values of neighbouring observations.

Selection can be based on the differences between tree and stand level characteristics of the target tree and the neighbours.

In the estimation of chosen character for a given tree the differences across all reference trees are calculated and the estimate is formed using the chosen nearest neighbours.

Neighbours are chosen from a database of previously measured tree and stand level observations.

Subsumes many methods and variations on methods.



Introduction

Nearest neighbour methods (NN, k-NN)

1. The commonly used Euclidean or squared Euclidean distance

$$d_{ij} = \sum_{l=1}^{p} (x_{il} - x_{jl})^2$$

2. Weighted Euclidean or squared weighted Euclidean distance

$$d_{ij} = \sum_{l=1}^{p} c_l (x_{il} - x_{jl})^2$$

3. Manhattan or weighted Manhattan distance

$$d_{ij} = \sum_{l=1}^{p} c_l \, |(x_{il} - x_{jl})|$$

4. Mahalanobis distance

$$d_{ij}^2 = (X_i - X_j)\beta \sum\nolimits_{zz}^{-1} \beta' (X_i - X_j)'$$

Introduction

Most Similar Neighbour method (MSN or k-MSN)

Consept based on canonical correlation analysis between independent and dependent variables.

$$d_{ij}^2 = \underset{1\times p}{(X_i - X_j)}\,\underset{p\times p}{\Gamma\Lambda^2\Gamma'}\,\underset{p\times 1}{(X_i - X_j)'}$$

Canonical correlation formulation gives the possibility to use only the first significant canonical variates – with regression formulation the full-rank coefficient matrix must be used.

Introduction

Generalized additive models (GAM)

Extensions of generalized linear models.

Generalized additive models are a method of fitting a smooth relationship between two or more variables through a scatterplot of data points.

Parametrized just like GLMs, but some predictors can be modelled non-parametrically; the linear function of the predictor values may be replaced by non-parametric function obtained by applying a scatterplot smoother to the scatterplot of partial residuals.

Generalized additive models (GAM) have the form:

$$y = \beta_0 + \sum_{j=1}^{p} f_j(X_j) + \varepsilon$$

Introduction

Only underlying assumption is that the functions are additive and that the components are smooth.

Probability distribution of the response variable must still be specified and in this sense, GAMs are more aptly named semi-parametric models.

Estimation procedure for a GAM requires iterative approximation in order to find the optimal estimates

Estimation is based on combination of a local scoring algorithm and backfitting algorithm.

What has been examined?

- Local non-parametric growth estimates for Kuusamo in Finnish Lapland
- Localization of growth estimates using non-parametric imputation methods
- Comparison of different non-parametric growth imputation methods in the presence of correlative observations
- Estimating individual tree diameter and height increment simultaneously using non-parametric imputation

Localization of growth estimates using non-parametric imputation methods

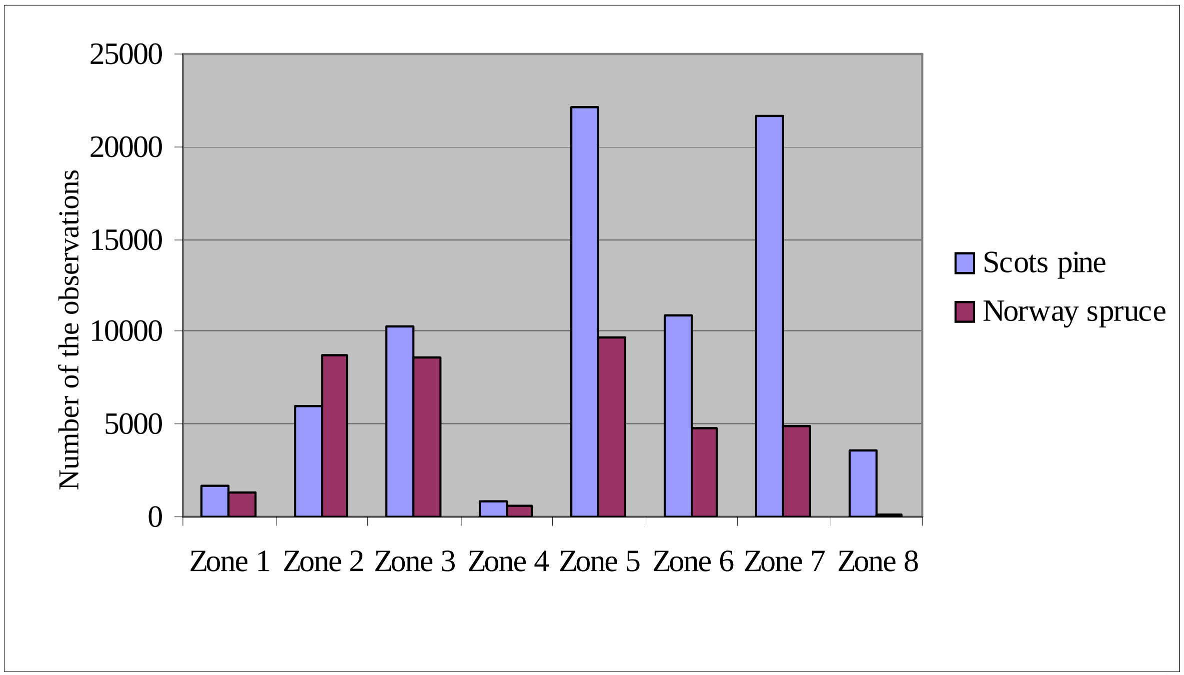
Why?

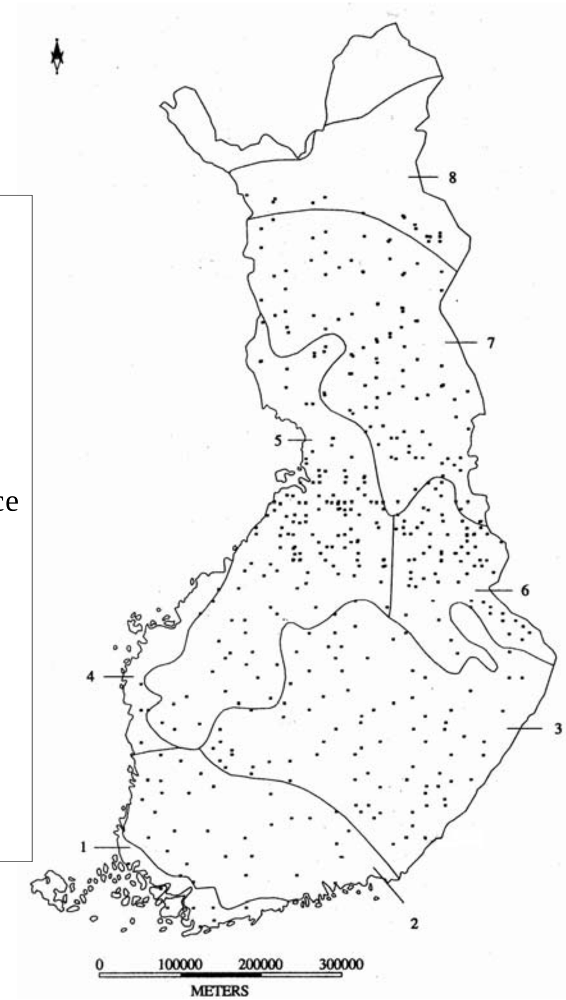
The typical models used in forest management planning situations are national models, which give accurate results in larger areas, but for a given area or a stand the models may give large overestimates or underestimates

Examination of different non-parametric imputation methods to reduce regional biases in growth estimates

Comparison of localized estimates to estimates obtained with non-spatial imputation and also with estimates from a parametric growth model

Localization of growth estimates …



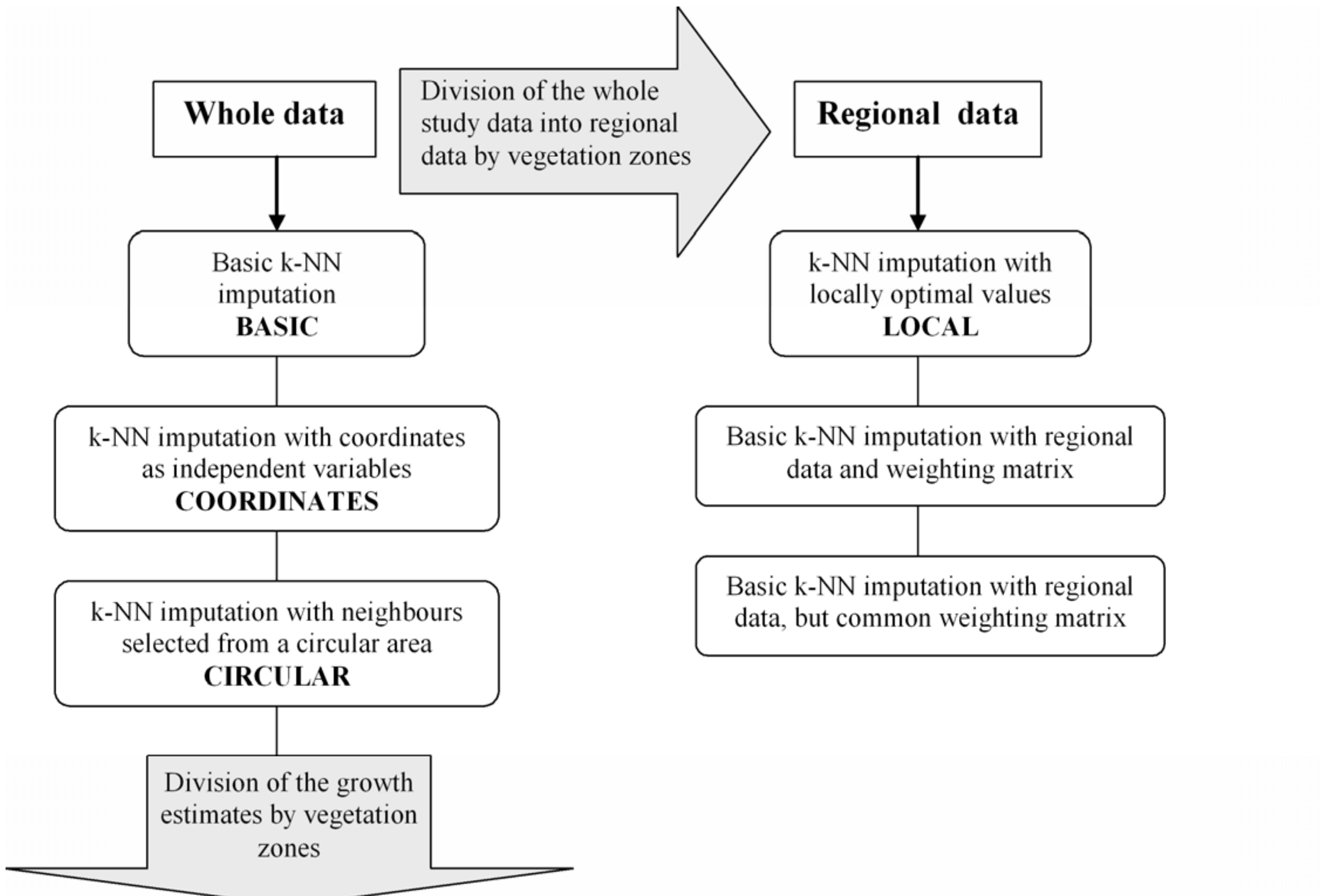


JOENSUUN YLIOPISTO

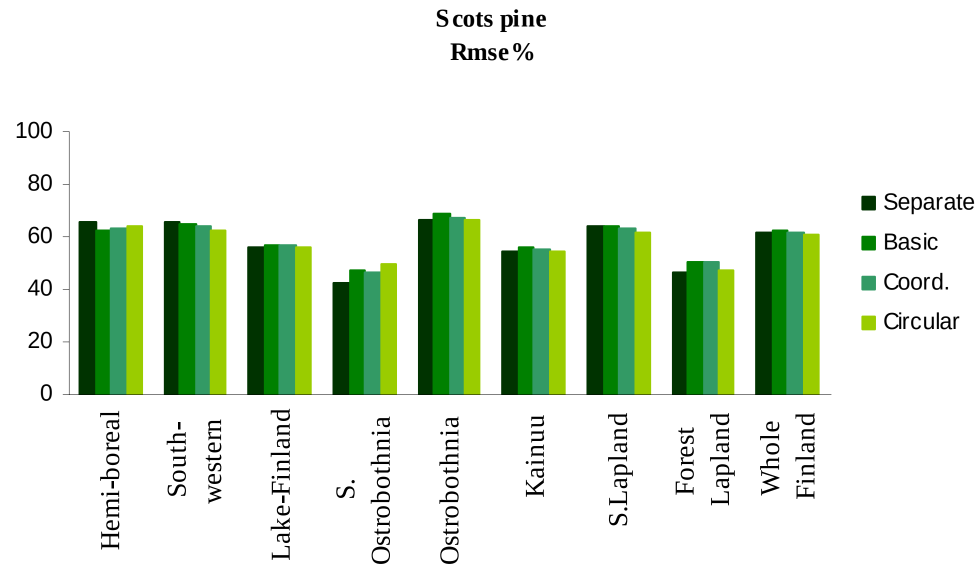
Localization of growth estimates….

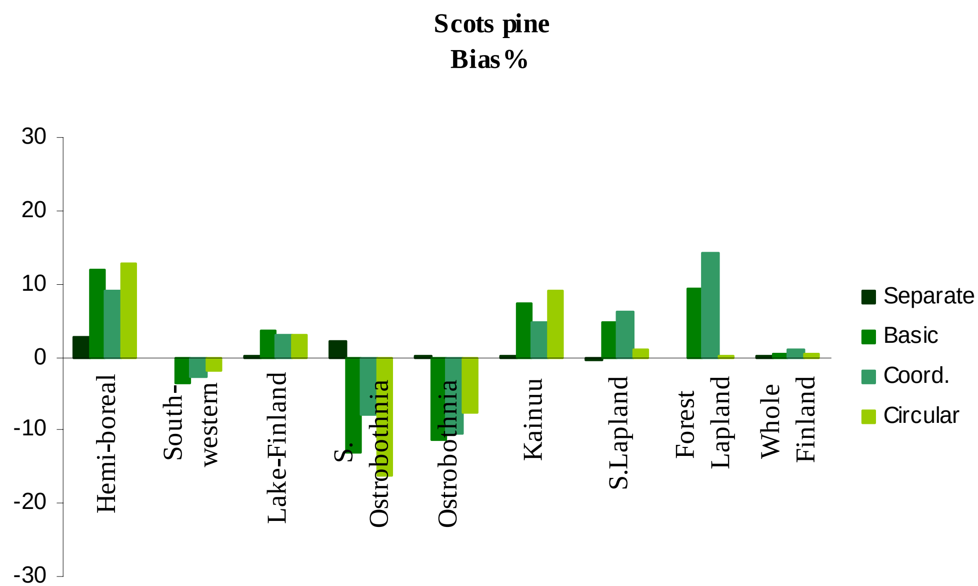
Localization was obtained through a variety of methods:

- Using spatial coordinates as independent variables
- Restricting the selection of neighbours to a circular area around the target tree
- Restricting the selection of neighbours to a local database

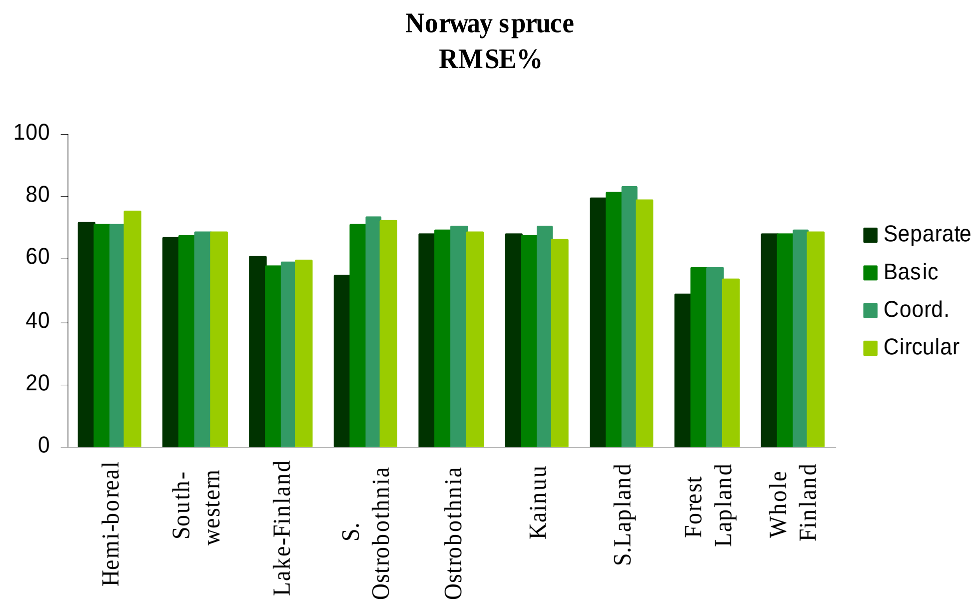


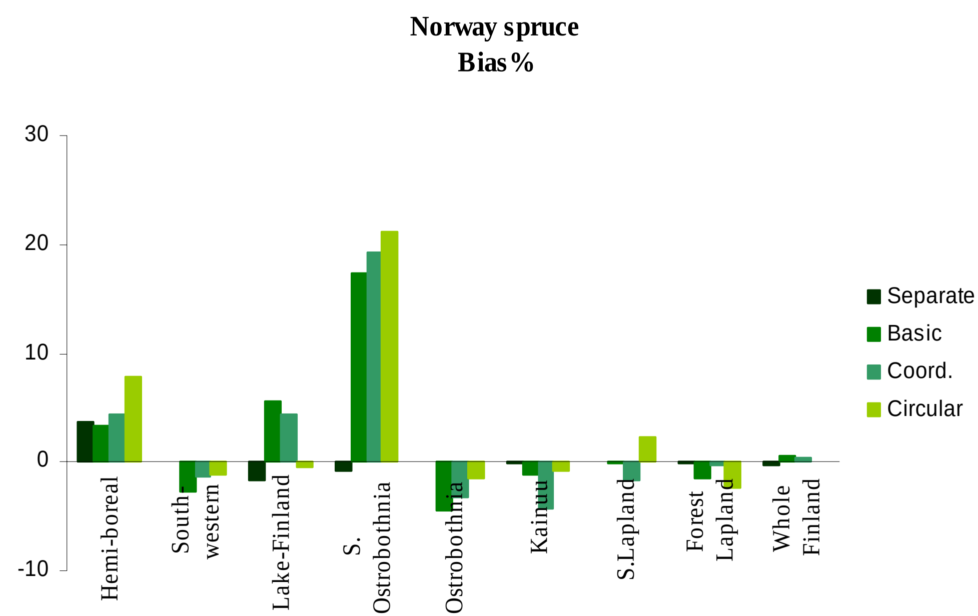
Accuracy of the 5-year tree-level growth estimates obtained by different methods, by vegetation zones:





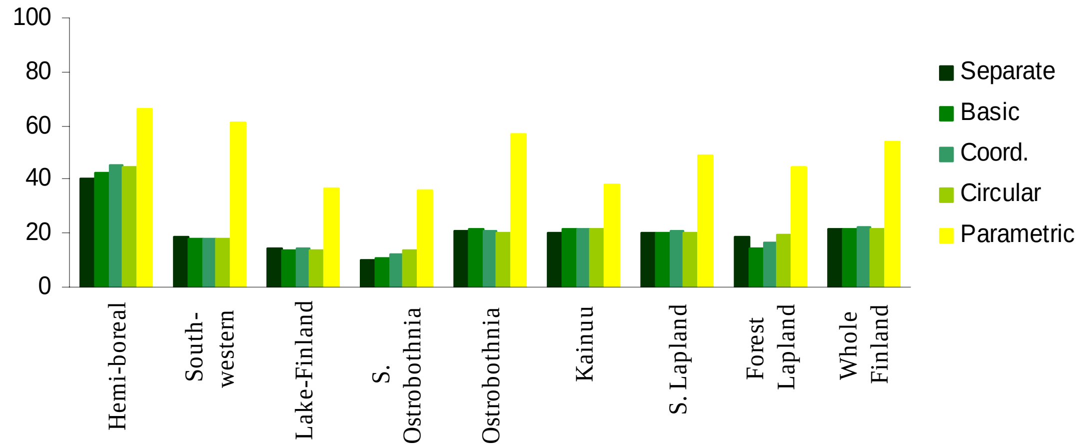
Accuracy of the 5-year tree-level growth estimates obtained by different methods, by vegetation zones:



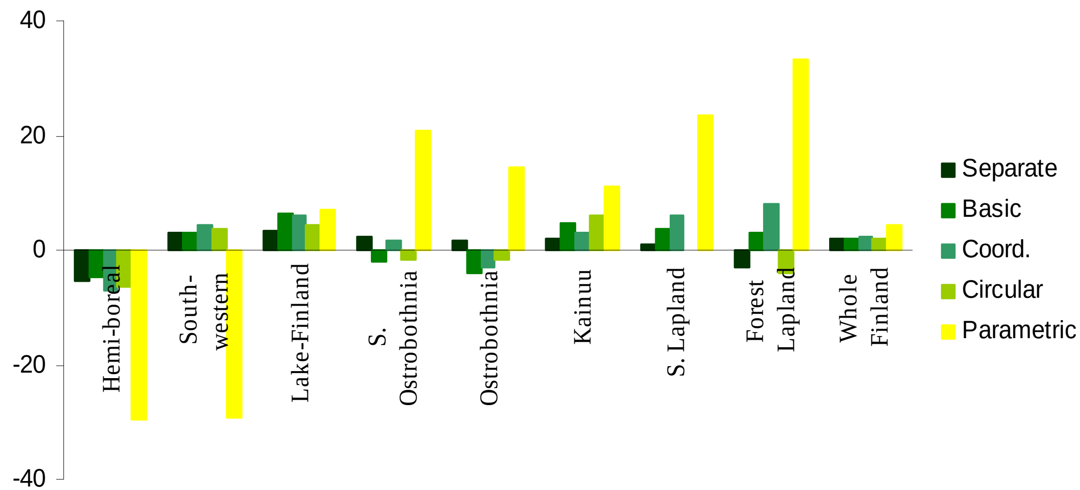


Accuracy of the 5-year stand-level growth estimates obtained by different methods, by vegetation zones:

Stand-level growth
Rmse%



Stand-level growth
Bias%



XQLYHUVLW\ RI MRHQVXX
IDFXOW\ RI IRUHVWU\

JOENSUUN YLIOPISTO

Localization of growth estimates ….

The localization did not reduce the regional biases compared to the basic non-spatial imputation

The basic k-nn imputation

- can find the nearest neighbours close enough
- had the temperature sum as independent variable, which can be seen as a way of localizing and might diminish the difference
- uses the whole data and usually better matches could be found with increasing sample size

It might be too hard to find the neighbours, at least for the exceptional observations, if the amount of possible neighbours is reduced

The biases for the exceptional observations were larger for the localized models

The most promising alternative to localize was the use of moving geographical areas.

Localization of growth estimates …

Compared to parametric model:

- In general, the non-parametric models performed well at thin and dense stands, while the parametric model produced very large variation of residuals in dense stands in many regions.

- The mean biases in all of the regions were close to each other, while the differences in mean bias of growth estimates with parametric model were over 25 % between the regions.

Comparison of different non-parametric growth imputation methods in the presence of correlative observations

Why?

Observations from the same stand as the target tree are usually excluded from the pool of possible nearest neighbours.

However, the nearest neighbours may still all be selected from one particular stand, if the stand-level variables contain much weight in the distance function.

Earlier studies indicated that some methods are better at tree level and some methods at stand level.

Stand-level or regioanal results may be affected, if the estimates are formed with too many neighbours from the same correlative observations.

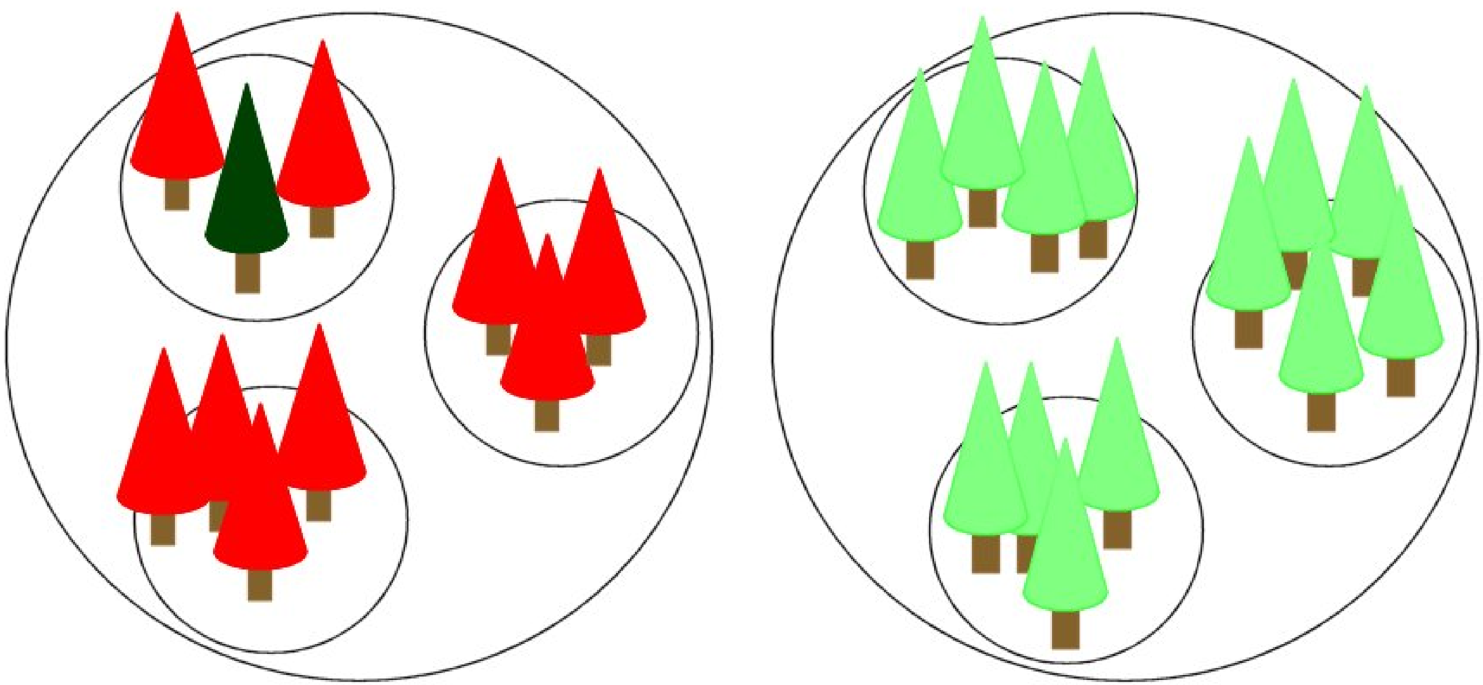
The errors of all individual trees may be parallel, if all neighbours are selected from a stand in dry site, while the target tree is growing in fresh site, for example.

Comparison of different non-parametric growth imputation …

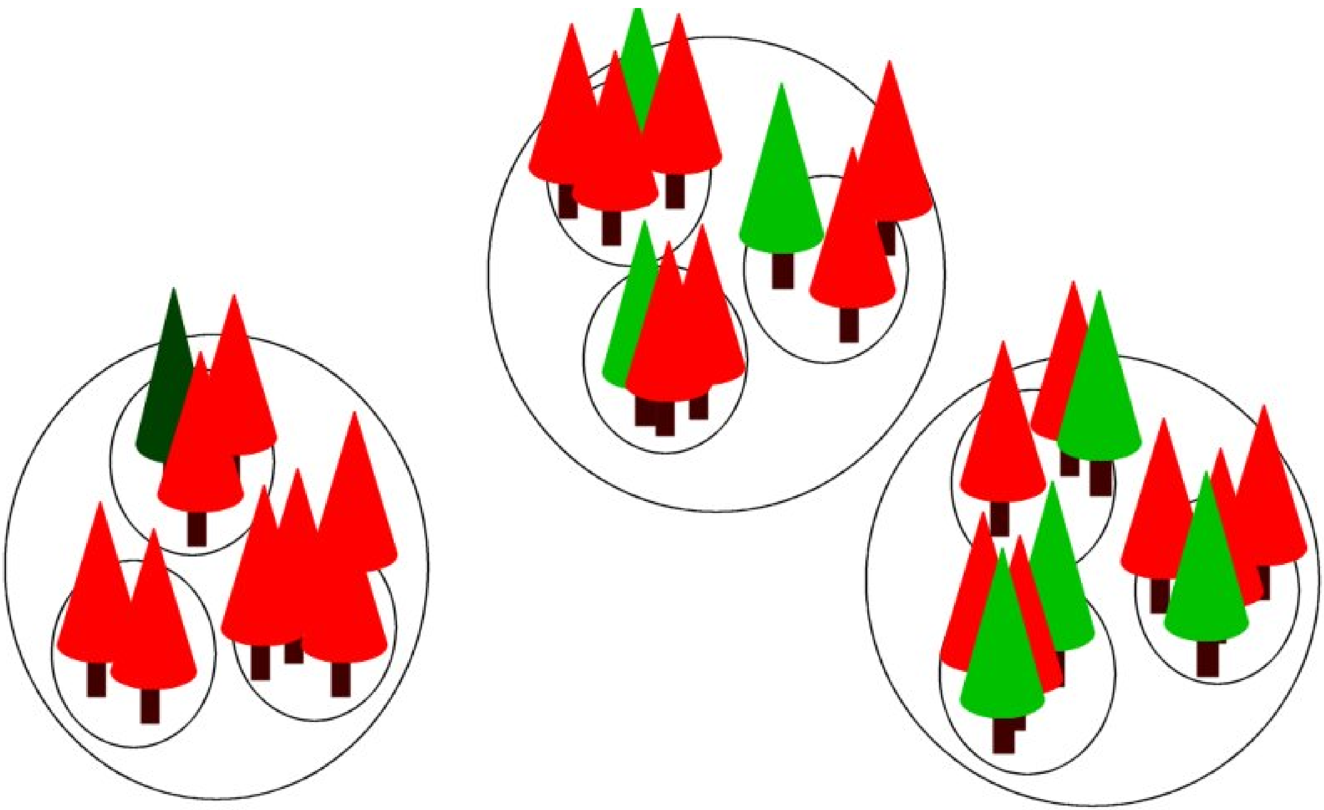
Methods

- The k-Most Similar Neighbour imputation
- Basic k-nearest neighbour imputation with similarity of the trees measured by using squared euclidean distance
- Generalized additive models
- Examination was carried out by using different kind of restrictions to the pool of possible nearest neighbours

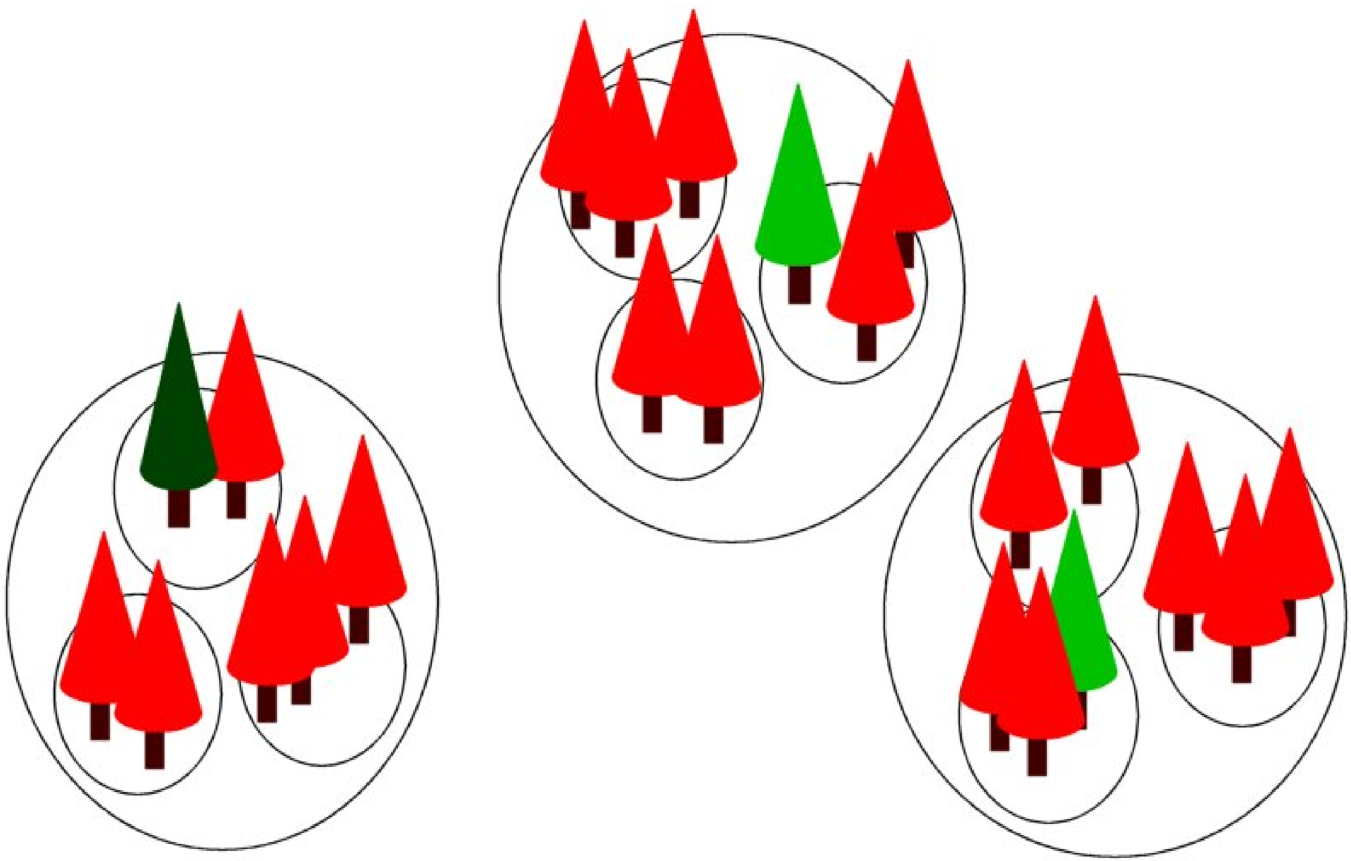
The usual situation: neighbours from the same stand as the target tree are excluded from the pool of possible nearest neighbours . Otherwise all the neighbours can be selected from a one particular plot or stand.



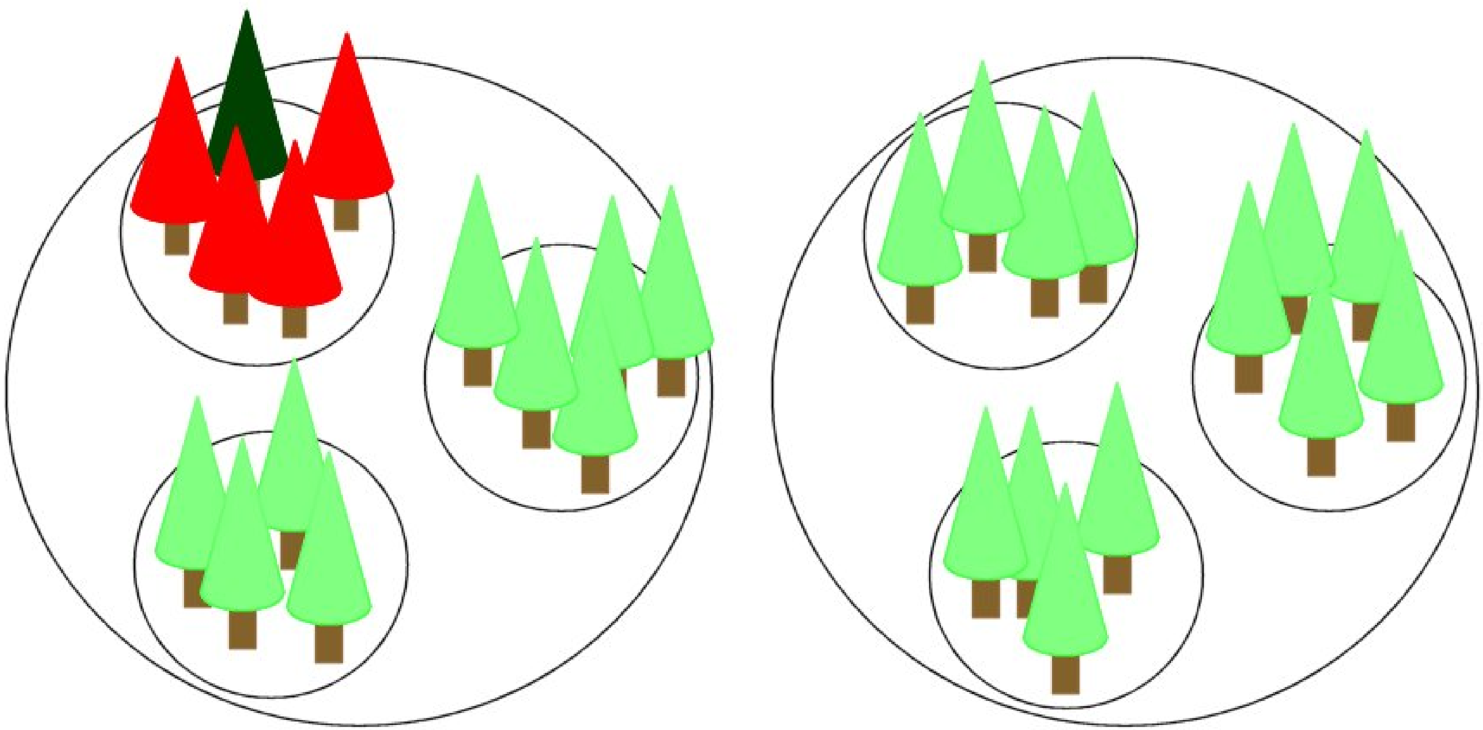
One per plot restriction: neighbours from the same stand as the target tree are excluded and only one neighbour per each INKA plot can be selected.



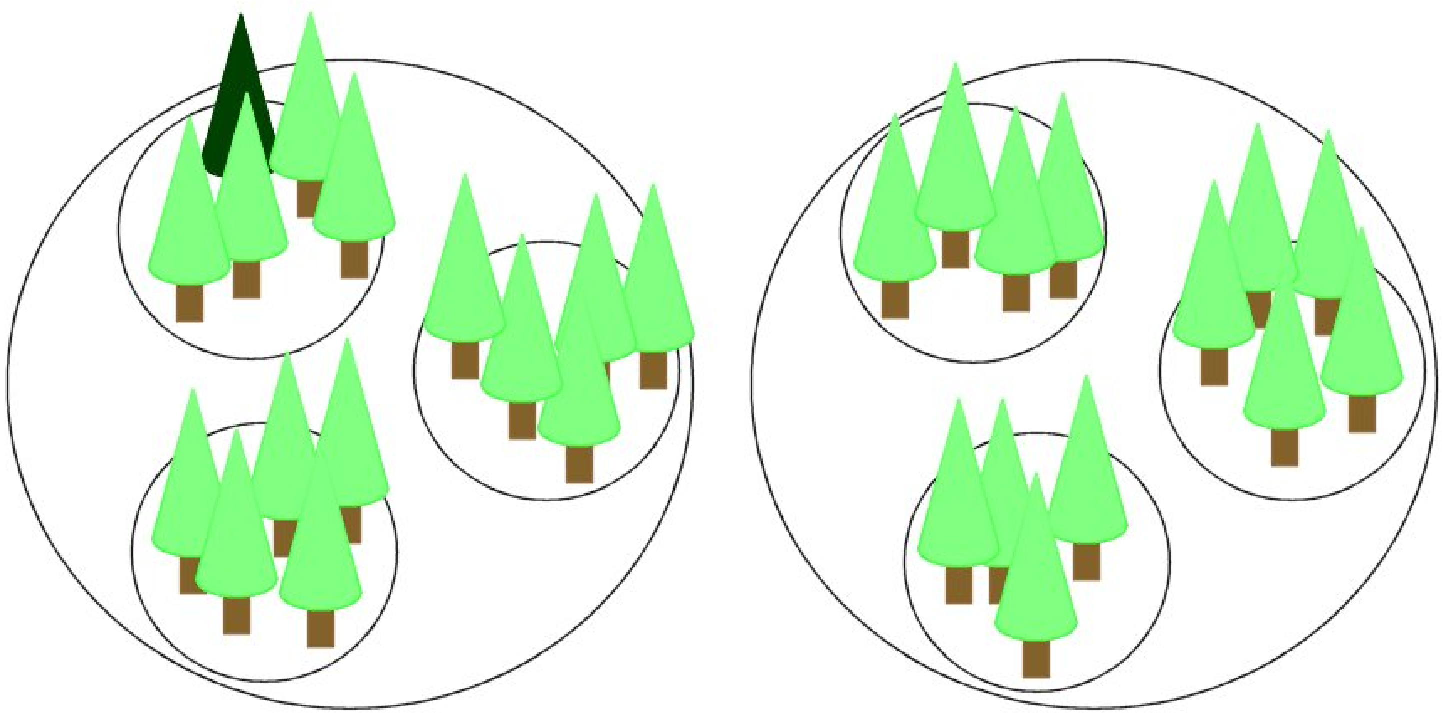
One per stand restriction: neighbours from the same stand as the target tree are excluded and only one neighbour per each INKA stand can be selected.



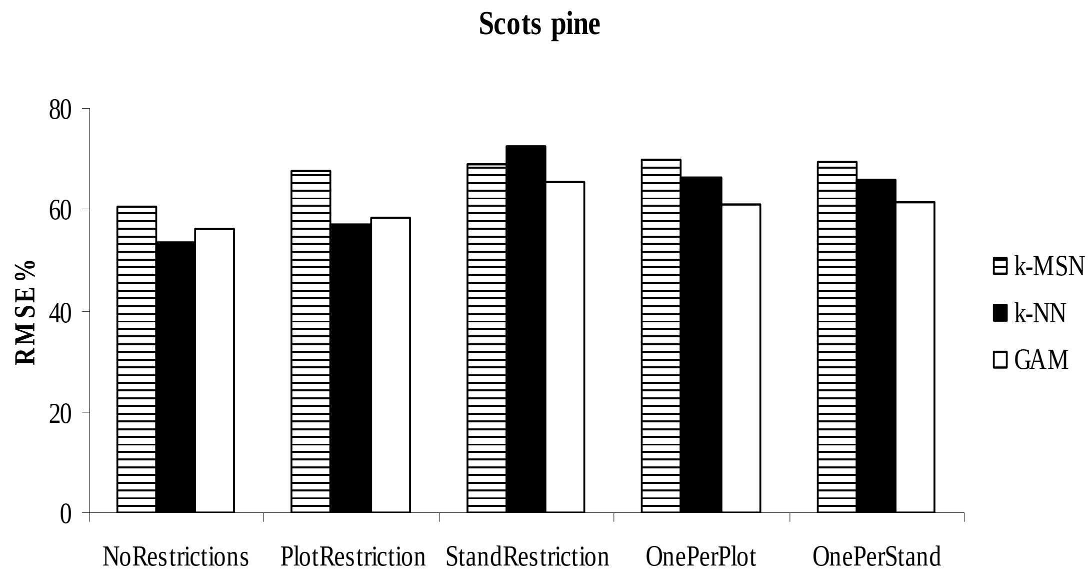
Plot restriction: neighbours from the same plot as the target tree are excluded. Otherwise all the neighbours can be selected from a one plot or stand.

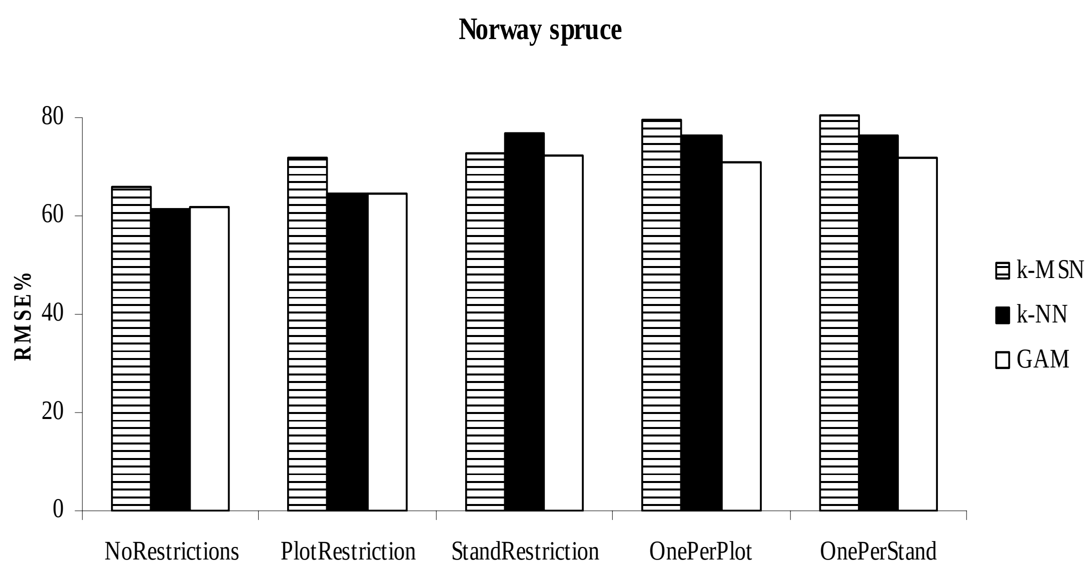


No restrictions: neighbours can be from the same stand and all of the neighbours can be from one plot or stand.

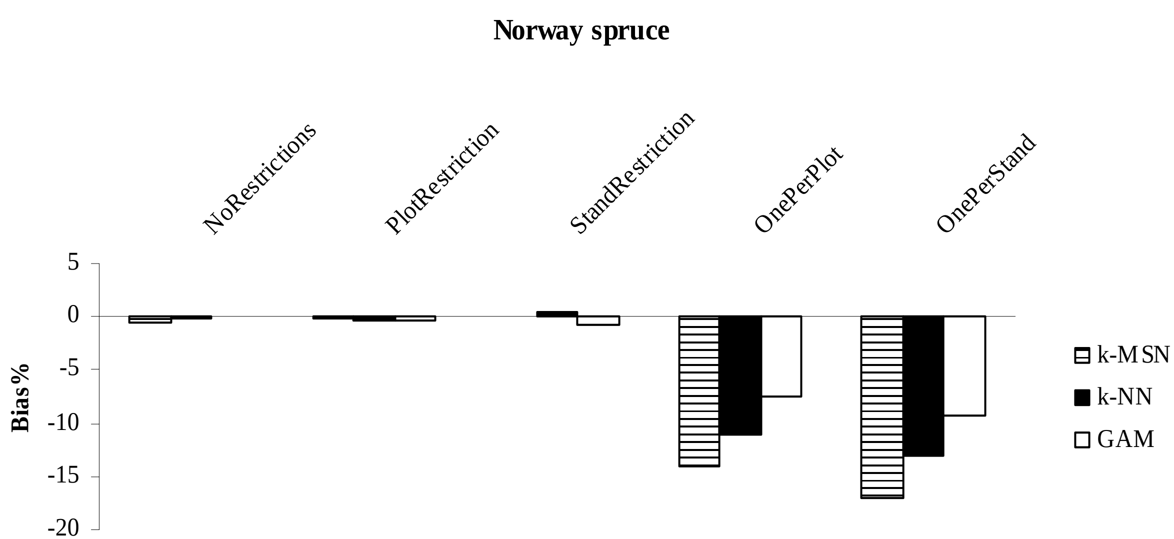


Accuracy of the tree-level results

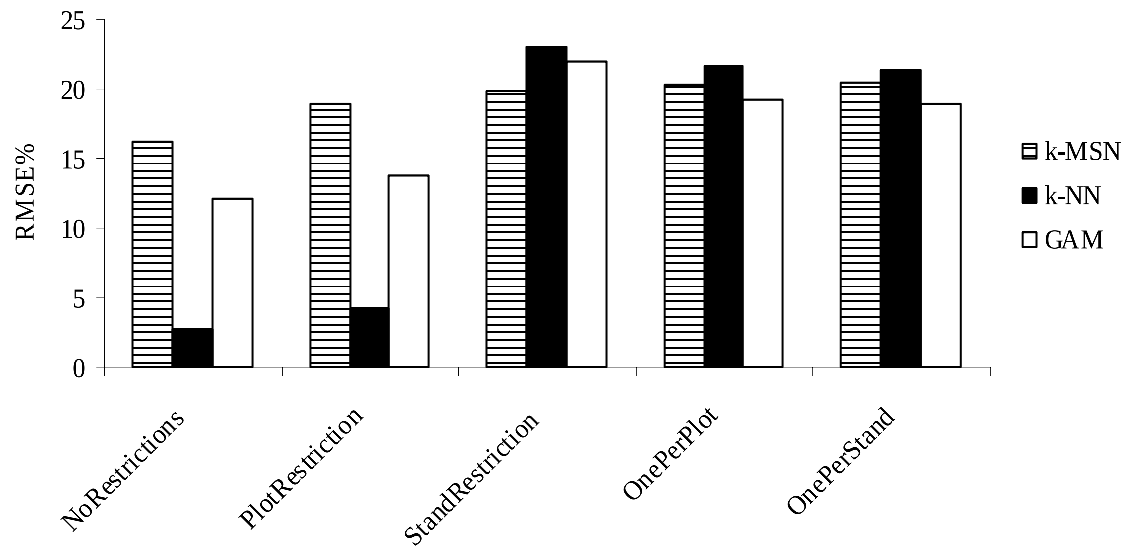


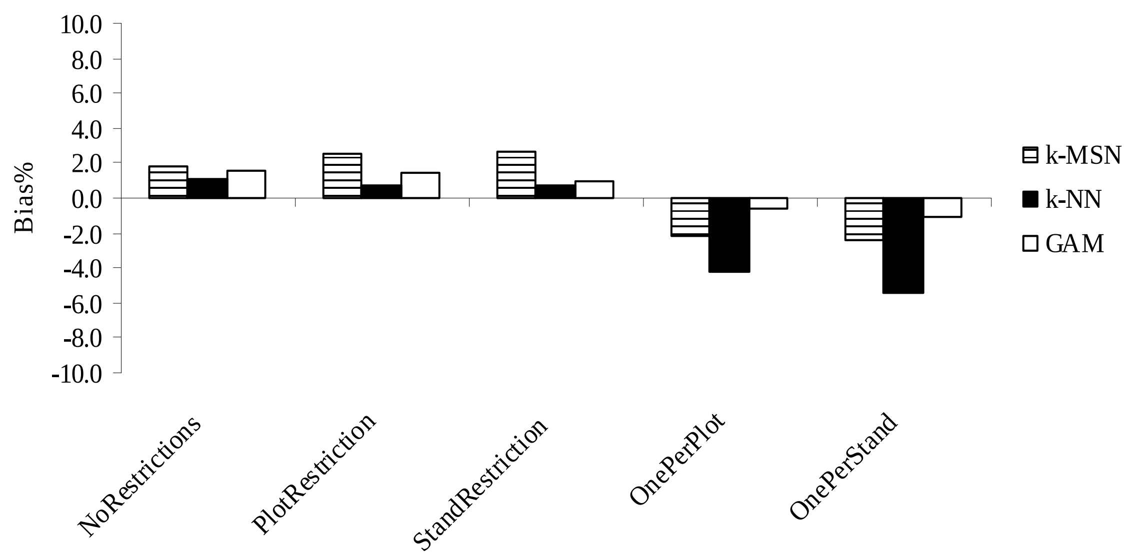


Accuracy of the tree-level results



Accuracy of the stand-level results





Comparison of different non-parametric growth imputation…

Correlative observations did not have any significant effect on the selection of the best possible neighbourhood size in any method.

Seemed to have similar effects in this and the earlier studies.

Results indicated some effects of the correlative observations:

- The correlative observations may cause that the tree-level errors are parallel and thus diminish the accuracy at the stand level
- Stand-level and regional results were not improved by including many neighbours from one stand, if it was not the target tree stand
- Restricting the amount of mutually correlated neighbours would be appropriate when considering the accuracy of stand-level or regional growth

Estimating individual tree diameter and height increment simultaneously using non-parametric imputation

Why?

- Non-parametric methods offer a possibility to easily predict diameter growth and height growth simultaneously
- To compare k-MSN imputation with parametric models also in an independent test data
- To examine forecasting of growth for a long period with k-MSN imputation and compare the results with parametric models
- To test how large the reference data should be to get accurate growth predictions locally

Estimating individual tree diameter and height increment…

Modelling data was INKA

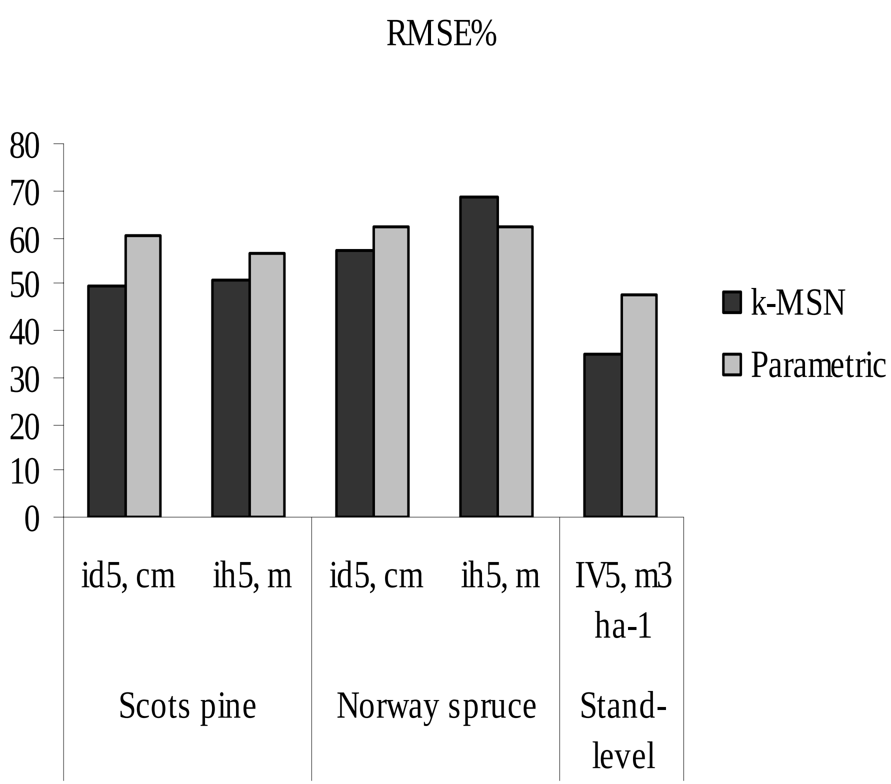
- 476 stands
- 12 438 Scots pines
- 3 692 Norway spruces

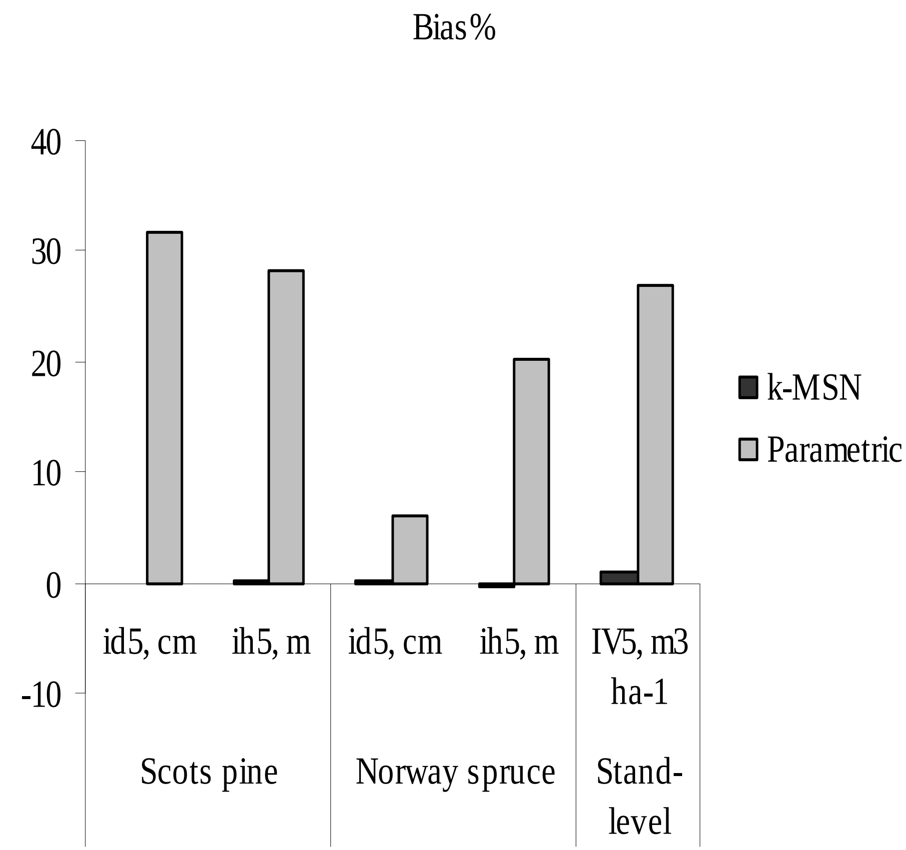
Kuusamo-data was used as a independent test data

- 71 stands
- 941 Scots pines
- 367 Norway spruces

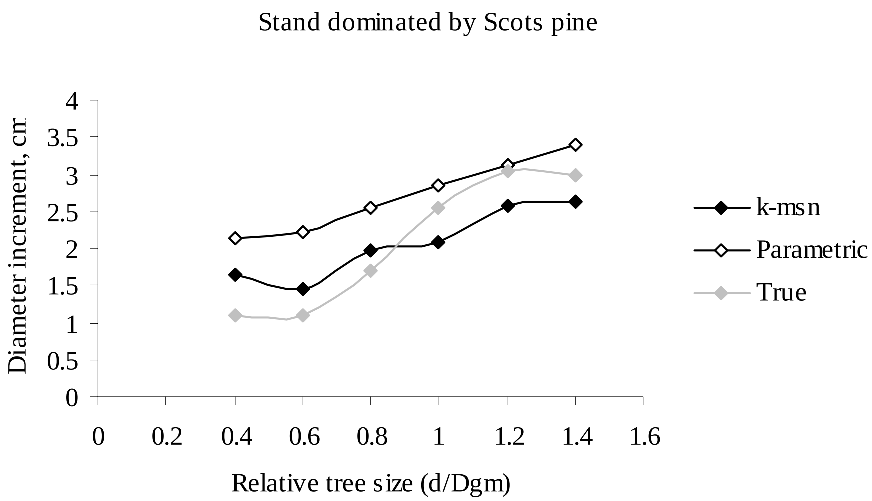
JOENSUUN YLIOPISTO

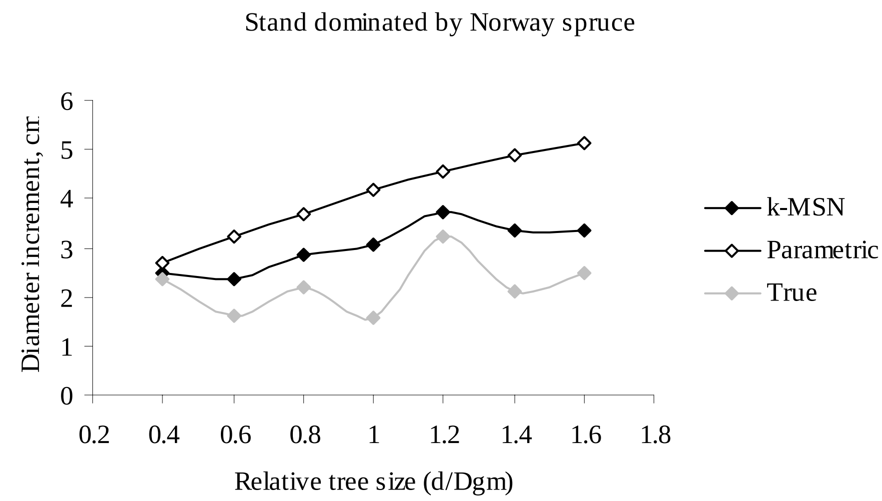
Accuracy of the methods in the modelling data

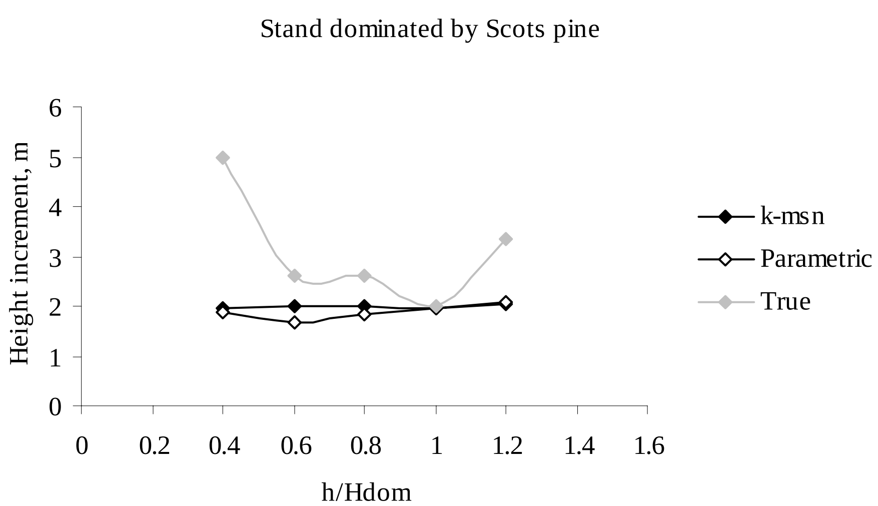


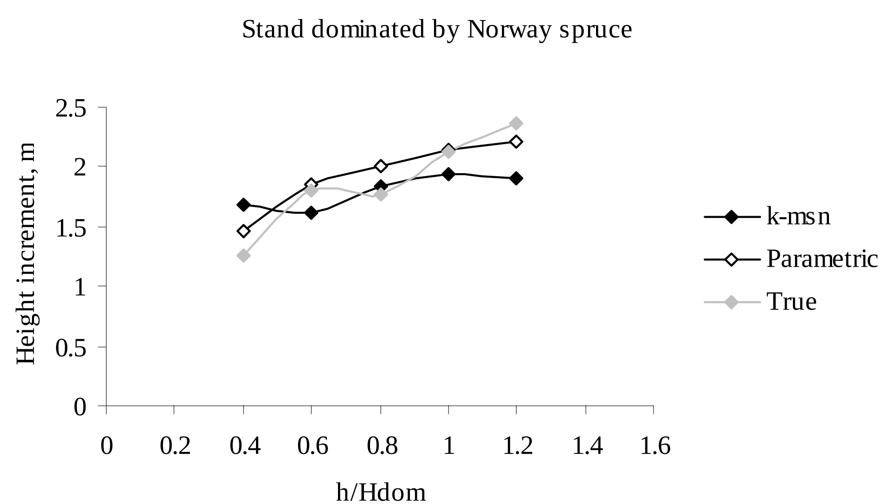


Performance of the different methods within stands



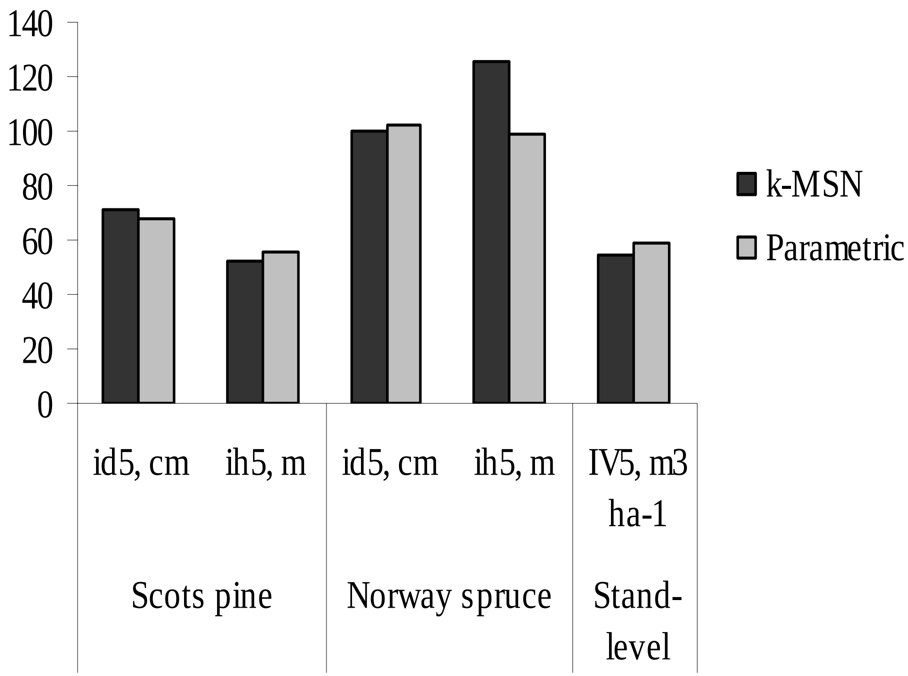




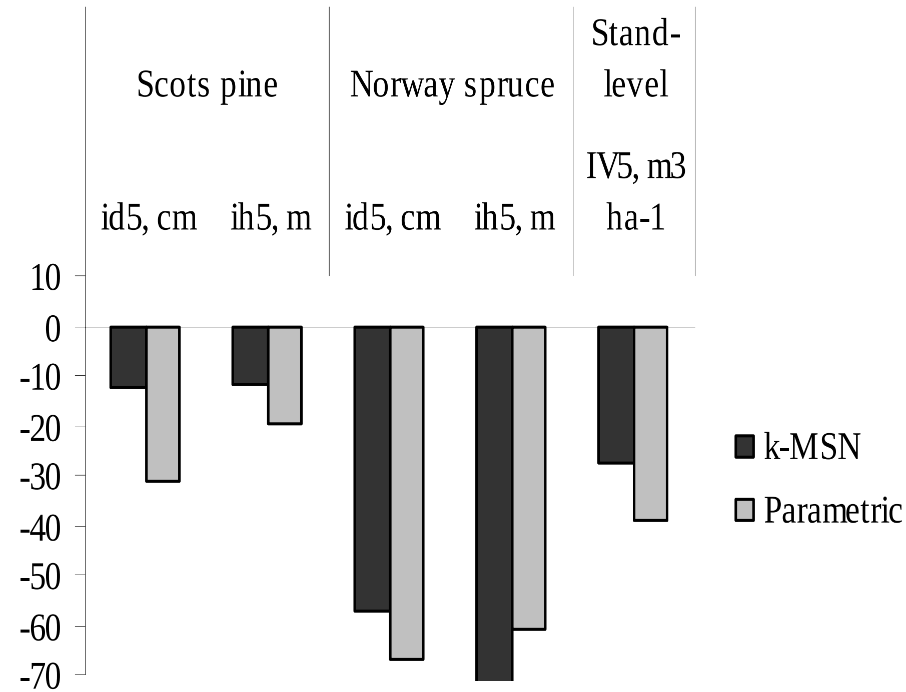


Accuracy of the methods in the Kuusamo data

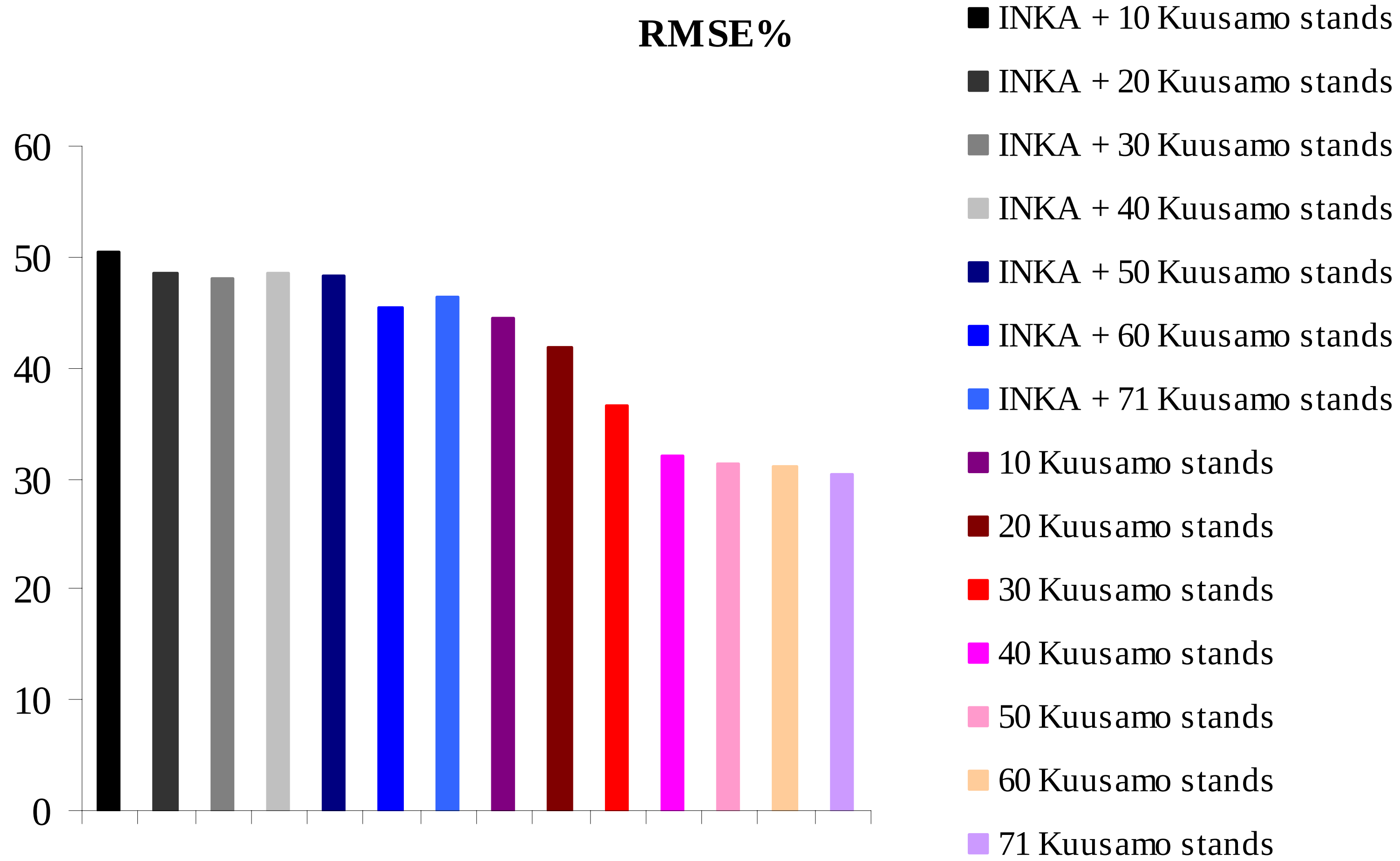
RMSE%



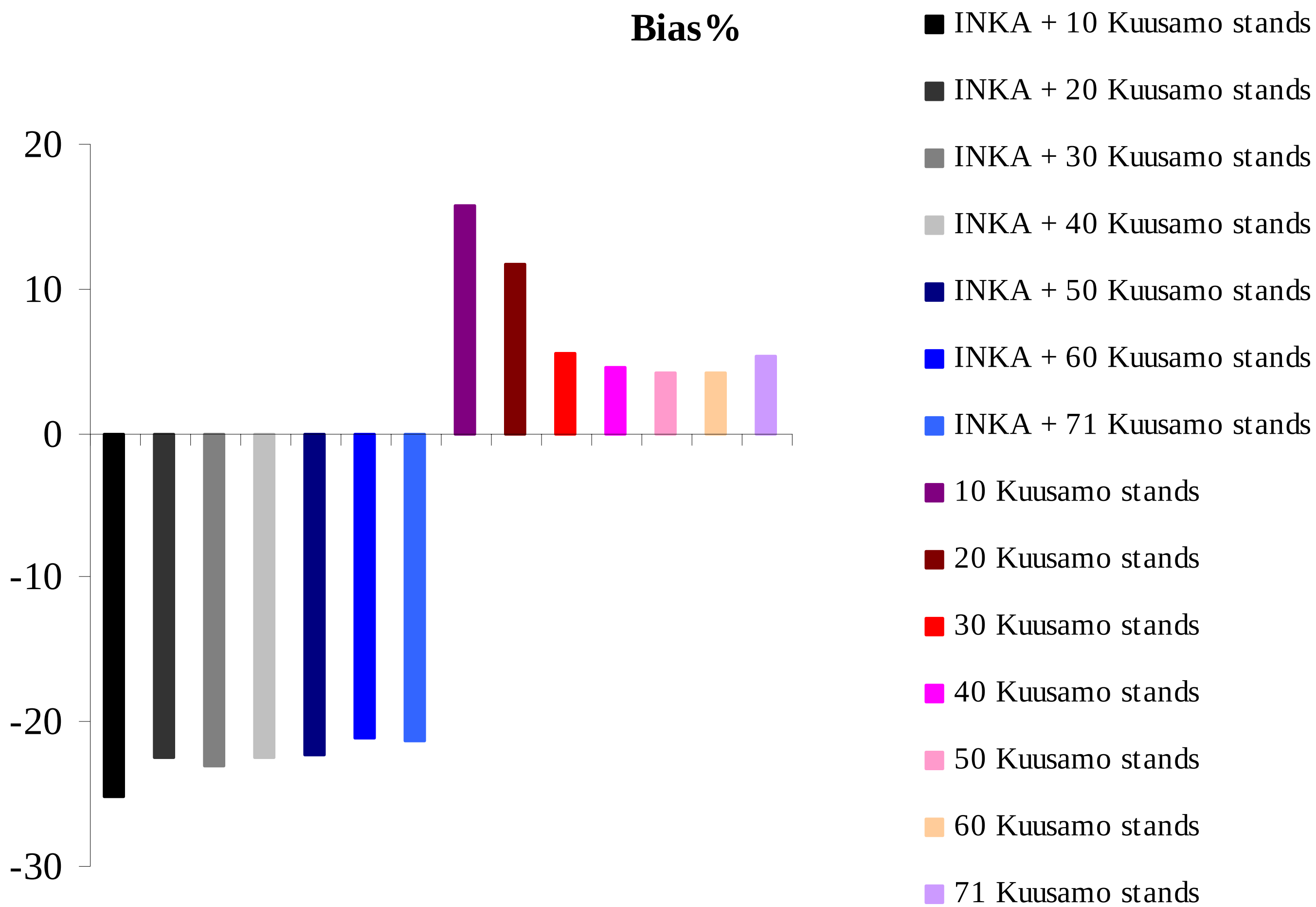
Bias%



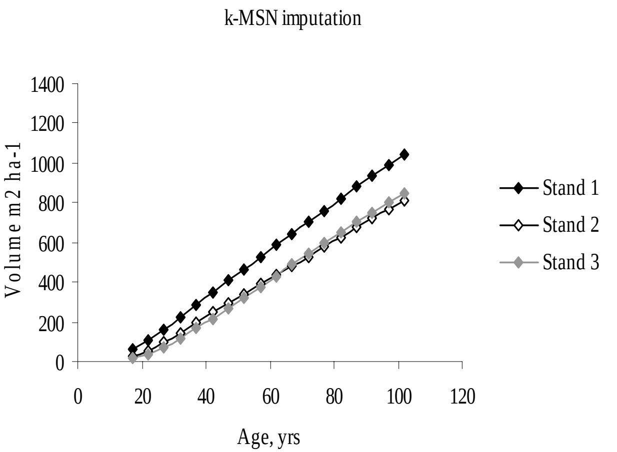
Effect of the size of local reference data on the accuracy of stand-level results

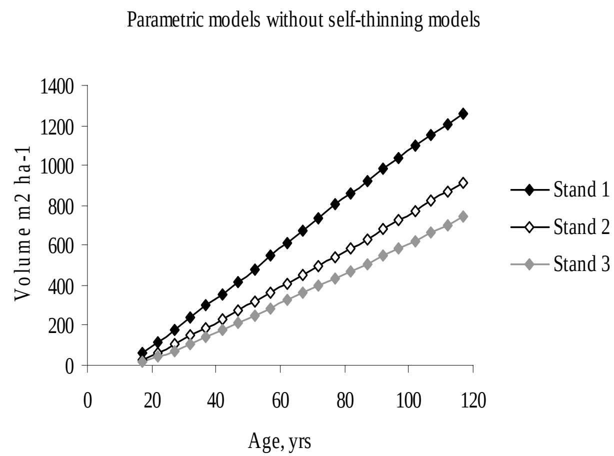


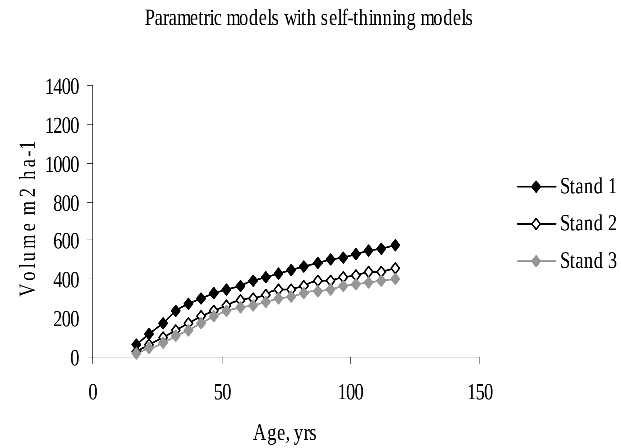
Effect of the size of local reference data on the accuracy of stand-level results

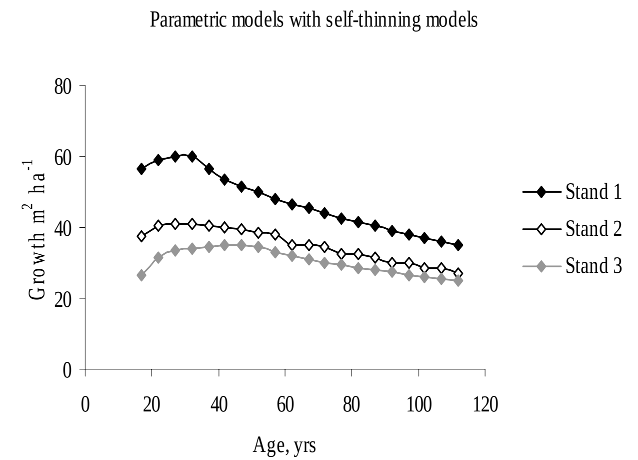


Forecasting growth with k-MSN and parametric models









Forecasting growth with k-MSN and parametric models