# Factors affecting Norway spruce and Scots pine regeneration result in private owned forests in Finland

Kankaanhuhta, Ville. Finnish forest Research Institute, Suonenjoki Research Station.

## Introduction

There has been developed a new low-cost field inventory method for private owned forests' regeneration quality control in Finland. A similar kind of method has also been applied in the forest company UPM-Kymmene's own forests and the regeneration quality improvement results have been promising. The goal of the project was to model the factors affecting forest regeneration results for Norway spruce planting and Scots pine direct seeding.

## Material and methods

### *Hierarchical levels of data*

The data consists of the inventory years 2000-2002 (Figure 1). This is only a subset of the inventory data selected from the years 2000-2005 in order to find appropriate methodology to analyze the topics in interest. The inventories were carried out in separate areas in different years, so the data is not longitudinal in classical sense. Other regeneration chains than Norway spruce planting and Scots pine direct seeding are excluded from the data. The Norway spruce planting data set consists of 1962 regeneration areas with 16 030 sample plots. The Scots pine direct seeding data set consists of 757 regeneration areas with 12 300 sample plots. Because the inventories continued in two FOAs' area in two years the year was not used as a hierarchical level. See the structure of the data in final analysis in Figure 1.

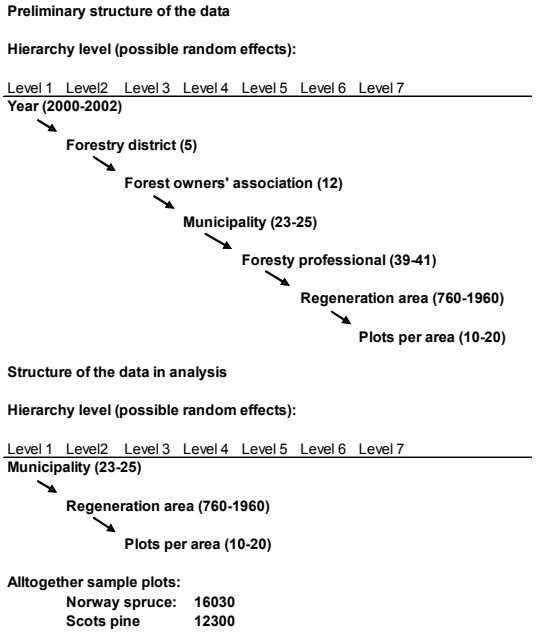


Figure 1. Structure of the inventory data.

The inventories have been carried out in five forest districts (hierarchy level 2). From one to three Forest Owners' Associations (FOAs) has volunteered from each district to the study. The amount of FOAs with reasonable land area is 12 (hierarchy level 3). The FOAs consists of one or more municipalities. For Norway spruce planting data set there are about 25 municipalities and for Scots pine direct seeding data set about 23 municipalities (hierarchy level 4). The estimation of the authors is that knowing the organizational structure and geographical conditions of the FOAs, probably the municipality level will be applied instead of FOA level. In addition to the municipality level there is forestry professionals' operational level within the municipality (hierarchy level 5). However the operational areas may overlap and thus this operational level will probably be ignored in the final models, but are anyway included in the data sets. In addition to the hierarchy levels mentioned above, the regeneration areas consist of 15-20 sample plots. Depending on the phenomenon studied and computational capacity available a regeneration area can also be handled as a hierarchical level 6.

## *Key dependent and independent variables*

The key dependent variable was the amount seedlings with growth potential (called later good seedlings) at sample plot (Poisson/normal distribution) .The following independent variables were included to the data set for both Norway spruce planting and Scots pine direct seeding:

Regeneration area level: a) who as carried out the regeneration (0,1,..3), b) land area of regeneration area (hectares) and c) seedling type used (0, 1,… 3).

Sample plot level: a) Site type (0,1,…6), b) Stoniness (0,1), c) Wetness (0,1), d) Soil type (0,1,…4), e) Soil preparation method (0,1,…5), f) Regeneration method (0,1,…6), g) Tree species (0,1,…5) h) amount of planted seedlings by tree species and naturally regenerated seedlings by tree species (data stored altogether in 11 variables), i) average height of good seedlings and j) dominant height of sprouts. The amount of seedlings acts also as dependent variables in subproject a) and mainly independent variables in subproject b).

## *Statistical methods*

The key dependent variable was count of good seedlings at the sample plot. The distribution of the dependent variable was at first supposed to be Poisson. This knowledge proved to be true for Scots pine direct seeding data set and generalized linear mixed models were applied in the analysis. However, for Norway spruce planting this knowledge changed as underdispersion of the data set was observed. The Norway spruce data set was analyzed with linear mixed models. The restricted iterative least squares (RIGLS) method was applied in the likelihood estimation for both data sets. For Norway spruce the model fit was mainly estimated applying the significant differences of deviances applying X2 distribution as well as inspection of residuals. For Scots pine extra Poisson with 2$^{nd}$ order Penalized Least squares was applied in the likelihood estimations. The model fit for Scots pine was tested comparing the differences of variances. Also Wald tests and inspection of residuals were applied.

# Results

## *Norway spruce planting*

Three level linear mixed model was created. The hierarchy levels were municipality (k), regeneration area (j) and sample plot level (i). The fixed effects of the model were the values of site preparation and which actor had carried out the regeneration job.

$$\text{good\_seedlings}_{ijk} \sim N(XB,\ \Omega)$$

$$\begin{aligned}\text{good\_seedlings}_{ijk} = {} & \beta_{0ijk}\text{cons} + 0.291(0.105)\text{carried\_out\_by\_1}_{ijk} + 0.161(0.104)\text{carried\_out\_by\_2}_{ijk} + \\ & -0.414(0.518)\text{carried\_out\_by\_3}_{jk} + 0.744(0.094)\text{site\_prep\_2}_{ijk} + \\ & 0.801(0.091)\text{site\_prep\_3}_{ijk} + 0.779(0.102)\text{site\_prep\_4}_{ijk} + \\ & 0.879(0.262)\text{site\_prep\_5}_{ijk}\end{aligned}$$

$$\beta_{0ijk} = 2.568(0.145) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\left[v_{0k}\right] \sim N(0,\ \Omega_v) \ :\ \Omega_v = \left[0.180(0.064)\right]$$

$$\left[u_{0jk}\right] \sim N(0,\ \Omega_u) \ :\ \Omega_u = \left[0.602(0.033)\right]$$

$$\left[e_{0ijk}\right] \sim N(0,\ \Omega_e) \ :\ \Omega_e = \left[1.945(0.022)\right]$$

$$-2*loglikelihood(IGLS\ Deviance) = 57951.250(16030\ \text{of}\ 16030\ \text{cases in use})$$

The variances of the various models were compared against each other. In the following there is a summary of the variables included to the model.

- 58034.750 – 57951.250 = 183.50, p=0,0000..*10-E42 (Model without Site preparation)
- 57960.130-57951.250=8.8789, p=0.0028849 (Model without the knowledge, who carried out the job).

Seedling type was dropped out of the model because of the insignificant effect on the varaince:

- 57951.250-57950.44=0.80859, p=0.36854

The residuals of the different levels (municipality and regeneration area level) were compared. There are some examples in the figures 2, 3 and 4. Some possible outlayers were detected, but they were not excluded from the data.

The municipality level had about 7 % of the whole variation and the regeneration area level 22 % respectively. The comparison of the deviances of the models with or without municipality as a random effect very statistically significant (p=0.000..10-E19).

The magnitude and interpretation of the different values of independent variables semed quite logical, although the single alues are not discussed here on more detail.

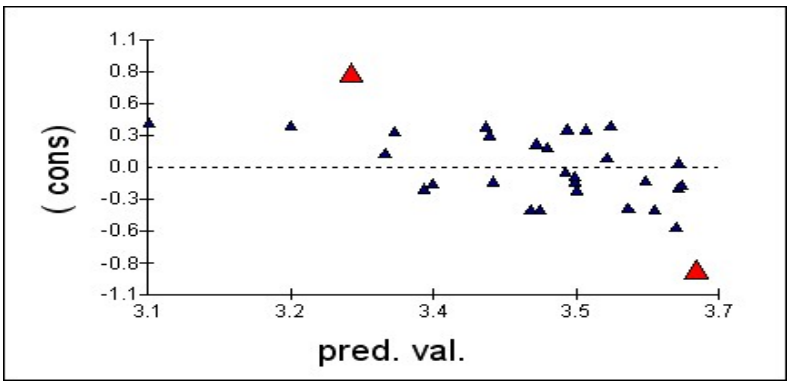


Figure 2. Norway spruce. Standardised residuals vs.fixed part prediction at the municipality level.

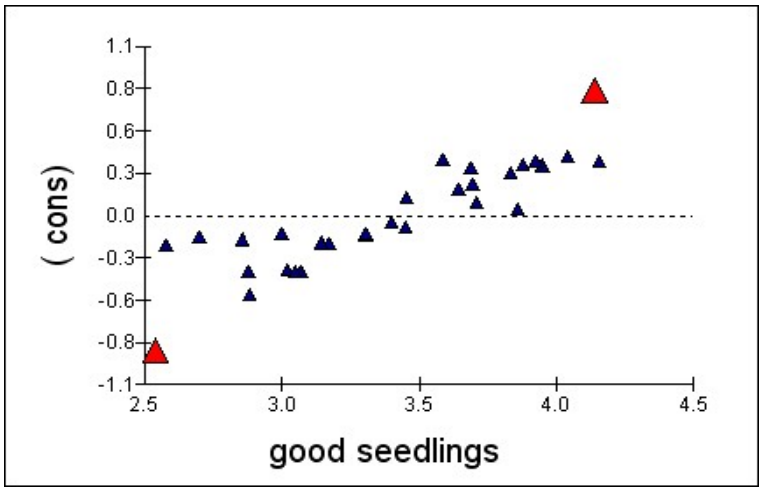


Figure 3. Norway spruce. Standardised residuals vs. good seedlings at the municipality level.

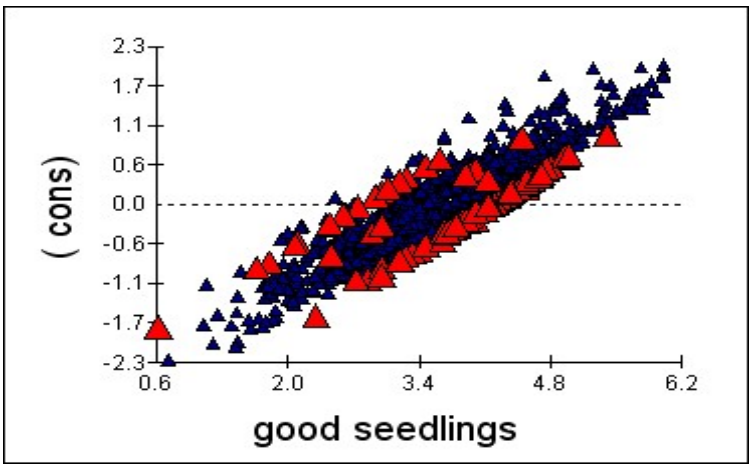


Figure 4. Norway spruce standardised residuals at the regeneration area level.

## *Scots pine direct seeding*

Three level linear mixed model was created. The hierarchy levels were municipality (k), regeneration area (j) and sample plot level (i). The fixed effects of the model were year , soil type and site type.

$$\text{good\_seedlings}_{ijk} \sim \text{Poisson}(\pi_{ijk})$$

$$\log(\pi_{ijk}) = \beta_{0jk}\text{cons}_{ijk} + -0.302(0.152)\text{year\_2001}_{jk} + -0.216(0.136)\text{year\_2002}_{jk} + 0.032(0.060)\text{soiltype\_1}_{ijk} + 0.011(0.058)\text{soiltype\_2}_{ijk} + -0.191(0.064)\text{soiltype\_3}_{ijk} + -0.302(0.066)\text{soiltype\_4}_{ijk} + 1.136(0.315)\text{sitetype\_2}_{ijk} + 1.681(0.309)\text{sitetype\_3}_{ijk} + 1.945(0.309)\text{sitetype\_4}_{ijk} + 2.079(0.311)\text{sitetype\_5}_{ijk} + 1.607(0.386)\text{sitetype\_6}_{ijk}$$

$$\beta_{0jk} = 0.155(0.336) + v_{0k} + u_{0jk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim \text{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} 0.031(0.013) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.242(0.014) \end{bmatrix}$$

$$\text{var}(\text{good\_seedlings}_{ijk} | \pi_{ijk}) = 2.786(0.037)\pi_{ijk}$$

The variance and its' standard deviation of different model types were compared. In the following there are the variances of the models with more overdispersion.

| Independent variables Included to the model | Variance(Std. Dev.) |
|---|---|
| Year + soiltype | 2.841 (0.037) |
| Year + site type | 2.812 (0.037) |
| Year + soil type + site type | 2.786 (0.037) |

The residuals of the different levels (municipality and regeneration area level) were compared. There are some examples in the figures 5, 6 and 7. Some possible outlayers were detected, but they were not excluded from the data. However, there will be a need to make further logical checks for the data in the future.

The municipality level had about 11 % of the whole variation.

The effect of different values of the various independent variables were quite logical to the previous kkowledge of their distributions.

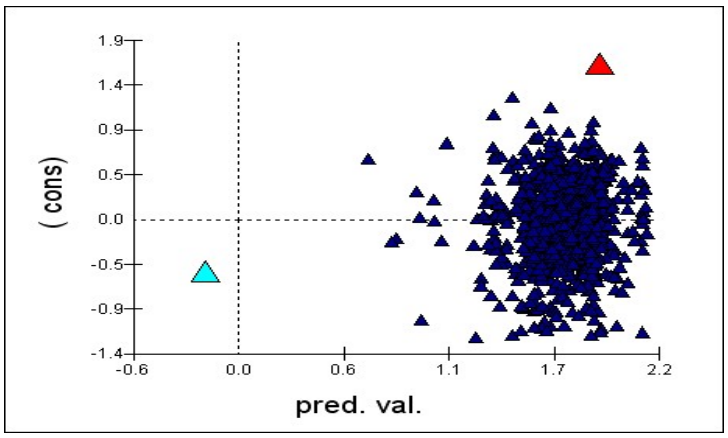


Figure 5. Scots pine direct seeding. Regeneration area level. Standardised residuals vs. fixed part prediction.

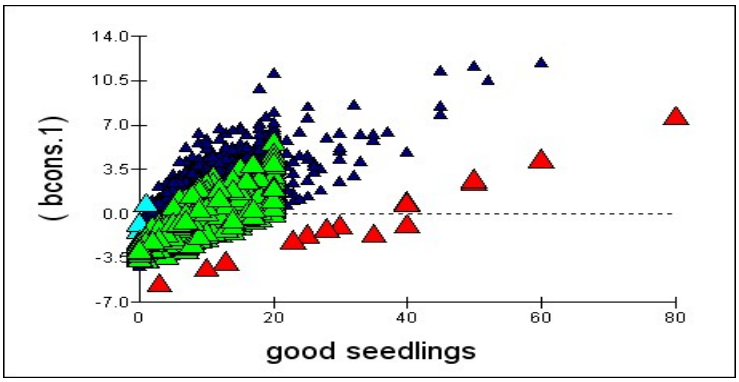


Figure 6. Scots pine direct seeding. Sample plot level. Standarsed residuals vs.good seedlings.

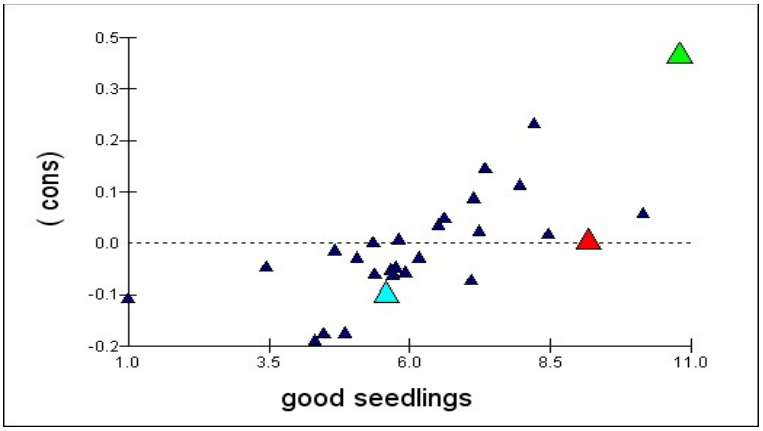


Figure 7. Scots pine direct seeding. Municipality level. Standardised residuals vs. good seedlings.

## Discussion

The purpose of the study is to develop analysis methodology applicable to the silvicultural process quality control described above. According to the experience from the project the linear mixed models and generalised mixed models seem to be the most obvious solution.

The variation between municipalities proved to be significant as noticed before. The magnitudes of the various independent variables were not prsented more thoroughly, but their effect proved to be similar to many previous studies.

Because of the two relatively raw data sets and lack of prior knowledge of the real distributions of the phenomenon under study the Bayesian MCMC based methods were only tested, not included to the final results. The applicability of the above mentioned methods are considered in the future as need for more complex data structures, i.e. nesting and cross-classification arise as well as if there is need for imputational methodology with smaller sub-datasets of the future projects.

## Acknowledgments

The author is most grateful for Henrik, William and Rasmus of excellent lectures and great assistance as well as for all the participants of the course for creating great positive learning experience. The author is also most thankful for the Finnish Forest Research Institute (Timo Saksa) for providing the data for the project work.