Factors affecting need for early non-commercial thinning in private owned forests in Finland

Oss Janis.

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 the need for early non-commercial thinning observed in the inventories with the regeneration chains mentioned above (main responsible author Janis Oss).

Description of the data

Hierarchical levels

The data consists of the inventory years 2000-2002 (See hierarchy level 1, in 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 are carried out in separate areas in different years, so the data is not longitudinal in classical sense. The Norway spruce planting data set consists of 1962 regeneration areas with 16 030 sample plots.

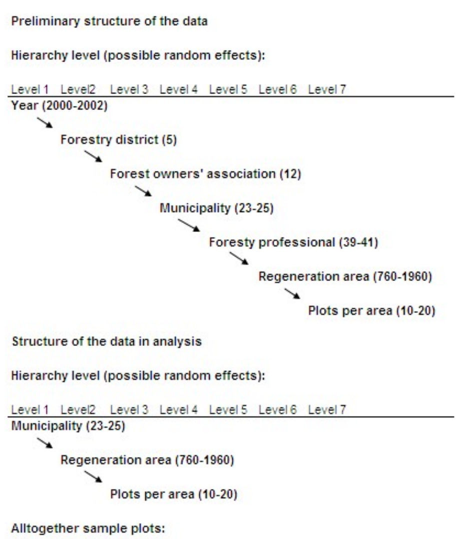


Figure 1. Structure of the data.

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 variables

Subproject:

The key dependent variable is the observed need for early non-commercial thinning. The need is coded as follows: 0 = no need for thinning, 1 = need for thinning in 3-5 years. The model is planned to be developed in polytomous logistic regression applying the comparisons/differences of the pairs 0 – 1 (Bernoulli distribution).

Key independent variables

The following independent variables will be included:

Sample plot level: a) Site type (0,1,…4), b) Stoniness (0,1), c) Wetness (0,1), d) Soil type (0,1,…4), e) i) average of good seedlings and mainly independent variables in project.

Expectations of the results

Is fairly new and undiscovered area to be studied with this kind of inventory data? There are hypothesis and studies of the effect of single dependent variables also included in the inventory data, but the most interesting point waiting for answer is to which extent we are able to estimate the need for early non-commercial thinning in the regeneration area inventories at the age of 3-5? What are the best factors and interactions estimating the need for early non-commercial thinning?

Methodology

The binomial distribution is used when there are exactly two mutually exclusive outcomes of a trial. These outcomes are appropriately labeled "success" and "failure". The binomial distribution is used to obtain the probability of observing x successes in N trials, with the probability of success on a single trial denoted by p. The binomial distribution assumes that p is fixed for all trials.

Logistic regression model are more complicated to fit then normal model using the IGLS estimation method in MLwiN

Event of work

In this project the depended variable has with binomial structure (new_treat). The variable shows, this true or not (0/1) to make treat after 2-5 years.

Graphic (histogram)*: new_treat1*

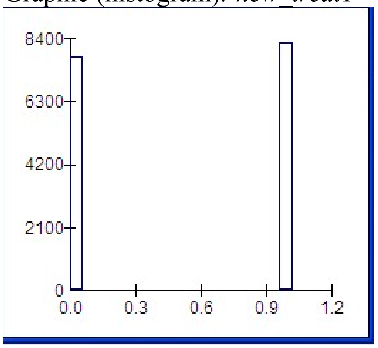


Figure 2. Histogram of variable new_treat1

Simple Logistic Regression Model

This is fixed effects simple logistic regression models analogous to, for the normal responses, linear regression models.

I will start by considering a simple logistic regression model accounting for the bad seeding total of the municipality only.

The model has build with as follows facts:

- response variable - *new_treat (binomial);*
- number of levels: *3;*
- level identifiers: *municipality, reg_area_id, plot_id;*
- distributional assumption: *Binomial;*
- predictor variable: *bad_total;*

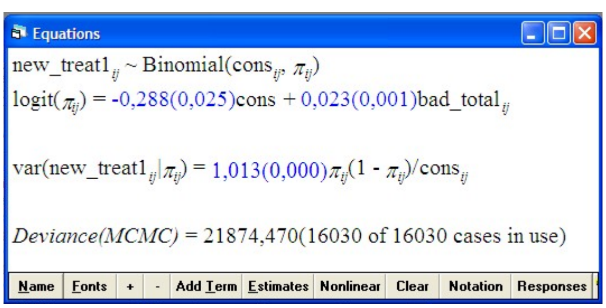


Figure 3. Simple logistic regression model accounting for the bed_seeding total of the municipality.

Here we see that the intercept term is -0.288, which, corresponds to the average plot. Transform into probability scale *CALC B1=ALOG(-0.288)*. Probabilities of is new treat of 0.428.

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC
22210.66	22209.66	0.99	22211.65 (Model without bad_total)
21874.47	21872.46	2.01	21876.49 (Model with bad_total)

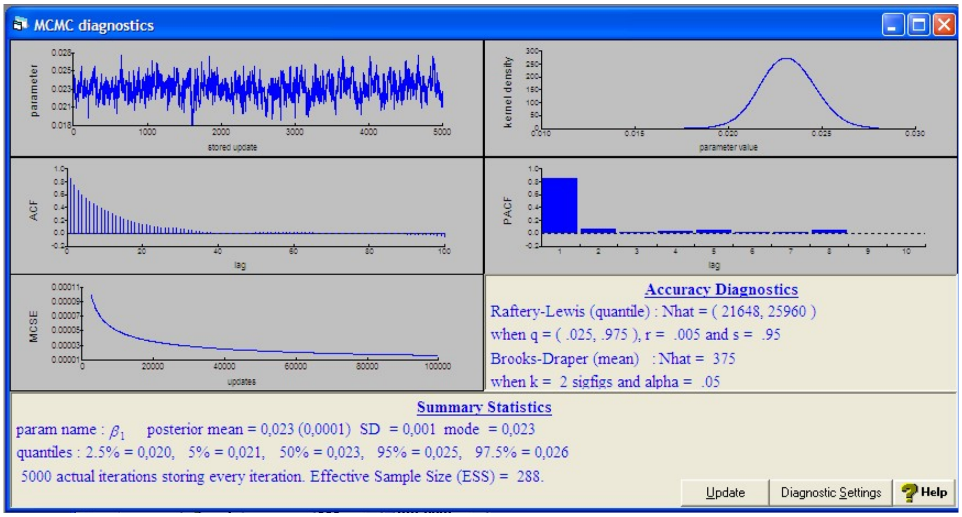


Figure 4. The diagnostics for the "bad_total" effect

Model run for 15,000 iterations.

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC
21874.49	21872.46	2.03	21876.52

Consider a simple logistic regression model accounting for the site type total of the municipality only.

The model has build with as follows facts:

- response variable - *new_treat (binomial);*
- number of levels: *3;*
- level identifiers: *municipality, reg_area_id, plot_id;*
- distributional assumption: *Binomial;*
- predictor variable: *sitetype_1...4;*

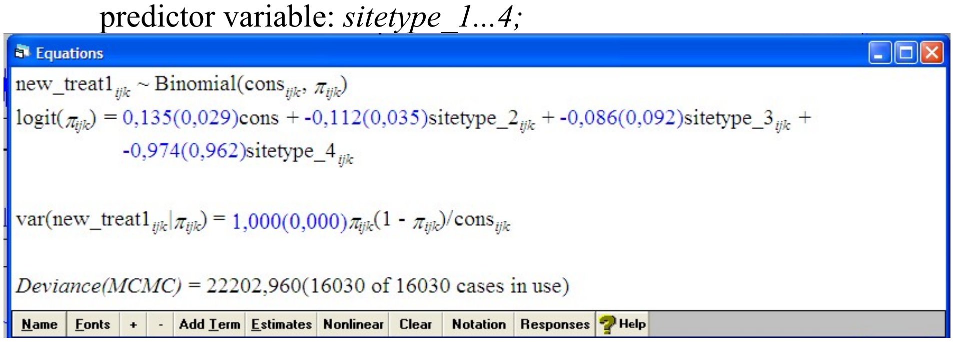


Figure 5. Simple logistic regression model accounting for the sitetype_1...4 total of the municipality.

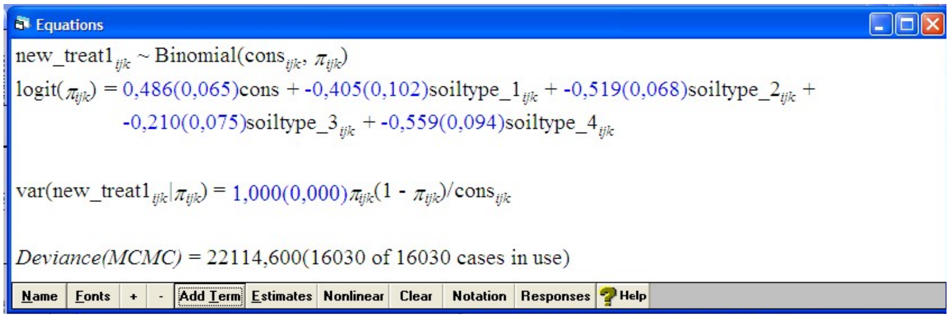
Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC	
22202.96	22198.80	4.15	22207.11	(model without bad_total)
21859.12	21854.32	4.81	21863.93	(Model with bad_total)

Consider a simple logistic regression model accounting for the soil type total of the municipality only.

The model has build with as follows facts:

response variable - *new_treat (binomial);*
number of levels: *3;*
level identifiers: *municipality, reg_area_id, plot_id;*
distributional assumption: *Binomial;*
predictor variable: *sitetype_1...4;*



$$\text{new\_treat1}_{ijk} \sim \text{Binomial}(\text{cons}_{ijk}, \pi_{ijk})$$
$$\text{logit}(\pi_{ijk}) = 0{,}486(0{,}065)\text{cons} + -0{,}405(0{,}102)\text{soiltype\_1}_{ijk} + -0{,}519(0{,}068)\text{soiltype\_2}_{ijk} + -0{,}210(0{,}075)\text{soiltype\_3}_{ijk} + -0{,}559(0{,}094)\text{soiltype\_4}_{ijk}$$
$$\text{var}(\text{new\_treat1}_{ijk}|\pi_{ijk}) = 1{,}000(0{,}000)\pi_{ijk}(1 - \pi_{ijk})/\text{cons}_{ijk}$$

Deviance(MCMC) = 22114,600(16030 of 16030 cases in use)

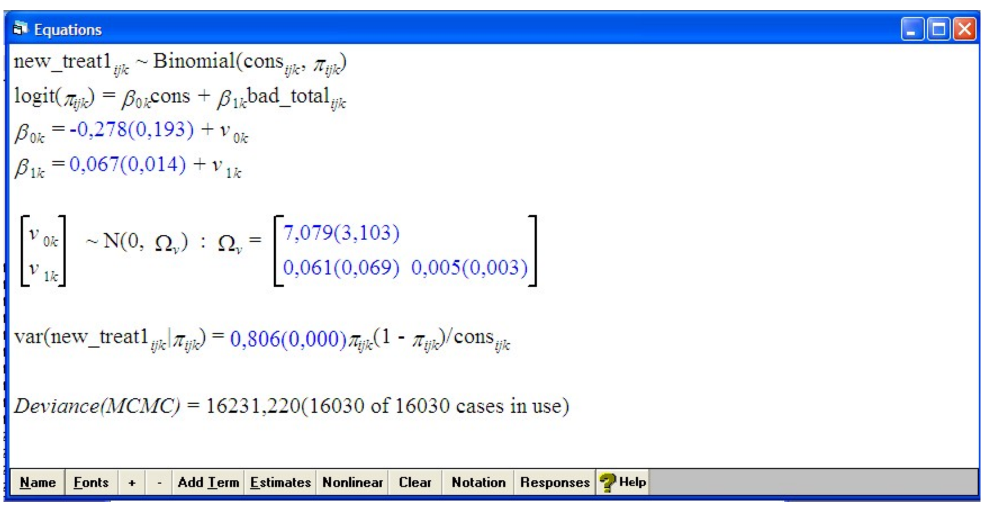
Figure 6. Simple logistic regression model accounting for the soiltype_1...4 total of the municipality.

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC
22114.60	22109.57	5.02	22119.62 (Model without bad_total)
21807.54	21801.60	5.94	21813.48 (Model with bad_total)

Random effects logistic regression model

We can add random effect to model to account for different probabilities of new treat for regeneration area in where they are.



$$\text{new\_treat1}_{ijk} \sim \text{Binomial}(\text{cons}_{ijk}, \pi_{ijk})$$
$$\text{logit}(\pi_{ijk}) = \beta_{0k}\text{cons} + \beta_{1k}\text{bad\_total}_{ijk}$$
$$\beta_{0k} = -0{,}278(0{,}193) + v_{0k}$$
$$\beta_{1k} = 0{,}067(0{,}014) + v_{1k}$$
$$\begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 7{,}079(3{,}103) & \\ 0{,}061(0{,}069) & 0{,}005(0{,}003) \end{bmatrix}$$
$$\text{var}(\text{new\_treat1}_{ijk}|\pi_{ijk}) = 0{,}806(0{,}000)\pi_{ijk}(1 - \pi_{ijk})/\text{cons}_{ijk}$$

Deviance(MCMC) = 16231,220(16030 of 16030 cases in use)

Figure 7. Random effect in the use and bad_total

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC
16231.22	16185.31	45.91	16277.13 (Model with random effect)
22210.66	22209.66	0.99	22211.65 (Model without random effect)
16702.57	16674.33	28.24	16730.81 (Model with random effect on use)

17423.17 17395.71 27.46 17450.62 (Model with random effect on bad_total)

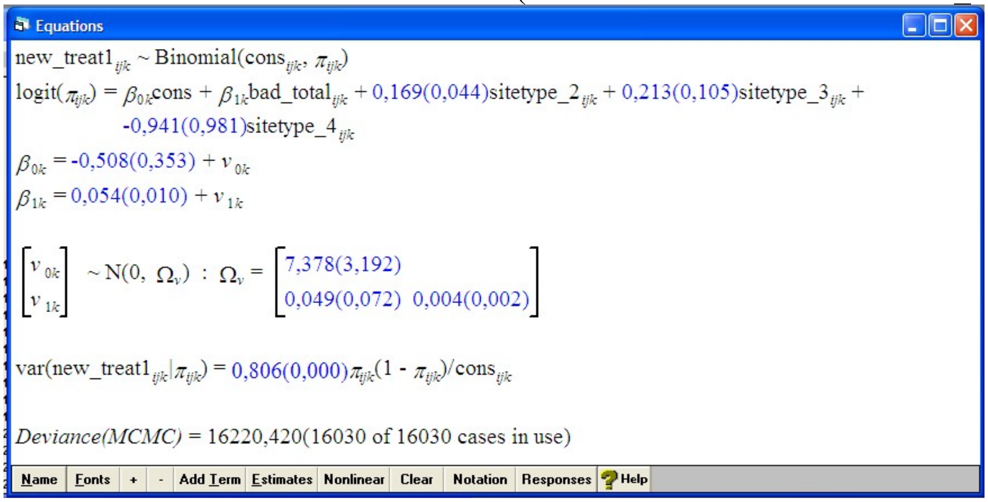


Figure 8. Site type and random effect on the use and bad_total

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC	
16220.42	16170.48	49.94	16270.37	(Model with sitetype1...4 and random effect on the use and bad_total)
21859.46	21854.32	5.14	21864.60	(Model with sitetype1...4 and without random effect on the use and bad_total)
16194.08	16144.22	49.86	16243.94	(Model with soiltype1...4 and random effect on the use and bad_total)
21808.02	21801.57	6.45	21814.46	(Model with soiltype1...4 and without random effect on the use and bad_total)

Conclusions

In this project I analyzed several models to get the one of the best. From the result I conclude:

1. Difference in new_trea between 0 (not necessary) and 1 (necessary in 2-5 years) are minimal - 450 from 16030 observations.
2. Difference between Simple Logistic regression and models with random effect are:

Bayesian Deviance Information Criterion (DIC)

Dbar	D(thetabar)	pD	DIC	
16231.22	16185.31	45.91	16277.13	(Model with random effect)
22210.66	22209.66	0.99	22211.65	(Model without random effect)

The better model is when using random effects.

3. Differences between models where use sitetyp un soiltype are minimal

Dbar	D(thetabar)	pD	DIC	
16220.42	16170.48	49.94	16270.37	(Model with sitetyp)
16194.08	16144.22	49.86	16243.94	(Model with soiltyp)