Application of mixed-effect models in forestry

Lauri Mehtätalo[1]

[1] University of Eastern Finland, School of Computing & School of Forest Sciences

May 31, 2013

Types of forest datasets

- Forest datasets are usually hierarchical e.g.
  - needles within branches
  - branches within trees
  - **trees within sample plots or aerial images**
  - sample plots within forest stands
  - forest stand within regions
  - **repeated observations of trees in successive years or on different images**
  - ...

Types of forest datasets

- Forest datasets are usually hierarchical e.g.
  - needles within branches
  - branches within trees
  - **trees within sample plots or aerial images**
  - sample plots within forest stands
  - forest stand within regions
  - **repeated observations of trees in successive years or on different images**
  - ...
- Also crossed grouping structures are common
  - Tree increments for different calendar years
  - Trees or forest stands on aerial images

Types of forest datasets

- Forest datasets are usually hierarchical e.g.
  - needles within branches
  - branches within trees
  - **trees within sample plots or aerial images**
  - sample plots within forest stands
  - forest stand within regions
  - **repeated observations of trees in successive years or on different images**
  - ...
- Also crossed grouping structures are common
  - Tree increments for different calendar years
  - Trees or forest stands on aerial images

These datasets are naturally modeled using random effect models.

Why random effects?

- Using mixed-effects models with hierarchical datasets result in
  1. More reliable inference on the model parameters
  2. Possibility to compute the predictions at different levels of the dataset.
  3. Estimates of covarainces between observations

Why random effects?

- Using mixed-effects models with hierarchical datasets result in
    1. More reliable inference on the model parameters
    2. Possibility to compute the predictions at different levels of the dataset.
    3. Estimates of covarainces between observations
- If the main interest is the inference (e.g. the effects of certain medical treatments on individuals) the first property is more important.

Why random effects?

- Using mixed-effects models with hierarchical datasets result in
  1. More reliable inference on the model parameters
  2. Possibility to compute the predictions at different levels of the dataset.
  3. Estimates of covaraince between observations
- If the main interest is the inference (e.g. the effects of certain medical treatments on individuals) the first property is more important.
- If the main interest is prediction, then greatest benefit may arise from the possibility to make predictions at different levels of hierarchy.

Why random effects?

- Using mixed-effects models with hierarchical datasets result in
  1. More reliable inference on the model parameters
  2. Possibility to compute the predictions at different levels of the dataset.
  3. Estimates of covarainces between observations
- If the main interest is the inference (e.g. the effects of certain medical treatments on individuals) the first property is more important.
- If the main interest is prediction, then greatest benefit may arise from the possibility to make predictions at different levels of hierarchy.
- The prediction is possible also for groups from outside the modeling data either (i) using the fixed part of the model or (ii) using predicted random effects based on some measurement data from the group. Even one observation is enough.

Topic of this presentation

I will first present a simple linear mixed-effects model and some extensions of it. Thereafter, I will demonstrate and discuss the use of mixed-effects models in four different forestry situations. The main benefit of mixed-effects models arsing either from prediction (P), inference (I) or estimated covariances (C).

- Using a previously fitted linear mixed-effects model for tree height prediction (P)

Topic of this presentation

I will first present a simple linear mixed-effects model and some extensions of it. Thereafter, I will demonstrate and discuss the use of mixed-effects models in four different forestry situations. The main benefit of mixed-effects models arsing either from prediction (P), inference (I) or estimated covariances (C).

- Using a previously fitted linear mixed-effects model for tree height prediction (P)
- Using a linear mixed-effect model with crossed grouping structure to predict a treatment-free response in a dataset of a thinning experiment (P).

Topic of this presentation

I will first present a simple linear mixed-effects model and some extensions of it. Thereafter, I will demonstrate and discuss the use of mixed-effects models in four different forestry situations. The main benefit of mixed-effects models arsing either from prediction (P), inference (I) or estimated covariances (C).

- Using a previously fitted linear mixed-effects model for tree height prediction (P)
- Using a linear mixed-effect model with crossed grouping structure to predict a treatment-free response in a dataset of a thinning experiment (P).
- Using nonlinear mixed-effect-models to analyse the previously extracted tree-level thinning effects (I)

Topic of this presentation

I will first present a simple linear mixed-effects model and some extensions of it. Thereafter, I will demonstrate and discuss the use of mixed-effects models in four different forestry situations. The main benefit of mixed-effects models arsing either from prediction (P), inference (I) or estimated covariances (C).

- Using a previously fitted linear mixed-effects model for tree height prediction (P)
- Using a linear mixed-effect model with crossed grouping structure to predict a treatment-free response in a dataset of a thinning experiment (P).
- Using nonlinear mixed-effect-models to analyse the previously extracted tree-level thinning effects (I)
- Using a multivariate linear mixed-effects model system with crossed grouping structure to estimate the variance-covariance structure of repeated aerial observations of tree reflectance to aid in species classification (C).

Simple linear mixed-effects model (LMM)

Let y_{ki} be the observed response for individual i in group k, and let x_{ki} be a fixed predictor. In a linear mixed-effects model, one may have both fixed (population level) parameters and random parameters, e.g.,

$$y_{ki} = a + bx_{ki} + \alpha_k + \epsilon_{ki} \,,$$

where we usually assume that $\alpha_k \sim N(0, \sigma_k^2)$ and $\epsilon_{ki} \sim N(0, \sigma^2)$. a and b are the fixed parameters.

- The model allows population level predictions $\widetilde{y} = \widehat{a} + \widehat{b}x_{ki}$,

Simple linear mixed-effects model (LMM)

Let y_{ki} be the observed response for individual i in group k, and let x_{ki} be a fixed predictor. In a linear mixed-effects model, one may have both fixed (population level) parameters and random parameters, e.g.,

$$y_{ki} = a + bx_{ki} + \alpha_k + \epsilon_{ki} ,$$

where we usually assume that $\alpha_k \sim N(0, \sigma_k^2)$ and $\epsilon_{ki} \sim N(0, \sigma^2)$. a and b are the fixed parameters.

- The model allows population level predictions $\widetilde{y} = \widehat{a} + \widehat{b}x_{ki}$,
- group-level predictions $\widetilde{y}_k = \widehat{a} + \widetilde{\alpha}_k + \widehat{b}x_{ki}$, where $\widetilde{\alpha}_k$ is the predited random effect (BLUP),

Simple linear mixed-effects model (LMM)

Let y_{ki} be the observed response for individual i in group k, and let x_{ki} be a fixed predictor. In a linear mixed-effects model, one may have both fixed (population level) parameters and random parameters, e.g.,

$$y_{ki} = a + bx_{ki} + \alpha_k + \epsilon_{ki} ,$$

where we usually assume that $\alpha_k \sim N(0, \sigma_k^2)$ and $\epsilon_{ki} \sim N(0, \sigma^2)$. a and b are the fixed parameters.

- The model allows population level predictions $\widetilde{y} = \widehat{a} + \widehat{b}x_{ki}$,
- group-level predictions $\widetilde{y}_k = \widehat{a} + \widetilde{\alpha}_k + \widehat{b}x_{ki}$, where $\widetilde{\alpha}_k$ is the predited random effect (BLUP),
- and corresponding residuals

Extensions of the simple LMM

- One may have also random slope

$$y_{ki} = a + bx_{ki} + \alpha_k + \beta_k x_{ki} + \epsilon_{ki}$$

where $(\alpha_k, \beta_k)' \sim MVN(0, \boldsymbol{D})$.

Extensions of the simple LMM

- One may have also random slope

$$y_{ki} = a + bx_{ki} + \alpha_k + \beta_k x_{ki} + \epsilon_{ki}$$

  where $(\alpha_k, \beta_k)' \sim MVN(0, \boldsymbol{D})$.
- For two nested groups, one may specify

$$y_{kti} = f(x_{kti}, \boldsymbol{b}) + \alpha_k + \alpha_{kt} + \epsilon_{kti}$$

  with $\alpha_k \sim N(0, \sigma_k^2)$ and $\alpha_{kt} \sim N(0, \sigma_{kt}^2)$

Extensions of the simple LMM

- One may have also random slope

$$y_{ki} = a + bx_{ki} + \alpha_k + \beta_k x_{ki} + \epsilon_{ki}$$

  where $(\alpha_k, \beta_k)' \sim MVN(0, \boldsymbol{D})$.
- For two nested groups, one may specify

$$y_{kti} = f(x_{kti}, \boldsymbol{b}) + \alpha_k + \alpha_{kt} + \epsilon_{kti}$$

  with $\alpha_k \sim N(0, \sigma_k^2)$ and $\alpha_{kt} \sim N(0, \sigma_{kt}^2)$
- For data with two crossed goups, one may specify

$$y_{kt} = f(x_{kt}; \boldsymbol{b}) + \alpha_k + \alpha_t + \epsilon_{kt} \, ,$$

  with $\alpha_k \sim N(0, \sigma_k^2)$ and $\alpha_t \sim N(0, \sigma_t^2)$

Extension of simple LMM (continued)

- For nonlinear responses one may specify

$$y_{ki} = f(x_{ki}; \boldsymbol{B}_{ki}) + \epsilon_{ki} \ ,$$

where

$$\boldsymbol{B}_{ki} = \boldsymbol{X}_{ki}\boldsymbol{b} + \boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$$

specify the parameters of the nonlinear function using fixed part $\boldsymbol{X}_{ki}\boldsymbol{b}$ and random part $\boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$; $\boldsymbol{\beta}_k \sim MVN(0, \boldsymbol{D})$.

Extension of simple LMM (continued)

- For nonlinear responses one may specify

$$y_{ki} = f(x_{ki}; \boldsymbol{B}_{ki}) + \epsilon_{ki} \ ,$$

where

$$\boldsymbol{B}_{ki} = \boldsymbol{X}_{ki}\boldsymbol{b} + \boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$$

specify the parameters of the nonlinear function using fixed part $\boldsymbol{X}_{ki}\boldsymbol{b}$ and random part $\boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$; $\boldsymbol{\beta}_k \sim MVN(0, \boldsymbol{D})$.

- A bivariate LMM may be specified by

$$\begin{aligned} y1_{ki} &= f1(x_{ki}; \boldsymbol{b}1) + \alpha1_k + \epsilon1_{ki} \\ y2_{ki} &= f2(x_{ki}; \boldsymbol{b}2) + \alpha2_k + \epsilon2_{ki} \end{aligned}$$

where $(\alpha1_k, \alpha2_k)' \sim N(0, \boldsymbol{D})$ and $(\epsilon1_k, \epsilon2_k)' \sim N(0, \boldsymbol{R})$.

Extension of simple LMM (continued)

- For nonlinear responses one may specify

$$y_{ki} = f(x_{ki}; \boldsymbol{B}_{ki}) + \epsilon_{ki} \,,$$

  where

$$\boldsymbol{B}_{ki} = \boldsymbol{X}_{ki}\boldsymbol{b} + \boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$$

  specify the parameters of the nonlinear function using fixed part $\boldsymbol{X}_{ki}\boldsymbol{b}$ and random part $\boldsymbol{Z}_{ki}\boldsymbol{\beta}_k$; $\boldsymbol{\beta}_k \sim MVN(0, \boldsymbol{D})$.
- A bivariate LMM may be specified by

$$\begin{aligned} y1_{ki} &= f1(x_{ki}; \boldsymbol{b}1) + \alpha1_k + \epsilon1_{ki} \\ y2_{ki} &= f2(x_{ki}; \boldsymbol{b}2) + \alpha2_k + \epsilon2_{ki} \end{aligned}$$

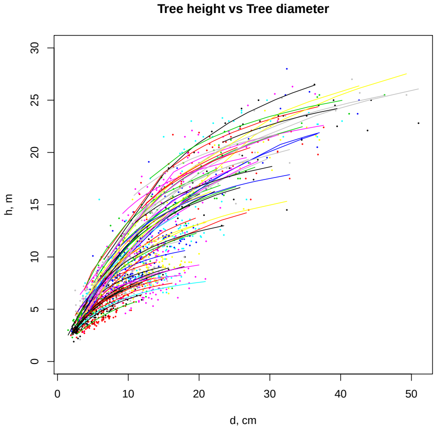
  where $(\alpha1_k, \alpha2_k)' \sim N(0, \boldsymbol{D})$ and $(\epsilon1_k, \epsilon2_k)' \sim N(0, \boldsymbol{R})$.
- Combinations are also possible, but fitting algorithms are not necessarily available.

Case 1: Prediction of tree height on diameter
Utilizing a prediction from a linear mixed-effects model with two nested levels of grouping

- Lappi, J. 1997. A longitudinal analysis of height/diameter curves. For. Sci. 43. 555–570.
- Mehtätalo, L. 2004. A longitudinal height-diameter model for Norway spruce in Finland. Can. J. For. Res. 34(1): 131-140.
- Mehtätalo, L. 2005a. Height-diameter models for Scots pine and birch in Finland. Silva Fennica 39(1): 55-66.

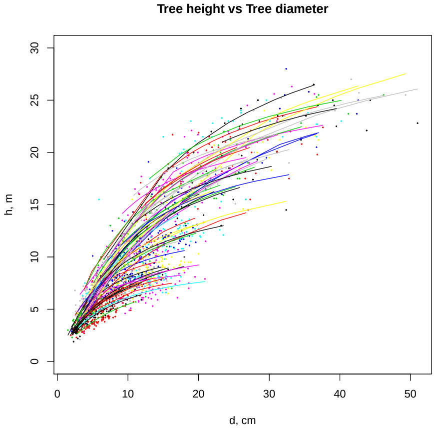
Why an H-D model?

- H-D relationship varies much among sample plots, but height measurement is time-consuming.



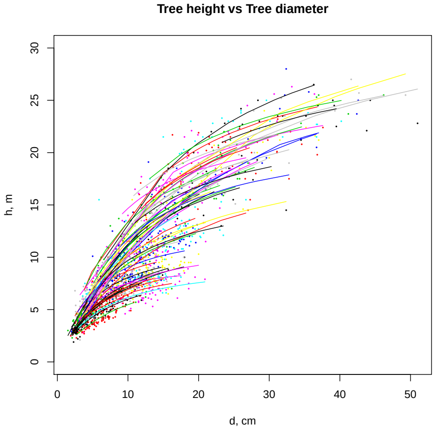
Why an H-D model?

- H-D relationship varies much among sample plots, but height measurement is time-consuming.
- In a forest inventory, diameter is usully tallied for all trees of a sample plot, whereas height is measured only for 0 – 5 trees per plot.



Why an H-D model?

- H-D relationship varies much among sample plots, but height measurement is time-consuming.
- In a forest inventory, diameter is usully tallied for all trees of a sample plot, whereas height is measured only for 0 – 5 trees per plot.



If a previously fitted H-D model is available, it can be localized, or calibrated, for the new plot by predicting the random effects using the sampled tree heights.

The Height-Diameter model

The logarithmic heigth H_{kti} for tree i in stand k at time t with diameter D_{kti} at the breast height is expressed by

$$\ln(H_{kti}) = a(DGM_{kt}) + \alpha_k + \alpha_{kt} + (b(DGM_{kt}) + \beta_k + \beta_{kt})D_{kti} + \epsilon_{kti} \, ,$$

where $a(DGM_{kt})$ and $b(DGM_{kt})$ are known fixed functions of plot-specific mean diameter DGM_{kt}, $(\alpha_k, \beta_k)'$ and $(\alpha_{kt}, \beta_{kt})'$ are the plot and measurement occasion -level random effects with varainces (correlations)

$$\mathrm{var}\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0.108^2 & (0.269) \\ 0.0028 & 0.0958^2 \end{bmatrix} \quad \mathrm{var}\begin{bmatrix} \alpha_{kt} \\ \beta_{kt} \end{bmatrix} = \begin{bmatrix} 0.0168^2 & (-0.681) \\ -0.0003 & 0.0223^2 \end{bmatrix}$$

and ϵ_{kti} are independent normal residuals with $\mathrm{var}(\epsilon_{kti}) = 0.401^2\left(\max(D_{kti}, 7.5)\right)^{-1.068}$

The stand level mixed-effects model

The sample tree heights of a new stand can be described by

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}\,,$$

where
$\boldsymbol{y}$ includes the observed sample tree heights,
$\boldsymbol{\mu}$ is the fixed part,
$\boldsymbol{\beta} = \begin{pmatrix} \alpha_k & \beta_k & \alpha_{k1} & \beta_{k1} & \alpha_{k2} & \beta_{k2} & \ldots \end{pmatrix}'$ includes the random effects,
$\boldsymbol{Z}$ is the corresponding design matrix, and
$\boldsymbol{\epsilon}$ includes the residuals.
We denote $\mathrm{var}(\boldsymbol{\beta}) = \boldsymbol{D}$ and $\mathrm{var}(\boldsymbol{\epsilon}) = \boldsymbol{R}$.

Prediction of random effects

The variances and covariances between random effects and observed heights can be written as

$$\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{y} \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{DZ}' \\ \boldsymbol{ZD} & \boldsymbol{ZDZ}' + \boldsymbol{R} \end{bmatrix} \right)$$

The Empirical Best Linear Unbiased Predictor (EBLUP) of random effects is

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{DZ}'(\boldsymbol{ZDZ}' + \boldsymbol{R})^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) .$$

and the variance of prediction errors is

$$\mathrm{var}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{D} - \boldsymbol{DZ}'(\boldsymbol{ZDZ}' + \boldsymbol{R})^{-1}\boldsymbol{ZD}$$

Example

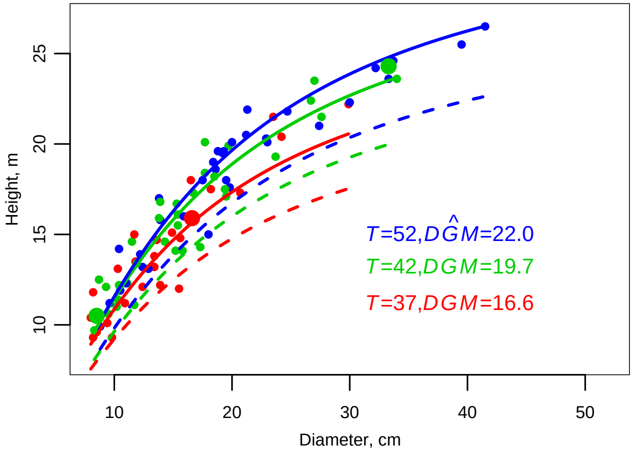
Height of one tree was measured 5 years ago and 2 trees at the current year. The matrices and vectors are

$$\boldsymbol{\mu} = \begin{bmatrix} 2.59 \\ 2.11 \\ 2.99 \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} 2.77 \\ 2.35 \\ 3.19 \end{bmatrix}$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & -0.36 & 1 & -0.36 & 0 & 0 \\ 1 & -1.22 & 0 & 0 & 1 & -1.22 \\ 1 & 0.058 & 0 & 0 & 1 & 0.058 \end{bmatrix} \quad \boldsymbol{R} = \begin{bmatrix} 0.008 & 0 & 0 \\ 0 & 0.016 & 0 \\ 0 & 0 & 0.004 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha_k \\ \beta_k \\ \alpha_{k1} \\ \beta_{k1} \\ \alpha_{k2} \\ \beta_{k2} \end{bmatrix} \quad \boldsymbol{D} = \begin{bmatrix} 0.0118 & 0.0028 & 0 & 0 & 0 & 0 \\ 0.0028 & 0.0092 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0003 & 0.0004 & 0 & 0 \\ 0 & 0 & 0.0004 & 0.0005 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0003 & 0.0004 \\ 0 & 0 & 0 & 0 & 0.0004 & 0.0005 \end{bmatrix}$$

Uncalibrated and calibrated predictions



dashed=fixed part only; solid= calibrated (fixed+random)

Case 2: Extracting effects of silvicultural thinnings

Utilizing a prediction from a linear mixed-effects model with crossed tree and calendar year effects

Mehtätalo, L., Peltola, H., Kilpeläinen, A. and Ikonen, V.-P. 2013. The effect of thinning on the basal area growth of Scots Pine: a longitudinal analysis using nonlinear mixed-effects model. Submitted manuscript.

Why thinning effects?

- Forest managers use silvicultural thinnings to decrease the competition of neighboring trees and, consequently, to increase the growth rate of the remaining trees for faster production of sawtimber.

Why thinning effects?

- Forest managers use silvicultural thinnings to decrease the competition of neighboring trees and, consequently, to increase the growth rate of the remaining trees for faster production of sawtimber.
- To understand the dynamics of thinning, one may wish to analyse the effect of thinnings on tree growth.

Why thinning effects?

- Forest managers use silvicultural thinnings to decrease the competition of neighboring trees and, consequently, to increase the growth rate of the remaining trees for faster production of sawtimber.
- To understand the dynamics of thinning, one may wish to analyse the effect of thinnings on tree growth.
- However, the growth is affected also by other factors, especially by the site productivity, tree age, and annual weather.

Why thinning effects?

- Forest managers use silvicultural thinnings to decrease the competition of neighboring trees and, consequently, to increase the growth rate of the remaining trees for faster production of sawtimber.
- To understand the dynamics of thinning, one may wish to analyse the effect of thinnings on tree growth.
- However, the growth is affected also by other factors, especially by the site productivity, tree age, and annual weather.
- Mixed-effects models can be used to model out these nuisance effects.

Study material

- Thinning experiment sample plots were established in naturally generated Scots pine stands at the age of $\sim$ 25 years in Mekrijärvi, Finland in 1986.

Study material

- Thinning experiment sample plots were established in naturally generated Scots pine stands at the age of $\sim$ 25 years in Mekrijärvi, Finland in 1986.
- One of the four following thinning treatments were applied to each plot: No thinning (I, Control), light (II), moderate (III), and heavy (IV) thinnings.

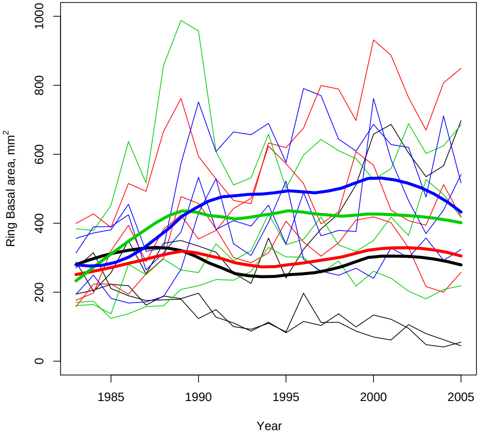
Study material

- Thinning experiment sample plots were established in naturally generated Scots pine stands at the age of $\sim$ 25 years in Mekrijärvi, Finland in 1986.
- One of the four following thinning treatments were applied to each plot: No thinning (I, Control), light (II), moderate (III), and heavy (IV) thinnings.
- 88 trees were felled in 2006, and the complete time series of diameter increments between 1983 and 2006 was measured for each tree using an X-ray densiometer.

Study material

- Thinning experiment sample plots were established in naturally generated Scots pine stands at the age of $\sim$ 25 years in Mekrijärvi, Finland in 1986.
- One of the four following thinning treatments were applied to each plot: No thinning (I, Control), light (II), moderate (III), and heavy (IV) thinnings.
- 88 trees were felled in 2006, and the complete time series of diameter increments between 1983 and 2006 was measured for each tree using an X-ray densiometer.
- The diameter growths were transformed to basal area growths, because $Volume \sim Diameter^2 Height$)

The raw data



- THICK: treatment-specific trends
- THIN: 12 randomly selected trees
- One can see
  - (Age trend)
  - climate-related year effects
  - tree effects

I (control) - black; II (light) - red
III (moderate) - green; IV (heavy) - blue

Modeling the non-thinned response

- A dataset without thinning treatments was produced by including from the original data
  - The control treatment for whole follow-up period
  - The thinned treatments until the year of thinning (1986)

Modeling the non-thinned response

- A dataset without thinning treatments was produced by including from the original data
  - The control treatment for whole follow-up period
  - The thinned treatments until the year of thinning (1986)
- A linear mixed effect model with random year and tree effects was fitted to the unthinned data

$$y_{kt} = f(T_{kt}; \boldsymbol{b}) + \alpha_k + \alpha_t + \epsilon_{kt} \quad (1)$$

where y_{kt} is the basal area growth of tree k at year t,
$f(T_{kt}; \boldsymbol{b})$ is the age trend (modeled using a spline),
α_k is a NID tree effect,
α_t is a NID year effect and
ϵ_{kt} is a NID residual.

Extracting the thinning effects

- Using the estimated age trend and BLUP's of year and tree effects, the growth without thinning, $\widetilde{y}_{kt}$ was predicted for treatments II -IV after the thinning year.

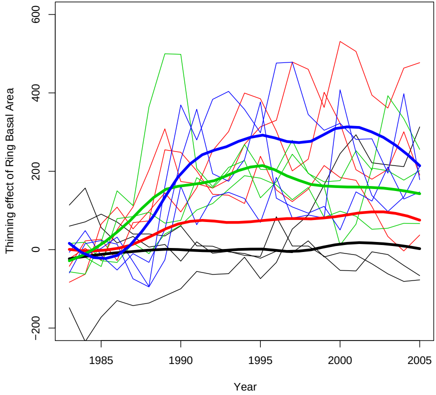
Extracting the thinning effects

- Using the estimated age trend and BLUP's of year and tree effects, the growth without thinning, $\widetilde{y}_{kt}$ was predicted for treatments II -IV after the thinning year.
- The pure thinning effects were estimated by subtracting the prediction from the observed growth

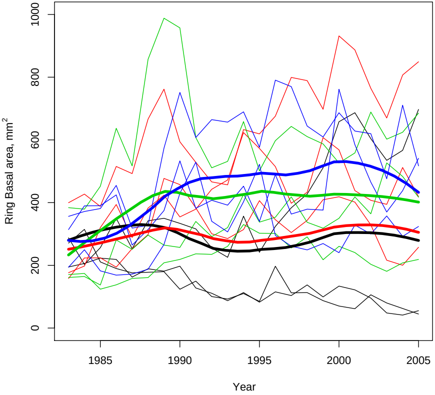
$$d_{kt} = y_{kt} - \widetilde{y}_{kt} \tag{2}$$

The estimated thinning effects

Extracted thinning effects



Raw data



Line color specifies treatment (I:black, II: red, III: green IV: blue). Thick lines show the treatment-specific mean trends; thin lines show 12 randomly selected trees.

Case 3: Modelling thinning effects uisng NLME's

A nonlinear model to analyze the effect of thinning intensity and tree size on the dynamics of tree-level thinning effect.

Mehtätalo, L., Peltola, H., Kilpeläinen, A. and Ikonen, V.-P. 2013. The effect of thinning on the basal area growth of Scots Pine: a longitudinal analysis using nonlinear mixed-effects model. Submitted manuscript.

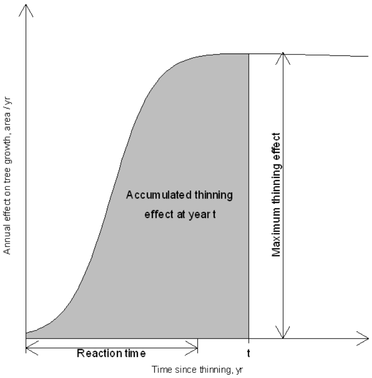
Modeling the thinning effects

- The thinning effects seem to switch on during a short time called **Reaction time** and stabilize thereafter at a level of **Maximum thinning effect**.
- To explore what predictors control these two parameters, the thinning effects of the thinnend treatments 2-4 were modeled using a nonlinear mixed-effects model.
- The random effects were used to take into account the data hierarchy for more reliable inference.

Nonlinear mixed-effects model for thinning effect

The thinning effect of tree *k* at time *t* was modeled using a logistic curve

$$d_{kt} = \frac{M_k}{1+\exp\left(4-8\frac{x_{kt}}{R_k}\right)} + e_{kt}$$



- d_{kt} - thinning effect
- x_{kt} - time since thinning
- $M_k = \mu_0 + \mu_1 T_2 + \mu_2 T_3 + \mu_4 x_{kt} + m_k$ - maximum thinning effect
- $T_2, \ldots, T_3$ - treatments
- $R_k = \rho_0 + \rho_1 z_k + r_k$ - reaction time
- z_k - standardized diameter
- $\begin{bmatrix} m_k \\ r_k \end{bmatrix} \sim MVN(\mathbf{0}, \boldsymbol{D}_{2x2})$
- e_{kt} - normal heteroscedastic residual with AR(1) structure within a tree.

The fitted model

- The reaction time was 6 years. It did not significantly vary among treatments but was shorter for large trees.

Fixed parameters	Estimate	s.e.	p-value
μ_0	112.8	23.29	0.0000
μ_1	91.91	30.45	0.0026
μ_2	169.2	32.14	0.0000
μ_3	-3.214	1.006	0.0014
ρ_0	5.749	0.4458	0.0000
ρ_1	-1.461	0.4568	0.0014
Random parameters			
var(r_k)	93.012		
var(m_k)	2.0852		
cor(r_k,m_k)	0.203		
Residual			
σ^2	8.157*10-4		
δ_1	8.746*104		
δ_2	1.886		
δ_3	0.5888		

The fitted model

- The reaction time was 6 years. It did not significantly vary among treatments but was shorter for large trees.
- The maximum thinning effect increased with thinning intensity, being 282 *mm*/*yr* for treatment IV, which indicates a 87% increase in the basal area growth compared to the control.

Fixed parameters	Estimate	s.e.	p-value
μ_0	112.8	23.29	0.0000
μ_1	91.91	30.45	0.0026
μ_2	169.2	32.14	0.0000
μ_3	-3.214	1.006	0.0014
ρ_0	5.749	0.4458	0.0000
ρ_1	-1.461	0.4568	0.0014
Random parameters			
var(r_k)	93.012		
var(m_k)	2.0852		
cor(r_k,m_k)	0.203		
Residual			
σ^2	8.157*10-4		
δ_1	8.746*104		
δ_2	1.886		
δ_3	0.5888		

Case 4: Modelling tree-level reflectance on aerial images

A multivariate linear mixed-effects model with crossed grouping structure was used to analyze the reflectance of forest trees on overlapping aerial images.

Korpela Ilkka, Mehtätalo Lauri, Seppänen Anne, Markelin Lauri. Tree species classification using directional reflectance anisotropy signatures in multiple aerial images. Submitted.

Motivation

- The reflectance (color) of a tree on an image can be used to classify tree species

Motivation

- The reflectance (color) of a tree on an image can be used to classify tree species
- However, the viewing direction with respect to sunlight affects the spectral characteristics of a tree.

Motivation

- The reflectance (color) of a tree on an image can be used to classify tree species
- However, the viewing direction with respect to sunlight affects the spectral characteristics of a tree.
- This effect is species-specific

Motivation

- The reflectance (color) of a tree on an image can be used to classify tree species
- However, the viewing direction with respect to sunlight affects the spectral characteristics of a tree.
- This effect is species-specific
- Therefore, observing a certain tree from multiple directions (=images) may provide more accurate species classification than an observation on one aerial image only.

Study material

- 20 partially overlapping aerial images of a forest area were taken.

Study material

- 20 partially overlapping aerial images of a forest area were taken.
- The raw data was postprocessed to provide (atmospherically corrected) reflectance data on four channels: RED, GRN, BLU and NIR.

Study material

- 20 partially overlapping aerial images of a forest area were taken.
- The raw data was postprocessed to provide (atmospherically corrected) reflectance data on four channels: RED, GRN, BLU and NIR.
- $N = 15188$ dominant trees discernible in 2-7 images formed the reference tree data (5914 Scots pines, 7105 Norway spruces, 2169 Birches)

Study material

- 20 partially overlapping aerial images of a forest area were taken.
- The raw data was postprocessed to provide (atmospherically corrected) reflectance data on four channels: RED, GRN, BLU and NIR.
- $N = 15188$ dominant trees discernible in 2-7 images formed the reference tree data (5914 Scots pines, 7105 Norway spruces, 2169 Birches)
- Individual trees on different images were using automatically matched.

Study material

- 20 partially overlapping aerial images of a forest area were taken.
- The raw data was postprocessed to provide (atmospherically corrected) reflectance data on four channels: RED, GRN, BLU and NIR.
- $N = 15188$ dominant trees discernible in 2-7 images formed the reference tree data (5914 Scots pines, 7105 Norway spruces, 2169 Birches)
- Individual trees on different images were using automatically matched.
- The individual pixels within tree crowns were divided to sunlit and self-shaded pixels. The mean reflectances in these parts were analyzed separately -> a system of 8 models (4 channels, shaded and sunlit) for each of the three tree species.

Structure of aerial image data on a forest

- Observations from a given image are similar due to e.g. the properties of the atmosphere at the time of imaging and the atmospheric correction.

Structure of aerial image data on a forest

- Observations from a given image are similar due to e.g. the properties of the atmosphere at the time of imaging and the atmospheric correction.
- Repeated measurements of a certain tree are correlated due to tree-specific properties.

Structure of aerial image data on a forest

- Observations from a given image are similar due to e.g. the properties of the atmosphere at the time of imaging and the atmospheric correction.
- Repeated measurements of a certain tree are correlated due to tree-specific properties.
- The model for each response and tree species has the following structure

$$y_{it} = f(\boldsymbol{x}_{it}|\boldsymbol{b}) + \alpha_i + \alpha_t + \epsilon_{it} ,$$

where i and t refer to image and tree effects, respectively. σ_i^2 and σ_t^2 are the corresponding variances. The predictors are trigonometric transformations of the horizontal and vertical viewing and Sun angles.

Structure of aerial image data on a forest

- Observations from a given image are similar due to e.g. the properties of the atmosphere at the time of imaging and the atmospheric correction.
- Repeated measurements of a certain tree are correlated due to tree-specific properties.
- The model for each response and tree species has the following structure

$$y_{it} = f(\boldsymbol{x}_{it}|\boldsymbol{b}) + \alpha_i + \alpha_t + \epsilon_{it} ,$$

  where i and t refer to image and tree effects, respectively. σ_i^2 and σ_t^2 are the corresponding variances. The predictors are trigonometric transformations of the horizontal and vertical viewing and Sun angles.
- The random effects at different levels of grouping are independent, therefore

$$\begin{aligned} var(y_{it}) &= \sigma_i^2 + \sigma_t^2 + \sigma^2 \\ cov(y_{it}, y_{i't'}) &= 0 \\ cov(y_{it}, y_{it'}) &= \sigma_i^2 \\ cov(y_{it}, y_{i't}) &= \sigma_t^2 \end{aligned}$$

The multivariate model

The multivariate model for a tree species is

$$\begin{aligned} y1_{it} &= f1(\boldsymbol{x}_{it}|\boldsymbol{b}1) + \alpha 1_i + \alpha 1_t + \epsilon 1_{it} \\ y2_{it} &= f2(\boldsymbol{x}_{it}|\boldsymbol{b}2) + \alpha 2_i + \alpha 2_t + \epsilon 2_{it} \\ &\vdots \\ y8_{it} &= f8(\boldsymbol{x}_{it}|\boldsymbol{b}8) + \alpha 8_i + \alpha 8_t + \epsilon 8_{it} \end{aligned}$$

or simply

$$\boldsymbol{y}_{it} = \boldsymbol{f}(\boldsymbol{x}_{it}|\boldsymbol{b}) + \boldsymbol{\alpha}_i + \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_{it}$$

where the responses 1-8 refer to the sunlit and self-shaded pixels of the four channels and

- $(\alpha 1_i, \alpha 2_i, \ldots, \alpha 8_i)' = \boldsymbol{\alpha}_i \sim MVN(0, \boldsymbol{A}_{8\times 8})$ include the random image-effects
- $(\alpha 1_t, \alpha 2_t, \ldots, \alpha 8_t)' = \boldsymbol{\alpha}_t \sim MVN(0, \boldsymbol{B}_{8\times 8})$ include the random tree-effects
- $(\epsilon 1_{it}, \epsilon 2_{it}, \ldots, \epsilon 8_{it})' = \boldsymbol{\epsilon}_{it} \sim MVN(0, \boldsymbol{E}_{8\times 8})$ include the random vector residuals

The multivariate model

The multivariate model for a tree species is

$$\begin{aligned} y1_{it} &= f1(\boldsymbol{x}_{it}|\boldsymbol{b}1) + \alpha 1_i + \alpha 1_t + \epsilon 1_{it} \\ y2_{it} &= f2(\boldsymbol{x}_{it}|\boldsymbol{b}2) + \alpha 2_i + \alpha 2_t + \epsilon 2_{it} \\ &\vdots \\ y8_{it} &= f8(\boldsymbol{x}_{it}|\boldsymbol{b}8) + \alpha 8_i + \alpha 8_t + \epsilon 8_{it} \end{aligned}$$

or simply

$$\boldsymbol{y}_{it} = \boldsymbol{f}(\boldsymbol{x}_{it}|\boldsymbol{b}) + \boldsymbol{\alpha}_i + \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_{it}$$

where the responses 1-8 refer to the sunlit and self-shaded pixels of the four channels and

- $(\alpha 1_i, \alpha 2_i, \ldots, \alpha 8_i)' = \boldsymbol{\alpha}_i \sim MVN(0, \boldsymbol{A}_{8\times 8})$ include the random image-effects
- $(\alpha 1_t, \alpha 2_t, \ldots, \alpha 8_t)' = \boldsymbol{\alpha}_t \sim MVN(0, \boldsymbol{B}_{8\times 8})$ include the random tree-effects
- $(\epsilon 1_{it}, \epsilon 2_{it}, \ldots, \epsilon 8_{it})' = \boldsymbol{\epsilon}_{it} \sim MVN(0, \boldsymbol{E}_{8\times 8})$ include the random vector residuals
- Now

$$\begin{aligned} var(\boldsymbol{y}_{it}) &= \boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E} \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{i't'}) &= 0 \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{it'}) &= \boldsymbol{A} \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{i't}) &= \boldsymbol{B} \end{aligned}$$

The multivariate model

The multivariate model for a tree species is

$$\begin{aligned} y1_{it} &= f1(\boldsymbol{x}_{it}|\boldsymbol{b}1) + \alpha1_i + \alpha1_t + \epsilon1_{it} \\ y2_{it} &= f2(\boldsymbol{x}_{it}|\boldsymbol{b}2) + \alpha2_i + \alpha2_t + \epsilon2_{it} \\ &\vdots \\ y8_{it} &= f8(\boldsymbol{x}_{it}|\boldsymbol{b}8) + \alpha8_i + \alpha8_t + \epsilon8_{it} \end{aligned}$$

or simply

$$\boldsymbol{y}_{it} = \boldsymbol{f}(\boldsymbol{x}_{it}|\boldsymbol{b}) + \boldsymbol{\alpha}_i + \boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_{it}$$

where the responses 1-8 refer to the sunlit and self-shaded pixels of the four channels and

- $(\alpha1_i, \alpha2_i, \ldots, \alpha8_i)' = \boldsymbol{\alpha}_i \sim MVN(0, \boldsymbol{A}_{8\times8})$ include the random image-effects
- $(\alpha1_t, \alpha2_t, \ldots, \alpha8_t)' = \boldsymbol{\alpha}_t \sim MVN(0, \boldsymbol{B}_{8\times8})$ include the random tree-effects
- $(\epsilon1_{it}, \epsilon2_{it}, \ldots, \epsilon8_{it})' = \boldsymbol{\epsilon}_{it} \sim MVN(0, \boldsymbol{E}_{8\times8})$ include the random vector residuals
- Now

$$\begin{aligned} var(\boldsymbol{y}_{it}) &= \boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E} \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{i't'}) &= 0 \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{it'}) &= \boldsymbol{A} \\ cov(\boldsymbol{y}_{it}, \boldsymbol{y}_{i't}) &= \boldsymbol{B} \end{aligned}$$

Estimated variance components (covariances not shown)

Variance components, real data, 200 000 observations (%)

	sunlit	shade	sunlit	shade	sunlit	shade	sunlit	shade
Fixed (Xβ)-%	33	11	32	13	45	29	7	-0
Tree-%	42	42	43	41	18	13	62	64
Image-%	4	12	5	14	27	46	6	2
Residual-%	21	35	20	32	10	13	25	34
Total	100	100	100	100	100	100	100	100

* Fixed part: The anisotropy trends explained SL >> SS, BLU > GRN > RED > NIR. In NIR, anisotropy is low.

* Tree-effect: The correlations are strong, both for SL and SS. A bright tree is bright across views and bands. In NIR > 60% of variance explained!!

* Image-effect: Substantial in BLU, SS > SL. Includes effects from solar elevation changes (07-09 GMT), atmospheric correction errors.

Ilkka Korpela, Oct 2012

The use in classification

- Let $\boldsymbol{y}_{it}$ be an observed vector (length=8) of the reflectances of one tree t on the 8 channels on one image i. The squared Mahalanobis distance between $\boldsymbol{y}_{it}$ and $\boldsymbol{\mu}_{it}$ is

$$d_{it}^2 = (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})'(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E})^{-1}(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})$$

  This distance takes into account the correlation of reflectance among different channels, and is (at least under multivariate normality of the reflectance data) in a way optimal for single tree on single image.

The use in classification

- Let $\boldsymbol{y}_{it}$ be an observed vector (length=8) of the reflectances of one tree t on the 8 channels on one image i. The squared Mahalanobis distance between $\boldsymbol{y}_{it}$ and $\boldsymbol{\mu}_{it}$ is

$$d_{it}^2 = (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})'(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E})^{-1}(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})$$

  This distance takes into account the correlation of reflectance among different channels, and is (at least under multivariate normality of the reflectance data) in a way optimal for single tree on single image.

- For multiple images, the squared Mahalanobis distance between $\boldsymbol{y}_{\cdot t}$ and $\boldsymbol{\mu}_{\cdot t}$ is

$$d_{\cdot t}^2 = (\boldsymbol{y}_{\cdot t} - \boldsymbol{\mu}_{\cdot t})'\boldsymbol{D}_{\cdot t}^{-1}(\boldsymbol{y}_{\cdot t} - \boldsymbol{\mu}_{\cdot t})\,,$$

  where $\boldsymbol{y}_{\cdot t} = (\boldsymbol{y}_{1t}', \ldots, \boldsymbol{y}_{mt})$ is an observed vector (with length of $8m$) of the reflectances of tree t on the 8 channels of m images. The $8m \times 8m$ variance-covariance matrix is

$$\boldsymbol{D}_{\cdot t} = \begin{bmatrix} \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} & \boldsymbol{B} & \ldots & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} & & \boldsymbol{B} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{B} & \boldsymbol{B} & \ldots & \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} \end{bmatrix}$$

  This distance takes into account the correlation arising from the common tree effects

The use in classification

- Let $\boldsymbol{y}_{it}$ be an observed vector (length=8) of the reflectances of one tree t on the 8 channels on one image i. The squared Mahalanobis distance between $\boldsymbol{y}_{it}$ and $\boldsymbol{\mu}_{it}$ is

$$d_{it}^2 = (\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})'(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E})^{-1}(\boldsymbol{y}_{it} - \boldsymbol{\mu}_{it})$$

  This distance takes into account the correlation of reflectance among different channels, and is (at least under multivariate normality of the reflectance data) in a way optimal for single tree on single image.

- For multiple images, the squared Mahalanobis distance between $\boldsymbol{y}_{\cdot t}$ and $\boldsymbol{\mu}_{\cdot t}$ is

$$d_{\cdot t}^2 = (\boldsymbol{y}_{\cdot t} - \boldsymbol{\mu}_{\cdot t})'\boldsymbol{D}_{\cdot t}^{-1}(\boldsymbol{y}_{\cdot t} - \boldsymbol{\mu}_{\cdot t}),$$

  where $\boldsymbol{y}_{\cdot t} = (\boldsymbol{y}_{1t}', \ldots, \boldsymbol{y}_{mt})$ is an observed vector (with length of $8m$) of the reflectances of tree t on the 8 channels of m images. The $8m \times 8m$ variance-covariance matrix is

$$\boldsymbol{D}_{\cdot t} = \begin{bmatrix} \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} & \boldsymbol{B} & \ldots & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} & & \boldsymbol{B} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{B} & \boldsymbol{B} & \ldots & \boldsymbol{A}+\boldsymbol{B}+\boldsymbol{E} \end{bmatrix}$$

  This distance takes into account the correlation arising from the common tree effects

- Extension to many trees and images would be possible as well.

Discussion

- Mixed-effects models are useful tools for analyzing grouped datasets in different contexts.

Discussion

- Mixed-effects models are useful tools for analyzing grouped datasets in different contexts.
- The benefit from the use of mixed-effects models depends on the application, but may be related to
  - inference (Case 3: Modelling the thinnig effect),
  - prediction (Case 1: H-D, Case 2: Extraction of thinning effect), or
  - estimated variance-covariance structure of the data (Case 4: Species classification).

Discussion

- Mixed-effects models are useful tools for analyzing grouped datasets in different contexts.
- The benefit from the use of mixed-effects models depends on the application, but may be related to
  - inference (Case 3: Modelling the thinnig effect),
  - prediction (Case 1: H-D, Case 2: Extraction of thinning effect), or
  - estimated variance-covariance structure of the data (Case 4: Species classification).
- The prediction of random effects for a **new** group is a powerful tool to localize models afterwards using very limited datasets.

Discussion

- Mixed-effects models are useful tools for analyzing grouped datasets in different contexts.
- The benefit from the use of mixed-effects models depends on the application, but may be related to
  - inference (Case 3: Modelling the thinnig effect),
  - prediction (Case 1: H-D, Case 2: Extraction of thinning effect), or
  - estimated variance-covariance structure of the data (Case 4: Species classification).
- The prediction of random effects for a **new** group is a powerful tool to localize models afterwards using very limited datasets.
- I wonder if other fields than forestry have or could have similar applications.

Thank you for your interest . . .

Thank you for your interest . . .

. . . and if you got interested in Joensuu, you may apply this open Senior researcher position at UEF