Overview of mixed-effects models

Lauri Mehtätalo[1]

[1] University of Eastern Finland, School of Computing

Machine learning course at UEF
August 14, 2017

Outline

Grouped datasets

- Empirical datasets are usually grouped e.g.
  - Repeated voice samples from the same person,
  - Trees within sample plots or aerial images,
  - repeated observations of trees (e.g., in successive years or on different images)
  - ...

Grouped datasets

- Empirical datasets are usually grouped e.g.
  - Repeated voice samples from the same person,
  - Trees within sample plots or aerial images,
  - repeated observations of trees (e.g., in successive years or on different images)
  - ...
- A dataset may also have multiple nested or grouped levels
  - Repeated measurements of trees within sample plots (nested)
  - Tree increments for different calendar years (crossed)
  - Repeated measurements of persons within devices (nested / crossed)

Grouped datasets

- Empirical datasets are usually grouped e.g.
  - Repeated voice samples from the same person,
  - Trees within sample plots or aerial images,
  - repeated observations of trees (e.g., in successive years or on different images)
  - ...
- A dataset may also have multiple nested or grouped levels
  - Repeated measurements of trees within sample plots (nested)
  - Tree increments for different calendar years (crossed)
  - Repeated measurements of persons within devices (nested / crossed)

These groups often constitute a sample from a population of groups, and are therefore naturally modeled using mixed-effect models.

Mixed effect models can be used for several purposes

1. To take into account the lack of dependence in a statistical analysis.

Mixed effect models can be used for several purposes

1. To take into account the lack of dependence in a statistical analysis.
2. To make improved group-level predictions of the variable of interest.

Mixed effect models can be used for several purposes

1. To take into account the lack of dependence in a statistical analysis.
2. To make improved group-level predictions of the variable of interest.
3. To estimate the variance-covariance structure of the data to be used e.g. for classification.

Mixed effect models can be used for several purposes

1. To take into account the lack of dependence in a statistical analysis.
2. To make improved group-level predictions of the variable of interest.
3. To estimate the variance-covariance structure of the data to be used e.g. for classification.

(2) and (3) are especially interesting in prediction and classification problems, which are considered in machine learning.

Linear mixed-effect model with random constant

$$y_{ij} \quad = \quad \boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + \epsilon_{ij} \,,$$

where

- y_{ij} is the observed response for individual j, $j = 1, \ldots, n_i$ in group i, $i = 1, \ldots, M$.
- $\boldsymbol{x}_{ij}$ is a vector of fixed predictors,
- $\boldsymbol{\beta}$ includes the fixed parameters,
- b_i are random group effects for groups i,
- ϵ_{ij} is residual error for individual ij.

Linear mixed-effect model with random constant

$$y_{ij} \quad = \quad \boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + \epsilon_{ij} \,,$$

where

- y_{ij} is the observed response for individual j, $j = 1, \ldots, n_i$ in group i, $i = 1, \ldots, M$.
- $\boldsymbol{x}_{ij}$ is a vector of fixed predictors,
- $\boldsymbol{\beta}$ includes the fixed parameters,
- b_i are random group effects for groups i,
- ϵ_{ij} is residual error for individual ij.
- We assume $b_i \sim N(0, \sigma_b^2)$ (i.i.d); $\epsilon_{ij} \sim N(0, \sigma^2)$ (i.i.d); b_i are independent of ϵ_{ij}.

Linear mixed-effect model with random constant

$$y_{ij} \quad = \quad \boldsymbol{\beta}'\boldsymbol{x}_{ij} + b_i + \epsilon_{ij} \,,$$

where

- y_{ij} is the observed response for individual j, $j = 1, \ldots, n_i$ in group i, $i = 1, \ldots, M$.
- $\boldsymbol{x}_{ij}$ is a vector of fixed predictors,
- $\boldsymbol{\beta}$ includes the fixed parameters,
- b_i are random group effects for groups i,
- ϵ_{ij} is residual error for individual ij.
- We assume $b_i \sim N(0, \sigma_b^2)$ (i.i.d); $\epsilon_{ij} \sim N(0, \sigma^2)$ (i.i.d); b_i are independent of ϵ_{ij}.
- Model parameters are $\boldsymbol{\beta}$, σ_b^2, and σ^2. Also group effects b_i can be predicted.

Linear mixed-effect model with random constant

$$y_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + \epsilon_{ij} \,,$$

where

- y_{ij} is the observed response for individual j, $j = 1, \ldots, n_i$ in group i, $i = 1, \ldots, M$.
- $\boldsymbol{x}_{ij}$ is a vector of fixed predictors,
- $\boldsymbol{\beta}$ includes the fixed parameters,
- b_i are random group effects for groups i,
- ϵ_{ij} is residual error for individual ij.
- We assume $b_i \sim N(0, \sigma_b^2)$ (i.i.d); $\epsilon_{ij} \sim N(0, \sigma^2)$ (i.i.d); b_i are independent of ϵ_{ij}.
- Model parameters are $\boldsymbol{\beta}$, σ_b^2, and σ^2. Also group effects b_i can be predicted.
- Can be seen as a marginal model $y_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + e_{ij}$, where $\mathrm{var}(y_{ij}) = \sigma_b^2 + \sigma^2$ and $\mathrm{cov}(y_{ij}, y_{ij'}) = \sigma_b^2$.

Parameter estimation

- The (restricted) likelihood for the marginal model $y_{ij} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + e_{ij}$ is easy to write to get (RE)ML estimates of parameters σ_b^2 and σ^2, and GLS/REML/ML estimates of $\boldsymbol{\beta}$. The REML is based on multivariate normality of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ but the resulting estimators have nice properties in non-normal data as well.

Parameter estimation

- The (restricted) likelihood for the marginal model $y_{ij} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + e_{ij}$ is easy to write to get (RE)ML estimates of parameters σ_b^2 and σ^2, and GLS/REML/ML estimates of $\boldsymbol{\beta}$. The REML is based on multivariate normality of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ but the resulting estimators have nice properties in non-normal data as well.
- The random group effects can be predicted using Best Linear Unbiased Predictor (BLUP)

$$\widetilde{b_i} = \frac{\sigma_b^2}{\frac{1}{n_i}\sigma^2 + \sigma_b^2}(\bar{y}_i - \bar{\boldsymbol{\beta}'\boldsymbol{x}}_{ij})$$

where $\bar{y}_i$ and $\bar{\boldsymbol{\beta}'\boldsymbol{x}}_{ij}$ are the means of the n_i observed values and fixed-part predictions for the group in question.

Parameter estimation

- The (restricted) likelihood for the marginal model $y_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + e_{ij}$ is easy to write to get (RE)ML estimates of parameters σ_b^2 and σ^2, and GLS/REML/ML estimates of $\boldsymbol{\beta}$. The REML is based on multivariate normality of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ but the resulting estimators have nice properties in non-normal data as well.
- The random group effects can be predicted using Best Linear Unbiased Predictor (BLUP)

$$\widetilde{b}_i = \frac{\sigma_b^2}{\frac{1}{n_i}\sigma^2 + \sigma_b^2}(\bar{y}_i - \bar{\boldsymbol{\beta}' \boldsymbol{x}}_{ij})$$

  where $\bar{y}_i$ and $\bar{\boldsymbol{\beta}' \boldsymbol{x}}_{ij}$ are the means of the n_i observed values and fixed-part predictions for the group in question.
- The prediction variance is

$$\mathrm{var}\left(\widetilde{b}_i - b_i\right) = \left(\frac{\sigma_b^2}{\frac{1}{n_i}\sigma^2 + \sigma_b^2}\right)\frac{\sigma^2}{n_i}$$

Parameter estimation

- The (restricted) likelihood for the marginal model $y_{ij} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + e_{ij}$ is easy to write to get (RE)ML estimates of parameters σ_b^2 and σ^2, and GLS/REML/ML estimates of $\boldsymbol{\beta}$. The REML is based on multivariate normality of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ but the resulting estimators have nice properties in non-normal data as well.
- The random group effects can be predicted using Best Linear Unbiased Predictor (BLUP)

$$\widetilde{b}_i = \frac{\sigma_b^2}{\frac{1}{n_i}\sigma^2 + \sigma_b^2}(\bar{y}_i - \bar{\boldsymbol{\beta}'\boldsymbol{x}}_{ij})$$

  where $\bar{y}_i$ and $\bar{\boldsymbol{\beta}'\boldsymbol{x}}_{ij}$ are the means of the n_i observed values and fixed-part predictions for the group in question.
- The prediction variance is

$$\mathrm{var}\left(\widetilde{b}_i - b_i\right) = \left(\frac{\sigma_b^2}{\frac{1}{n_i}\sigma^2 + \sigma_b^2}\right)\frac{\sigma^2}{n_i}$$

- In practice, we use Empirical BLUP where the unknown $\boldsymbol{\beta}$, σ_b^2 and σ^2 are replaced by their numerical estimates.

Some notes on prediction

- Mixed-effects allows *group-level prediction* where the predicted random effect is used.

[1] For forestry-related examples, see Lappi 1986, 1991, 1997, Lappi and Bailey 1988

Some notes on prediction

- Mixed-effects allows *group-level prediction* where the predicted random effect is used.
- If no measurements of y are available from the group in question, the BLUP of random effect is its expected value 0, the prediction for *a typical group*

[1] For forestry-related examples, see Lappi 1986, 1991, 1997, Lappi and Bailey 1988

Some notes on prediction

- Mixed-effects allows *group-level prediction* where the predicted random effect is used.
- If no measurements of y are available from the group in question, the BLUP of random effect is its expected value 0, the prediction for *a typical group*
- Group-level predictions utilize the observed values of the response from the group in question. For groups not present in the modeling data, the typical-group prediction is the best one can get, unless local calibration data from the group in question are available for prediction of the random effects.

[1] For forestry-related examples, see Lappi 1986, 1991, 1997, Lappi and Bailey 1988

Some notes on prediction

- Mixed-effects allows *group-level prediction* where the predicted random effect is used.
- If no measurements of *y* are available from the group in question, the BLUP of random effect is its expected value 0, the prediction for *a typical group*
- Group-level predictions utilize the observed values of the response from the group in question. For groups not present in the modeling data, the typical-group prediction is the best one can get, unless local calibration data from the group in question are available for prediction of the random effects.
- Prediction of random-effect for an previously fitted model provides a highly useful application, which has a Bayesian flavour. [1]

[1] For forestry-related examples, see Lappi 1986, 1991, 1997, Lappi and Bailey 1988

Example 1: Stemwood volume of eucalyptus trees

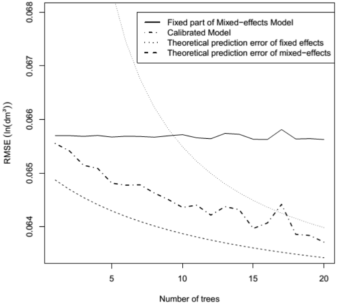
Model

$$\ln(v_{ij}) = \beta_0 + \beta_1 \ln(dbh_{ij}) + \beta_2 \ln(h_{ij}) + b_i + \epsilon_{ij}$$

was fitted for the volume of Eucalyptus trees j on farms i, using a stem analysis data of 1434 stems from 15 farms (de Souza Vismara et al. 2016).

The parameter estimates for random part were $\widehat{\sigma_b^2} = 0.18^2$ and $\widehat{\sigma^2} = 0.62^2$.

Therefore, some benefit may be obtained by prediction of random effects, as shown below.



More advanced mixed-effects models

- One may have other random effects than just constant:

$$y_{ij} = \beta' \boldsymbol{x}_{ij} + \boldsymbol{b}_i' \boldsymbol{z}_{ij} + \epsilon_{ij}$$

where $\boldsymbol{z}_{ij}$ includes $\boldsymbol{x}_{ij}$ or part of it, and $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ (i.i.d).

More advanced mixed-effects models

- One may have other random effects than just constant:

$$y_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + \boldsymbol{b}_i' \boldsymbol{z}_{ij} + \epsilon_{ij}$$

  where $\boldsymbol{z}_{ij}$ includes $\boldsymbol{x}_{ij}$ or part of it, and $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ (i.i.d).
- For two nested groups, we specify

$$y_{ijk} = \boldsymbol{\beta}' \boldsymbol{x}_{ijk} + \boldsymbol{a}_i' \boldsymbol{z}_{ijk}^{(a)} + \boldsymbol{c}_{ij}' \boldsymbol{z}_{ijk}^{(c)} + \epsilon_{ijk}$$

  where $\boldsymbol{z}_{ijk}^{(a)}$ includes $\boldsymbol{x}_{ijk}$ or part of it, and $\boldsymbol{z}_{ijk}^{(c)}$ includes $\boldsymbol{z}_{ijk}^{(a)}$ or part of it, and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_{ij} \sim N(0, \boldsymbol{D}_c)$ (i.i.d).

More advanced mixed-effects models

- For two crossed groups [2], we specify

$$y_{ijk} = \beta' \boldsymbol{x}_{ijk} + \boldsymbol{a}_i' \boldsymbol{z}_{ijk}^{(a)} + \boldsymbol{c}_j' \boldsymbol{z}_{ijk}^{(c)} + \epsilon_{ijk}$$

where $\boldsymbol{z}_{ijk}^{(a)}$ and $\boldsymbol{z}_{ijk}^{(c)}$ includes $\boldsymbol{x}_{ijk}$ or part of it and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_j \sim N(0, \boldsymbol{D}_c)$ (i.i.d).

[2] e.g. Mehtätalo et al 2014, Korpela et al. 2014

[3] e.g. Lappi 1991, Mehtätalo 2005, Lappi et al. 2006, Maltamo et al 2012. Korpela et al 2014

More advanced mixed-effects models

- For two crossed groups [2], we specify

$$y_{ijk} = \boldsymbol{\beta}'\boldsymbol{x}_{ijk} + \boldsymbol{a}_i'\boldsymbol{z}_{ijk}^{(a)} + \boldsymbol{c}_j'\boldsymbol{z}_{ijk}^{(c)} + \epsilon_{ijk}$$

where $\boldsymbol{z}_{ijk}^{(a)}$ and $\boldsymbol{z}_{ijk}^{(c)}$ includes $\boldsymbol{x}_{ijk}$ or part of it and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_j \sim N(0, \boldsymbol{D}_c)$ (i.i.d).

- A bivariate LMM (with single level of grouping) may be specified by [3]

$$\begin{aligned} y1_{ij} &= \boldsymbol{\beta}'\boldsymbol{x}1_{ij} + \boldsymbol{b}1_i'\boldsymbol{z}1_{ij} + \epsilon1_{ij} \\ y2_{ij} &= \boldsymbol{\beta}'\boldsymbol{x}2_{ij} + \boldsymbol{b}2_i'\boldsymbol{z}2_{ij} + \epsilon2_{ij} \end{aligned}$$

where $(\boldsymbol{b}1_k', \boldsymbol{b}2_i')' \sim N(0, \boldsymbol{D})$ (iid) and $(\epsilon1_k, \epsilon2_k)' \sim N(0, \boldsymbol{R})$.

[2] e.g. Mehtätalo et al 2014, Korpela et al. 2014

[3] e.g. Lappi 1991, Mehtätalo 2005, Lappi et al. 2006, Maltamo et al 2012. Korpela et al 2014

More advanced mixed-effects models

- For two crossed groups [2], we specify

$$y_{ijk} = \beta' \boldsymbol{x}_{ijk} + \boldsymbol{a}_i' \boldsymbol{z}_{ijk}^{(a)} + \boldsymbol{c}_j' \boldsymbol{z}_{ijk}^{(c)} + \epsilon_{ijk}$$

  where $\boldsymbol{z}_{ijk}^{(a)}$ and $\boldsymbol{z}_{ijk}^{(c)}$ includes $\boldsymbol{x}_{ijk}$ or part of it and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_j \sim N(0, \boldsymbol{D}_c)$ (i.i.d).
- A bivariate LMM (with single level of grouping) may be specified by [3]

$$\begin{aligned} y1_{ij} &= \beta' \boldsymbol{x}1_{ij} + \boldsymbol{b}1_i' \boldsymbol{z}1_{ij} + \epsilon 1_{ij} \\ y2_{ij} &= \beta' \boldsymbol{x}2_{ij} + \boldsymbol{b}2_i' \boldsymbol{z}2_{ij} + \epsilon 2_{ij} \end{aligned}$$

  where $(\boldsymbol{b}1_k', \boldsymbol{b}2_i')' \sim N(0, \boldsymbol{D})$ (iid) and $(\epsilon 1_k, \epsilon 2_k)' \sim N(0, \boldsymbol{R})$.
- The assumption of constant error variance can also be relaxed using variance functions/ correlation structures.

[2] e.g. Mehtätalo et al 2014, Korpela et al. 2014

[3] e.g. Lappi 1991, Mehtätalo 2005, Lappi et al. 2006, Maltamo et al 2012. Korpela et al 2014

More advanced mixed-effects models

- For two crossed groups [2], we specify

$$y_{ijk} = \boldsymbol{\beta}' \boldsymbol{x}_{ijk} + \boldsymbol{a}_i' \boldsymbol{z}_{ijk}^{(a)} + \boldsymbol{c}_j' \boldsymbol{z}_{ijk}^{(c)} + \epsilon_{ijk}$$

  where $\boldsymbol{z}_{ijk}^{(a)}$ and $\boldsymbol{z}_{ijk}^{(c)}$ includes $\boldsymbol{x}_{ijk}$ or part of it and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_j \sim N(0, \boldsymbol{D}_c)$ (i.i.d).
- A bivariate LMM (with single level of grouping) may be specified by [3]

$$\begin{aligned} y1_{ij} &= \boldsymbol{\beta}' \boldsymbol{x}1_{ij} + \boldsymbol{b}1_i' \boldsymbol{z}1_{ij} + \epsilon1_{ij} \\ y2_{ij} &= \boldsymbol{\beta}' \boldsymbol{x}2_{ij} + \boldsymbol{b}2_i' \boldsymbol{z}2_{ij} + \epsilon2_{ij} \end{aligned}$$

  where $(\boldsymbol{b}1_k', \boldsymbol{b}2_i')' \sim N(0, \boldsymbol{D})$ (iid) and $(\epsilon1_k, \epsilon2_k)' \sim N(0, \boldsymbol{R})$.
- The assumption of constant error variance can also be relaxed using variance functions/ correlation structures.
- Parameter estimation can be based on (RE)ML/GLS.

[2] e.g. Mehtätalo et al 2014, Korpela et al. 2014

[3] e.g. Lappi 1991, Mehtätalo 2005, Lappi et al. 2006, Maltamo et al 2012. Korpela et al 2014

More advanced mixed-effects models

- For two crossed groups [2], we specify

$$y_{ijk} = \boldsymbol{\beta}'\boldsymbol{x}_{ijk} + \boldsymbol{a}_i'\boldsymbol{z}^{(a)}_{ijk} + \boldsymbol{c}_j'\boldsymbol{z}^{(c)}_{ijk} + \epsilon_{ijk}$$

  where $\boldsymbol{z}^{(a)}_{ijk}$ and $\boldsymbol{z}^{(c)}_{ijk}$ includes $\boldsymbol{x}_{ijk}$ or part of it and $\boldsymbol{a}_i \sim N(0, \boldsymbol{D}_a)$ (i.i.d) and $\boldsymbol{c}_j \sim N(0, \boldsymbol{D}_c)$ (i.i.d).
- A bivariate LMM (with single level of grouping) may be specified by [3]

$$\begin{aligned} y1_{ij} &= \boldsymbol{\beta}'\boldsymbol{x}1_{ij} + \boldsymbol{b}1_i'\boldsymbol{z}1_{ij} + \epsilon 1_{ij} \\ y2_{ij} &= \boldsymbol{\beta}'\boldsymbol{x}2_{ij} + \boldsymbol{b}2_i'\boldsymbol{z}2_{ij} + \epsilon 2_{ij} \end{aligned}$$

  where $(\boldsymbol{b}1_k', \boldsymbol{b}2_i')' \sim N(0, \boldsymbol{D})$ (iid) and $(\epsilon 1_k, \epsilon 2_k)' \sim N(0, \boldsymbol{R})$.
- The assumption of constant error variance can also be relaxed using variance functions/ correlation structures.
- Parameter estimation can be based on (RE)ML/GLS.
- Prediction of random effect is based on the general formulation of BLUP.

[2] e.g. Mehtätalo et al 2014, Korpela et al. 2014

[3] e.g. Lappi 1991, Mehtätalo 2005, Lappi et al. 2006, Maltamo et al 2012. Korpela et al 2014

BLUP - the general case

- Consider random vector $\mathbf{h}$ which is partitioned as follows:

$$\mathbf{h} = \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}$$

and has the following mean and variance:

$$\begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix} \sim \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_{12} \\ \mathbf{V}'_{12} & \mathbf{V}_2 \end{pmatrix} \right]$$

- Consider a situation where the value of $\mathbf{h}_2$ has been observed and one wants to predict the value of unobserved vector $\mathbf{h}_1$.
- The Best Linear Unbiased Predictor (BLUP) of $\mathbf{h}_1$ is

$$BLUP(\mathbf{h}_1) = \widetilde{\mathbf{h}_1} = \mu_1 + \mathbf{V}_{12}\mathbf{V}_2^{-1}(\mathbf{h}_2 - \mu_2) \tag{1}$$

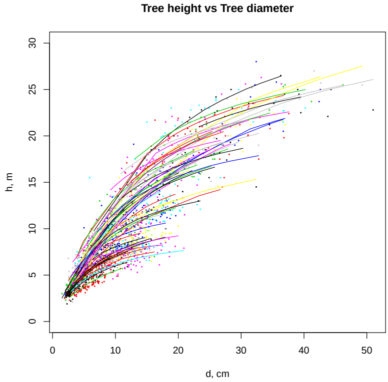
- The prediction variance is

$$\mathrm{var}(\widetilde{\mathbf{h}_1} - \mathbf{h}_1) = \mathbf{V}_1 - \mathbf{V}_{12}\mathbf{V}_2^{-1}\mathbf{V}'_{12} \tag{2}$$

- If $\mathbf{h}$ has multivariate normal distribution, BLUP is BP.
- If the mean and variances are estimates, the resulting estimator is called Estimated or empirical BLUP (EBLUP).

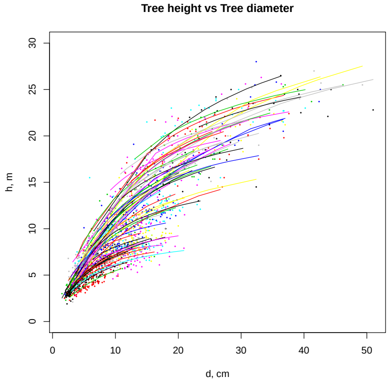
Example 2: A longitudinal H-D model

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



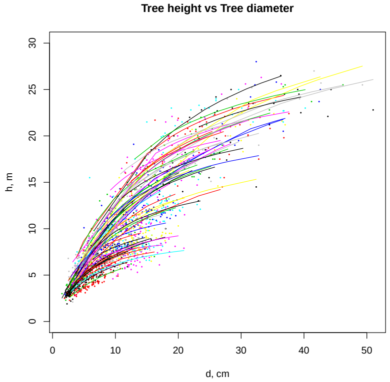
Example 2: A longitudinal H-D model

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



Example 2: A longitudinal H-D model

- H-D relationship varies much among groups (sample plots), but height measurement is time-consuming.
- In a forest inventory, diameter is usually 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_{ijk} for tree k in stand i at time j with transformed diameter D_{ijk} at the breast height is expressed by [4]

$$\begin{aligned}\ln(H_{ijk}) &= \beta_0(DGM_{ij}) + a_i^{(1)} + c_{ij}^{(1)} + (\beta_1(DGM_{kt}) + a_i^{(2)} + c_{ij}^{(2)})D_{ijk} + \epsilon_{ijk} \\ &= \beta_0(DGM_{ij}) + \beta_1(DGM_{kt})D_{ijk} + a_i^{(1)} + a_i^{(2)}D_{ijk} + c_{ij}^{(1)} + c_{ij}^{(2)}D_{ijk} + \epsilon_{ijk}\,,\end{aligned}$$

where

- $\beta_0(DGM_{ij})$ and $\beta_1(DGM_{ij})$ are known fixed functions of plot-specific mean diameter DGM_{ij},
- $\boldsymbol{a} = (a_i^{(1)}, a_i^{(2)})'$ are plot-level random effects
- $\boldsymbol{c} = (c_{ij}^{(1)}, c_{ij}^{(2)})'$ are measurement occasion -level random effects

[4] Mehtätalo 2004, 2005

The Height-Diameter model

The logarithmic heigth H_{ijk} for tree k in stand i at time j with transformed diameter D_{ijk} at the breast height is expressed by [4]

$$\begin{aligned}\ln(H_{ijk}) &= \beta_0(DGM_{ij}) + a_i^{(1)} + c_{ij}^{(1)} + (\beta_1(DGM_{kt}) + a_i^{(2)} + c_{ij}^{(2)})D_{ijk} + \epsilon_{ijk} \\ &= \beta_0(DGM_{ij}) + \beta_1(DGM_{kt})D_{ijk} + a_i^{(1)} + a_i^{(2)}D_{ijk} + c_{ij}^{(1)} + c_{ij}^{(2)}D_{ijk} + \epsilon_{ijk}\,,\end{aligned}$$

where

- $\beta_0(DGM_{ij})$ and $\beta_1(DGM_{ij})$ are known fixed functions of plot-specific mean diameter DGM_{ij},
- $\boldsymbol{a} = (a_i^{(1)}, a_i^{(2)})'$ are plot-level random effects
- $\boldsymbol{c} = (c_{ij}^{(1)}, c_{ij}^{(2)})'$ are measurement occasion -level random effects
- The variances (correlations) were estimated to be

$$\mathrm{var}(\boldsymbol{a}_i) = \begin{bmatrix} 0.108^2 & (0.269) \\ 0.0028 & 0.0958^2 \end{bmatrix} \qquad \mathrm{var}(\boldsymbol{c}_{ij}) = \begin{bmatrix} 0.0168^2 & (-0.681) \\ -0.0003 & 0.0223^2 \end{bmatrix}$$

[4] Mehtätalo 2004, 2005

The Height-Diameter model

The logarithmic heigth H_{ijk} for tree k in stand i at time j with transformed diameter D_{ijk} at the breast height is expressed by [4]

$$
\begin{aligned}
\ln(H_{ijk}) &= \beta_0(DGM_{ij}) + a_i^{(1)} + c_{ij}^{(1)} + (\beta_1(DGM_{kt}) + a_i^{(2)} + c_{ij}^{(2)})D_{ijk} + \epsilon_{ijk} \\
&= \beta_0(DGM_{ij}) + \beta_1(DGM_{kt})D_{ijk} + a_i^{(1)} + a_i^{(2)}D_{ijk} + c_{ij}^{(1)} + c_{ij}^{(2)}D_{ijk} + \epsilon_{ijk}\,,
\end{aligned}
$$

where

- $\beta_0(DGM_{ij})$ and $\beta_1(DGM_{ij})$ are known fixed functions of plot-specific mean diameter DGM_{ij},
- $\boldsymbol{a} = (a_i^{(1)}, a_i^{(2)})'$ are plot-level random effects
- $\boldsymbol{c} = (c_{ij}^{(1)}, c_{ij}^{(2)})'$ are measurement occasion -level random effects
- The variances (correlations) were estimated to be

$$
\mathrm{var}(\boldsymbol{a}_i) = \begin{bmatrix} 0.108^2 & (0.269) \\ 0.0028 & 0.0958^2 \end{bmatrix} \qquad \mathrm{var}(\boldsymbol{c}_{ij}) = \begin{bmatrix} 0.0168^2 & (-0.681) \\ -0.0003 & 0.0223^2 \end{bmatrix}
$$

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

[4] Mehtätalo 2004, 2005

The stand level mixed-effects model

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

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i ,$$

where
$\boldsymbol{y}_i$ includes the observed sample tree heights,
$\boldsymbol{X}_i\boldsymbol{\beta}$ is the fixed part,
$\boldsymbol{b}_i = \begin{pmatrix} a_i^{(1)} & a_i^{(2)} & c_{i1}^{(1)} & c_{i1}^{(2)} & c_{i2}^{(1)} & c_{i2}^{(2)} & \ldots \end{pmatrix}'$ includes the random effects,
$\boldsymbol{Z}_i$ is the random part design matrix of the group, and
$\boldsymbol{\epsilon}_i$ includes the residuals.
We denote $\mathrm{var}(\boldsymbol{b}_i) = \boldsymbol{D}$ and $\mathrm{var}(\boldsymbol{\epsilon_i}) = \boldsymbol{R_i}$.

Prediction of random effects

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

$$\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{y}_i \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_i\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{D}\boldsymbol{Z}_i' \\ \boldsymbol{Z}_i\boldsymbol{D} & \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i \end{bmatrix} \right)$$

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

$$\widetilde{\boldsymbol{b}_i} = \boldsymbol{D}\boldsymbol{Z}_i'(\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \,.$$

and the variance of prediction errors is

$$\mathrm{var}(\widetilde{\boldsymbol{b}_i} - \boldsymbol{b}_i) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i'(\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i)^{-1}\boldsymbol{Z}_i\boldsymbol{D}$$

Example

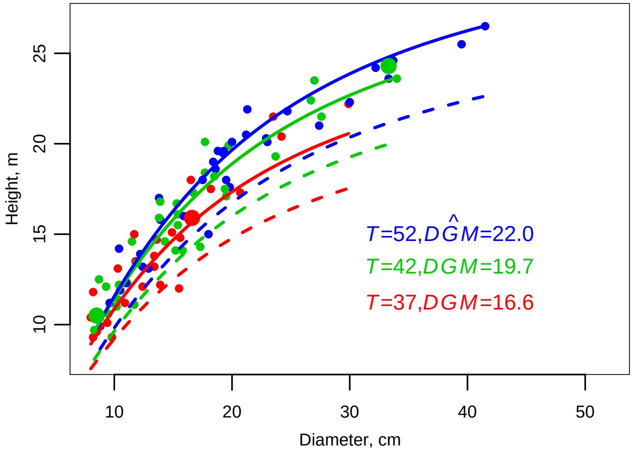
Height of one tree was measured 5 years ago and 2 trees at the current year for plot i. The matrices and vectors are

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

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

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha_i \\ \beta_i \\ \alpha_{i1} \\ \beta_{i1} \\ \alpha_{i2} \\ \beta_{i2} \end{bmatrix} \quad \boldsymbol{D} = \begin{bmatrix} 0.0118 & 0.0028 & 0 & 0 & 0 & 0 \\ 0.0028 & 0.0092 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0003 & 0.0004 & 0 & 0 \\ 0 & 0 & 0.0004 & 0.0005 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0003 & 0.0004 \\ 0 & 0 & 0 & 0 & 0.0004 & 0.0005 \end{bmatrix}$$

Uncalibrated and calibrated predictions



Dashed shows prediction based on fixed part. Three trees (large symbols) were used to predict the random effects to get plot-level predictions (solid)

Example 3: 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. [5]
- The reflectance (color) of a tree on an image can be used to classify tree species

[5] Korpela et al. 2014

Example 3: 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. [5]
- 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.

[5] Korpela et al. 2014

Example 3: 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. [5]
- 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

[5] Korpela et al. 2014

Example 3: 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. [5]
- 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.

[5] Korpela et al. 2014

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 wavelengths: 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 wavelengths: 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 wavelengths: 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)
- Trees of different images were 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 wavelengths: 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)
- Trees of different images were 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 (crossed) structure

$$y_{ij} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + b_i + c_j + \epsilon_{ij} ,$$

where i and j refer to image and tree effects, respectively. σ_i^2 and σ_j^2 are the corresponding variances. The predictors $\boldsymbol{x}_{ij}$ 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 (crossed) structure

$$y_{ij} = \boldsymbol{\beta}' \boldsymbol{x}_{ij} + b_i + c_j + \epsilon_{ij} ,$$

  where i and j refer to image and tree effects, respectively. σ_i^2 and σ_j^2 are the corresponding variances. The predictors $\boldsymbol{x}_{ij}$ are trigonometric transformations of the horizontal and vertical viewing and Sun angles.
- The random effects at different levels of grouping are independent, therefore

$$\begin{aligned} var(y_{ij}) &= \sigma_i^2 + \sigma_j^2 + \sigma^2 \\ cov(y_{ij}, y_{i'j'}) &= 0 \\ cov(y_{ij}, y_{ij'}) &= \sigma_i^2 \\ cov(y_{ij}, y_{i'j}) &= \sigma_j^2 \end{aligned}$$

The multivariate model

The multivariate model for a tree species is

$$\begin{aligned} y1_{ij} &= \boldsymbol{\beta}1' x1_{ij} + b1_i + c1_j + \epsilon1_{ij} \\ &\vdots \\ y8_{ij} &= \boldsymbol{\beta}8' x8_{ij} + b8_i + c8_j + \epsilon8_{ij} \end{aligned}$$

or simply

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{x}_{ij} + \boldsymbol{b}_i + \boldsymbol{c}_j + \boldsymbol{\epsilon}_{ij}$$

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

- $(b1_i, \ldots, b8_i)' = \boldsymbol{b}_i \sim N(0, \boldsymbol{A}_{8\times8})$ include the random image-effects
- $(c1_j, \ldots, c8_j)' = \boldsymbol{c}_j \sim N(0, \boldsymbol{B}_{8\times8})$ include the random tree-effects
- $(\epsilon1_{ij}, \ldots, \epsilon8_{ij})' = \boldsymbol{\epsilon}_{ij} \sim N(0, \boldsymbol{E}_{8\times8})$ include the residual errors

The multivariate model (continued)

- 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 (continued)

- Now

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

- Model fitting (based on REML/ML/GLS) yields $\widehat{\beta}$, $\widehat{\boldsymbol{A}}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{E}}$

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 $\boldsymbol{y}_{ij}$ be an observed vector (length=8) of the reflectances of one tree j on the 8 channels on one image i. The squared Mahalanobis distance between $\boldsymbol{y}_{ij}$ and $\boldsymbol{\mu}_{ij}$ is
$$d_{ij}^2 = (\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{ij})'(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{E})^{-1}(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{ij})$$
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 (continued)

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

$$d_{\cdot j}^2 = (\boldsymbol{y}_{\cdot j} - \boldsymbol{\mu}_{\cdot j})' \boldsymbol{D}_{\cdot j}^{-1} (\boldsymbol{y}_{\cdot j} - \boldsymbol{\mu}_{\cdot j}) \,,$$

where $\boldsymbol{y}_{\cdot j} = (\boldsymbol{y}_{1j}', \ldots, \boldsymbol{y}_{mj})$ is an observed vector (with length of $8m$) of the reflectances of tree j on the 8 channels of m images. The $8m \times 8m$ variance-covariance matrix is

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

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

The use in classification (continued)

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

$$d_{\cdot j}^2 = (\boldsymbol{y}_{\cdot j} - \boldsymbol{\mu}_{\cdot j})' \boldsymbol{D}_{\cdot j}^{-1} (\boldsymbol{y}_{\cdot j} - \boldsymbol{\mu}_{\cdot j}) \,,$$

  where $\boldsymbol{y}_{\cdot j} = (\boldsymbol{y}_{1j}', \ldots, \boldsymbol{y}_{mj})$ is an observed vector (with length of $8m$) of the reflectances of tree j on the 8 channels of m images. The $8m \times 8m$ variance-covariance matrix is

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

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

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

Example 4: Eucalyptus volumes on two rotations

A bivariate volume model

$$\begin{aligned} \ln(v_{1ij}) &= \boldsymbol{\beta}_1' \boldsymbol{x}_{1ij} + b_i^{(1)} + \epsilon_{1ij} \\ \ln(v_{2ij}) &= \boldsymbol{\beta}_2' \boldsymbol{x}_{2ij} + b_i^{(2)} + \epsilon_{2ij} \end{aligned}$$

was used for rotations 1 and 2 of Eucalyptus plantations [6].
The parameter estimates for random part were

$$\widehat{\mathrm{var}} \begin{pmatrix} b_i^{(1)} \\ b_i^{(2)} \end{pmatrix} = \begin{pmatrix} 0.0192^2 & 0.0005170176 \\ 0.0005170176 & 0.0272^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{H} \end{pmatrix}$$

and

$$\widehat{\mathrm{var}} \begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} = \begin{pmatrix} 0.0624 & 0 \\ 0 & 0.0596^2 \end{pmatrix}$$

[6] de Souza Vismara, Mehtatalo and Batista 2016

Example 4: Eucalyptus volumes on two rotations

A bivariate volume model

$$\begin{aligned} \ln(v_{1ij}) &= \boldsymbol{\beta}_1' \boldsymbol{x}_{1ij} + b_i^{(1)} + \epsilon_{1ij} \\ \ln(v_{2ij}) &= \boldsymbol{\beta}_2' \boldsymbol{x}_{2ij} + b_i^{(2)} + \epsilon_{2ij} \end{aligned}$$

was used for rotations 1 and 2 of Eucalyptus plantations [6].
The parameter estimates for random part were

$$\widehat{\mathrm{var}} \begin{pmatrix} b_i^{(1)} \\ b_i^{(2)} \end{pmatrix} = \begin{pmatrix} 0.0192^2 & 0.0005170176 \\ 0.0005170176 & 0.0272^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{H} \end{pmatrix}$$

and

$$\widehat{\mathrm{var}} \begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} = \begin{pmatrix} 0.0624 & 0 \\ 0 & 0.0596^2 \end{pmatrix}$$

The correlation between random effects is high (0.99), therefore both models could be calibrated by using observations of first rotation only.

[6] de Souza Vismara, Mehtatalo and Batista 2016

Example 4: Eucalyptus volumes on two rotations

A bivariate volume model

$$\begin{aligned} \ln(v_{1ij}) &= \beta_1' \boldsymbol{x}_{1ij} + b_i^{(1)} + \epsilon_{1ij} \\ \ln(v_{2ij}) &= \beta_2' \boldsymbol{x}_{2ij} + b_i^{(2)} + \epsilon_{2ij} \end{aligned}$$

was used for rotations 1 and 2 of Eucalyptus plantations [6].
The parameter estimates for random part were

$$\widehat{\mathrm{var}} \begin{pmatrix} b_i^{(1)} \\ b_i^{(2)} \end{pmatrix} = \begin{pmatrix} 0.0192^2 & 0.0005170176 \\ 0.0005170176 & 0.0272^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{H} \end{pmatrix}$$

and

$$\widehat{\mathrm{var}} \begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} = \begin{pmatrix} 0.0624 & 0 \\ 0 & 0.0596^2 \end{pmatrix}$$

The correlation between random effects is high (0.99), therefore both models could be calibrated by using observations of first rotation only.
The error variance is high compared to that of random effects, $\rightarrow$ calibration effects will be only modest.

[6] de Souza Vismara, Mehtatalo and Batista 2016

BLUP in this case

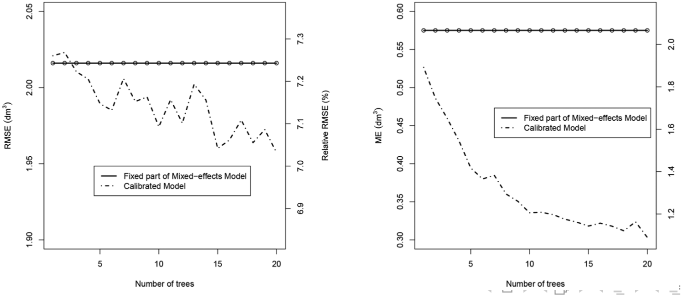
We have now

$$\begin{bmatrix} \boldsymbol{b}_i \\ \ln \boldsymbol{v}_{1i} \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{C}\boldsymbol{Z}_{1i}' \\ \boldsymbol{Z}_{1i}\boldsymbol{C}' & \boldsymbol{Z}_{1i}\mathrm{var}(b_i^{(1)})\boldsymbol{Z}_{1i}' + \boldsymbol{R}_{1i} \end{bmatrix} \right)$$

Leading to EBLUP:

$$\widetilde{\boldsymbol{b}_i} = \boldsymbol{C}\boldsymbol{Z}_{1i}' \left(\boldsymbol{Z}_{1i}\mathrm{var}(b_i^{(1)})\boldsymbol{Z}_{1i}' + \boldsymbol{R}_{1i} \right)^{-1} (\ln \boldsymbol{v}_{1i} - \boldsymbol{X}_{1i}\boldsymbol{\beta}_1) \ .$$

etc..



Example 5: System of mixed-effects model for aerial forest inventory

- Airborne Laser Scanners (ALS) provide information on the 3*D*- structure of forest
- Majority of large individual trees can be detected from an ALS point cloud
- Point cloud characteristics can be assigned to field-measured tree characteristics to estimate a system of predictive models for tree characteristics, such as stem volume, height, diameter, crown base height, dead crown height.
- These tree-specific characteristics are correlated within a forest stand
- Also the stand effects are correlated across models
- These correlations can be utilized to predict the random effects of a mixed-effects model for a given stand for all 5 models using even one observation of one characteristics only
- Enables improved predictions of hard-to-measure characteristics by using easy-to-measure characteristics. [7]

[7] Maltamo et al 2012

The model

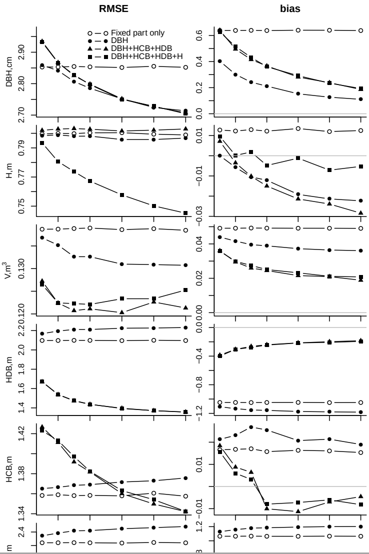
The model includes a system of 5 mixed-effects models of form for tree i in stand k:

$$
\begin{aligned}
y1_{ki} &= a1 + b1x1_{ki} + \ldots + \alpha 1_k + \beta 1_k x1_{ki} + \epsilon 1_{ki} \\
y2_{ki} &= a2 + b1x2_{ki} + \ldots + \alpha 2_k + \beta 2_k x2_{ki} + \epsilon 2_{ki} \\
&\vdots \\
y5_{ki} &= a5 + b5x5_{ki} + \ldots + \alpha 5_k + \beta 5_k x5_{ki} + \epsilon 5_{ki}
\end{aligned}
$$

where the fixed parts are as with the previous mixed-effects models and include the ALS-based predictors.

- The assumptions on the random effects and residuals are $(\alpha 1_k, \beta 1_k, \alpha 2_k, \beta 2_k, \ldots, \alpha 5_k, \beta 5_k)' \sim MVN(0, \boldsymbol{D}_{10x10})$, and $(\epsilon 1_{k1}, \epsilon 2_{ki}, \ldots, \epsilon 5_{ki}) \sim MVN(0, \boldsymbol{R}_{5x5})$
- The intended use of the model is prediction applying the random effects.
- The previously presented principles were used to predict the random effects of the model system by using 1-10 sample trees per stand and 3 different measurement strategies

Results


RMSE
bias
Fixed part only
DBH
DBH+HCB+HDB
DBH+HCB+HDB+H
DBH,cm
H,m
V,m³
HDB,m
HCB,m

Summary

- Random-effect prediction is a widely applicable tool for modeling grouped data where groups constitute a sample from a population of groups.
- Random effects may be justified for many different purposes, and modeling procedures should be adopted for the purpose of modeling.
  - local predictions through random effects.
  - statistical inference in grouped datasets
  - variance partitioning
  - estimation of variance-covariance structure for classification

Questions

1. Compute $\widetilde{\boldsymbol{b}}_i$ and $(\widetilde{\boldsymbol{b}}_i - \boldsymbol{b}_i)$ using the matrices of example 2. Use the computed value of $\widetilde{\boldsymbol{b}}_i$ to recover the relationship between Diameter and Height for the two points in time of the example. Compare to the figure shown in the notes.
2. Consider your own area of interest and describe such a problem where mixed-effect models could be used for group-specific prediction or classification.

References

- de Souza Vismara, E., Mehtätalo, L., and Batista, J.L.F. 2016. Linear mixed-effects models and calibration applied to volume models in two rotations of Eucalyptus grandis plantations. Can. J. For. Res. 46(1): 132-141
- Korpela, I., Mehtätalo, L., Markelin, L., Seppänen, A. and Kangas, A. 2014.Tree species identification in aerial image data using directional reflectance signatures. Silva Fenn. 48(3): 1087
- Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. Communicationes Instituti Forestalis Fenniae 134. 69 p.
- Lappi, J. and Bailey, R. L. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. For. Sci 34: 907–927.
- Lappi, J. 1991. Calibration of height and volume equations with random parameters. For. Sci. 37(3): 781-801.
- Lappi, J. 1997. A longitudinal analysis of height/diameter curves. For. Sci. 43. 555–570.
- Lappi, J., Mehtätalo, L. and Korhonen, K.T. 2006. Generalizing sample tree information. in Kangas, A. and Maltamo, M (editors) Forest inventory - methodology & applications. Springer.
- 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.
- Mehtätalo, L. 2005b. Localizing a Predicted Diameter Distribution Using Sample Information For. Sci. 51(4):292Ű303.
- Mehtätalo, L., Peltola, H., Kilpeläinen, A. and Ikonen, V.-P. 2014. The effect of thinning on the basal area growth of Scots Pine: a longitudinal analysis using nonlinear mixed-effects model. For.Sci. 60(4):636-644
- Maltamo, M., Mehtätalo, L., Vauhkonen, J. and Packalén P. 2012. Predicting and calibrating tree attributes by means of airborne laser scanning and field measurements. Can. J. For. Res. 42: 1896-1907

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.

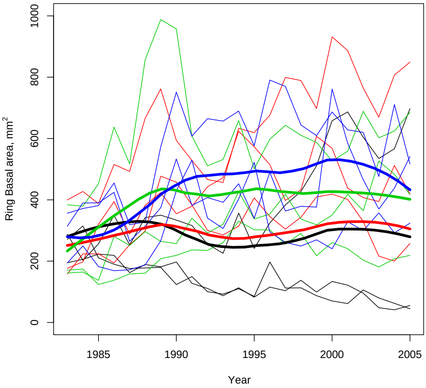
Study material

- 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



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

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

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

$$y_{kt} = f(T_{kt}; \boldsymbol{b}) + \alpha_k + \alpha_t + \epsilon_{kt} \quad (3)$$

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

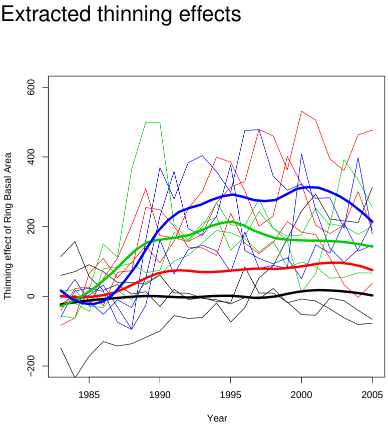
- Using the estimated age trend and BLUP's of year and tree effects, the growth without thinning, $\widetilde{y}_{kt}$ was predicted for treatments II -IV after the thinning year.

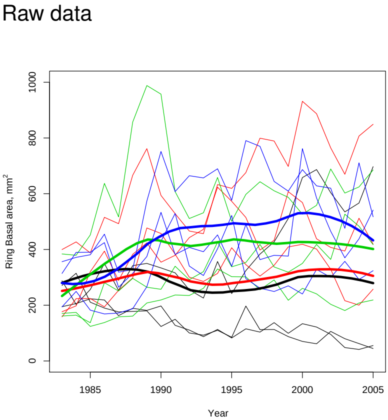
Extracting the thinning effects

- 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

$$d_{kt} = y_{kt} - \widetilde{y}_{kt} \tag{4}$$

The estimated thinning effects





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 using 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.

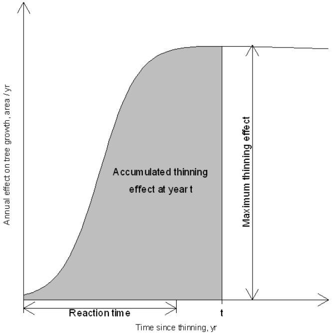
Modeling the thinning effects

- 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		