# Mixed-effect models for nonlinear natural processes

Lauri Mehtätalo[1] Anna Laine[2] Heli Peltola[3] Eeva-Stiina Tuittila[3]
Antti Kilpeläinen[3]

[1]University of Eastern Finland, School of Computing

[2]University of Oulu, Department of Biology

[3]University of Eastern Finland, School of Forest Sciences

IUFRO 125th Anniversary Congress, Freiburg
September 18-22, 2017

# Outline

# Models for a nonlinear relationship

There are two options to model a nonlinear relationship between a response variable $y$ and predictor (vector) $\boldsymbol{x}$:

1. The linear model

$$y_i = \beta_0 + \beta_1 f_1(\boldsymbol{x}_i) + \ldots + \beta_K f_K(\boldsymbol{x}_i) + e_i \,,$$

where $\beta_1, \ldots, \beta_K$ are parameters to be estimated, $f_1(\boldsymbol{x}_i), \ldots, f_K(\boldsymbol{x}_i)$ are nonlinear transformations of predictors $\boldsymbol{x}_i$, $y_i$ is the (possibly transformed) response, and $e_i$ is residual error for sampling unit $i$.

- Often $f_k()$'s are one-to-one nonlinear functions of a single component of $\boldsymbol{x}_i$, such as logarithmic and power transformations or spline components[1].
- Term linear in the model refers to linearity in $\beta_k$'s, not in $\boldsymbol{x}_i$.
- Can successfully model any nonlinear relationship between $y$ and $\boldsymbol{x}$ through transformations $f_k()$.

[1] Harrell F., 2001: Regression modeling strategies. Springer.

# Models for a nonlinear relationship

There are two options to model a nonlinear relationship between a response variable $y$ and predictor (vector) $\boldsymbol{x}$:

1. The linear model

$$y_i = \beta_0 + \beta_1 f_1(\boldsymbol{x}_i) + \ldots + \beta_K f_K(\boldsymbol{x}_i) + e_i \,,$$

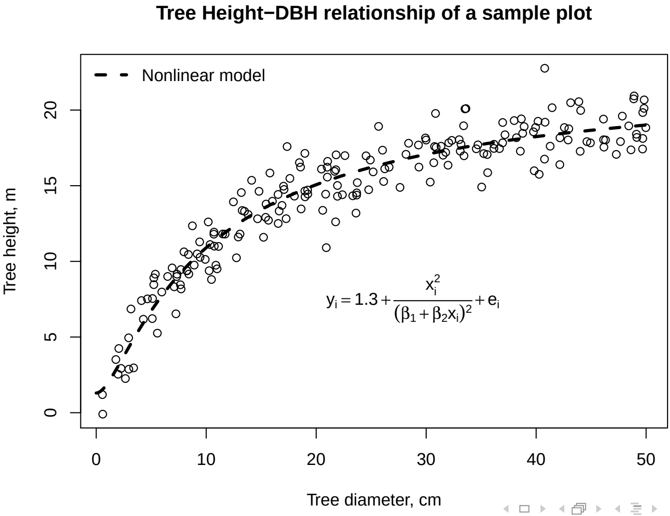
   where $\beta_1, \ldots, \beta_K$ are parameters to be estimated, $f_1(\boldsymbol{x}_i), \ldots, f_K(\boldsymbol{x}_i)$ are nonlinear transformations of predictors $\boldsymbol{x}_i$, $y_i$ is the (possibly transformed) response, and $e_i$ is residual error for sampling unit $i$.
   - Often $f_k()$'s are one-to-one nonlinear functions of a single component of $\boldsymbol{x}_i$, such as logarithmic and power transformations or spline components[1].
   - Term linear in the model refers to linearity in $\beta_k$'s, not in $\boldsymbol{x}_i$.
   - Can successfully model any nonlinear relationship between $y$ and $\boldsymbol{x}$ through transformations $f_k()$.
2. The nonlinear model

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + e_i \,,$$

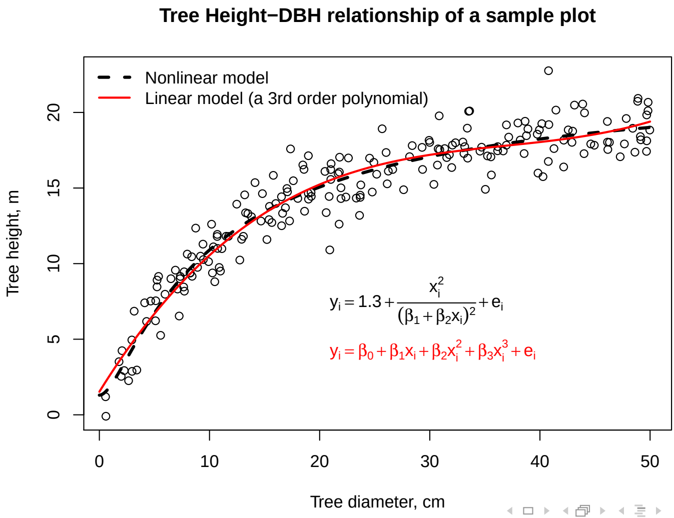
   - Where $f()$ is a function of predictors and model parameters.

[1] Harrell F., 2001: Regression modeling strategies. Springer.

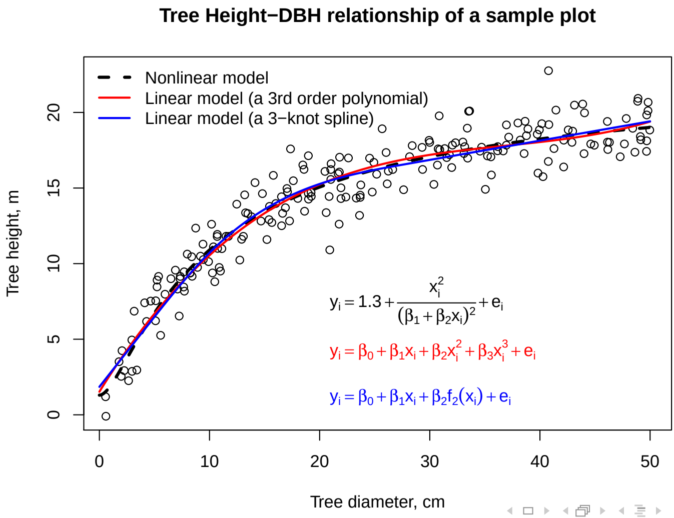
# Illustration



# Illustration



# Illustration



# Why nonlinear models?

No improvement to model fit is expected with nonlinear models compared to the linear model. Why then to use a nonlinear model?

- Building the model on the subject-matter theory on the process is elegant (recall what Göran Ståhl said on Monday),
- leads to parameters that are as such of interest and have interesting interpretations, and
- allows in-depth analyses on the effects of different predictors (e.g. treatments or continuous predictors) on the parameters of the process.
- The models are also more robust in extrapolation and
- parameter-parsimonious.

# Nonlinear mixed-effect model formulation

- In a nonlinear fixed-effects model, one of the predictors ($t_i$) is often a primary predictor of the process (e.g. time in a growth model, or photosynthetically active radiation (PAR) in a model of net photosynthesis). The other secondary predictors $\boldsymbol{x}_i$ describe the variability in the primary parameters of the process. This leads to model

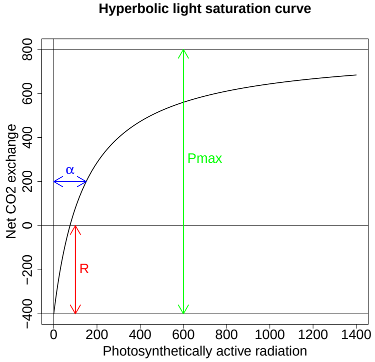
$$y_i = f(t_i; \boldsymbol{\alpha}_i) + e_i \,,$$

where $\boldsymbol{\alpha} = (\alpha_i^{(1)}, \ldots, \alpha_i^{(K)})$ and $\alpha_i^{(k)} = \boldsymbol{\beta}_k' \boldsymbol{x}_i^{(k)}$ for $k = 1, \ldots, K$.

# Nonlinear mixed-effect model formulation

- In a nonlinear fixed-effects model, one of the predictors ($t_i$) is often a primary predictor of the process (e.g. time in a growth model, or photosynthetically active radiation (PAR) in a model of net photosynthesis). The other secondary predictors $\boldsymbol{x}_i$ describe the variability in the primary parameters of the process. This leads to model
$$y_i = f(t_i; \boldsymbol{\alpha}_i) + e_i \,,$$
where $\boldsymbol{\alpha} = (\alpha_i^{(1)}, \ldots, \alpha_i^{(K)})$ and $\alpha_i^{(k)} = \boldsymbol{\beta}_k' \boldsymbol{x}_i^{(k)}$ for $k = 1, \ldots, K$.
- The nonlinear mixed-effect model for a single level of grouping is defined as
$$y_{ij} = f(t_{ij}; \boldsymbol{\alpha}_{ij}) + e_{ij} \,,$$
where the linear sub-models of $\alpha_{ij}^{(k)}$ include also random group effects:
$$\alpha_{ij}^{(k)} = \boldsymbol{\beta}_k' \boldsymbol{x}_{ij}^{(k)} + \boldsymbol{b}_i^{(k)\prime} \boldsymbol{z}_{ij}^{(k)} \,.$$
We assume $\boldsymbol{b}_i = (\boldsymbol{b}_i^{(1)\prime}, \ldots, \boldsymbol{b}_i^{(K)\prime})' \sim N(\boldsymbol{0}, \boldsymbol{D})$ and $e_{ij} \sim N(0, \sigma_{ij}^2)$ with an appropriate variance function. Also spatial or temporal dependence of residual errors can be modeled parametrically. The extension to multiple levels is straightforward.

# The model for the process




$$f(PAR; Pmax, R, \alpha) = -R + \frac{Pmax \times PAR}{\alpha + PAR}$$

where

- *PAR*: photosynthetically active radiation
- *Pmax*: Maximum gross $CO_2$ exchange
- *R*: Respiration
- $\alpha$: *PAR* at the 50% of gross $CO_2$ exchange
- $Pmax, R, \alpha > 0$

# Research question and data

- The effect of some categorical predictors (treatments) on the parameter *Pmax*, *R* and $\alpha$.
- The nuisance caused by varying Leaf Area Index (LAI), air temperature, and soil temperature on these parameters should also be taken into account (and this was the most interesting part to me).

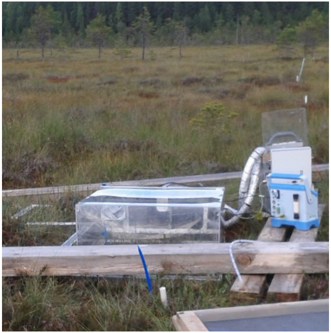


Photo: Juho Kettunen.

# Research question and data

- The effect of some categorical predictors (treatments) on the parameter *Pmax*, *R* and $\alpha$.
- The nuisance caused by varying Leaf Area Index (LAI), air temperature, and soil temperature on these parameters should also be taken into account (and this was the most interesting part to me).

- The data are collected using chamber measurements. At a given time, the net $CO_2$ exchange has been measured at 1-7 different levels of *PAR*.
- A total of 210 plots. Each plot is monitored for two years, with 5-10 measurements per year.
- LAI is also monitored by counting all plants of the plot and measuring the mean area of the leaves in surrounding plants.

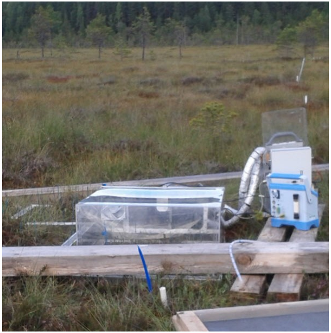
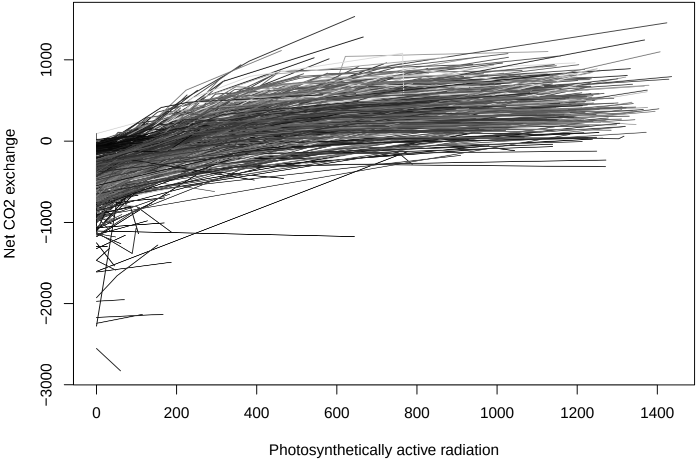


Photo: Juho Kettunen.

# The raw data


1000
0
−1000
−2000
−3000
Net CO2 exchange
0
200
400
600
800
1000
1200
1400
Photosynthetically active radiation

# The model

$$y_{ij} = -R_{ij} + \frac{Pmax_{ij} \times PAR_{ij}}{\alpha_{ij} + PAR_{ij}}$$

where

$$\log(R_{ij}) = \boldsymbol{\beta}_R' \boldsymbol{x}_{ij}^{(R)\prime} + b_i^{(R)}$$

$$\log(Pmax_{ij}) = \boldsymbol{\beta}_P' \boldsymbol{x}_{ij}^{(P)\prime} + b_i^{(P)}$$

$$\log(\alpha_{ij}) = \mu_\alpha + b_i^{(\alpha)}$$

We start with the model where the predictor vectors include treatments only.

# The model

$$y_{ij} = -R_{ij} + \frac{Pmax_{ij} \times PAR_{ij}}{\alpha_{ij} + PAR_{ij}}$$

where

$$\log(R_{ij}) = \boldsymbol{\beta}_R' \boldsymbol{x}_{ij}^{(R)\prime} + b_i^{(R)}$$

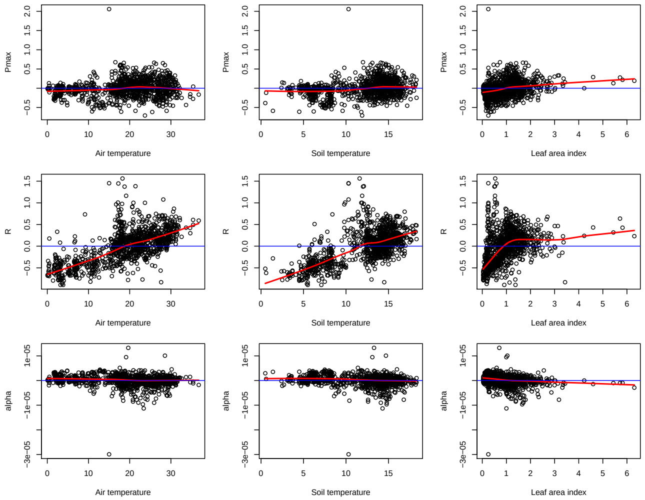
$$\log(Pmax_{ij}) = \boldsymbol{\beta}_P' \boldsymbol{x}_{ij}^{(P)\prime} + b_i^{(P)}$$

$$\log(\alpha_{ij}) = \mu_\alpha + b_i^{(\alpha)}$$

We start with the model where the predictor vectors include treatments only.

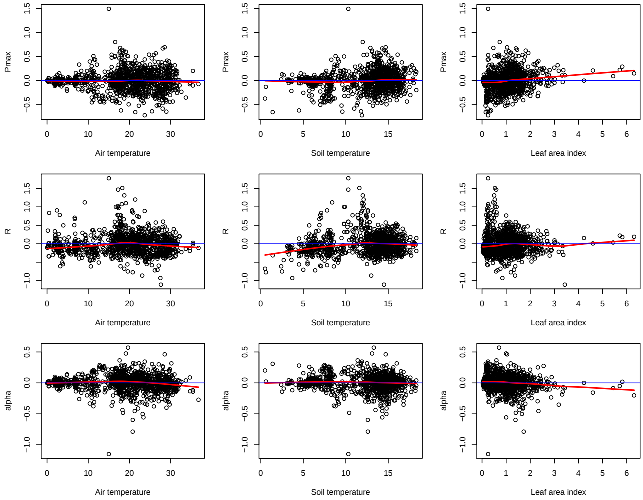
Let us see whether the predicted random effects have trends with respect to the plot-specific candidate predictors.

# The random effect on potential predictors + a lowess curve



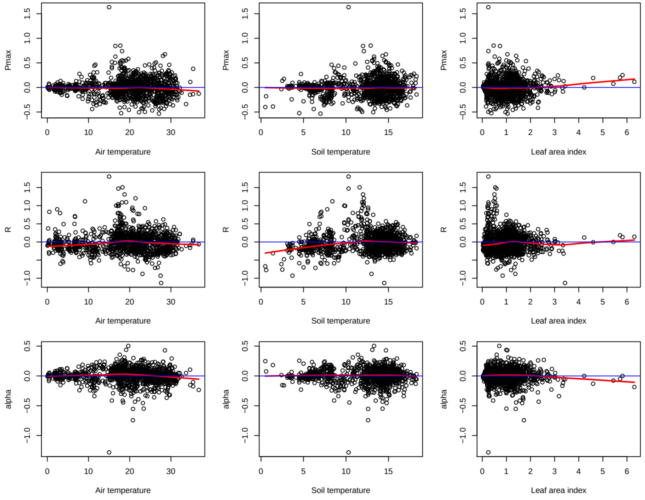

Include log($LAI$) and air $T$ to log($R$) and a polynomial air $T$ to log($Pmax$)

# The random effects on potential predictors + a lowess curve



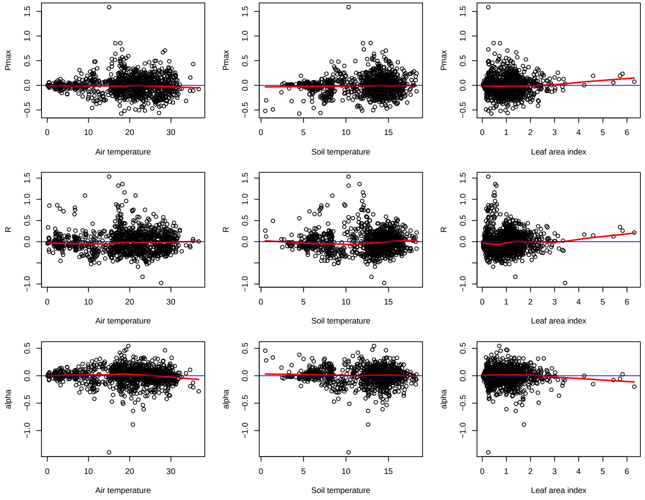

Include $\log(1 - \exp(-LAI))$ to model the self-shading LAI of leaves in $\log(Pmax)$.

# The random effects on potential predictors + a lowess curve




Soil temperature still seems to affect on Respiration, add it

# The random effects on potential predictors + a lowess curve




Seems to be quite ok.

# The final model

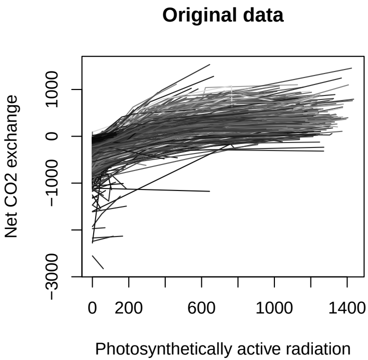
```
mod4co2<-nlme(nee~valovaste(par,lP,lR,la),
       fixed=list(lP~1+Year+treat1+treat2+group+tcham+I(tcham^2)+llai2,
                  lR~1+Year+treat1+treat2+group+tcham+log(lai)+pmin(t15,10),
                  la~1),
       random=list(occasion=pdDiag(lP+lR+la~1)),
       data=co2dat,
       start=c(log(1000),0,0,0,0,0,0,0,0,
               log(500),0,0,0,0,0,0,0,0,
               log(300)),
       verbose=TRUE)
```

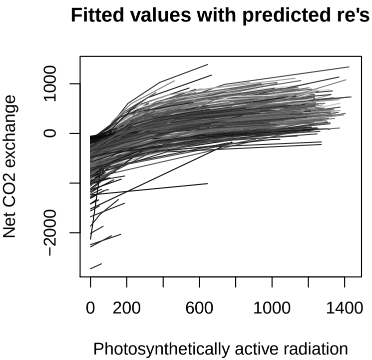
0.0716/(2*0.00146)=24.5

```
Random effects:
 Formula: list(lP ~ 1, lR ~ 1, la ~ 1)
 Level: occasion
 Structure: Diagonal
        lP.(Intercept) lR.(Intercept)        la Residual
StdDev:       0.262121      0.3063666 0.3239001 78.14503

Fixed effects: list(lP ~ 1 + Year + treat1 + treat2 + group +
                       Value Std.Error   DF   t-value p-value
lP.(Intercept)      6.471230 0.12974689 2241  49.87580  0.0000
lP.Year2013         0.152669 0.02444492 2241   6.24544  0.0000
lP.treat12         -0.004140 0.02765285 2241  -0.14970  0.8810
lP.treat13         -0.061529 0.03045692 2241  -2.02020  0.0435
lP.treat21         -0.055281 0.02328987 2241  -2.37362  0.0177
lP.group2          -0.097561 0.02371712 2241  -4.11354  0.0000
lP.tcham            0.071658 0.00994940 2241   7.20220  0.0000
lP.I(tcham^2)      -0.001465 0.00021403 2241  -6.84522  0.0000
lP.llai2            0.418375 0.03855223 2241  10.85215  0.0000
lR.(Intercept)      4.050615 0.10769026 2241  37.61357  0.0000
lR.Year2013         0.248282 0.02461425 2241  10.08690  0.0000
lR.treat12         -0.046315 0.03004181 2241  -1.54168  0.1233
lR.treat13          0.282602 0.02936815 2241   9.62274  0.0000
lR.treat21         -0.157959 0.02393427 2241  -6.59970  0.0000
lR.group2          -0.362806 0.02344484 2241 -15.47487  0.0000
lR.tcham            0.030448 0.00210702 2241  14.45084  0.0000
lR.log(lai)         0.102099 0.02229508 2241   4.57942  0.0000
lR.pmin(t15, 10)    0.147759 0.01140993 2241  12.95000  0.0000
la                  5.740796 0.02891320 2241 198.55275  0.0000
```

# The raw data and fitted values



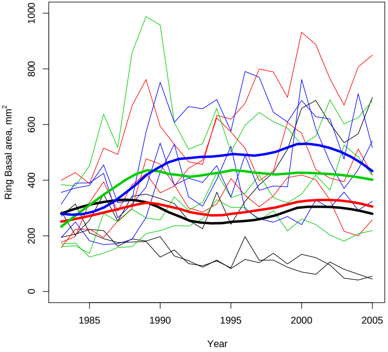



# Study material

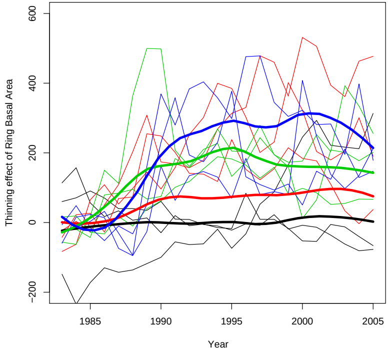
- Scots pine plots where one of the four following thinning treatments were applied to each plot in 1986: Control, Light, Moderate and Heavy.
- 88 trees were felled in 2006, and the diameter increments between 1983 and 2006 was measured for each tree using an X-ray densiometer.
- The diameter growths were transformed to basal area growths (assuming circular boles), because $\textit{Volume} \sim \textit{Diameter}^2 \textit{Height}$

# Raw data and the estimated thinning effects in Ring Basal Area (RBA)

The raw data



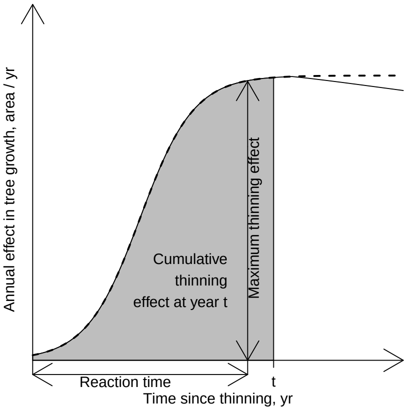

Extracted thinning effects [2]




Thick lines: the treatment-specific trends. Thin lines: 12 randomly selected trees.
Treatments: Control, Light, Moderate, and Heavy.

[2] Extraction was based on a mixed-effect model with crossed random effects, see Mehtätalo et al. (2014) for details.

# Nonlinear mixed-effects model for thinning effect




The thinning effect of tree $i$ at time $j$ was modeled using a logistic curve

$$d_{ij} = \frac{M_i}{1+\exp\left(4-8\frac{x_{ij}}{R_i}\right)} + e_{ij} \text{ where}$$

- $d_{ij}$ - thinning effect
- $x_{ij}$ - time since thinning
- $M_i = \mu_0 + \mu_1 T_2 + \mu_2 T_3 + \mu_4 x_{ij} + m_k$ - maximum thinning effect
- $T_2, \ldots, T_3$ - treatments
- $R_i = \rho_0 + \rho_1 z_i + r_i$ - reaction time
- $z_i$ - standardized diameter
- $\begin{bmatrix} m_i \\ r_i \end{bmatrix} \sim N(\mathbf{0}, \boldsymbol{D}_{2x2})$
- $e_{ij}$ - 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 |
|---|---|---|---|
| $\mu_0$ | 112.8 | 23.29 | 0.0000 |
| $\mu_1$ | 91.91 | 30.45 | 0.0026 |
| $\mu_2$ | 169.2 | 32.14 | 0.0000 |
| $\mu_3$ | -3.214 | 1.006 | 0.0014 |
| $\rho_0$ | 5.749 | 0.4458 | 0.0000 |
| $\rho_1$ | -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** | | | |
| $\sigma^2$ | 8.157*10-4 | | |
| $\delta_1$ | 8.746*104 | | |
| $\delta_2$ | 1.886 | | |
| $\delta_3$ | 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 |
|---|---|---|---|
| $\mu_0$ | 112.8 | 23.29 | 0.0000 |
| $\mu_1$ | 91.91 | 30.45 | 0.0026 |
| $\mu_2$ | 169.2 | 32.14 | 0.0000 |
| $\mu_3$ | -3.214 | 1.006 | 0.0014 |
| $\rho_0$ | 5.749 | 0.4458 | 0.0000 |
| $\rho_1$ | -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** | | | |
| $\sigma^2$ | 8.157*10-4 | | |
| $\delta_1$ | 8.746*104 | | |
| $\delta_2$ | 1.886 | | |
| $\delta_3$ | 0.5888 | | |

# Conclusions

- Nonlinear models are an elegant tool for processes where a theoretically justified function for the process is nonlinear with respect to its parameters is.

# Conclusions

- Nonlinear models are an elegant tool for processes where a theoretically justified function for the process is nonlinear with respect to its parameters is.
  - The models are both process-based and statistical..

# Conclusions

- Nonlinear models are an elegant tool for processes where a theoretically justified function for the process is nonlinear with respect to its parameters is.
  - The models are both process-based and statistical..
  - ..and random effects make them a compromise between Bayesian and frequentists models.

# Conclusions

- Nonlinear models are an elegant tool for processes where a theoretically justified function for the process is nonlinear with respect to its parameters is.
  - The models are both process-based and statistical..
  - ..and random effects make them a compromise between Bayesian and frequentists models.
- I especially like the formulation where the process is driven by a primary predictor and the parameters of this process are linear combinations of other predictors (recall also Sonja Vospernik's talk on Tuesday evening)
- If you want to play with CO2 data modeling, see data `foto` of R- package `lmfor`.

# References etc.


- 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 of tree-specific series using nonlinear mixed-effects model. For.Sci. 60(4):636-644
- Laine, A., Mehtätalo, L., Tolvanen, A., Tuittila, E.-S. Combined effect of drainage, rewetting and warming on peatland greenhouse gas fluxes and leaf area. In preparation.
- Mehtätalo, L. and Lappi, J. Forest Biometrics with examples in R. In preparation for Chapman&Hall / CSC. http://cs.uef.fi/ lamehtat/textbook.htm.


## Thank you for your interest!

lauri.mehtatalo@uef.fi
http://cs.uef.fi/ lamehtat/
https://www.researchgate.net/profile/Lauri_Mehtaetalo

Would like to see you in the International Biometric Conference, Barcelona, Spain in July 2018, where I will organize an invited session “Modeling grouped environmental data”