Height-Diameter curves from longitudinal data

Lauri Mehtätalo
lauri.mehtatalo@yale.edu

May 9, 2006, Yale School of F & ES

Outline of the presentation

The aim of this study

The aim was to estimate such a Height-Diameter model that could be used for prediction using the fixed predictors and height sample tree measurements from various points in time.
More specifically, the model should be able to

1. predict H-D curve when no sample tree heights are known,

The aim of this study

The aim was to estimate such a Height-Diameter model that could be used for prediction using the fixed predictors and height sample tree measurements from various points in time.
More specifically, the model should be able to

1. predict H-D curve when no sample tree heights are known,
2. use measured height sample trees to improve the prediction,

The aim of this study

The aim was to estimate such a Height-Diameter model that could be used for prediction using the fixed predictors and height sample tree measurements from various points in time.
More specifically, the model should be able to

1. predict H-D curve when no sample tree heights are known,
2. use measured height sample trees to improve the prediction,
3. use old height sample trees to improve the prediction of current point in time,

The aim of this study

The aim was to estimate such a Height-Diameter model that could be used for prediction using the fixed predictors and height sample tree measurements from various points in time.
More specifically, the model should be able to

1. predict H-D curve when no sample tree heights are known,
2. use measured height sample trees to improve the prediction,
3. use old height sample trees to improve the prediction of current point in time,
4. predict the H-D curve in the future,

Data

- Data includes (almost) randomly selected stands from all around the Finland
- We are using Norway spruces, which is a shade tolerant tree species

Data

- Data includes (almost) randomly selected stands from all around the Finland
- We are using Norway spruces, which is a shade tolerant tree species
- Data inclusdes 249 stands, 1-4 measurement occasions with 5 years intervals
- A total of 18056 height measurements, 3–49 trees per stand

Model for tree i in stand k at time t

Starting point is the Korf function

$$H = a\exp(-bD^{-c})$$

Model for tree i in stand k at time t

Starting point is the Korf function

$$H = a \exp(-bD^{-c})$$

I reparameterized D and linearized the model to get

$$\ln(H_{kti}) = A_{kt} + B_{kt}x_{kti} + \epsilon_{kti} \quad (1)$$

The reparameterization of D_{kti} was

$$x_{kti} = \frac{(D_{kti} + \lambda)^{-C} - (DGM + 10 + \lambda)^{-C}}{(30 + \lambda)^{-C} - (10 + \lambda)^{-C}},$$

where $\lambda = 7cm$ and $C = 1.564$.

Model for tree i in stand k at time t

Starting point is the Korf function

$$H = a\exp(-bD^{-c})$$

I reparameterized D and linearized the model to get

$$\ln(H_{kti}) = A_{kt} + B_{kt}x_{kti} + \epsilon_{kti} \quad (1)$$

The reparameterization of D_{kti} was

$$x_{kti} = \frac{(D_{kti} + \lambda)^{-C} - (DGM + 10 + \lambda)^{-C}}{(30 + \lambda)^{-C} - (10 + \lambda)^{-C}},$$

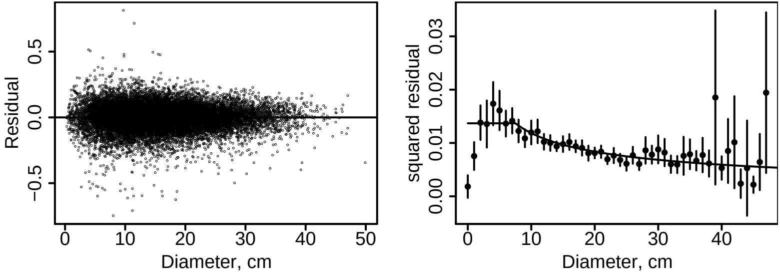
where $\lambda = 7cm$ and $C = 1.564$.

Interpretations

A is the expected logarithmic height of thickest trees in the stand
B is the difference in $\ln(H)$ between trees of diameters 30 and 10 cm

The variance function

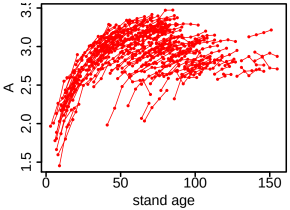
In a single stand, the variance of height decreases with increasing diameter

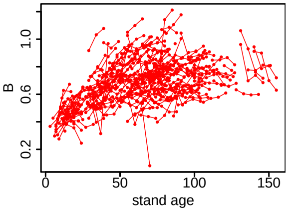


The variance function I used was

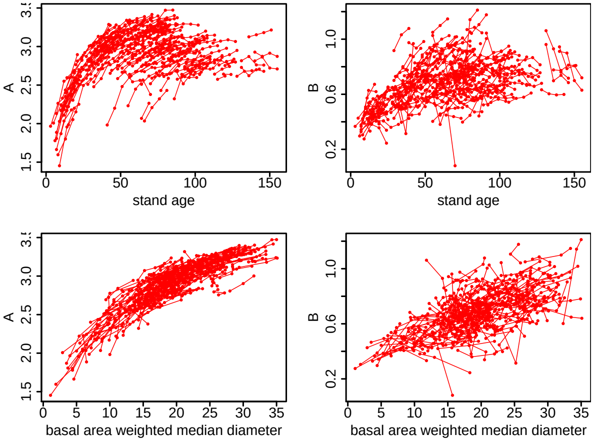
$$\mathrm{var}(\epsilon_{kti}) = \sigma^2 \left(\max(D_{kti}, 7.5) \right)^{-2\delta} \tag{2}$$

Separate fits





Separate fits



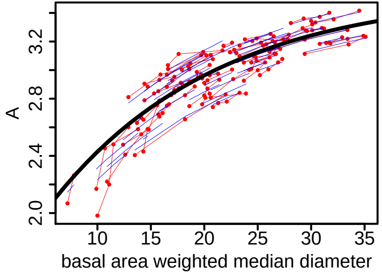
Trend functions

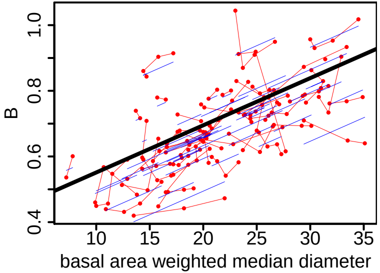
A: Champman-Richards

$$A_{kt} = p_{1a} + p_{2a}\left(1 - \exp(-p_{3a}DGM_{kt})\right)^{p_{4a}} + \alpha_k + \alpha_{kt} \quad (3)$$

B: linear

$$B_{kt} = p_{1b} + p_{2b}DGM_{kt} + \beta_k + \beta_{kt}$$





Complete model

I wrote the trend functions to model 1 to obtain

$$\begin{aligned} \ln(H_{kti}) &= p_{1a} + \alpha_k + \alpha_{kt} + p_{2a} z_{kt} \\ &\quad + (p_{1b} + \beta_k + \beta_{kt} + p_{2b} DGM_{kt}) x_{kti} + \epsilon_{kti} \,, \end{aligned}$$

where z_{kt} is the nonlinear part of trend function 3, which is treated as a transformation of DGM_{kt}:

$$z_{kt} = \left(1 - \exp(-0.0651 DGM_{kt})\right)^{0.999}$$

α_k and β_k are correlated, normal, stand specific random parameters, α_{kt} and β_{kt} are correlated, normal, measurement occasion specific random parameters, and ϵ_{kti} are normal tree level errors with variance as given in 2.

Estimates

The parameter estimates and their standard errors were as follows

	p_{1a}	p_{2a}	p_{1b}	p_{2b}
estimate	1.40	2.15	0.384	0.0157
standard error	0.026	0.035	0.017	0.001

$$
\begin{aligned}
\mathrm{var}\begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} &= \begin{bmatrix} 0.108^2 & (0.269) \\ 0.0028 & 0.0958^2 \end{bmatrix} \\
\mathrm{var}\begin{bmatrix} \alpha_{kt} \\ \beta_{kt} \end{bmatrix} &= \begin{bmatrix} 0.0168^2 & (-0.681) \\ -0.0003 & 0.0223^2 \end{bmatrix} \\
\mathrm{var}(\epsilon_{kti}) &= 0.401^2 \left(\max(D_{kti}, 7.5) \right)^{-1.068}
\end{aligned}
$$

Prediction of random effects

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

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zb} + \boldsymbol{\epsilon}\,,$$

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

Utilizing the model

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

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

The Empirical Best Linear Unbiased predictor of random effects is

$$\widehat{\mathbf{b}} = \mathbf{DZ}'(\mathbf{ZDZ}' + \mathbf{R})^{-1}(\mathbf{y} - \boldsymbol{\mu}) \,.$$

and the variance of prediction errors is

$$\mathrm{var}(\widehat{\mathbf{b}} - \mathbf{b}) = \mathbf{D} - \mathbf{DZ}'(\mathbf{ZDZ}' + \mathbf{R})^{-1}\mathbf{ZD}$$

Example

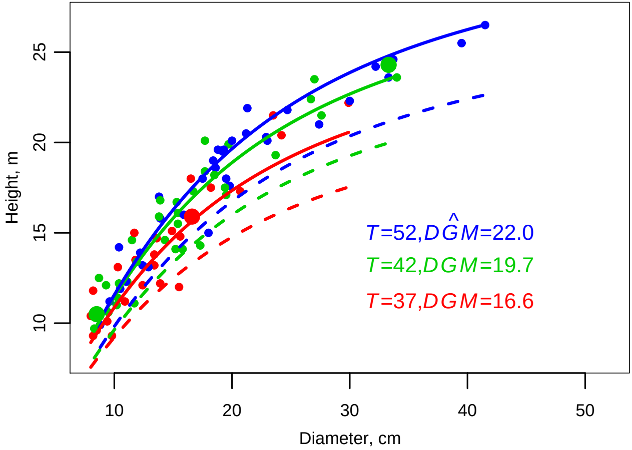
Heights of one tree was measured 5 years ago and 2 trees at this year. The matrices and vectors are

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

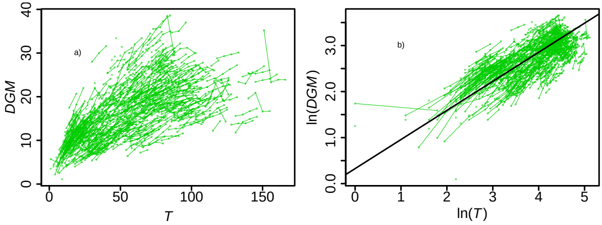
$$\mathbf{Z} = \begin{bmatrix} 1 & -0.36 & 1 & -0.36 & 0 & 0 \\ 1 & -1.22 & 0 & 0 & 1 & -1.22 \\ 1 & 0.058 & 0 & 0 & 1 & 0.058 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 0.008 & 0 & 0 \\ 0 & 0.016 & 0 \\ 0 & 0 & 0.004 \end{bmatrix}$$

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

Example


T=52,DĜM=22.0
T=42,DGM=19.7
T=37,DGM=16.6
Height, m
Diameter, cm
10
15
20
25
10
20
30
40
50

Model for DGM



The model for *DGM* was

$$\ln(DGM_{kt}) = u + v \ln T_{kt} + u_k + e_{kt} \; ,$$

where $u = 0.33$, $v = 0.63$, $\mathrm{var}(u_k) = 0.32^2$ and $\mathrm{var}(e_{kt}) = 0.069^2$.

Final remarks

- The mixed modeling for longitudinal data was a good approach in this context. It makes it possible to flexibly move forward and backward in time, and to utilize mesurements from different points in time in prediction.

Final remarks

- The mixed modeling for longitudinal data was a good approach in this context. It makes it possible to flexibly move forward and backward in time, and to utilize mesurements from different points in time in prediction.
- It is very important to plot predictions and residuals in different ways to find good model formulae. Here it helped me to find how parameter A developes over time.

Final remarks

- The mixed modeling for longitudinal data was a good approach in this context. It makes it possible to flexibly move forward and backward in time, and to utilize mesurements from different points in time in prediction.
- It is very important to plot predictions and residuals in different ways to find good model formulae. Here it helped me to find how parameter *A* developes over time.
- The development of *H*-*D* curves was best modeled as a function of a variable describing the mean (or median) size of trees in the stand, instead of stand age. That could be a good approach also in other models.

Final remarks

- The mixed modeling for longitudinal data was a good approach in this context. It makes it possible to flexibly move forward and backward in time, and to utilize mesurements from different points in time in prediction.
- It is very important to plot predictions and residuals in different ways to find good model formulae. Here it helped me to find how parameter A developes over time.
- The development of *H-D* curves was best modeled as a function of a variable describing the mean (or median) size of trees in the stand, instead of stand age. That could be a good approach also in other models.
- R-sowtware has many good properties that make this kind of analysis easy in R: function 'unique', packages 'nlme' and 'lme', possibility to handle several datasets and models at same time.