Modelling Height-Diameter Curves For Prediction

Lauri Mehtätalo, University of Eastern Finland

March 2, 2016

Mehtätalo L, de-Miguel S, and Gregoire, T.G. 2015. Modeling height-diameter curves for prediction. Canadian Journal of Forest Research, 45(7): 826-837, 10.1139/cjfr-2015-0054

Outline

Introduction

Height-Diameter (H-D) relationship is one of the first models one need to fit in forest inventories.

The aims of this study were

- To emphasize the differences between
    - marginal (population-averaged of one kind) and plot-specific H-D relationship and
    - simple and generalized relationship
- Explore the fit of 16 nonlinear functions for the H-D relationship in 28 different datasets of different tree species from different regions.
- Develop generalized models for four example datasets for demonstration purposes
- Produce easy-to-use R functions with sensible defaults for height imputation

Introduction

Height-Diameter (H-D) relationship is one of the first models one need to fit in forest inventories.

The aims of this study were

- To emphasize the differences between
  - marginal (population-averaged of one kind) and plot-specific H-D relationship and
  - simple and generalized relationship
- Explore the fit of 16 nonlinear functions for the H-D relationship in 28 different datasets of different tree species from different regions.
- Develop generalized models for four example datasets for demonstration purposes
- Produce easy-to-use R functions with sensible defaults for height imputation

Introduction

Height-Diameter (H-D) relationship is one of the first models one need to fit in forest inventories.

The aims of this study were

- To emphasize the differences between
  - marginal (population-averaged of one kind) and plot-specific H-D relationship and
  - simple and generalized relationship
- Explore the fit of 16 nonlinear functions for the H-D relationship in 28 different datasets of different tree species from different regions.
- Develop generalized models for four example datasets for demonstration purposes
- Produce easy-to-use R functions with sensible defaults for height imputation

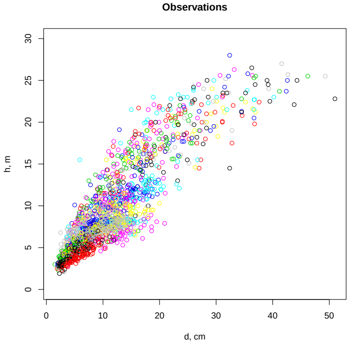
Introduction

Height-Diameter (H-D) relationship is one of the first models one need to fit in forest inventories.

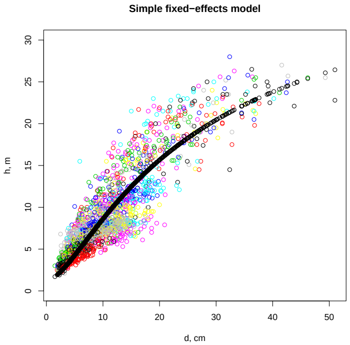
The aims of this study were

- To emphasize the differences between
  - marginal (population-averaged of one kind) and plot-specific H-D relationship and
  - simple and generalized relationship
- Explore the fit of 16 nonlinear functions for the H-D relationship in 28 different datasets of different tree species from different regions.
- Develop generalized models for four example datasets for demonstration purposes
- Produce easy-to-use R functions with sensible defaults for height imputation

A typical H-D dataset



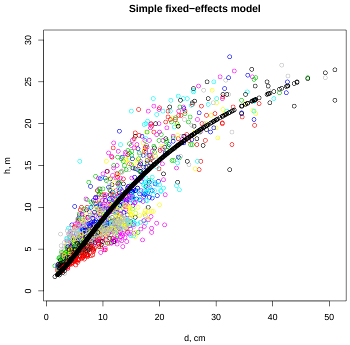
Marginal H-D relationship



$$h_{ij} = f(d_{ij};\beta) + e_{ij}$$

- An easy way to estimate the marginal relationship.
- The model is improperly formulated: the model ignores the grouped structure

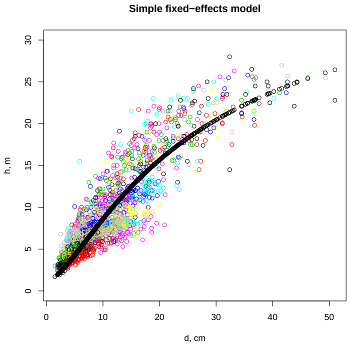
Marginal H-D relationship



$$h_{ij} = f(d_{ij};\beta) + e_{ij}$$

- An easy way to estimate the marginal relationship.
- The model is improperly formulated: the model ignores the grouped structure

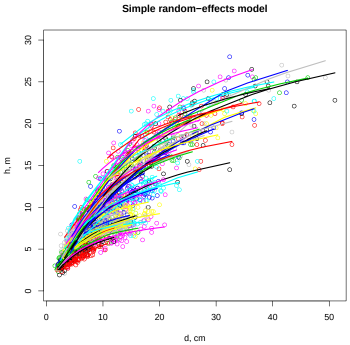
Marginal H-D relationship



$$h_{ij} = f(d_{ij}; \beta) + e_{ij}$$

- An easy way to estimate the marginal relationship.
- The model is improperly formulated: the model ignores the grouped structure

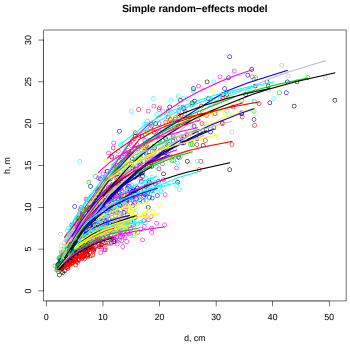
Plot-specific H-D relationship



$$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}, \text{ where } \beta_i = \mathbf{B} + \mathbf{b}_i \text{ and } \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

- An easy way to estimate the plot-specific relationship.
- Still improperly formulated model: the random effect mean depends on the mean diameter of the plot.

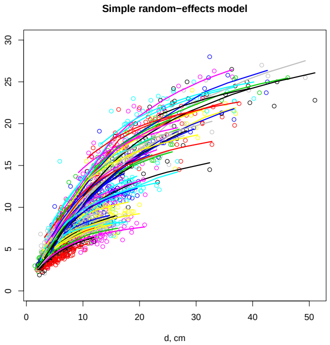
Plot-specific H-D relationship



$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}$, where $\beta_i = \mathbf{B} + \mathbf{b}_i$ and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$

- An easy way to estimate the plot-specific relationship.
- Still improperly formulated model: the random effect mean depends on the mean diameter of the plot.

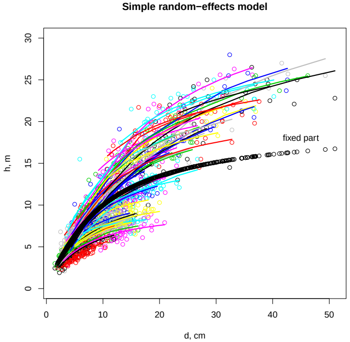
Plot-specific H-D relationship



$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}$, where $\beta_i = \mathbf{B} + \mathbf{b}_i$ and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$

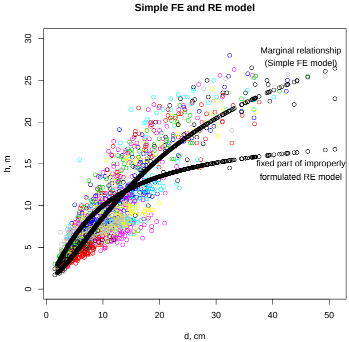
- An easy way to estimate the plot-specific relationship.
- Still improperly formulated model: the random effect mean depends on the mean diameter of the plot.

Plot-specific H-D relationship and fixed part



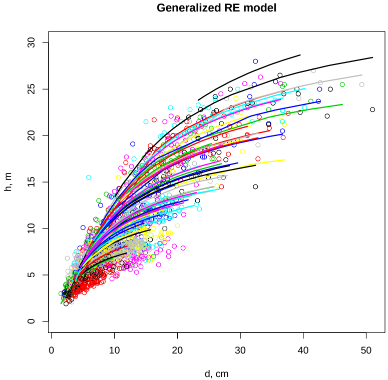
$$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}, \text{ where } \beta_i = \mathbf{B} + \mathbf{b}_i \text{ and } \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

Two fixed-effect predictions



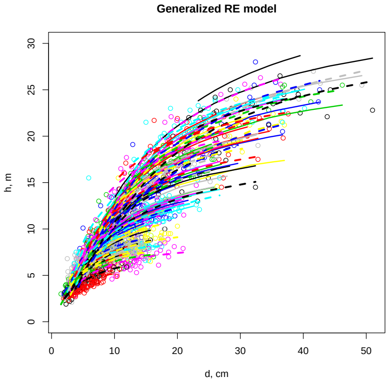
$$h_{ij} = f(d_{ij};\beta) + e_{ij}$$
$$h_{ij} = f(d_{ij};\beta_i) + e_{ij}, \text{ where } \beta_i = \mathbf{B} + \mathbf{b}_i \text{ and } \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

A generalized model: fixed part



$$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}, \text{ where } \beta_i = \beta' \mathbf{x}_i + \mathbf{b}_i \text{ and } \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

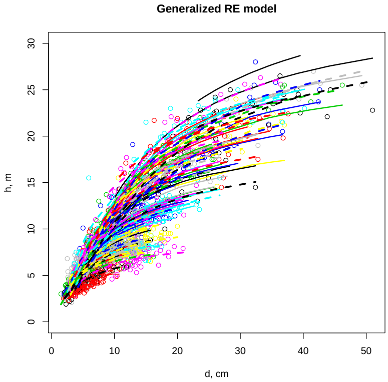
A generalized model: fixed + random part



$$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}, \text{ where } \beta_i = \beta' \mathbf{x}_i + \mathbf{b}_i \text{ and } \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

- Model properly formulated
- Could provide also marginal relationship (not shown)

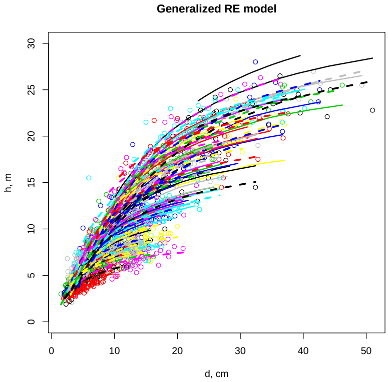
A generalized model: fixed + random part



$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}$, where $\beta_i = \beta' \mathbf{x}_i + \mathbf{b}_i$ and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$

- Model properly formulated
- Could provide also marginal relationship (not shown)

A generalized model: fixed + random part



$h_{ij} = f(d_{ij}; \beta_i) + e_{ij}$, where $\beta_i = \beta' \mathbf{x}_i + \mathbf{b}_i$ and $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$

- Model properly formulated
- Could provide also marginal relationship (not shown).

The applied functions

Table 2. The applied H-D functions.

Number	Function name	Equation	References
2-parameter functions			
1	Näslund	$H(D) = BH + \frac{D^2}{(aD + b)^2}$	Näslund (1937), Peschel (1938)
2	Curtis	$H(D) = BH + \frac{aD}{(1 + D)^b}$	Curtis (1967)
3	Schumacher	$H(D) = BH + a\exp(-bD^{-1})$	Schumacher (1939), Michailoff (1943), Curtis (1967)
4	Meyer	$H(D) = BH + a(1 - \exp(-bD))$	Meyer (1940), Curtis (1967)
5	Power	$H(D) = BH + aD^b$	Stoffels and van Soest (1953)
6	Michaelis–Menten	$H(D) = BH + aD/(b + D)$	Menten and Michaelis (1913), Huang et al. (1992)
7	Wykoff	$H(D) = BH + \exp(a - b(D + 1)^{-1})$	Wykoff et al. (1982)
3-parameter functions			
8	Prodan	$H(D) = BH + \frac{D^2}{aD^2 + bD + c}$	Strand (1959)
9	Logistic	$H(D) = BH + \frac{a}{1 + b\exp(-cD)}$	Pearl and Reed (1920), Huang et al. (1992)
10	Chapman-Richards	$H(D) = BH + a(1 - \exp(-bD))^c$	Richards (1959), Huang et al. (1992)
11	Weibull	$H(D) = BH + a(1 - \exp(-bD^c))$	Weibull (1951), Huang et al. (1992)
12	Gomperz	$H(D) = BH + a\exp(-b\exp(-cD))$	Gomperz (1825), Huang et al. (1992)
13	Sibbesen	$H(D) = BH + aD^{bD^{-c}}$	Sibbesen (1981), Huang et al. (1992)
14	Korf	$H(D) = BH + a\exp(-bD^{-c})$	Lundqvist (1957), Flewelling and de Jong (1994)
15	Ratkowsky	$H(D) = BH + a\exp\left(\frac{-b}{D + c}\right)$	Ratkowsky (1990), Huang et al. (1992)
16	Hossfeld IV	$H(D) = BH + \frac{a}{1 + \frac{1}{bD^c}}$	Peschel (1938)

Note: The references give the original reference and the first use in H–D modeling. Naming follows Zeide (1993) when applicable. H = tree height, D = tree diameter at breast height, BH = breast height, a, b, c = parameters of the equation.

Material

Table 1. Summary of the modeling datasets.

Data set	Latin name	Country	N	K	$\bar{n}_i$	d_{min}	$\bar{d}$	d_{max}	h_{min}	$\bar{h}$	h_{max}
Scots pine A	*Pinus Sylvestris*	Finland	4234	103	41	1.5	14.5	51.0	1.4	13.2	35.1
Norway spruce A	*Picea abies*	Finland	2513	51	49	2.9	17.2	57.0	2.1	13.7	29.8
Scots pine B	*Pinus Sylvestris*	Finland	1644	66	25	3.0	20.0	49.1	1.6	17.3	33.1
Norway spruce B	*Picea abies*	Finland	3020	66	46	0.9	11.5	52.3	1.4	9.9	33.2
Birch A	*Betula pendula. B. pubescens*	Finland	1673	72	23	1.6	8.7	48.8	1.8	10.0	29.8
Norway spruce C	*Picea abies*	Finland	1252	31	40	5.0	14.3	68.8	1.5	12.9	34.3
Turkish red pine	*Pinus brutia*	Syria. Lebanon	1283	114	11	5.0	27.3	96.9	3.5	13.7	35.1
Aleppo pine	*Pinus halepensis*	Spain	16378	1016	16	7.5	32.6	174.0	2.0	14.6	41.0
Canarian island pine	*Pinus canariensis*	Spain	7327	870	8	7.5	19.5	74.8	2.0	7.7	23.0
Loblolly pine 1	*Pinus taeda*	VA, USA	5634	99	57	1.3	13.9	34.3	1.5	10.9	23.8
Loblolly pine 2	*Pinus taeda*	VA, USA	4895	99	49	3.3	18.2	37.6	4.6	15.8	26.8
Loblolly pine 3	*Pinus taeda*	VA, USA	4171	99	42	5.1	20.8	42.9	4.9	18.8	31.4
Lodgepole pine 1	*Pinus contorta*	BC, Canada	10817	140	77	0.1	6.0	25.5	1.3	7.4	21.3
Lodgepole pine 2	*Pinus contorta*	BC, Canada	9336	141	66	0.3	8.9	31.0	1.3	8.7	22.8
Lodgepole pine 3	*Pinus contorta*	BC, Canada	5903	93	63	0.7	12.6	29.8	1.4	12.4	24.2
Eucalyptus clone	*Eucalyptus urograndis*	Brazil	1141	191	6	6.2	19.4	34.1	12.0	30.0	41.0
Blue gum A	*Eucalyptus globulus*	Bolivia	6554	50	131	0.1	3.8	18.2	1.4	6.0	19.1
Blue gum B1	*Eucalyptus globulus*	Bolivia	884	6	147	1.0	9.5	31.7	1.9	9.6	28.5
Blue gum B2	*Eucalyptus globulus*	Bolivia	1261	6	210	1.0	10.1	33.7	1.7	11.1	30.0
Centrolobium 1	*Centrolobium tomentosum*	Bolivia	2199	46	48	1.2	11.2	28.3	1.8	11.1	20.5
Centrolobium 2	*Centrolobium tomentosum*	Bolivia	2167	46	47	1.2	12.6	30.3	2.2	12.6	22.1
Centrolobium 3	*Centrolobium tomentosum*	Bolivia	2023	44	46	2.5	13.4	31.3	2.2	13.6	25.8
Brasilian firetree	*Schizolobium parahyba*	Bolivia	2631	46	57	0.8	8.8	33.2	1.4	8.7	27.0
Teak 1	*Tectona grandis*	Bolivia	4928	62	79	1.0	6.6	41.5	1.4	6.5	29.6
Teak 2	*Tectona grandis*	Bolivia	3444	43	80	1.0	8.1	44.3	1.4	7.9	29.5
Mixed tropical	Multi-species	Bolivia	15049	41	367	8.2	23.7	115.6	2.5	12.6	36.9
Balsa 1	*Ochroma pyramidale*	Bolivia	2943	53	56	0.8	8.7	19.8	1.5	8.6	21.7
Balsa 2	*Ochroma pyramidale*	Bolivia	715	23	31	1.5	9.9	22.7	1.4	9.8	17.5

Note: Whenever two datasets of same species have been used, a capital letter is used to denote different independent datasets and an Arabic number to denote different measurement occasions of the same dataset. *N*: the number of trees; *K*: the number of sample plots; $\bar{n}_i$: mean number of trees per plot; d_{min}, $\bar{d}$, d_{max}: the minimum, mean and maximum diameter, cm; h_{min}, $\bar{h}$, h_{max}: the minimum, mean and maximum height, m.

Ranking of the functions

Table 3. Evaluation of the simple two-parameter models according to the four criteria.

	Criteria	Näslund	Curtis	Schumacher	Meyer	Power	Mic.-Ment.	Wykoff
Mixed-effects	1st ranks	**12**	**7**	**8**	0	1	0	1
	Ranks 1–3	15	**26**	**20**	5	2	3	**16**
	Mean rank (sd)	**2.7 (1.5)**	**2.3 (1.2)**	**3.0 (1.9)**	4.7 (1.4)	5.9 (1.5)	5.8 (1.4)	3.6 (1.3)
	Conv. Prob's	**0**	**0**	**0**	**0**	**0**	1	**0**
Fixed-effects	1st ranks	**13**	**5**	0	**6**	2	3	0
	Ranks 1–3	**23**	**19**	8	**12**	6	**10**	8
	Mean rank (sd)	**1.9 (1.1)**	**2.8 (1.3)**	4.2 (1.5)	**2.5 (1.5)**	4.3 (2.0)	2.5 (1.5)	4.1 (1.1)
	Conv. Prob's	**4**	**3**	**5**	12	11	16	**3**

Note: The criteria are: 1st ranks is the number of first ranks among the datasets; ranks 1–3 gives the number of rankings among three best models; mean rank gives the mean rank of the model (the number in parentheses is the standard deviation of the ranks; Conv. Prob's gives the number of unsuccessful fits. The three best models according to each criteria are highlighted.

Table 4. Evaluation of the simple three-parameter models according to the four criteria.

	Criteria	Prodan	Logistic	Ch-Ri	Weibull	Gomperz	Sibbesen	Korf	Ratkowsky	Hossf. IV
Mixed-effects	1st ranks	**11**	**6**	1	0	4	0	0	**6**	0
	Ranks 1–3	**18**	12	10	8	**18**	0	2	**14**	2
	Mean rank (sd)	**3.0 (2.3)**	4.0 (2.2)	4.3 (1.8)	4.8 (1.7)	**2.8 (1.4)**	8.1 (1.3)	7.3 (1.7)	**3.6 (2.0)**	6.1 (1.3)
	Conv. Prob's	**0**	**0**	1	**0**	**0**	18	6	**0**	1
Fixed-effects	1st ranks	**3**	**3**	**4**	**3**	**3**	0	**3**	**3**	**4**
	Ranks 1–3	9	4	**13**	9	**11**	4	4	9	**10**
	Mean rank (sd)	3.3 (1.6)	5.3 (2.6)	**2.7 (1.4)**	3.3 (1.7)	3.9 (2.2)	3.5 (1.6)	4.1 (2.1)	**2.8 (1.6)**	**2.7 (1.5)**
	Conv. Prob's	**9**	**10**	**10**	13	**6**	22	15	16	13

Note: For notations, see Table 3.

The best-fitting functions for plot-specific relationship

Table 5. The best plot-specific fits of the 2- and 3- parameter models and the related RMSE in different datasets.

Dataset	Model name and RMSE (m): 2-parameter model		3-parameter model	
Scots pine A	Curtis	1.39	***Logistic***	***1.37***
Norway spruce A	Näslund	1.62	***Prodan***	***1.60***
Scots pine B	***Näslund***	***1.64***	Prodan	1.64
Norway spruce B	Näslund	1.27	***Prodan***	***1.21***
Birch	Näslund	1.97	***Logistic***	***1.92***
Norway spruce C	Näslund	2.01	***Gomperz***	***1.95***
Turkish red pine	Wykoff	1.95	***Prodan***	***1.95***
Canarian island pine	***Curtis***	***1.99***	Logistic	1.94
Aleppo pine	Näslund	0.97	***Prodan***	***0.97***
Loblolly pine 1	***Näslund***	***0.82***	Gomperz	0.82
Loblolly pine 2	***Schumacher***	***0.97***	Gomperz	0.97
Loblolly pine 3	**Schumacher**	**1.14**	*Prodan*	*1.13*
Lodgepole pine 1	Curtis	0.61	***Ratkowsky***	***0.61***
Lodgepole pine 2	Curtis	0.68	***Ratkowsky***	***0.68***
Lodgepole pine 3	Schumacher	0.85	***Ratkowsky***	***0.85***
Eugalyptus clone	***Schumacher***	***0.90***	Prodan	0.83
Blue gum A	Näslund	1.17	***Prodan***	***1.16***
Blue gum B1	***Näslund***	***2.27***	Ratkowsky	2.26
Blue gum B2	Näslund	2.29	***Logistic***	***2.20***
Centrolobium 1	**Curtis**	**1.26**	*Prodan*	*1.25*
Centrolobium 2	***Schumacher***	***1.28***	Ratkowsky	1.25
Centrolobium 3	***Schumacher***	***1.42***	Ratkowsky	1.39
Brasilian firetree	**Curtis**	**1.54**	*Prodan*	*1.53*
Teak 1	***Curtis***	***1.10***	Logistic	1.08
Teak 2	**Näslund**	**1.10**	*Gomperz*	*1.08*
Mixed tropical	Näslund	2.81	***Logistic***	***2.79***
Balsa 1	**Schumacher**	**1.23**	*Ratkowsky*	*1.23*
Balsa 2	**Schumacher**	**1.24**	*Prodan*	*1.24*

Note: The model with lower BIC value between the 2- and 3- parameter models is indicated by **boldface** and the model with lower AIC by *italics*.

Best generalized model for four datasets

- Scots Pine A/ Logistic:

(7) $$h_{ij} = 1.3 + \frac{(\alpha_0 + \alpha_1\bar{d}_i + a_i)}{1 + (\beta_0 + \beta_1\bar{d}_i + b_i)\exp[(\gamma_0 + \gamma_1\bar{d}_i + c_i)d_{ij}]} + e_{ij}$$

- Loblolly pine/ Näslund:

(8) $$h_{ij} = 1.3 + \frac{d_{ij}^2}{[(\alpha_0 + \alpha_1 \ln \bar{d}_i + a_i)d_{ij} + \beta_0 + b_i + \beta_1 \ln \bar{d}_i]^2} + e_{ij}$$

- Teak / Curtis:

(9) $$h_{ij} = 1.3 + \frac{(\alpha_0 + \alpha_1\bar{d}_i + a_i)d_{ij}}{(1 + d_{ij})^{\beta_0+\beta_1\bar{d}_i+b_i}} + e_{ij}$$

- Centrolobium / Schumacher:

(10) $$h_{ij} = 1.3 + (\alpha_0 + \alpha_1\bar{d}_i + a_i)\exp[(\beta_0 + b_i)d_{ij}^{-1}] + e_{ij}$$

- Some data sets had constant residual variance, others had increasing or decreasing as a function of fitted value.

Discussion and conclusions

- Differences between two and three-parameter models were slight in most cases. Two parameters was usually enough.
- Curtis' and Näslunds functions were most commonly the best functions for plot-specific H-D relationship
- The generalized models were very different in different datasets with respect to functional form, random effects and variance function.
- Functions for height imputation are available in R-package `lmfor`.

Discussion and conclusions

- Differences between two and three-parameter models were slight in most cases. Two parameters was usually enough.
- Curtis' and Näslunds functions were most commonly the best functions for plot-specific H-D relationship
- The generalized models were very different in different datasets with respect to functional form, random effects and variance function.
- Functions for height imputation are available in R-package `lmfor`.

Discussion and conclusions

- Differences between two and three-parameter models were slight in most cases. Two parameters was usually enough.
- Curtis' and Näslunds functions were most commonly the best functions for plot-specific H-D relationship
- The generalized models were very different in different datasets with respect to functional form, random effects and variance function.
- Functions for height imputation are available in R-package `lmfor`.

Discussion and conclusions

- Differences between two and three-parameter models were slight in most cases. Two parameters was usually enough.
- Curtis' and Näslunds functions were most commonly the best functions for plot-specific H-D relationship
- The generalized models were very different in different datasets with respect to functional form, random effects and variance function.
- Functions for height imputation are available in R-package `lmfor`.