Parameter Estimation of Base-age Invariant Site Index Models: Which Data Structure We Use?

Mingliang Wang, Bruce E. Borders, Dehai Zhao

Warnell School Of Forestry and Natural Resources
The University of Georgia

Base-age Invariant Site Index Models

- What? Bailey and Clutter (1974): dominant height (site index) curves do not depend on arbitrary choice of base age
- How? (Traditional, not Mixed Model)

(I): ADA (Algebraic Difference Approach): Bailey and Clutter (1974), GADA (generalized ADA): Cieszewski and Bailey (2000)

(II): Difference Equation Method (Clutter et al. 1983)

(G)ADA (Algebraic Difference Approach)

- Base Model: $$H = f(A;\theta) + \varepsilon$$

  where *H* is dominant height, *A* is age, θ is a vector of parameters, and ε is the error term

- (G)ADA Model: $$H = f(A;\chi_i, \beta) + \varepsilon$$

  where $(\chi_i, \beta) = \theta$, χ_i, specific to plot (subject)

  one-parameter (local, index) family (Garcia 1983)

Re-parameterization of (G)ADA Models

$$H = f(A; \chi_i, \beta) + \varepsilon$$

$$\chi_i = f^{-1}(A_s; S, \beta)$$

where site index (*S*), the expected height at base age (*A*s). Note: *S* can be taken as expected height at any arbitrary age, and is a expected-value PARAMETER, not input variable (as in base-age specific model).

$$H = f(A, A_s; S, \beta) + \varepsilon$$

Dummy Variable

$$\hat{H} = f(A, A_s, \hat{S}, \hat{\beta}) \longrightarrow \hat{S} = f^{-1}(A, A_s, \hat{H}, \hat{\beta}) \approx f^{-1}(A, A_s, H, \hat{\beta})$$

Difference Equation Method (Clutter et al. 1983)

- Base Model: $$H = f(A;\theta) + \varepsilon$$
- Local Parameter: $$H = f(A;\chi_i,\beta) + \varepsilon$$ $$\chi_i = f^{-1}(H_1 - \varepsilon_1, A_1;\beta)$$
- Difference Model: $$H_2 = f(H_1 - \varepsilon_1, A_1, A_2;\beta) + \varepsilon_2$$
- Measurement Error Model (Error-in-Variable, Fuller 1987)

$$\hat{H}_2 = f(\hat{H}_1, A_1, A_2;\beta)$$

$$H_1 = \hat{H}_1 + \varepsilon_1 \quad H_2 = \hat{H}_2 + \varepsilon_2$$

$$\rightarrow \quad \hat{H} = f(A, A_s, \hat{S}, \hat{\beta})$$

$$\hat{S} = f^{-1}(A, A_s, \hat{H}, \hat{\beta})$$

Two variables $\hat{H}_1$ and $\hat{H}_2$ are assumed to satisfy the functional relationship determined by *f*, but the true values of these variables can not be known as observations of the two variables (*H*1 and *H*2) are subject to measurement error and (or) natural variability, and even model error .

Two Fitting Approaches Depending on Two General Data Structures: Which One?

- Dummy Variable (Least Squares)

Data Structure (*H*, *A*)
(Unique Fitting, Unbiased)

$$H = f(A; \chi_i, \beta) + \varepsilon$$

- Difference Data Fitting: $H_2 = f(H_1 - \varepsilon_1, A_1, A_2; \beta) + \varepsilon_2$

Six Data Structure (*A*1, *H*1; *A*2, *H*2): I (longest non-descending), II (longest), III (non-overlapping and non-descending), IV (non-overlapping), V (all possible non-descending), and VI (all possible) (Borders 1988, Huang 1997)

Error-In-Variable (EIV) Fitting: Tang et al. 2001, Tang and Wang 2002, Wang et al. 2005 (unbiased)

Ordinary Least Squares (OLS) Fitting: biased (assuming ε_1 as zero)

An Empirical Comparison Study: Loblolly Pine Data

- Plantation Management Research Cooperative (PMRC) established a study in 1986 in the Southeastern USA to evaluate effects of improved genetics/vegetation control on yields of loblolly pine
- The level of genetic improvement at the time was first generation improvement
- Seedlings were planted in January 1987

Loblolly Pine Data: continued (I)

- 16 Locations in Piedmont, 15 Locations in Coastal Plain
- Eight 0.4 ac. treatment plots were included at each study installation

  (1) Unimproved stock, no vegetation control (UNC),

  (2) Unimproved stock, complete vegetation control (UCC),

  (3) Bulk lot improved stock, no vegetation control (BNC),

  (4) Bulk lot improved stock, complete vegetation control (BCC),

  (5) Replicate plot of one of the first four treatments,

  (6) Single family improved stock, no vegetation control (SNC),

  (7) Single family improved stock, complete control (SCC), and

  (8) Replicate plot of one of the single family treatments.

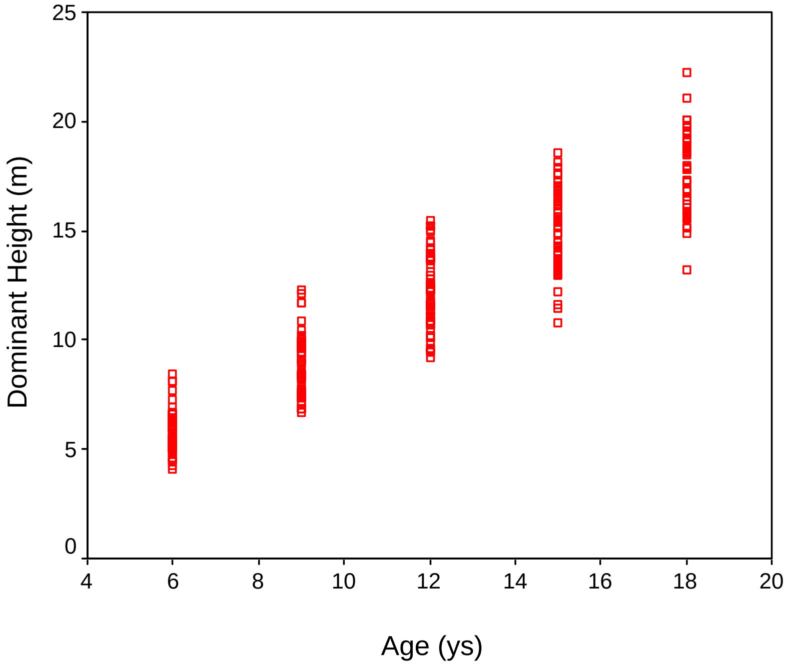
- For this analysis only treatments 3 and 6 are considered (no Veg control, single or bulk improved): previous studies show significant difference between improved and un-improved, between Veg. and no Veg.

Bulk Lot, No Veg Cntl - Age 13

Loblolly Pine Data: continued (II)

- Dominant height is defined as the average (measured) height of the trees of DBH size bigger than the median. Dominant Heights available at ages 6, 9, 12, 15, and 18 years

- 110 plots, randomly split into two data sets: model fitting and validation



Models Considered

No.	Dummy Variable Form	Difference Form
M1	$H = a_i(1-e^{-bA})^c + \varepsilon$	$H_2 = (H_1 - \varepsilon_1)\left(\frac{1-e^{-bA_2}}{1-e^{-bA_1}}\right)^c + \varepsilon_2$
M2	$H = a(1-e^{-b_i A})^c + \varepsilon$	$H_2 = a\{1-[1-((H_1-\varepsilon_1)/a)^{1/c}]^{A_2/A_1}\}^c + \varepsilon_2$
M3	$H = a(1-e^{-bA})^{c_i} + \varepsilon$	$H_2 = a[(H_1 - \varepsilon_1)/a]^{\frac{\ln(1-e^{-bA_2})}{\ln(1-e^{-bA_1})}} + \varepsilon_2$
M4	$H = e^{a+b/A+k_i e^{c/A}} + \varepsilon$	$H_2 = e^{a+b/A_2+[\ln(H_1-\varepsilon_1)-a-b/A_1]e^{c(1/A_2-1/A_1)}} + \varepsilon_2$
M5	$H = e^{\chi_i}(1-e^{-bA})^{c_1+c_2/\chi_i} + \varepsilon$	$H_2 = (H_1 - \varepsilon_1)\left(\frac{1-e^{-bA_2}}{1-e^{-bA_1}}\right)^{c_1+c_2/\chi_i} + \varepsilon_2$ $\chi_i = \frac{1}{2}\{[\ln(H_1-\varepsilon_1)-c_1 L] + \sqrt{[\ln(H_1-\varepsilon_1)-c_1 L]^2 - 4c_2 L}\}$ $L = \ln(1-e^{-bA_1})$
M6	$H = \frac{\alpha + \chi_i}{1+\frac{\beta}{\chi_i}A^{-c}} + \varepsilon$	$H_2 = (H_1 - \varepsilon_1)\left(\frac{1+\frac{\beta}{\chi_i}A_1^{-c}}{1+\frac{\beta}{\chi_i}A_2^{-c}}\right) + \varepsilon_2$ $\chi_i = \frac{1}{2}\{[(H_1-\varepsilon_1)-\alpha] + \sqrt{[(H_1-\varepsilon_1)-\alpha]^2 + 4\beta(H_1-\varepsilon_1)A_1^{-c}}\}$

Model Fitting

- Dummy Variable OLS, (H, A)
- EIV Difference Fitting (III): Matlab Program Published (Tang and Wang 2002, Ecological Modelling), III recommended by Wang et al. (2005)
- OLS Difference Fitting (III, IV, V, VI)

Six difference data structures illustrated with a height growth series of four points (h_1, t_1), (h_2, t_2), (h_3, t_3), (h_4, t_4)

I	II	III	IV	V	VI
$H_1\ A_1\ H_2\ A_2$	$H_1\ A_1\ H_2\ A_2$	$H_1\ A_1\ H_2\ A_2$	$H_1\ A_1\ H_2\ A_2$	$H_1\ A_1\ H_2\ A_2$	$H_1\ A_1\ H_2\ A_2$
$h_1\ t_1\ h_4\ t_4$	$h_1\ t_1\ h_4\ t_4$	$h_1\ t_1\ h_2\ t_2$	$h_1\ t_1\ h_2\ t_2$	$h_1\ t_1\ h_2\ t_2$	$h_1\ t_1\ h_2\ t_2$
	$h_4\ t_4\ h_1\ t_1$	$h_2\ t_2\ h_3\ t_3$	$h_2\ t_2\ h_1\ t_1$	$h_1\ t_1\ h_3\ t_3$	$h_1\ t_1\ h_3\ t_3$
		$h_3\ t_3\ h_4\ t_4$	$h_2\ t_2\ h_3\ t_3$	$h_1\ t_1\ h_4\ t_4$	$h_1\ t_1\ h_4\ t_4$
			$h_3\ t_3\ h_2\ t_2$	$h_2\ t_2\ h_3\ t_3$	$h_2\ t_2\ h_1\ t_1$
			$h_3\ t_3\ h_4\ t_4$	$h_2\ t_2\ h_4\ t_4$	$h_2\ t_2\ h_3\ t_3$
			$h_4\ t_4\ h_3\ t_3$	$h_3\ t_3\ h_4\ t_4$	$h_2\ t_2\ h_4\ t_4$
					$h_3\ t_3\ h_1\ t_1$
					$h_3\ t_3\ h_2\ t_2$
					$h_3\ t_3\ h_4\ t_4$
					$h_4\ t_4\ h_1\ t_1$
					$h_4\ t_4\ h_2\ t_2$
					$h_4\ t_4\ h_3\ t_3$

Model Evaluation

- Note that in the fitting stage the same model in dummy variable form and difference form are not comparable, as they in fact use different fit data
- Height Prediction Given Site Index (Observed Height at base age 18): $\hat{H} = f(A, A_s, S, \hat{\beta})$

$$MD_H = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{H_{ij} - \hat{H}_{ij}}{(n_1 + n_2 + \ldots + n_N)} \qquad RMSE_H = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(H_{ij} - \hat{H}_{ij})^2}{(n_1 + n_2 + \ldots + n_N)}}$$

- Site Index (base age 18) Prediction Given (Observed) Height: $\hat{S} = f^{-1}(A, A_s, H, \hat{\beta})$

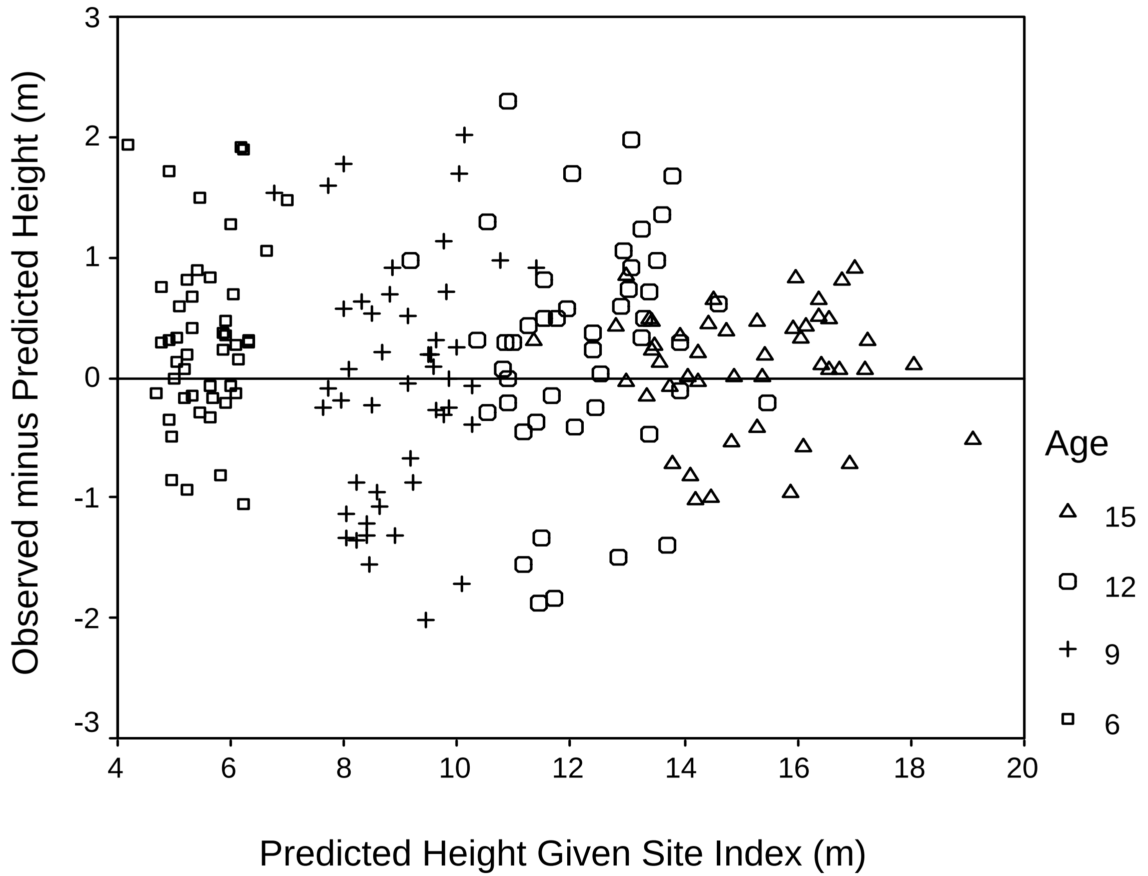
$$MD_S = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{S_{ij} - \hat{S}_{ij}}{(n_1 + n_2 + \ldots + n_N)} \qquad RMSE_S = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(S_{ij} - \hat{S}_{ij})^2}{(n_1 + n_2 + \ldots + n_N)}}$$

Results – Height Prediction Given Site Index

Method	No.	MD_H	$RMSE_H$	Sum rank	No.	MD_H	$RMSE_H$	Sum rank
Dummy	M1	0.16(6)	0.84(6)	12	M5	0.14(6)	0.83(6)	12
EIV III		0.10(3)	0.83(3)	6		0.08(4)	0.82(2)	6
III		0.06(1)	0.82(1)	**2**		0.04(1)	0.81(1)	**2**
IV		0.08(2)	0.82(1)	3		0.07(3)	0.82(2)	5
V		0.12(4)	0.83(3)	7		0.06(2)	0.82(2)	4
VI		0.13(5)	0.83(3)	8		0.10(5)	0.82(2)	7
Dummy	M2	0.11(5)	0.85(1)	6	M6	0.13(6)	0.83(5)	11
EIV III		0.04(3)	0.85(1)	4		0.08(4)	0.82(3)	7
III		-0.08(4)	0.88(4)	8		0.01(1)	0.81(1)	**2**
IV		0.02(1)	0.85(1)	**2**		0.05(3)	0.81(1)	4
V		-0.35(6)	1.11*(6)	12		-0.08(4)	0.84(6)	10
VI		-0.03(2)	0.89(5)	7		0.01(1)	0.82(3)	4

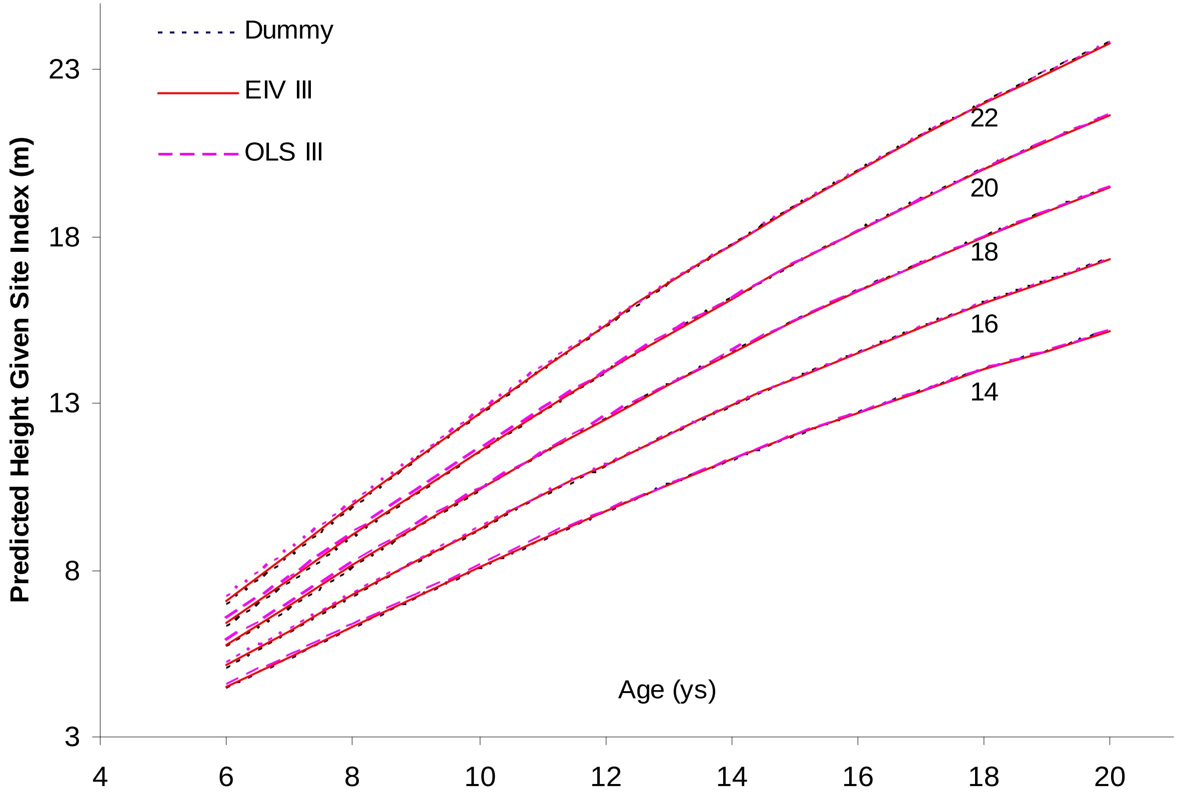
Results-Height Prediction (cont.)

- Generally Unbiased



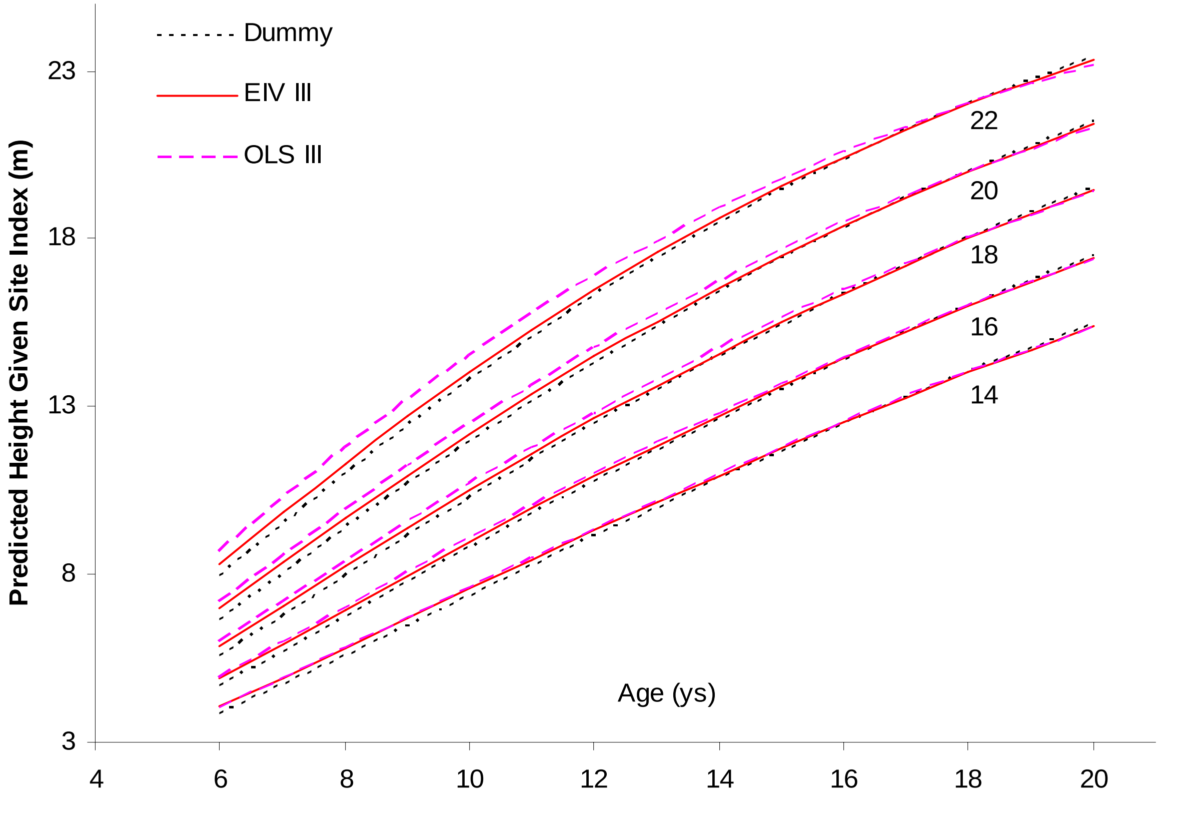
Results-Height Prediction (cont.)

- Similar Prediction Performance: Dummy, EIV III, OLS III



Predicted heights given site index (base age 18) using anamorphic Chapman-Richards model M1 fitted with three estimation methods

Results-Height Prediction (cont.)



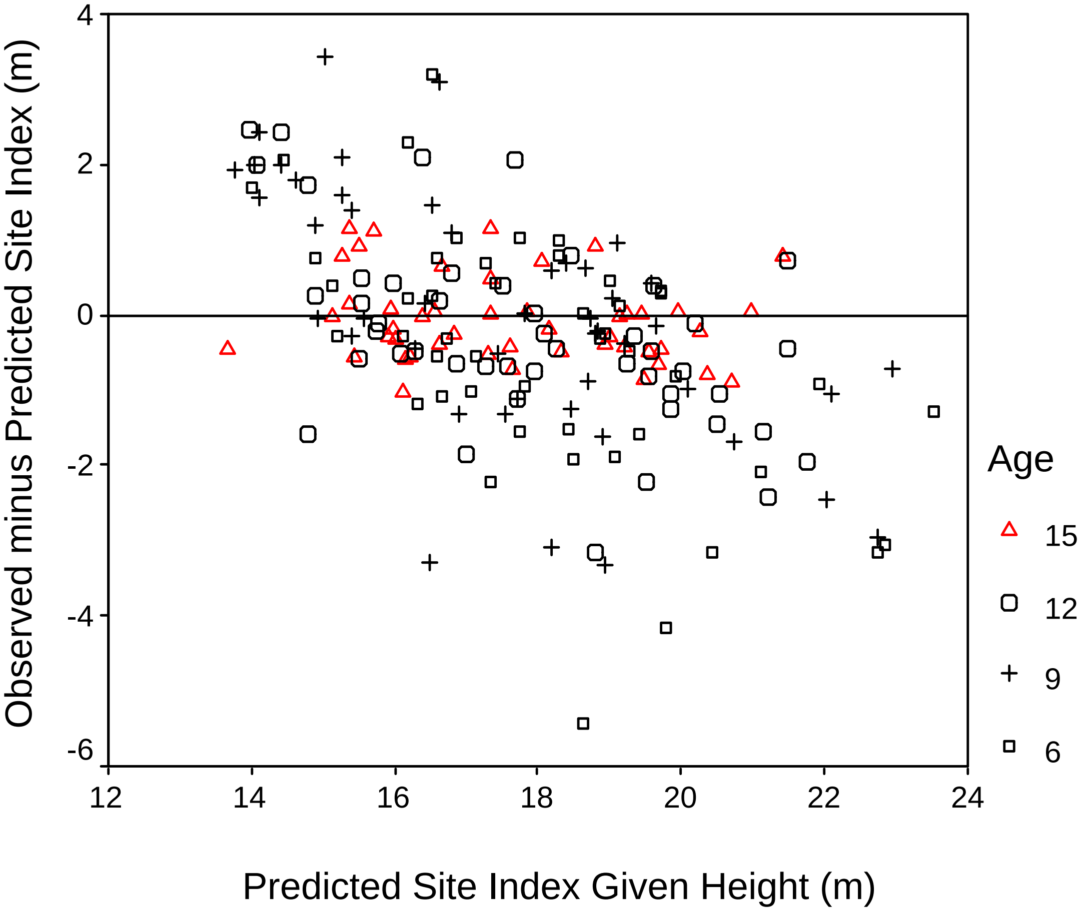
Predicted heights given site index (base age 18) using polymorphic (single asymptote) Chapman-Richards model M2 fitted with three estimation methods

Results – Site Index Prediction Given Height

Method	No.	MD_S	$RMSE_S$	Sum rank	No.	MD_S	$RMSE_S$	Sum rank
Dummy	M1	-0.36(6)	1.77*(6)	12	M5	-0.27(6)	1.50*(6)	12
EIV III		-0.23(4)	1.71*(5)	9		-0.16(4)	1.46*(4)	8
III		-0.12(1)	1.67*(1)	**2**		-0.07(1)	1.44*(1)	**2**
IV		-0.17(2)	1.69*(3)	5		-0.12(3)	1.45*(2)	5
V		-0.22(3)	1.68*(2)	6		-0.11(2)	1.45*(2)	4
VI		-0.26(5)	1.70*(4)	9		-0.18(5)	1.47*(5)	10
Dummy	M2	-0.16(5)	1.24*(6)	11	M6	-0.25(6)	1.48*(6)	12
EIV III		-0.06(3)	1.23*(5)	8		-0.15(5)	1.45*(5)	10
III		0.08(4)	1.20(3)	7		-0.03(1)	1.43*(2)	**3**
IV		-0.03(2)	1.22*(4)	6		-0.10(4)	1.44*(4)	8
V		0.28(6)	1.19(2)	8		0.07(3)	1.41*(1)	4
VI		-0.00(1)	1.18(1)	**2**		-0.06(2)	1.43*(2)	4

Results-Site Index Prediction (cont.)

- Generally Biased



Results-Site Index Prediction (cont.)

- But, Generally, Better Prediction Performance of Difference Methods (EIV III, OLS III) over Dummy Variable Method

- Generally, Better Prediction Performance of OLS III over EIV III

- Generally, OLS III > EIV III >Dummy

BUT!

Biased Estimation

Unbiased Estimation

Why?

Possible Explanation

- Height Prediction (given site index) and Site Index Prediction (given height) involve two processes:
  - Estimation of COMMON parameters:

    Dummy and EIV III are Unbiased Estimators of Common Parameters: $H = f(A; \chi_i, \beta) + \varepsilon \quad H_2 = f(H_1 - \varepsilon_1, A_1, A_2; \beta) + \varepsilon_2$

    OLS III Biased: $H_2 = f(H_1, A_1, A_2; \beta) + \varepsilon_2$
  - Prediction of Local (plot-specific) parameters (with ONE observed Site Index or Height), however, is under suspicion! (imagine the problem of estimating a sample mean with only one observation subject to ERROR)

$$\hat{\chi}_i = f^{-1}(H_0, A_0, \hat{\beta}) \qquad \hat{S} = f^{-1}(A_0, A_s, H_0, \hat{\beta})$$

[error in H_0!]

Summary/Conclusions

- We recommend the dummy variable method if the main concern is height prediction as it avoids the necessity of choosing the difference data structure for model fitting

- However, if we are interested in both height prediction and site index prediction, we recommend the difference method with data structure III (OLS or EIV).

Thank You for Your Attention!