# Height Imputation for the Acadian Variant of FVS (ACD)

Greg Johnson
Greg Johnson Biometrics LLC

2024-06-06

## Data

Data from the `ALL_DHT.csv` data set was used for this analysis. The data come from 4: FIA ME, Nova Scotia PSP, ERM ME, CTRN sources. Preliminary examination of the data revealed that the `CTRN` data behaved differently from the other sources. The `CTRN` study[1] focuses on commercial thinning and could reasonably expected to alter the diameter - height relationship. For the purposes of this analysis, it was dropped. The resulting data set includes 783104 observations distributed by species shown in Table 1.

In the analysis that follows, equations for individual species were limited to those species with > 2500 observations (set after inspecting the stability of the parameter estimates for species with low sample sizes). This left 774382 observations. For species not meeting that threshold, two groups: Other Conifers (`OC`), and Other Hardwoods (`OH`) were used.

Table 1: Height Observations by Species

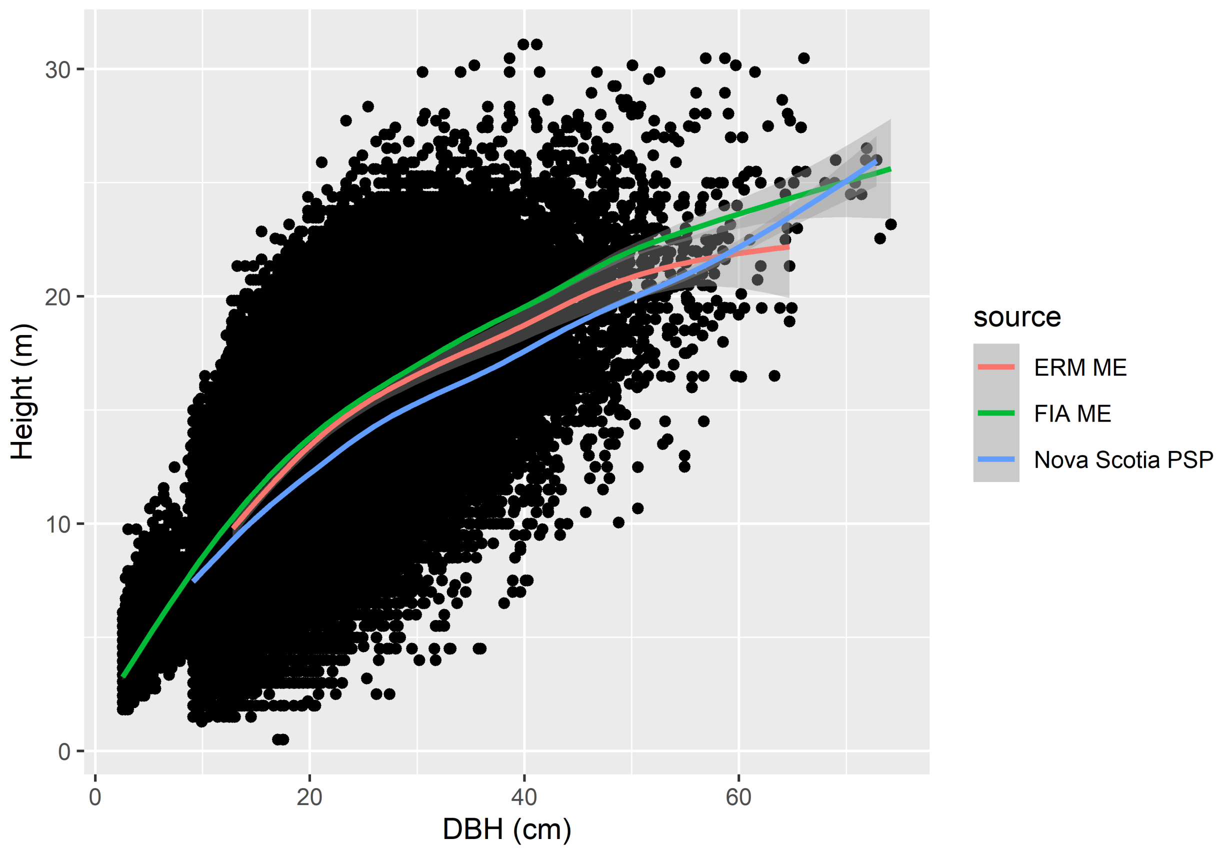
| FVS Sp | FIA Sp | N Observations |
|---|---|---|
| BF | 12 | 192357 |
| RS | 97 | 139907 |
| RM | 316 | 126834 |
| BS | 95 | 52292 |
| WS | 94 | 43596 |
| PB | 375 | 40746 |
| SM | 318 | 29398 |
| YB | 371 | 29163 |
| WP | 129 | 21536 |
| EH | 261 | 20403 |
| WC | 241 | 17124 |
| AB | 531 | 15336 |
| QA | 746 | 11045 |
| TA | 71 | 11039 |
| RO | 833 | 9278 |
| BT | 743 | 5494 |
| WA | 541 | 5119 |
| GB | 379 | 3715 |

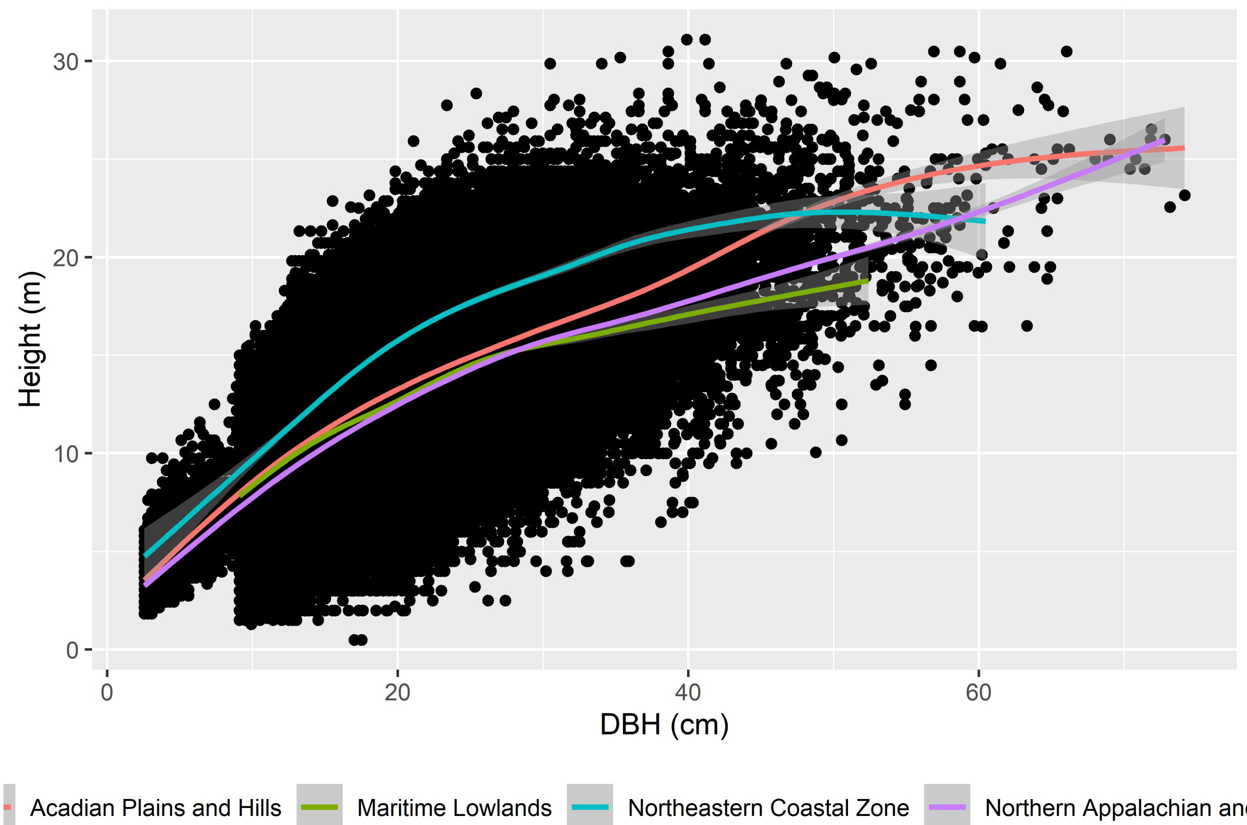
[1] Seymour, Robert S.; Meyer, Spencer R.; Wagner, Robert G. 2014. The cooperative forestry research unit Commercial Thinning Research Network–9-year results. In: Kenefic, Laura S.; Brissette, John C., comps. Penobscot Experimental Forest: 60 years of research and demonstration in Maine, 1950-2010. Gen. Tech. Rep. NRS-P-123. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 81-90.

| | | |
|---|---|---|
| RN | 125 | 2403 |
| JP | 105 | 1181 |
| HH | 701 | 981 |
| BA | 543 | 900 |
| BC | 762 | 763 |
| ST | 315 | 482 |
| BP | 741 | 431 |
| AE | 972 | 256 |
| PR | 761 | 215 |
| WO | 802 | 191 |
| BO | 837 | 162 |
| BW | 951 | 152 |
| GA | 544 | 129 |
| PP | 126 | 76 |
| AP | 660 | 74 |
| SE | 356 | 70 |
| SB | 372 | 66 |
| SV | 317 | 13 |
| AH | 391 | 10 |
| BN | 601 | 5 |
| SH | 407 | 4 |
| HT | 500 | 3 |
| SO | 806 | 3 |
| BE | 313 | 2 |
| PY | 744 | 2 |
| YP | 621 | 2 |
| BR | 823 | 1 |
| EC | 742 | 1 |
| NS | 91 | 1 |
| TM | 123 | 1 |

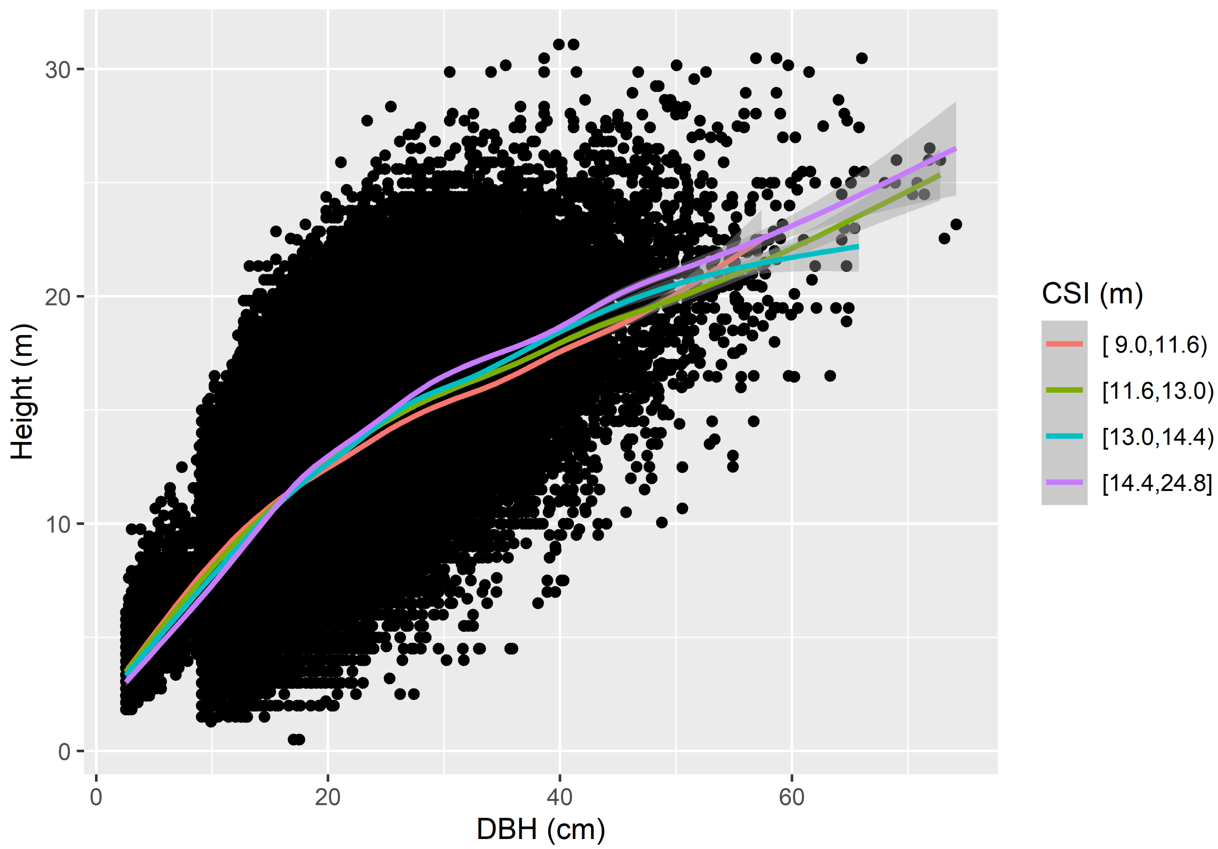
The objective of this analysis is to develop equations to predict height (`ht`) from tree and stand variables for trees missing `ht` observations. The primary explanatory variable is diameter at breast height (`dbh`) and its relationship to `ht` is shown below. It is well known that the shape of the `dbh-ht` relationship is different (typically flatter) within locations than among them. One challenge is to try to account for this shape change with additional information.

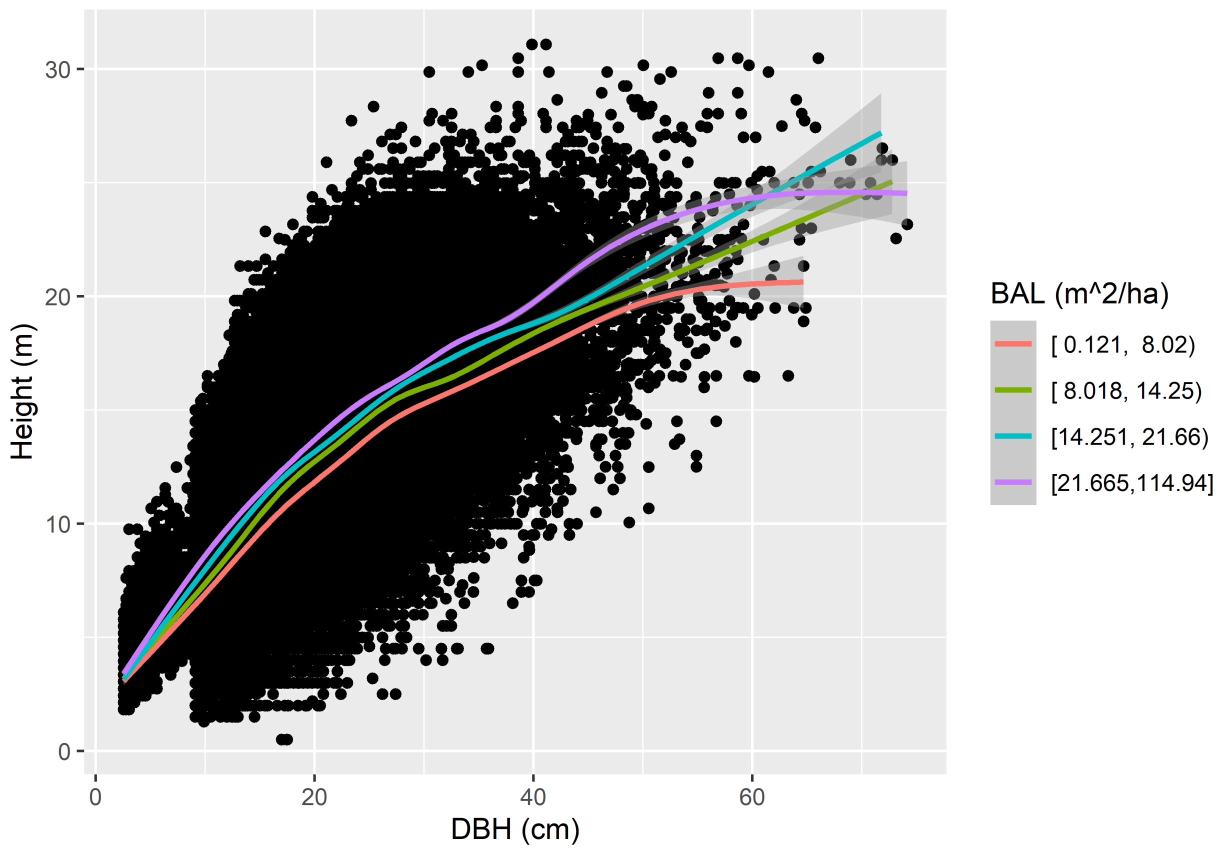
The graphs below show some distinct differences in the primary `dbh - ht` relationship among data sources and eco-regions.


Height (m)
30
20
10
0
0
20
40
60
DBH (cm)
source
ERM ME
FIA ME
Nova Scotia PSP

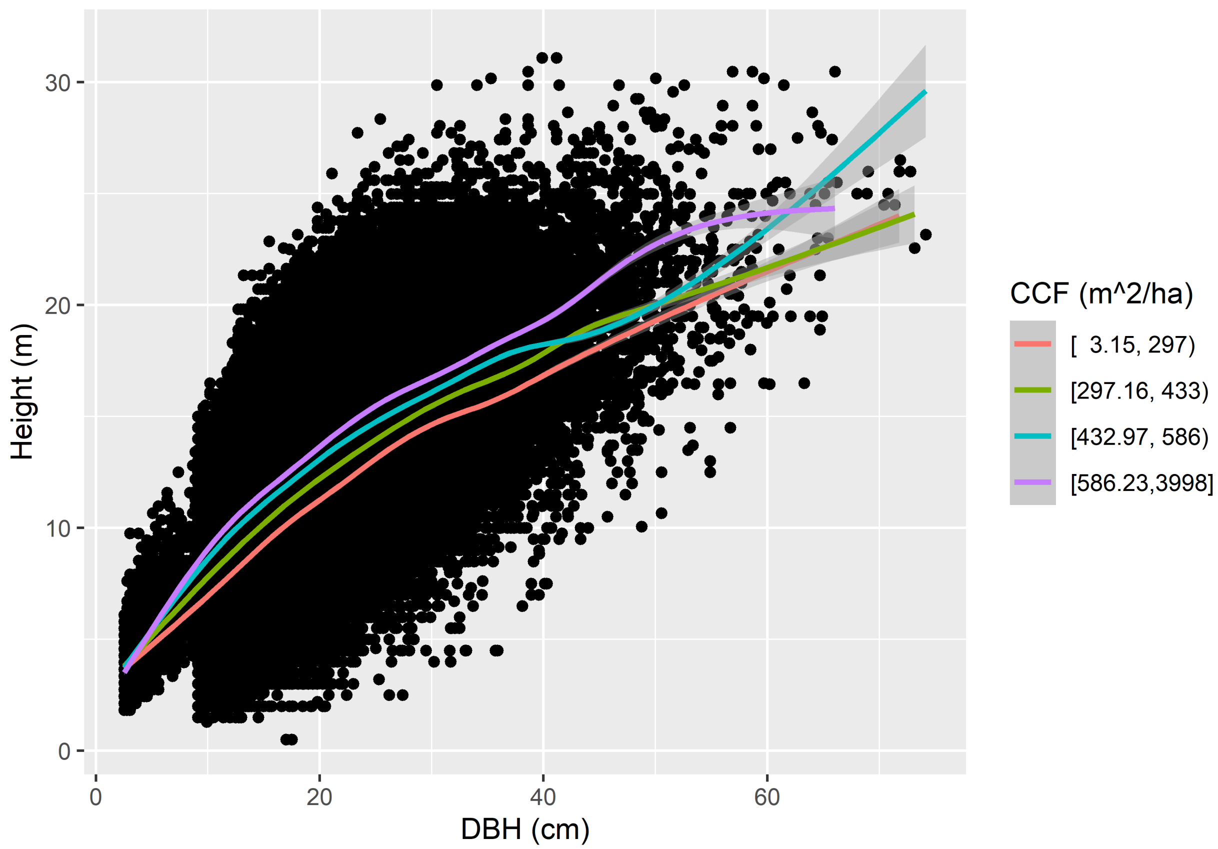



The following graphs show the overall relationship between `ht` and other variables (climate site index (`csi`), basal area in larger trees (`bal`), crown competition factor (`ccf`), and crown competition factor in larger trees (`ccfl`)). Of the these variables, `csi` appears to have the least influence on the `dbh` - `ht` relationship.

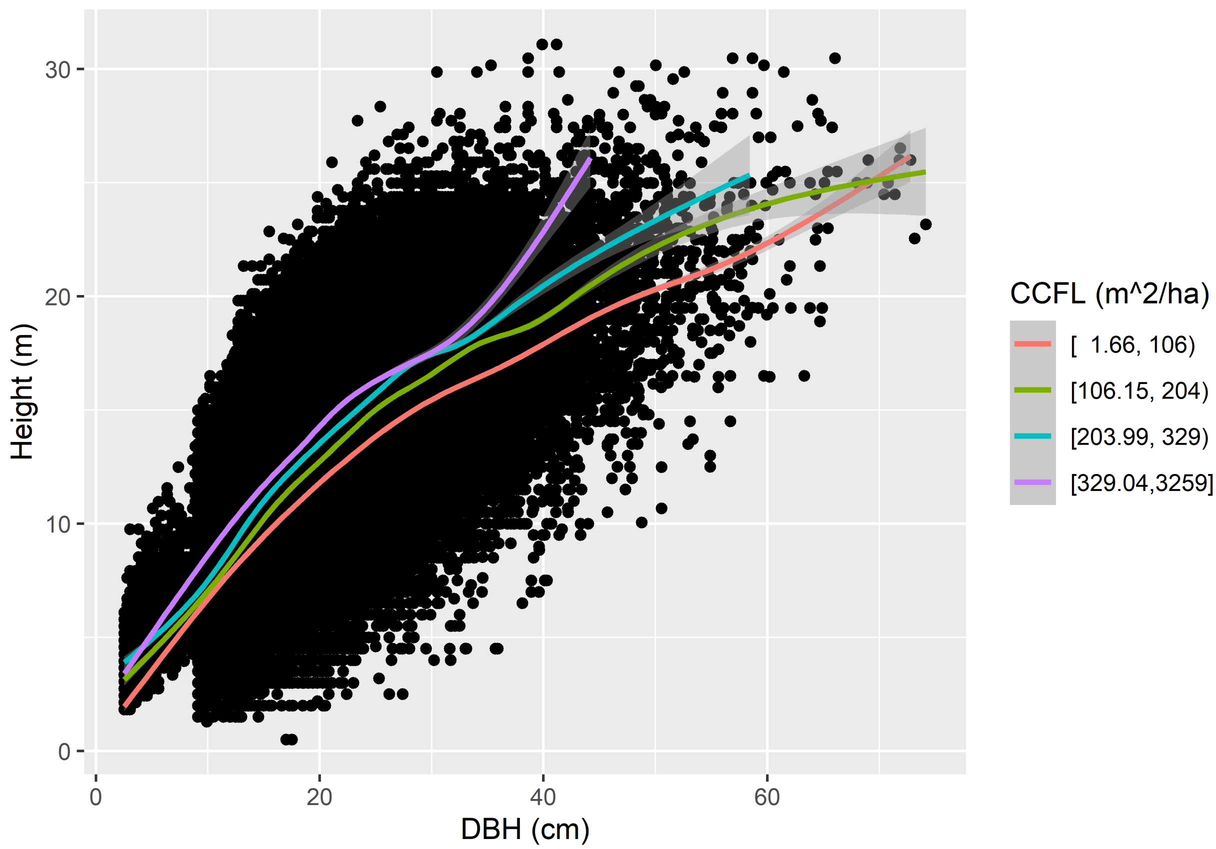

Height (m)
DBH (cm)
CSI (m)
[ 9.0,11.6)
[11.6,13.0)
[13.0,14.4)
[14.4,24.8]
0
10
20
30
0
20
40
60



30
20
10
0
0
20
40
60
Height (m)
DBH (cm)
BAL (m^2/ha)
[ 0.121,  8.02)
[ 8.018, 14.25)
[14.251, 21.66)
[21.665,114.94]



30
20
10
0
0
20
40
60
DBH (cm)
Height (m)
CCF (m^2/ha)
[ 3.15, 297)
[297.16, 433)
[432.97, 586)
[586.23,3998]



Height (m)
DBH (cm)
0
10
20
30
0
20
40
60
CCFL (m^2/ha)
[ 1.66, 106)
[106.15, 204)
[203.99, 329)
[329.04,3259]

## Existing Equation Performance

To provide a basis for comparison with an Alternative equation formulation, we fit the existing equation form to the data set and examine the residuals.

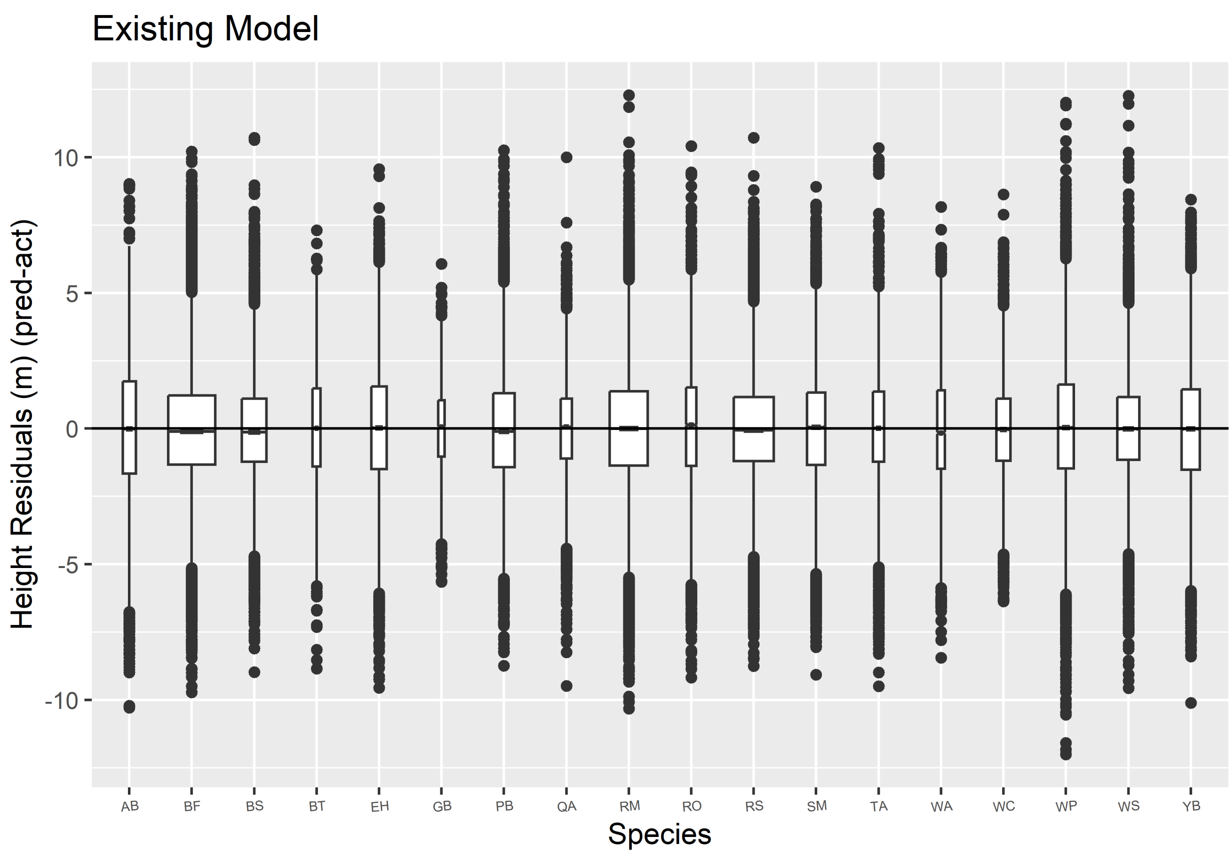
The existing equation is:

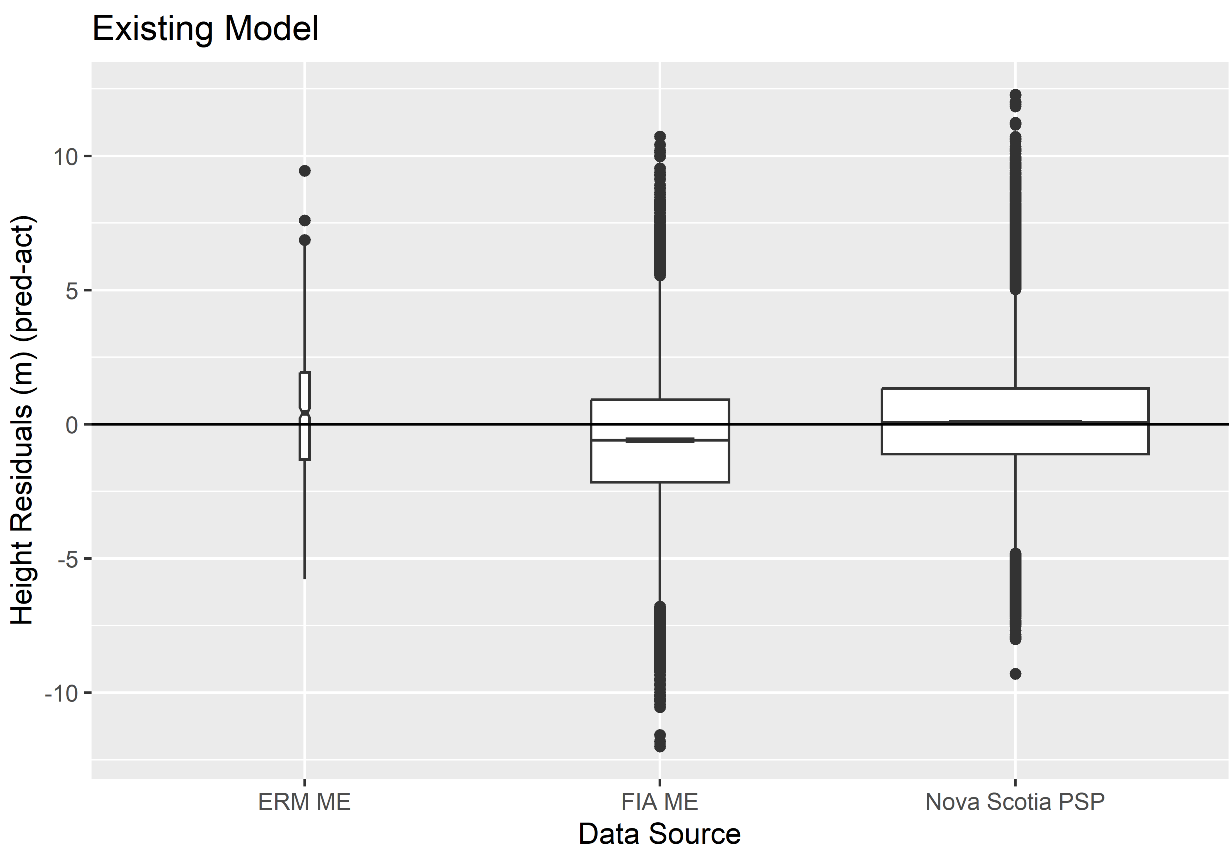
$$\hat{ht} = 1.37 + (\beta_0 + csi^{\beta_1})(1.0 - e^{-\beta_2 dbh})^{(\beta_3 + \beta_4 log(ccf+1.0) + \beta_5 bal))}$$

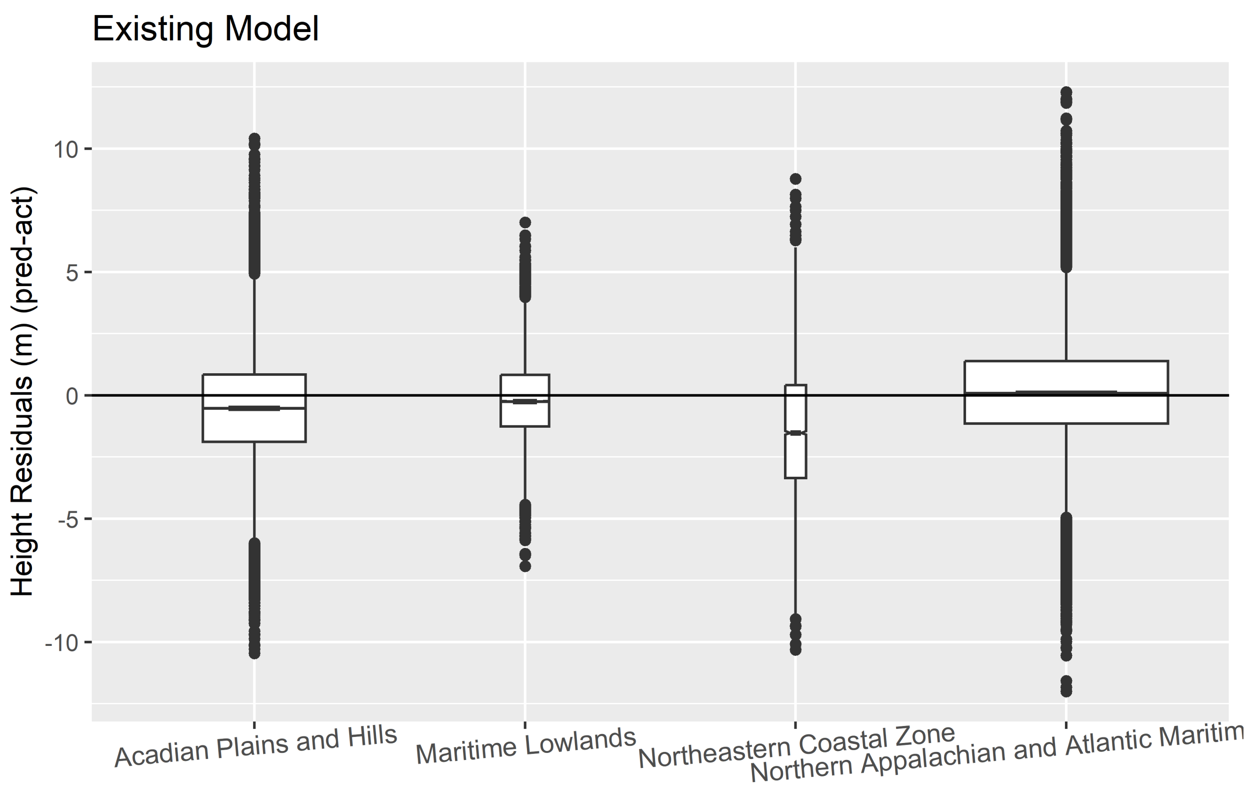
where $\beta_0$ and $\beta_3$ are treated as random effects.

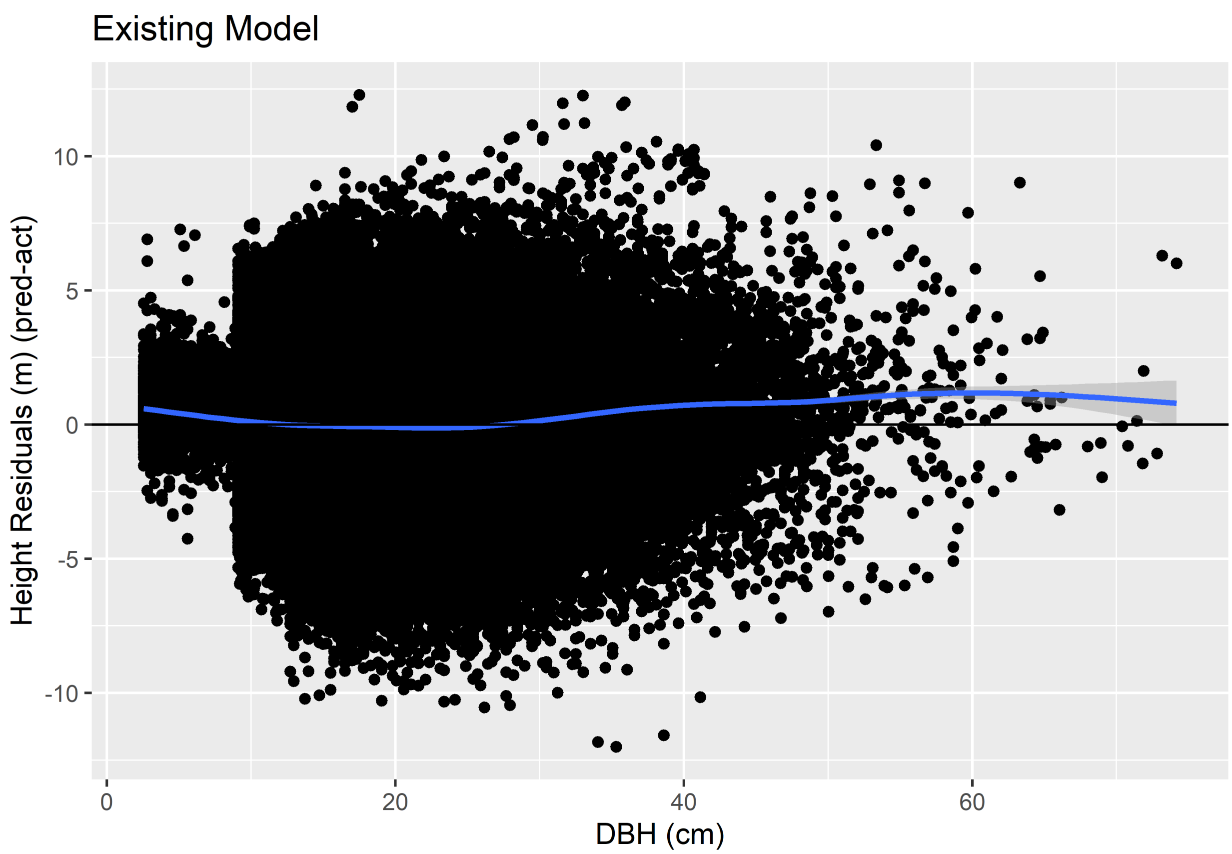
```
## Nonlinear mixed-effects model fit by maximum likelihood
##   Model: ht ~ 1.37 + (b0 + csi^b1) * (1 - exp(-b2 * dbh))^(b3 + b4 * log(ccf +      1) + b5 * bal)
##   Data: ht2
##       AIC     BIC   logLik
##   3289463 3289578 -1644721
##
## Random effects:
##  Formula: list(b0 ~ 1, b3 ~ 1)
##  Level: fvs
##  Structure: General positive-definite, Log-Cholesky parametrization
##          StdDev    Corr
## b0       4.6845329 b0
## b3       0.1108814 0.57
## Residual 2.0234391
##
## Fixed effects:  b0 + b1 + b2 + b3 + b4 + b5 ~ 1
##       Value Std.Error     DF   t-value p-value
## b0 33.58115 1.2009454 774359  27.96226       0
## b1  0.47065 0.0076540 774359  61.49038       0
## b2  0.01351 0.0003290 774359  41.06858       0
## b3  1.28089 0.0277260 774359  46.19821       0
## b4 -0.07243 0.0009723 774359 -74.49580       0
## b5 -0.00317 0.0000462 774359 -68.48670       0
##  Correlation:
##    b0     b1     b2     b3     b4
## b1  0.059
## b2 -0.381 -0.295
## b3  0.375 -0.104  0.323
## b4  0.362  0.311 -0.946 -0.326
## b5  0.351  0.270 -0.931 -0.299  0.837
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -6.07397462 -0.62475407  0.02298072  0.64453798  5.93316779
##
## Number of Observations: 774382
## Number of Groups: 18
```

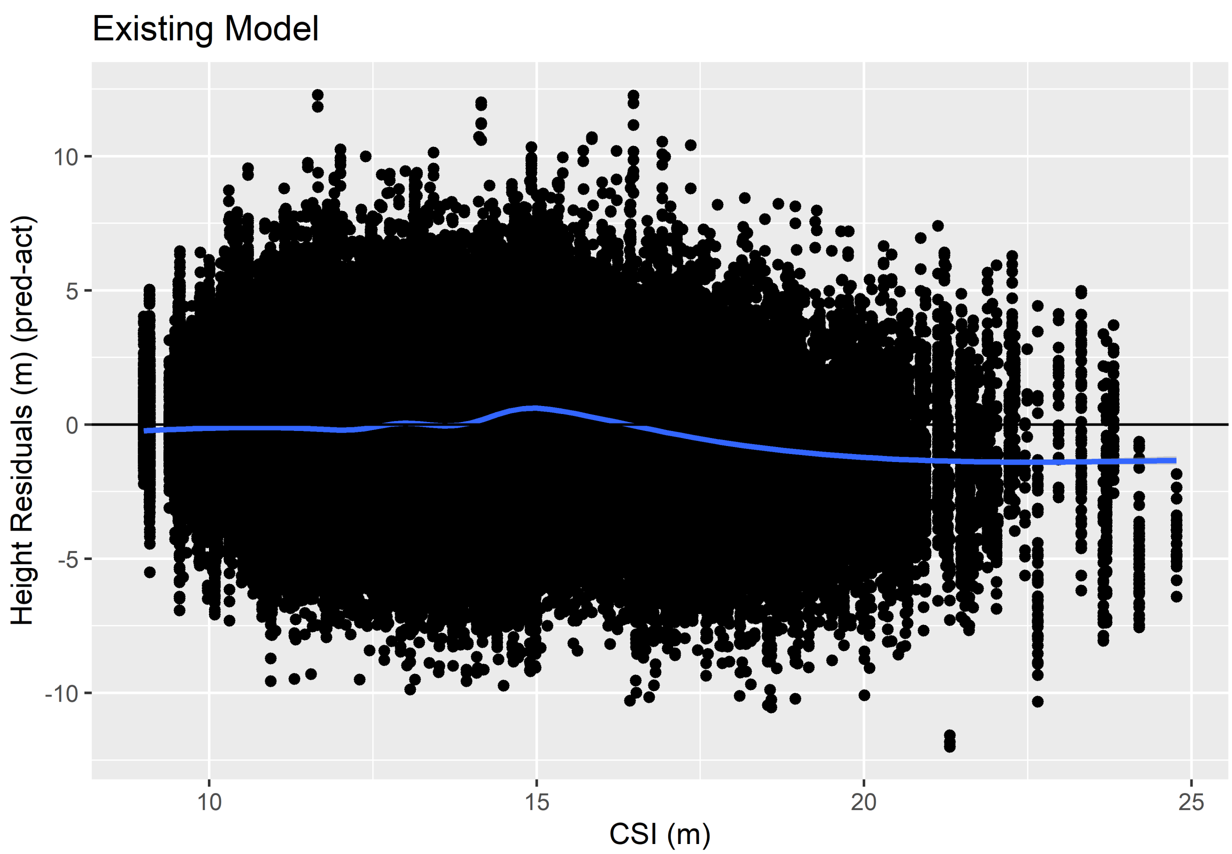
The residual analysis indicates that `ht` is over-predicted for trees under high levels of competition and larger trees. There is also a trend toward under-estimating `ht` for high `csi` values. One eco-region (Northeastern Coastal Zone) is under-predicted as well.

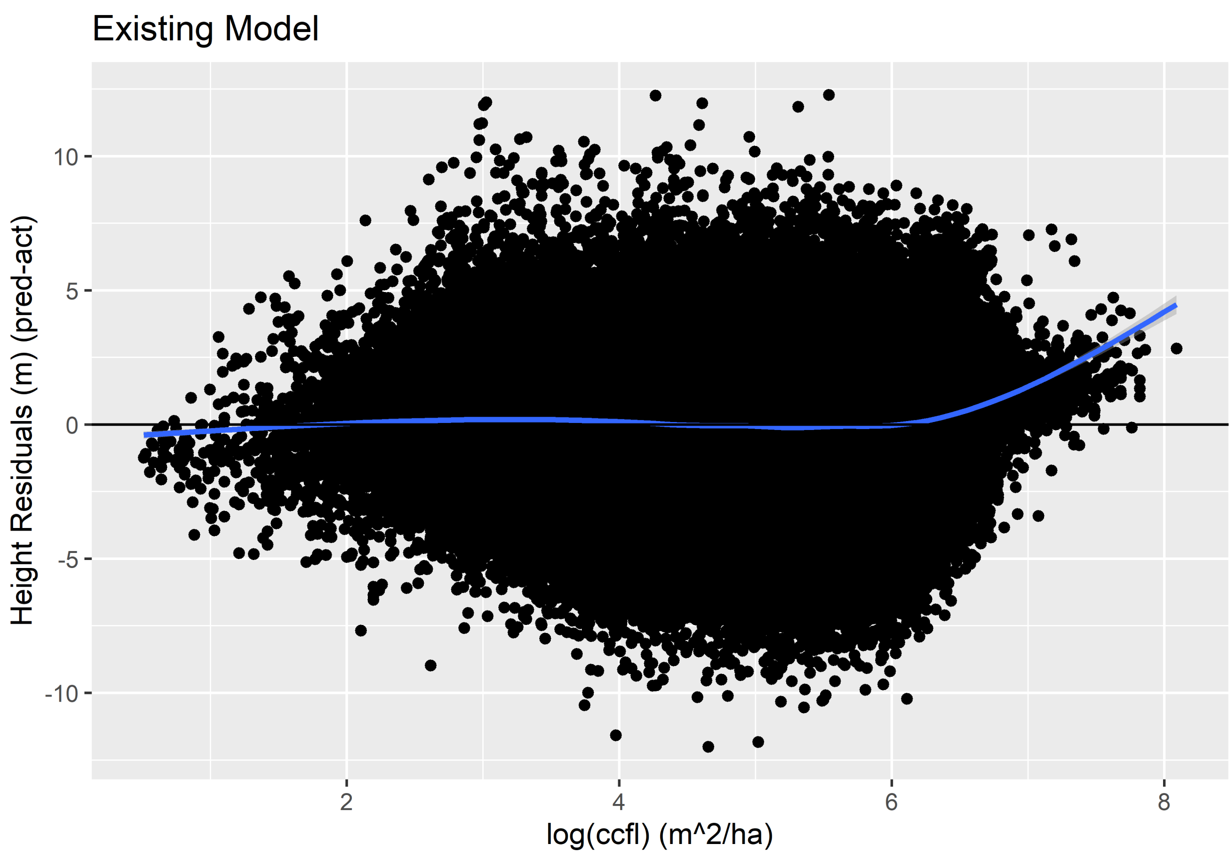

Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
AB
BF
BS
BT
EH
GB
PB
QA
RM
RO
RS
SM
TA
WA
WC
WP
WS
YB
Species

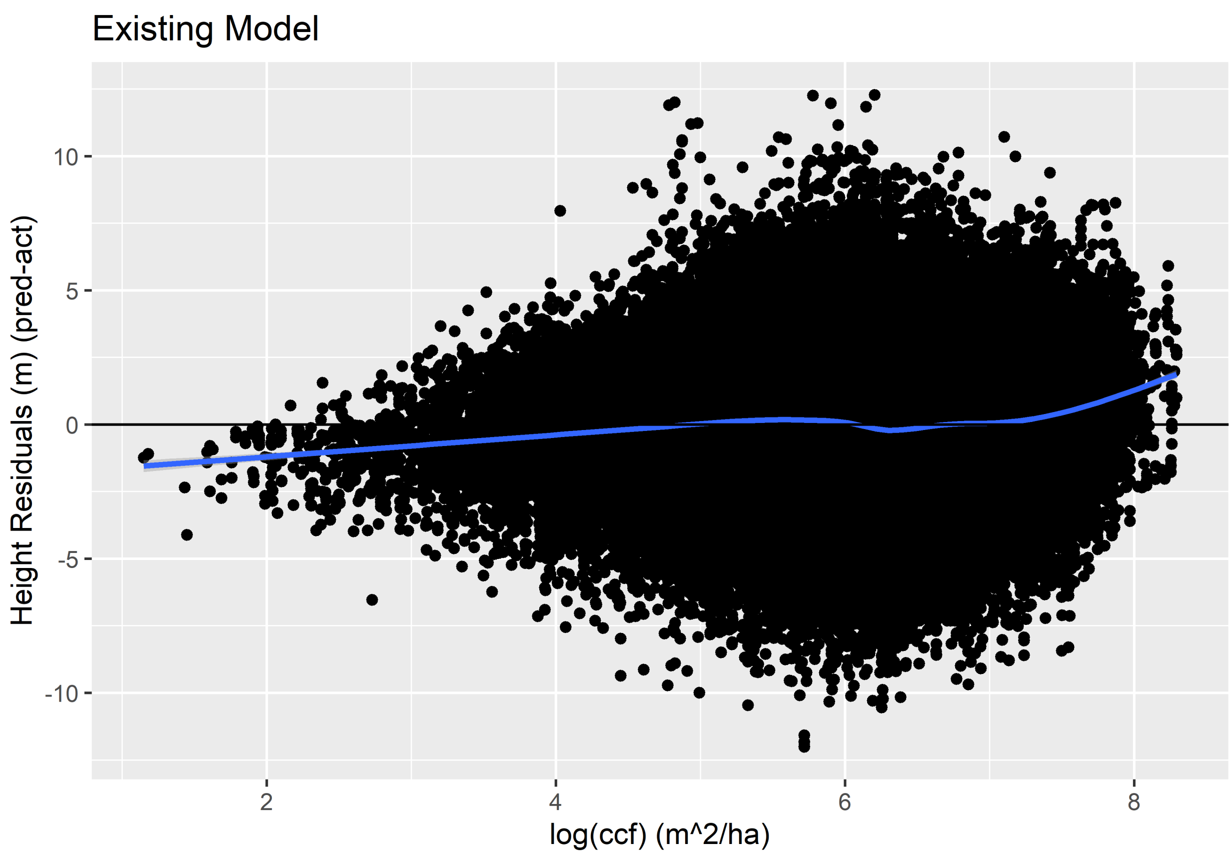

Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
ERM ME
FIA ME
Nova Scotia PSP
Data Source

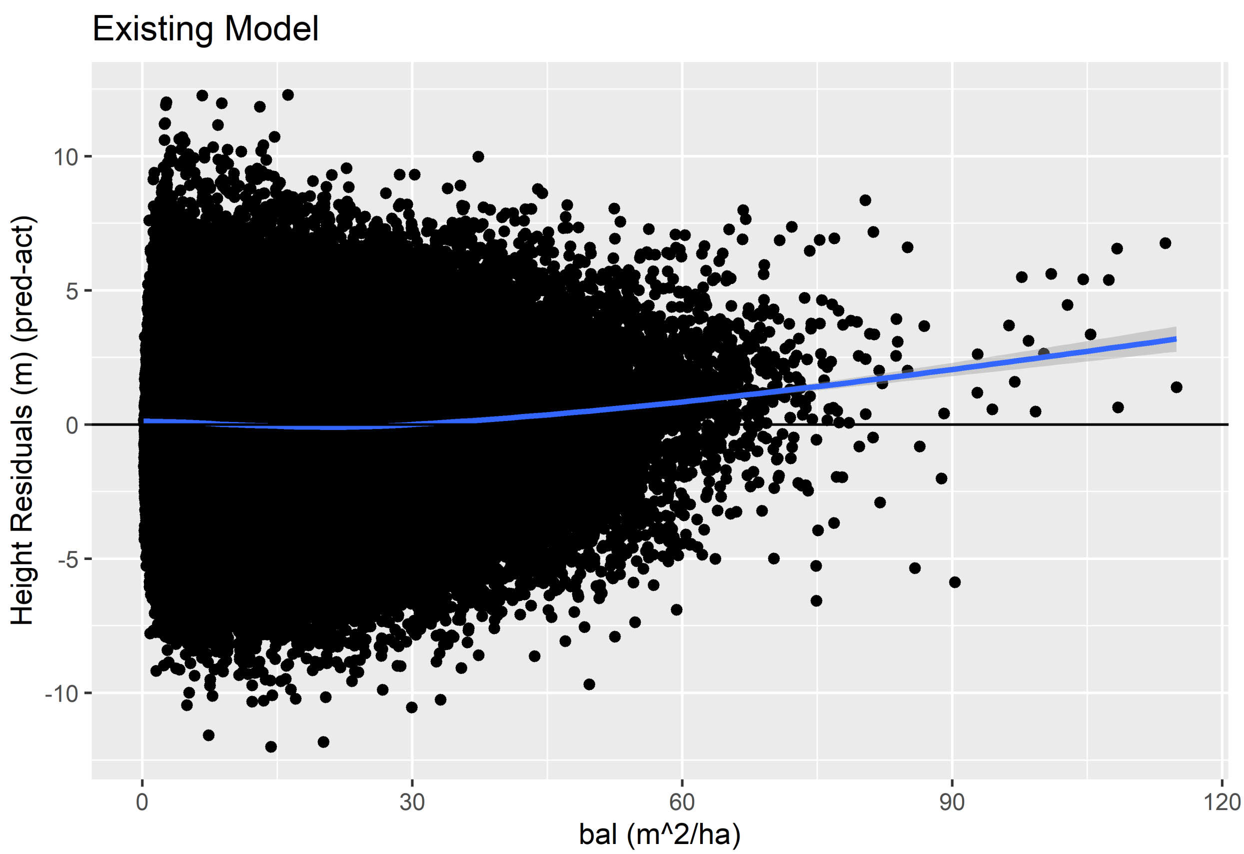

Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
Acadian Plains and Hills
Maritime Lowlands
Northeastern Coastal Zone
Northern Appalachian and Atlantic Maritim


Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
0
20
40
60
DBH (cm)


Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
10
15
20
25
CSI (m)


Existing Model
10
5
0
-5
-10
Height Residuals (m) (pred-act)
2
4
6
8
log(ccfl) (m^2/ha)


Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
2
4
6
8
log(ccf) (m^2/ha)


Existing Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
0
30
60
90
120
bal (m^2/ha)

## Alternative Equation

An alternative equation (`Alternative`) formulated to eliminate the residual trends noted for the Existing Equation (`Existing`) uses the following form:

$$\hat{ht} = 1.37 + (\beta_0 + \beta_{0a} Region)(1.0 - e^{-\beta_1 dbh - \beta_3 (bal+1.0)})^{\beta_2} log(ccf)^{\beta_4}$$

where `Region` is an indicator variable (0 if in Maine, 1 if in Nova Scotia).

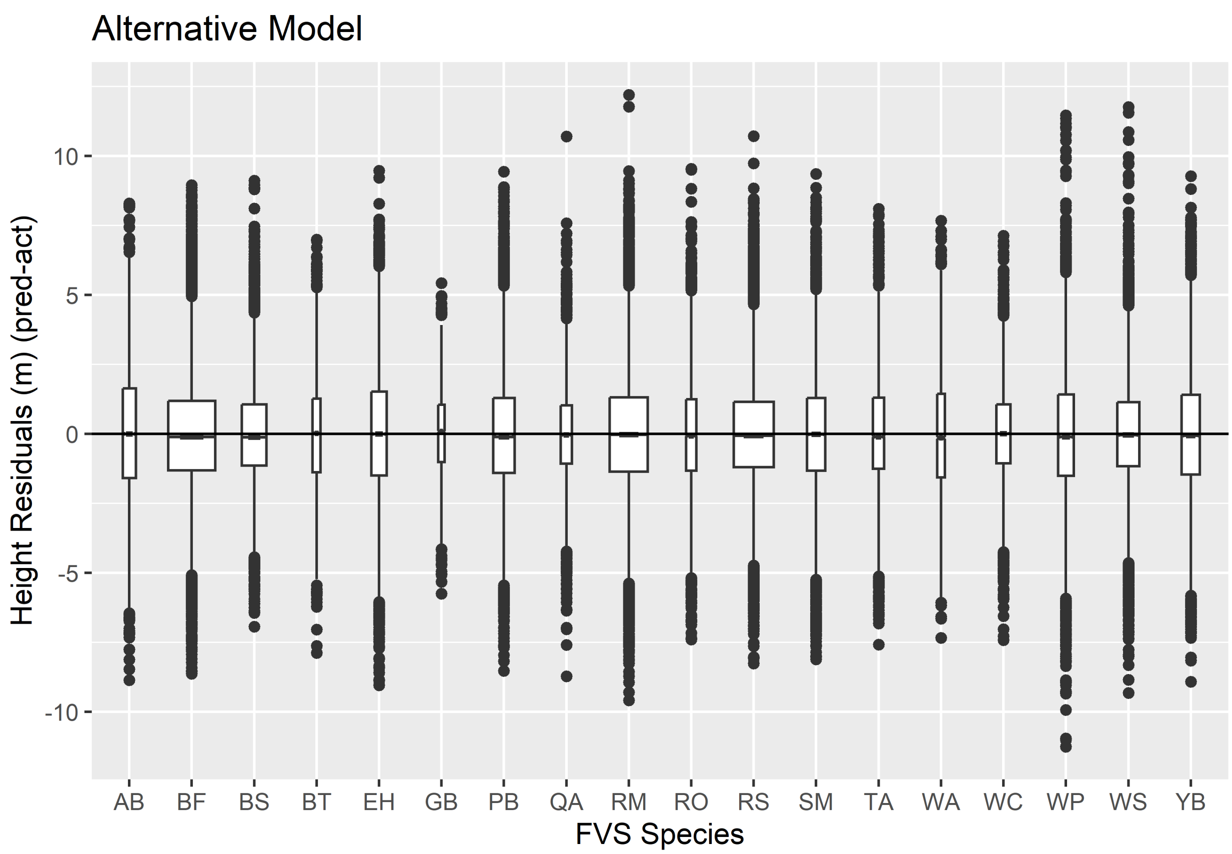
Species were fit individually (in contrast to `Existing` where a mixed model was used). Where a species occurred only in one `Region` the indicator variable was dropped from the equation.

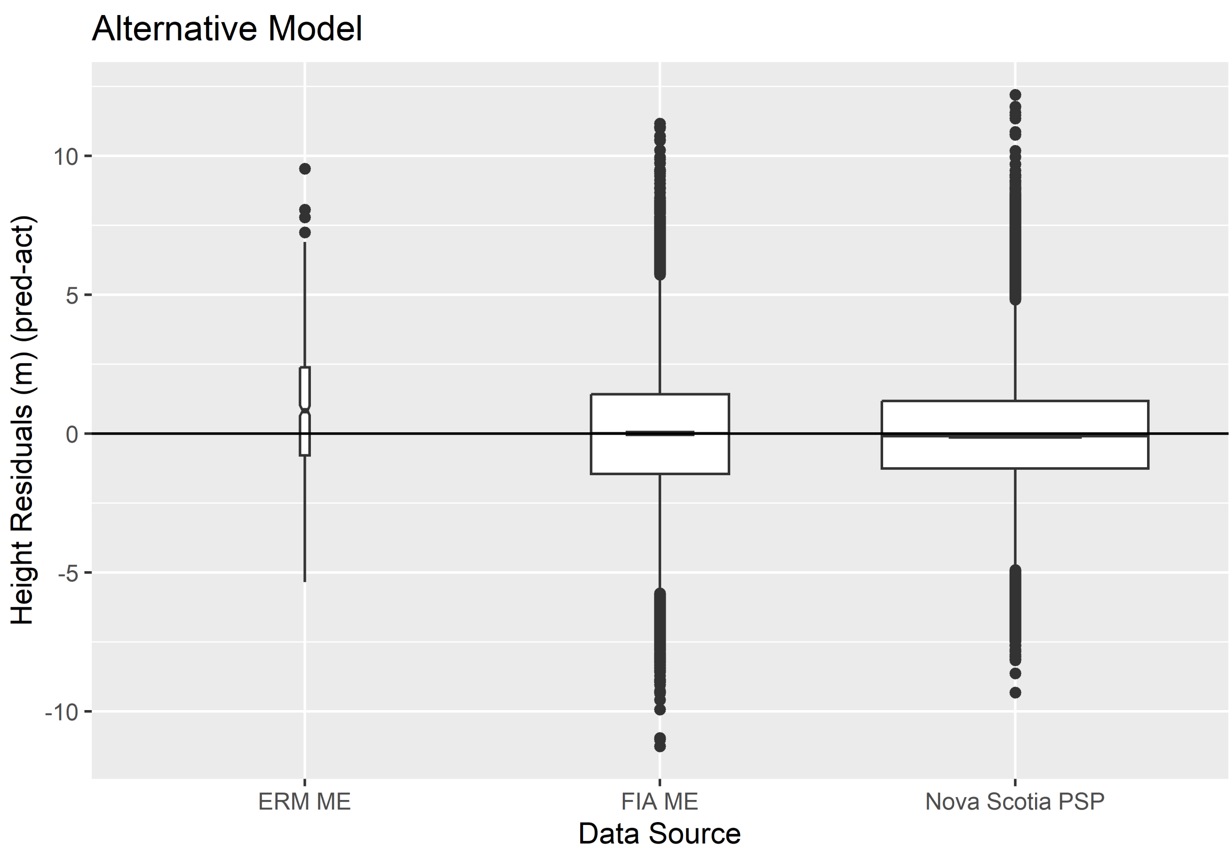
The parameter estimates are found in Table 2.

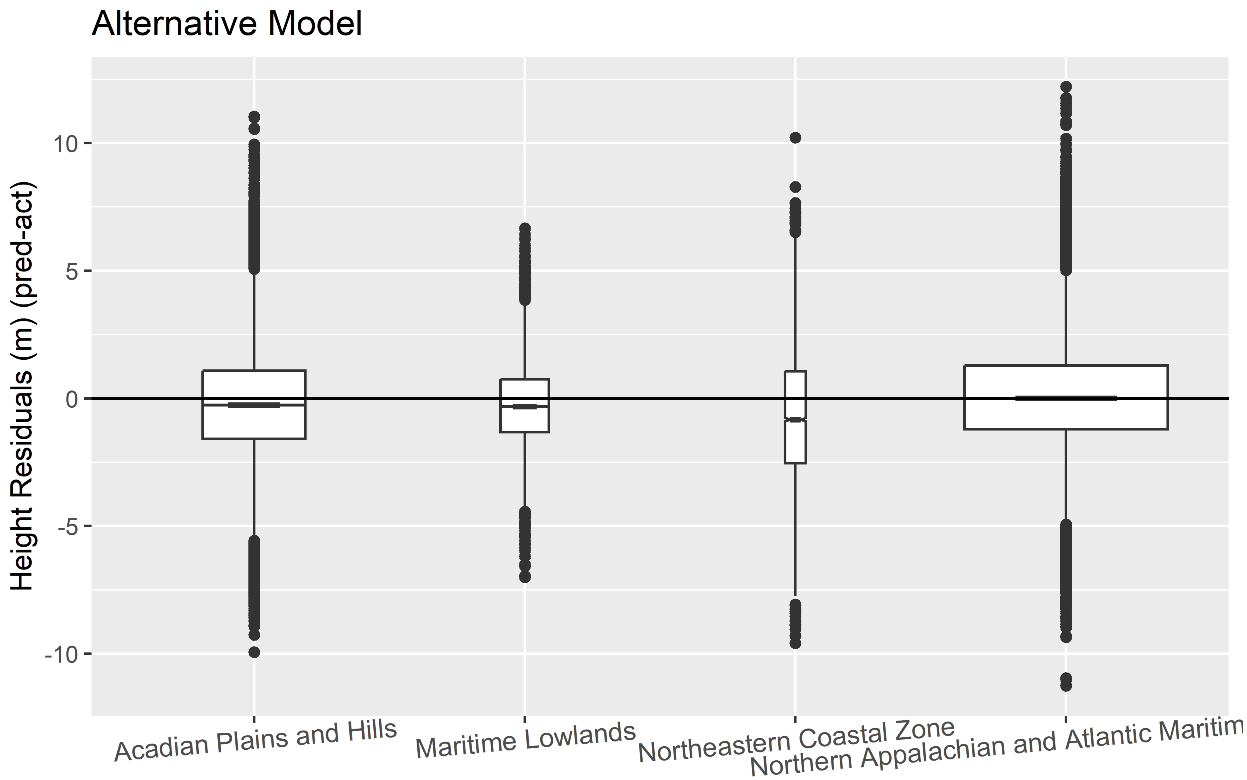
Table 2: Height Prediction Equation Parameter Estimates

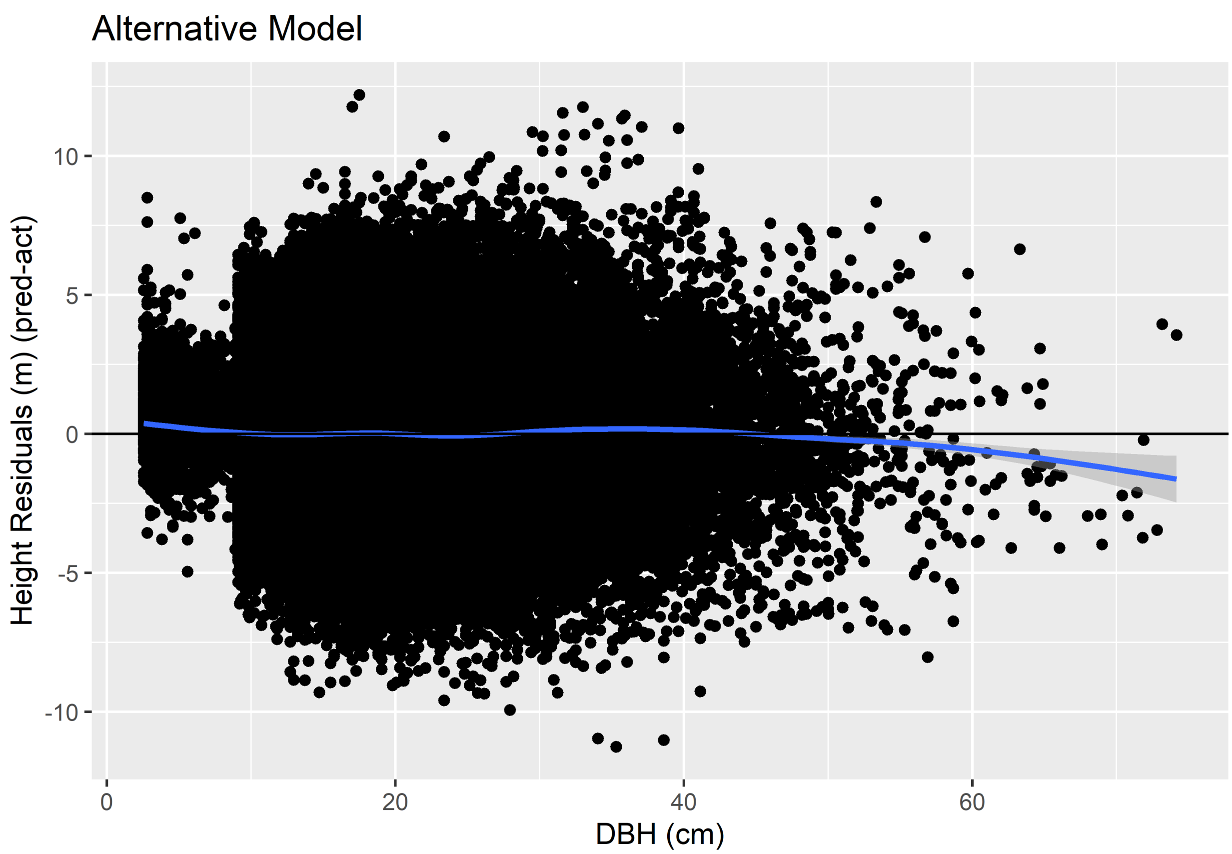
| FVS Sp | N Observations | MSE | b0 | b0a | b1 | b2 | b3 | b4 |
|---|---|---|---|---|---|---|---|---|
| BF | 192357 | 3.704206 | 5.131034 | -0.6880061 | 0.0435567 | 1.172035 | 0.0029563 | 0.7568912 |
| RS | 139907 | 3.288333 | 13.357887 | -0.6170510 | 0.0425798 | 1.455918 | 0.0050827 | 0.3313788 |
| RM | 126834 | 4.166832 | 5.737082 | -0.6022688 | 0.0552172 | 1.177891 | 0.0081403 | 0.6930164 |
| BS | 52292 | 2.905632 | 15.686720 | -2.5150059 | 0.0879415 | 2.137218 | 0.0105556 | 0.0948864 |
| WS | 43596 | 3.249584 | 9.299591 | -0.6976246 | 0.0334513 | 1.421107 | 0.0039226 | 0.6273081 |
| PB | 40746 | 4.250563 | 6.662807 | -0.2094344 | 0.0570106 | 1.282445 | 0.0070540 | 0.6049400 |
| SM | 29398 | 4.409973 | 14.242720 | -0.9400849 | 0.0537606 | 1.244795 | 0.0098851 | 0.2227569 |
| YB | 29163 | 4.799653 | 7.990101 | -0.9084566 | 0.0544340 | 1.079054 | 0.0060822 | 0.4570736 |
| WP | 21536 | 5.443705 | 13.376763 | -1.0314136 | 0.0386871 | 1.625443 | 0.0084621 | 0.3922846 |
| EH | 20403 | 5.249113 | 18.891738 | -0.9638330 | 0.0325306 | 1.411625 | 0.0045582 | 0.1664579 |
| WC | 17124 | 2.627591 | 13.518415 | 0.0000000 | 0.0562739 | 1.569955 | 0.0034078 | 0.1094524 |
| AB | 15336 | 5.371063 | 21.848180 | -4.7113294 | 0.0297698 | 1.035262 | 0.0031868 | 0.0905116 |
| QA | 11045 | 2.805391 | 13.824510 | -1.6174458 | 0.0482408 | 1.051316 | 0.0057278 | 0.3156300 |
| TA | 11039 | 4.241455 | 10.119560 | -1.7548756 | 0.0655958 | 1.788895 | 0.0126045 | 0.4773238 |
| RO | 9278 | 4.294340 | 7.689509 | -1.6548746 | 0.0581872 | 1.216263 | 0.0046263 | 0.5733108 |
| BT | 5494 | 3.848021 | 11.254140 | -1.6698431 | 0.0232875 | 0.805161 | 0.0032737 | 0.5691219 |
| WA | 5119 | 5.163908 | 14.565228 | -1.7494327 | 0.0444603 | 1.167474 | 0.0079207 | 0.3090990 |
| GB | 3715 | 2.212172 | 9.013788 | -0.3509257 | 0.0648630 | 1.191640 | 0.0065971 | 0.3960406 |
| OC | 498254 | 4.287191 | 15.703106 | -0.9006955 | 0.0298852 | 1.176950 | 0.0027325 | 0.2956746 |
| OH | 276128 | 5.056554 | 8.228732 | -0.9239308 | 0.0514263 | 1.086426 | 0.0063728 | 0.5046676 |

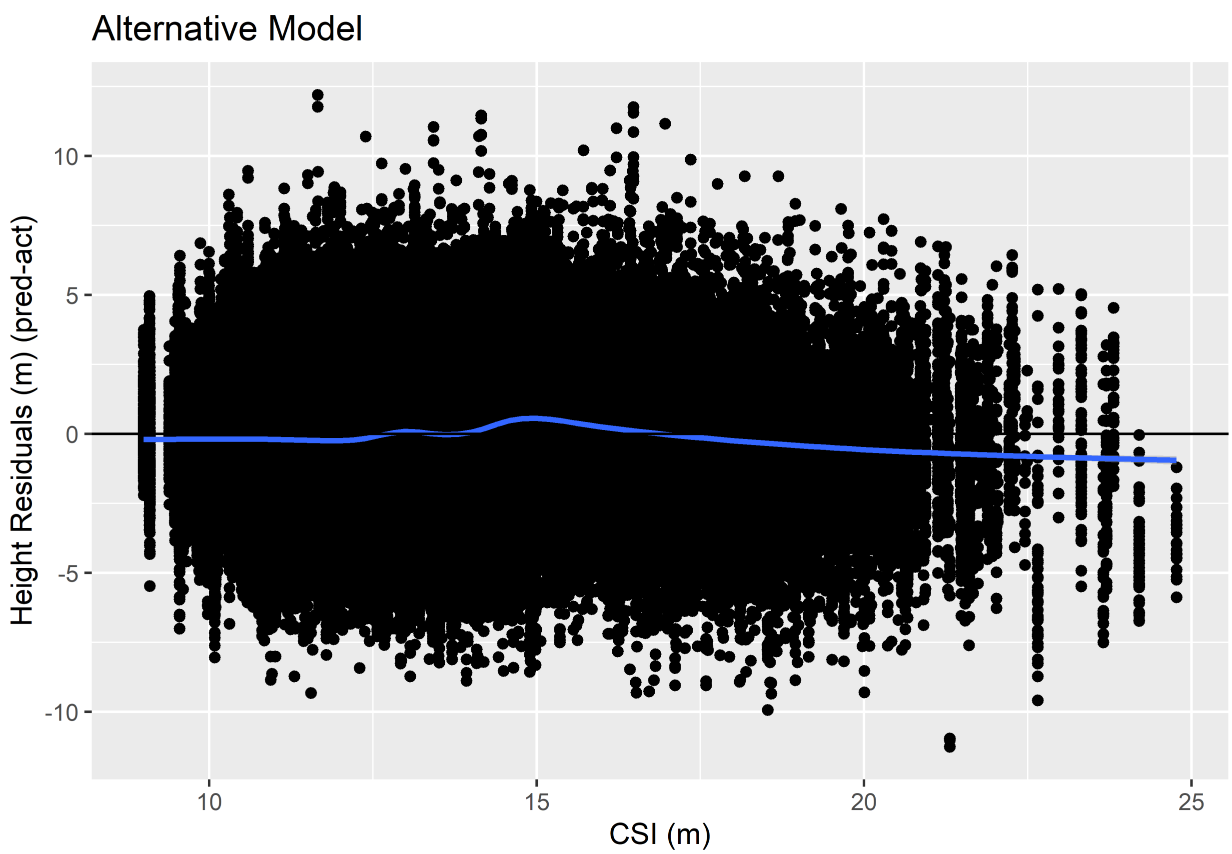
The `Alternative` corrects the large tree bias in `dbh` and the residuals trend in `csi` (interestingly without having `csi` in the model). The over-prediction bias for trees with high competition is attenuated but not eliminated.

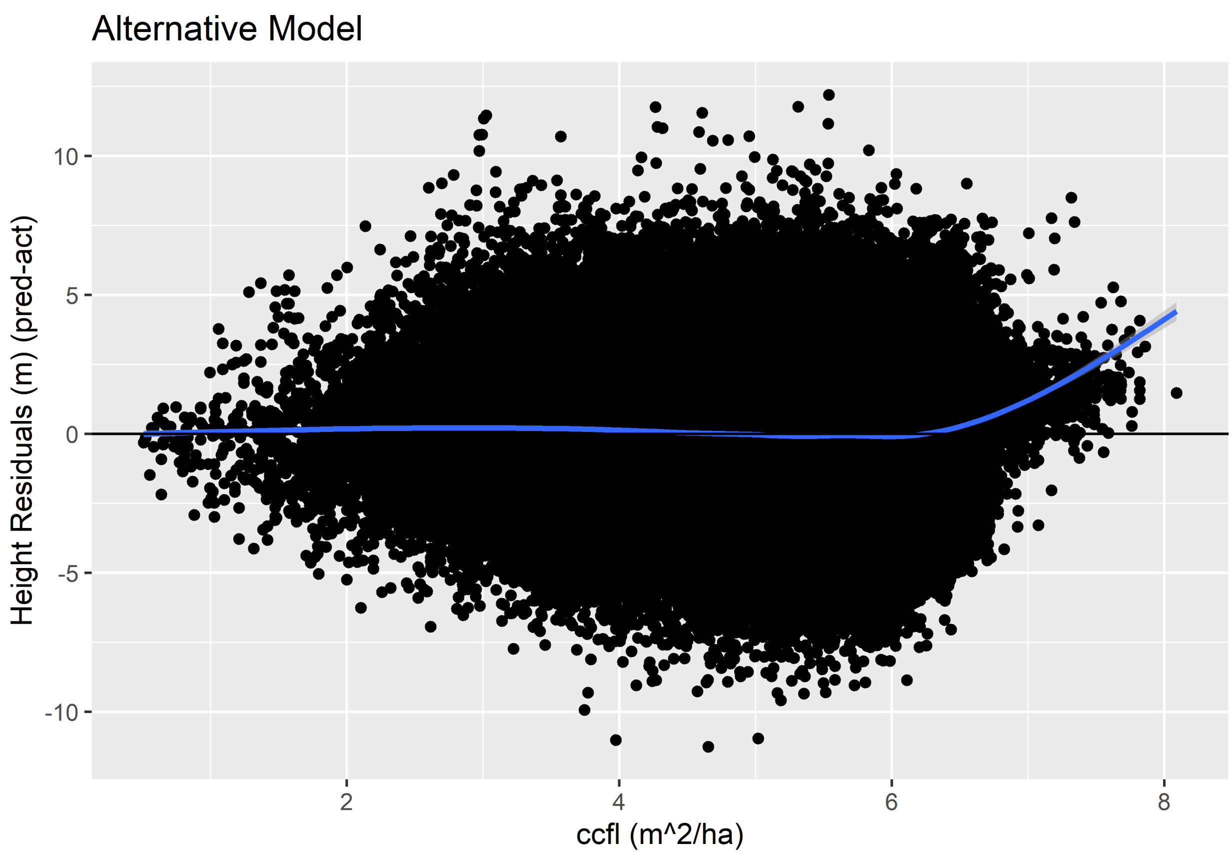

Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
AB
BF
BS
BT
EH
GB
PB
QA
RM
RO
RS
SM
TA
WA
WC
WP
WS
YB
FVS Species

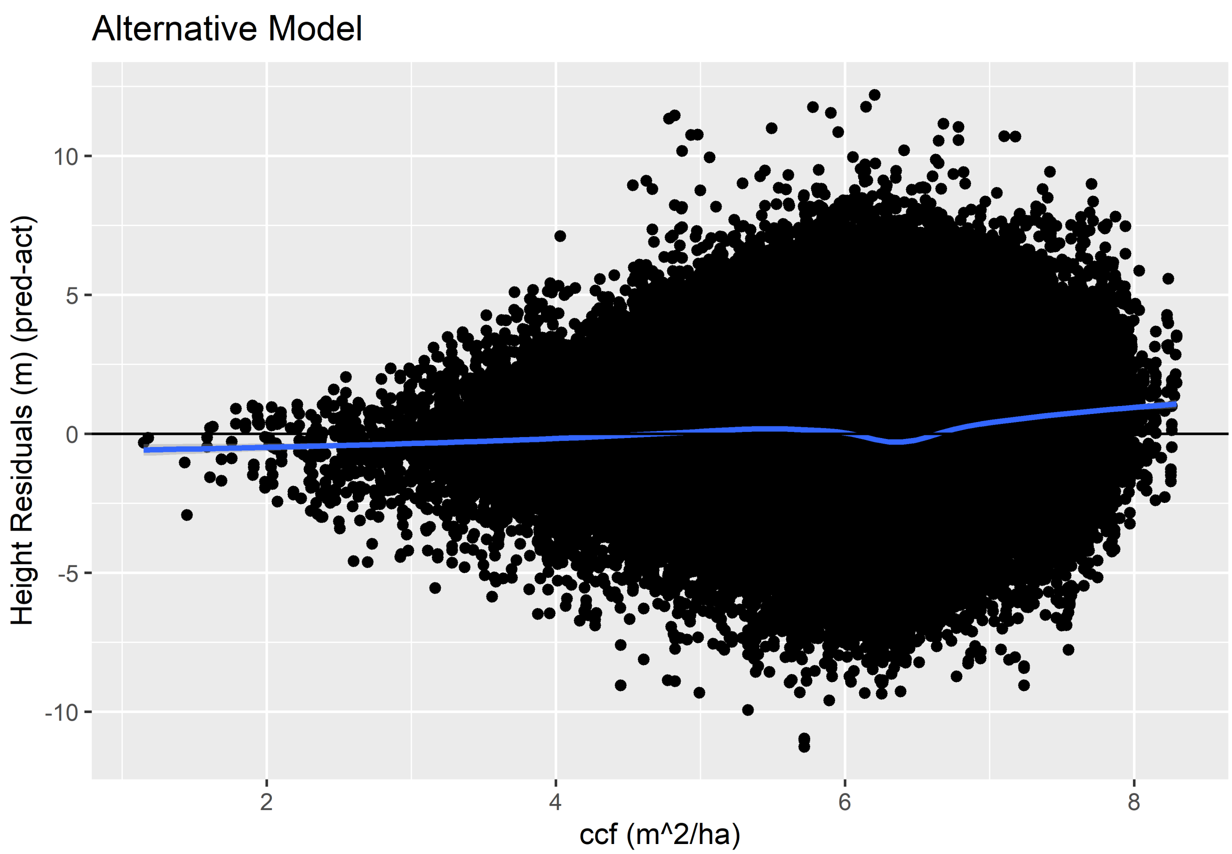

Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
ERM ME
FIA ME
Nova Scotia PSP
Data Source

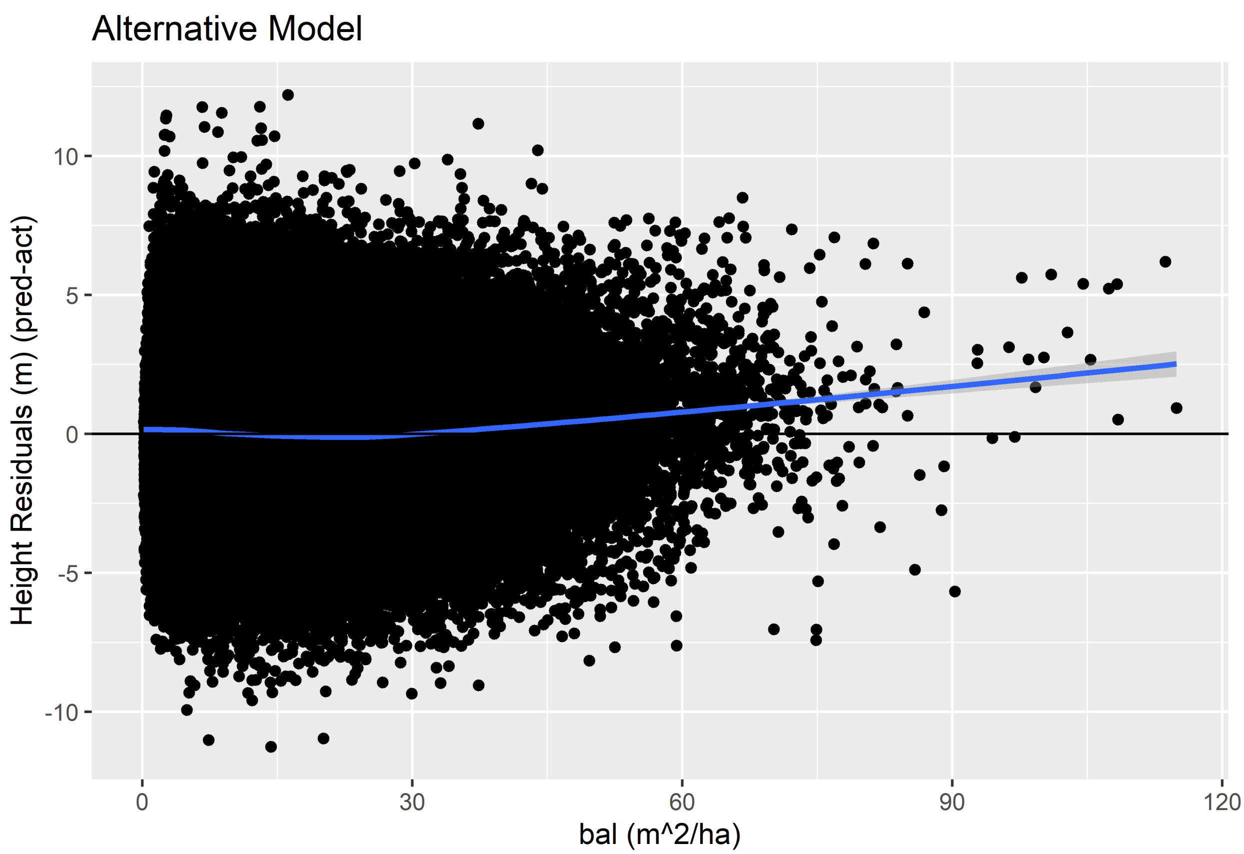

Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
Acadian Plains and Hills
Maritime Lowlands
Northeastern Coastal Zone
Northern Appalachian and Atlantic Maritim


Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
0
20
40
60
DBH (cm)


Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
10
15
20
25
CSI (m)


Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
2
4
6
8
ccfl (m^2/ha)

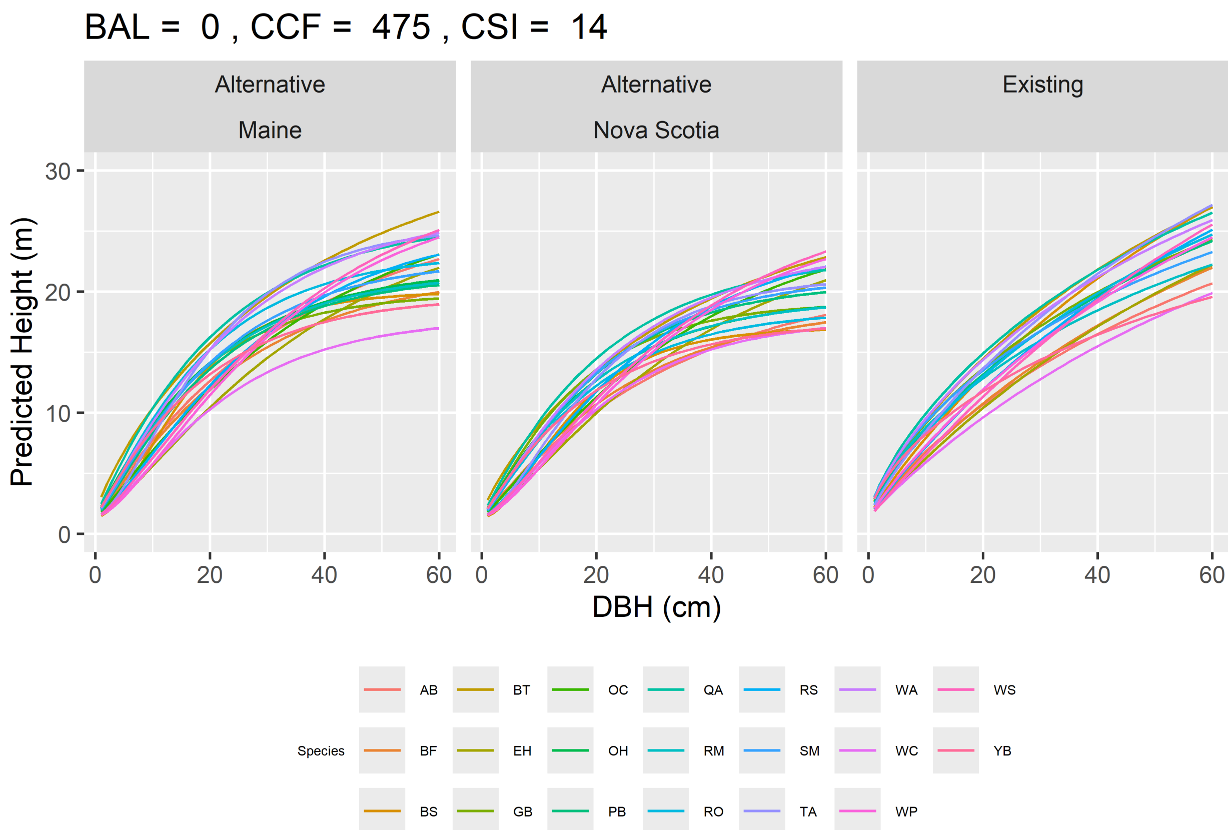

Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
2
4
6
8
ccf (m^2/ha)

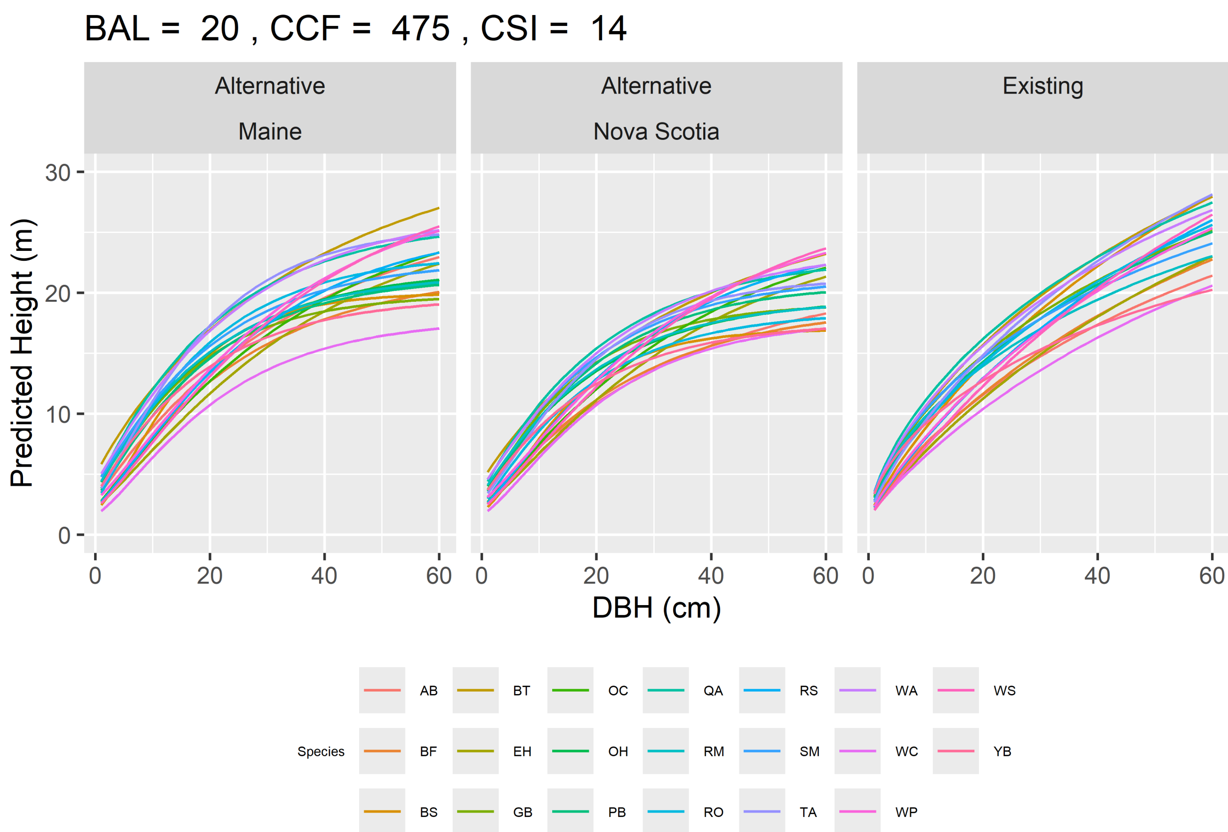


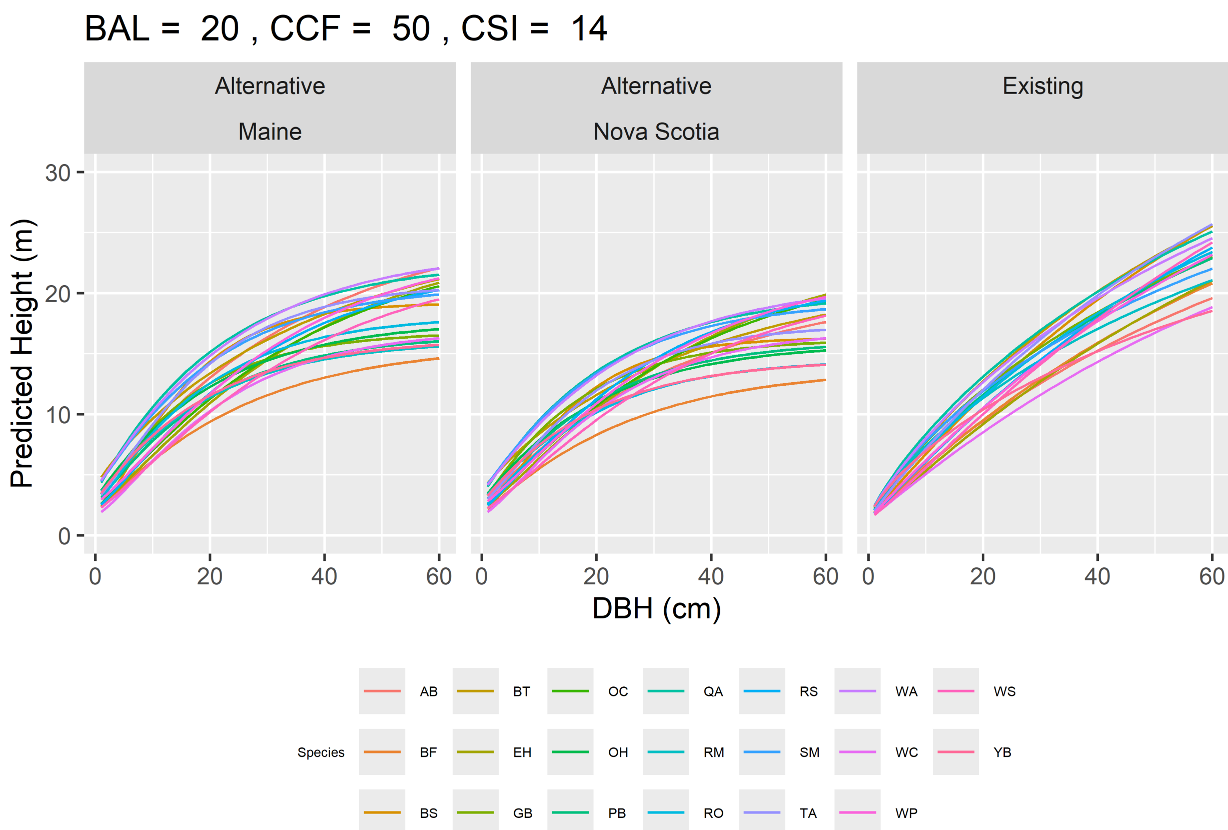
Alternative Model
Height Residuals (m) (pred-act)
10
5
0
-5
-10
0
30
60
90
120
bal (m^2/ha)

## Model Performance Comparison

Below we compare `Existing` to `Alternative` models for trees of differing competitive status. The `Alternative` produces shorter asymptotic heights over `dbh` than the `Existing`, which does not appear to have an upper bound on height. Nova Scotia (as a surrogate for the `ACD` New Brunswick region) attains shorter heights for a given `dbh` than the Maine region. Both the `Alternative` and `Existing` exhibit similar sensitivites to `bal` and `ccf`.




BAL = 20 , CCF = 475 , CSI = 14
Alternative
Maine
Alternative
Nova Scotia
Existing
Predicted Height (m)
30
20
10
0
0
20
40
60
DBH (cm)
Species
AB
BT
OC
QA
RS
WA
WS
BF
EH
OH
RM
SM
WC
YB
BS
GB
PB
RO
TA
WP


BAL = 20 , CCF = 50 , CSI = 14
Alternative
Maine
Alternative
Nova Scotia
Existing
Predicted Height (m)
30
20
10
0
0
20
40
60
DBH (cm)
Species
AB
BT
OC
QA
RS
WA
WS
BF
EH
OH
RM
SM
WC
YB
BS
GB
PB
RO
TA
WP