Toward Developing a Regional IFC Forest Productivity Model

Christopher Chase (Presenter)
Nicholas Crookston

Summary

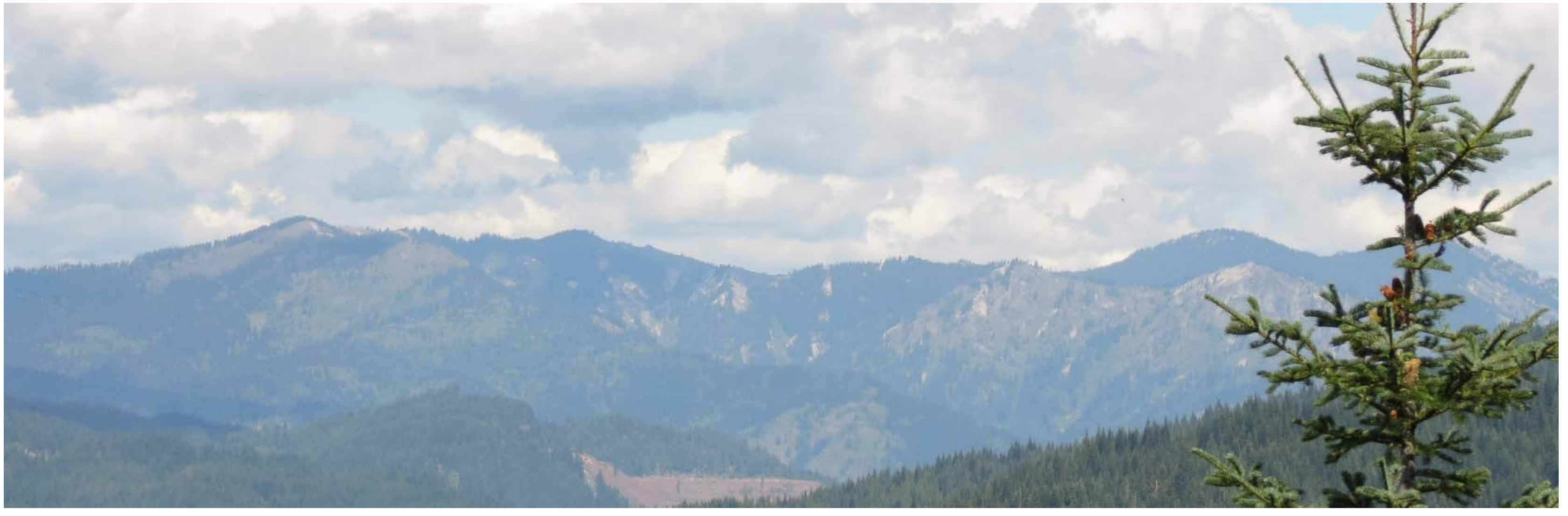
- Background
- Individual tree analysis
  - RF
  - Linear models
- Whole stand productivity
- Stand subset productivity
- Challenges
- Conclusions / future directions

Background

- What is Site Productivity?

- Productivity can be directly measured
  - Volume growth/acre/year

- Site Productivity is a little more tricky . . .
  - Volume growth/acre/year the site is capable of producing
    - Difficult to directly measure



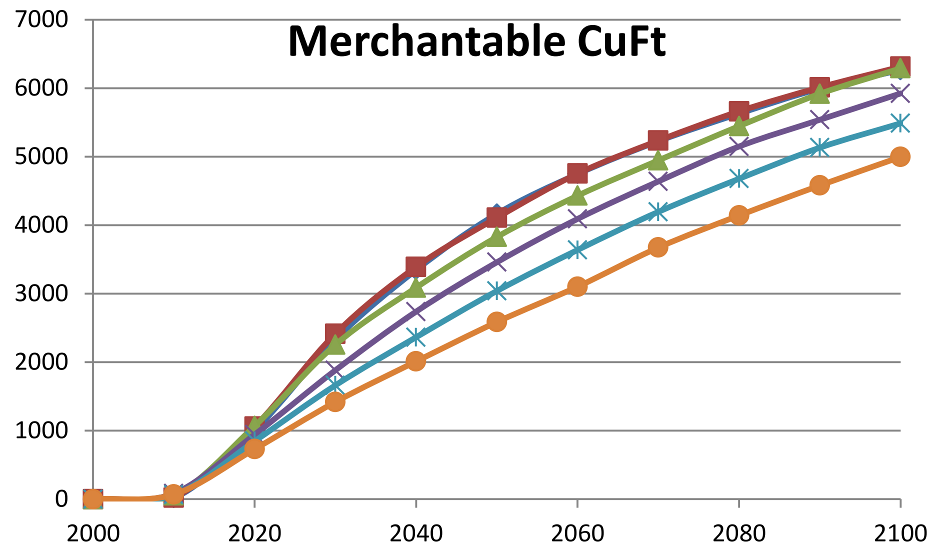
Background

- Proxies are often used for site productivity
  - Site index
  - Habitat type
- Site productivity is very useful
  - Growth and yield models
  - Land appraisals
  - Management strategies

Background

Growth and Yield Models

- Many use individual tree models
  - Diameter growth
  - Volume growth
  - FVS



- Whole stand models
  - Less common
  - Ultimate goal of G & Y models

The Dataset

- Repeated growth measurements (CFI)
- Primarily Northern ID
- Variety of owners
    - USFS
    - IDL
    - Potlatch
    - Etc.

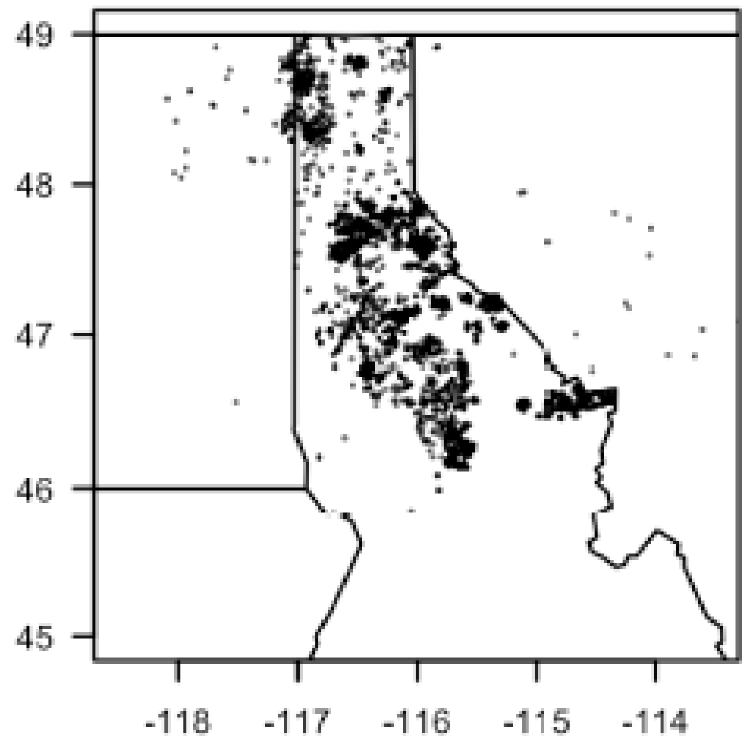


Figure 1: Sample plot locations

The Dataset

- 4589 Unique Stands / Plots
- Over 1,000,000 Individual tree observations
- Initially . . .
- Much is lost after data screening

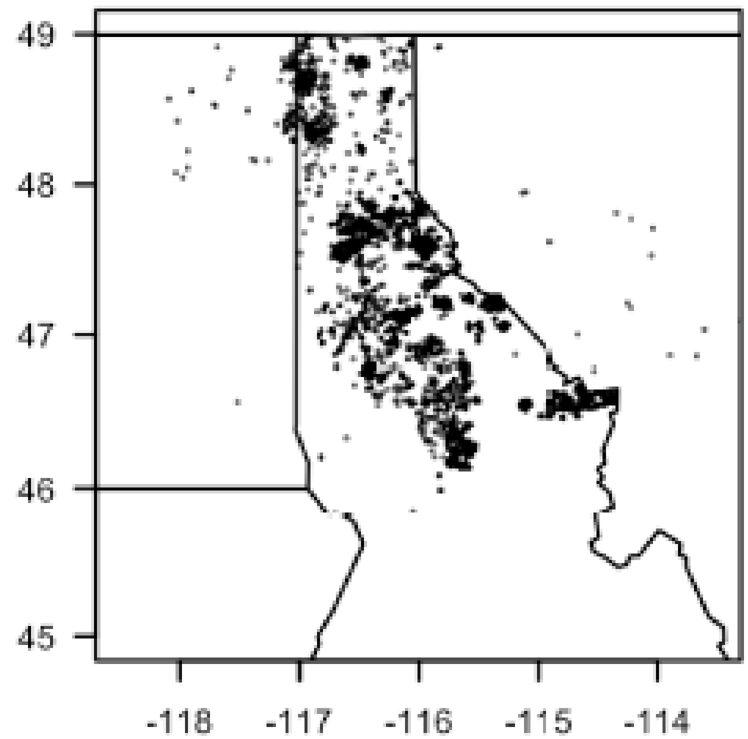


Figure 1: Sample plot locations

Phase 1: Individual Tree

- Response Variable
  - lnDDS
    - "…Logarithm of the change in squared inside-bark individual tree diameter over a period of 10 years…"
  - Used in FVS

- 110,852 individual tree observations

- Model selection methods
  - Random Forests (RF, Breiman 2001)
  - Best R^2

Random Forests Models

- Builds multiple regression trees
- Rank explanatory variables on influence / importance
- Examines:
  - Increase in Mean Squared Error
  - Node purity
    - Decrease in residual sum of squares
    - Averaged over all the trees

Random Forests Models

Four models

- rf1: all predictors were included
- rf2: just site predictors
- rf3: rf2 plus Species and cr.
- rf4: traditional variables used in FVS

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf1	0.84	DBH, cr, Species, soilcd, TopHt, baldbh, QMD, pratio, BAPctile, Slope
rf2	0.39	slca, slsa, soilcd, pratio, ffp, smrsprpb, map, Geology, gsp, winp
rf3	0.54	Species, cr, pratio, slca, slsa, winp, soilcd, map, adi, Geology
rf4	0.83	cr, Species, DBH, TopHt, elev, slca, baldbh, BAPctile, QMD, PtBAL

RF 1

- rf1: all predictors were included

Variable	Description
DBH	Diameter breast height
Species	Tree species
cr	Crown ratio
soilcd	Soil class names
pratio	Ratio of summer precipitation to total precipitation
TopHt	Height of the largest 40 trees per acre in the stand
baldbh	Interaction between DBH and Ratio of basal area in trees larger than subject tree (bal)
QMD	Quadratic mean diameter
BAPctile	Percentile point in the distribution of tree basal areas
Slope	Slope percentage

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf1	0.84	DBH, cr, Species, soilcd, TopHt, baldbh, QMD, pratio, BAPctile, Slope

Phase 1: Individual Tree

Explanatory Variables

- Species

Table 1: Species codes, names, and observation counts

FVSCode	Common Name	Name	Frequency
AF	subalpine fir	*Abies lasiocarpa*	3085
DF	Douglas-fir	*Pseudotsuga menziesii*	24769
ES	Engelmann spruce	*Picea engelmannii*	3383
GF	grand fir	*Abies grandis*	23314
LP	lodgepole pine	*Pinus contorta*	5520
MH	mountain hemlock	*Tsuga mertensiana*	1986
PP	ponderosa pine	*Pinus ponderosa*	2477
RC	western redcedar	*Thuja plicata*	13465
WH	western hemlock	*Tsuga heterophylla*	8686
WL	western larch	*Larix occidentalis*	7481
WP	western white pine	*Pinus monticola*	16701

Phase 1: Individual Tree

Soil class names (soilcd)

Table 2: Soil class names (soilcd) and observation counts

Soil class name	Frequency
Basalt	155
Glacial	1427
Loess	261
LoessBasalt	803
LoessMetaGranite	1001
Metased	426
Other	331
Seds	1180
VolcanicBasalt	166
VolcanicGlacial	7901
VolcanicLoess	2696
VolcanicLoessBasalt	2814
VolcanicLoessGlacial	8156
VolcanicLoessMetaGranite	3540
VolcanicLoessMetased	523
VolcanicLoessOther	239
VolcanicLoessQuartzite	1209
VolcanicMetaGranite	16208
VolcanicMetased	41667
VolcanicOther	408
VolcanicQuartzite	4919
VolcanicSeds	14822

RF2

- rf2: just site predictors

Variable	Description
slca	Slope * cos(pi/180*Aspect)
slsa	Slope * sin(pi/180*Aspect)
soilcd	Soil class names
pratio	Ratio of summer precip to total precip
ffp	Frost free period
smrsprpb	Summer spring precipitation balance
winp	Winter precipitation
map	Mean annual precipitation
Geology	Parent material
gsp	Growing season precipitation

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf2	0.39	slca, slsa, soilcd, pratio, ffp, smrsprpb, map, Geology, gsp, winp

RF 3

- rf3: rf2 plus Species and cr.

Variable	Description
slca	Slope * cos(pi/180*Aspect)
Species	Tree species
cr	Crown ratio
slsa	Slope * sin(pi/180*Aspect)
soilcd	Soil class names
pratio	Ratio of summer precip to total precip
winp	Winter precipitation
map	Mean annual precipitation
Geology	Parent material
adi	Annual dryness index

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf3	0.54	Species, cr, pratio, slca, slsa, winp, soilcd, map, adi, Geology

RF 3

- Geology

Table 3: Geology variable frequencies

Geology	Frequency
CaMetased	24464
Extrusive	4895
Glacial	10497
Intrusive	18375
Metasedimentary	44045
Other	8467
Sedimentary	109

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf3	0.54	Species, cr, pratio, slca, slsa, winp, soilcd, map, adi, Geology

RF 4

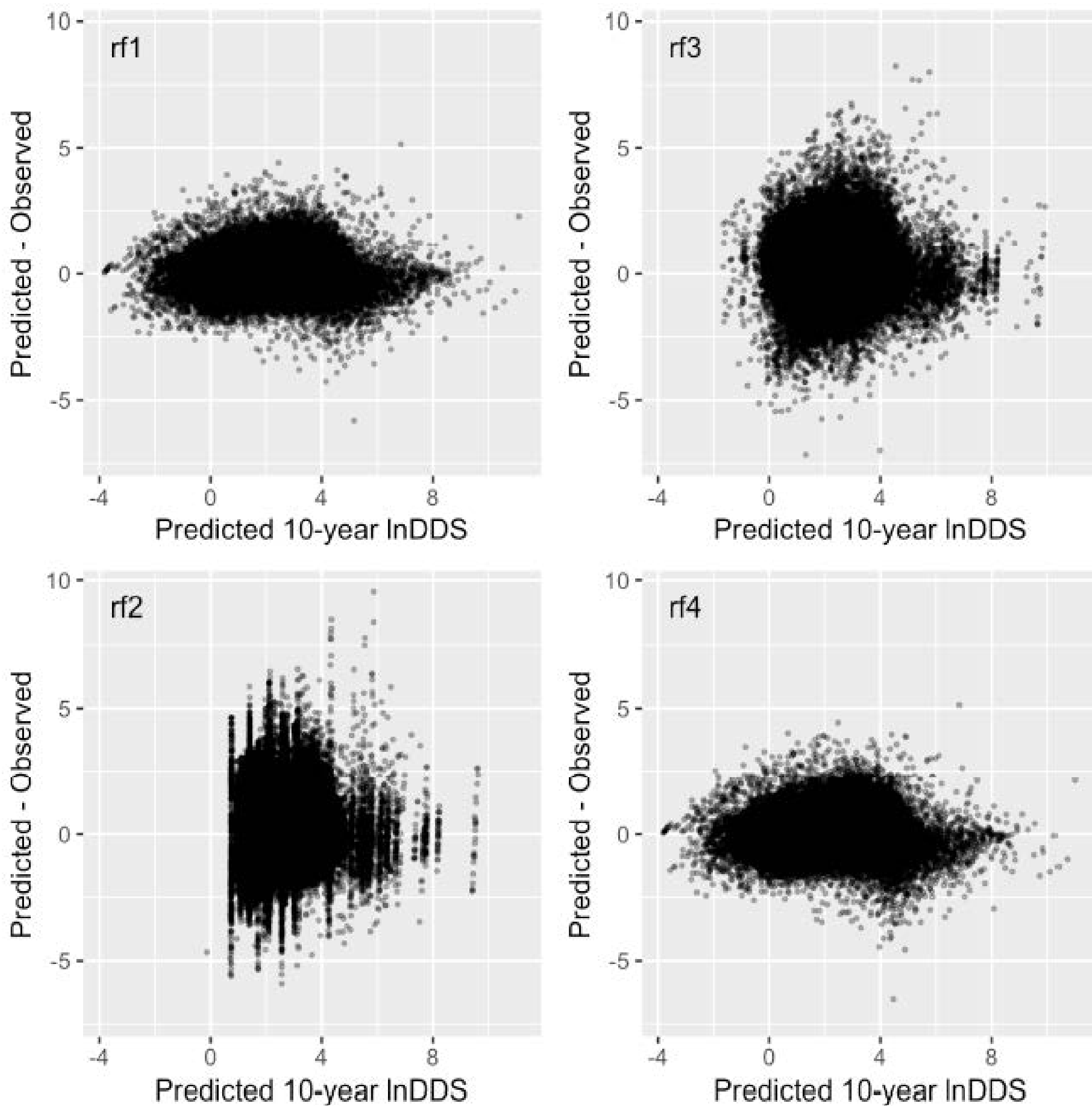
rf4: traditional variables used in FVS

Variable	Description
DBH	Diameter breast height
slca	Slope * cos(pi/180*Aspect)
Species	Tree species
cr	Crown ratio
TopHt	Height of the largest 40 trees per acre in the stand
elev	Elevation
baldbh	Interaction between DBH and Ratio of basal area in trees larger than subject tree (bal)
QMD	Quadratic mean diameter
BAPctile	Percentile point in the distribution of tree basal areas
PtBAL	Basal area per acre in larger trees measured on the subplot same as BAL if there is one plot

Table 9: Summary of the RF models

RF model	R2	Best 10 predictors
rf4	0.83	cr, Species, DBH, TopHt, elev, slca, baldbh, BAPctile, QMD, PtBAL

RF Model Summary



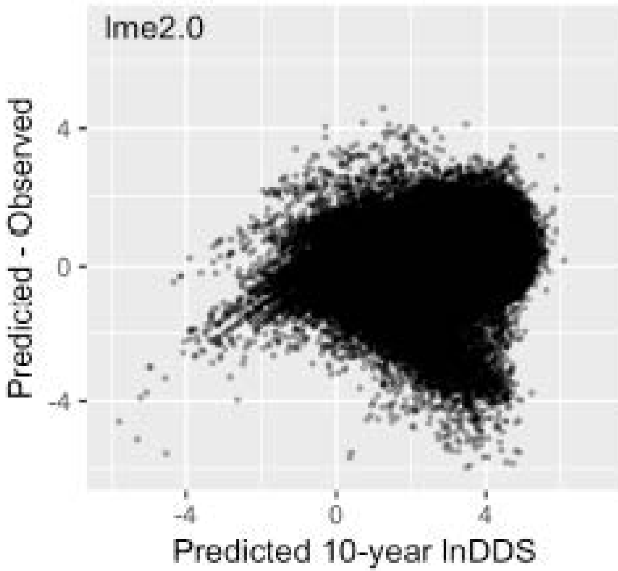
Phase 2: Individual Tree

- Linear and mixed effects models
- Selected based on RF models
- Best model R^2
- 2 best models presented here

Table 11: Two sequences of adding variables to the lme models[R2 is reported without random effect

R2	Predictors
	Sequence 1, adding tree metrics
0.65	`Species + soilcd + Species:cr + slca + slsa + pratio + map + Species:adi + I(log(DBH)) + I(DBH*DBH) + bal + Species:baldbh`
	Sequence 2, adding site metrics
0.64	`Species + Species:cr + I(log(DBH)) + I(DBH*DBH) + bal + Species:baldbh + soilcd + slca + slsa + pratio + map`

Phase 2: Individual Tree



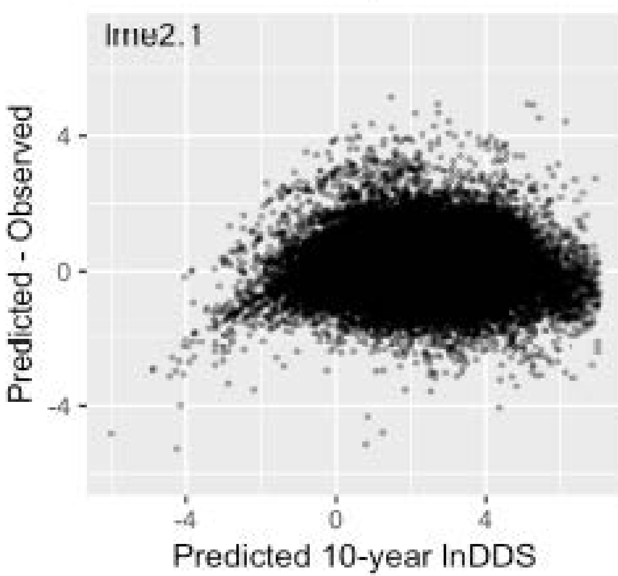


Table 10: Pseudo-R-square (R2) values for the linear models in table 8

Linear model	R2	R2 when predictions include random "StandID" effect
lm1	0.23	n/a
lme1	0.10	0.48
lm2	0.69	n/a
lme2	0.65	0.82
lme3	0.64	0.82
lme4	0.64	0.82

- All models tested here
- Random "StandID" effect explaining a lot of variation
  - Not explained by other variables

Individual Tree Conclusions

- Growth models can benefit from including soil and climate data
- Stand and tree characteristics often out-weight site characteristics
  - Site characteristics are indirectly incorporated in stand and tree measurements
- Quantifying inter-tree competition was difficult on large fixed area plots
  - “StandID” effect?
- Whole stand productivity could be explored. . .

Whole Stand Productivity

- Same data set
- Summarized and examined in different ways
- 4,308 stands initially
- 12% – 20% available after screening
    - Stands that were treated with fertilizer
    - Negative or 0 growth

Whole Stand Productivity

Data Summaries

- DBH, Height, Volume calculated for every tree
- Every measurement period
- Expanded to a per acre value
- Means for each stand to represent productivity
- Standardized to per year
  - Different measurement period lengths

Whole Stand Productivity

Model and Variable selection

- 3 parts
- First chose model with the best fit via AICC selection
- Removed insignificant variables ($\alpha < 0.1$)
- Tested interactions between significant variables

Whole Stand Productivity

- Whole stand volume growth (ft^3/acre/year)
- Including everything . . . $R^2 = 0.56$!

```
Response: mean.vol.growth.yr
                      Sum Sq  Df F value    Pr(>F)
(Intercept)              571   1  0.2718 0.6024906
ElevFt                 32129   1 15.2918 0.0001115 ***
cub.ft.acre            70483   1 33.5467 1.599e-08 ***
qmd                    38516   1 18.3319 2.425e-05 ***
logit.shade             8338   1  3.9684 0.0471717 *
ElevFt:DEM_SlopePct    21862   1 10.4052 0.0013802 **
DEM_SlopePct:tp.acre   27307   1 12.9969 0.0003593 ***
cub.ft.acre:SlopePct   36598   1 17.4190 3.823e-05 ***
cub.ft.acre:DF_SI      92848   1 44.1915 1.205e-10 ***
cub.ft.acre:tp.acre    19223   1  9.1494 0.0026797 **
qmd:Mat                12258   1  5.8341 0.0162520 *
qmd:DEM_SlopePct       30280   1 14.4120 0.0001742 ***
DEM_SlopePct:curt.rd   10578   1  5.0347 0.0254955 *
DF_SI:curt.rd          94144   1 44.8085 9.125e-11 ***
cub.ft.acre:curt.rd    46801   1 22.2754 3.472e-06 ***
qmd:curt.rd            33538   1 15.9626 7.942e-05 ***
DF_SI:sdi             104879   1 49.9178 9.293e-12 ***
cub.ft.acre:sdi        44641   1 21.2471 5.743e-06 ***
logit.shade:Mat        16244   1  7.7316 0.0057323 **
ElevFt:logit.shade      7880   1  3.7506 0.0536270 .
qmd:logit.shade        21468   1 10.2178 0.0015230 **
logit.shade:curt.rd    26262   1 12.4995 0.0004642 ***
logit.shade:sdi        25508   1 12.1405 0.0005587 ***
Residuals             705946 336
```

Whole Stand Productivity

- Got a little messy

- Just look at site factors . . . $R^2 = 0.09$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1000.98241  162.98553   6.142 1.45e-09 ***
dd5           -0.29277    0.05517  -5.306 1.55e-07 ***
smrsprpb    -407.96401   53.33885  -7.649 7.70e-14 ***
ElevFt        -0.07385    0.01726  -4.279 2.17e-05 ***
```

- Again, stand characteristics very important

- Predicting site productivity (ft^3/acre/year) is difficult
  - Particularly when only using site variables
    - Climate, topography, soils

- Other measures of productivity?

Other Measures of Site Productivity

Response variables

- Largest 10 trees per plot
- Largest 10 DF per plot
- Fastest 10 growing trees (height)

→

- Volume growth
- Height growth
- Diameter growth

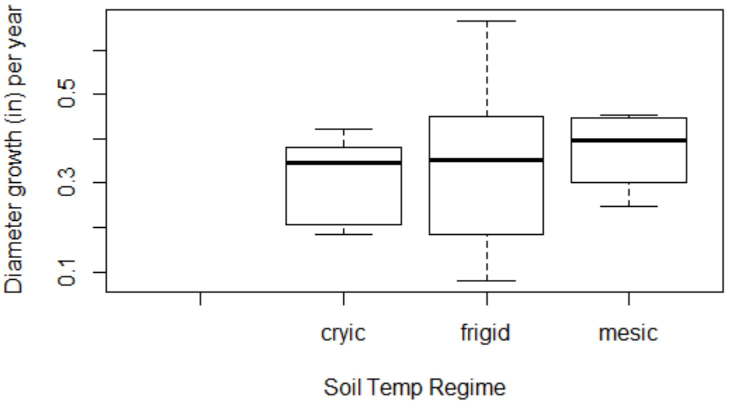
Include one stand explanatory variable

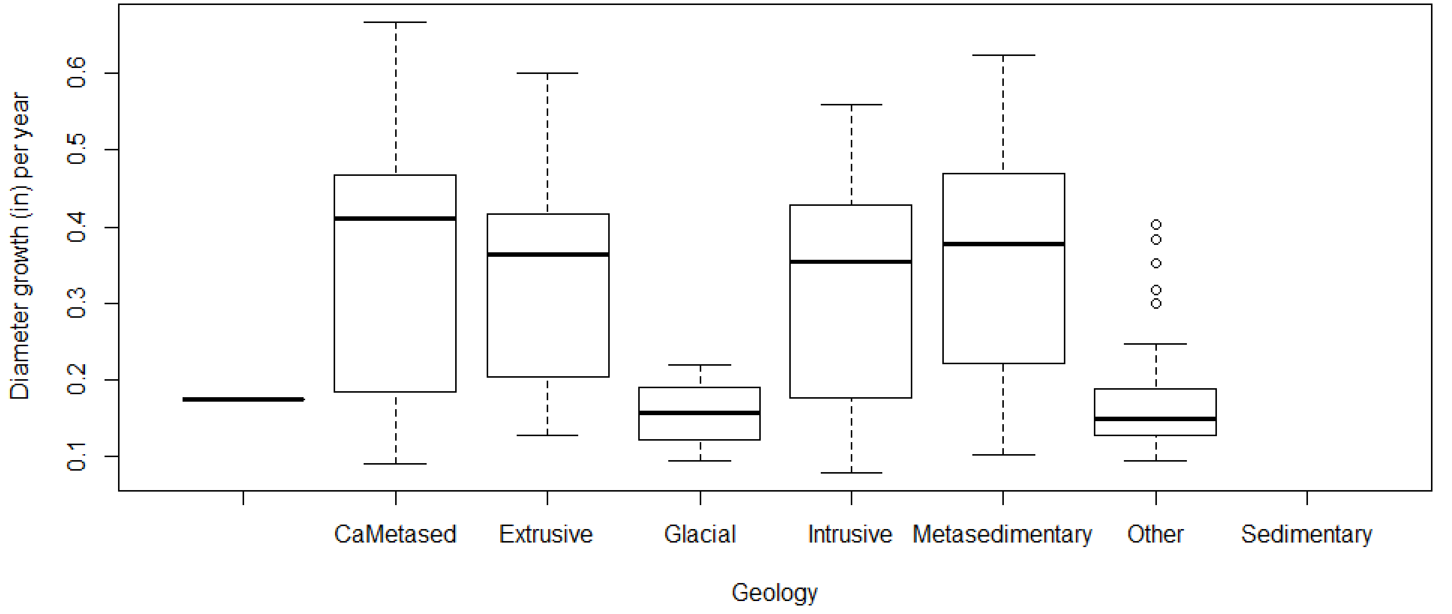
- QMD

Other Measures of Site Productivity

- 10 Largest trees per plot
  - Diameter growth $R^2 = 0.55$

```
Response: mean.dom.dg.yr
               Sum Sq  Df F value    Pr(>F)    
(Intercept)    0.1073   1 10.9170  0.001044 ** 
Geology        0.1572   6  2.6662  0.015153 *  
SoilTempRegime 0.0965   2  4.9085  0.007860 ** 
Mat            0.0964   1  9.8096  0.001871 ** 
DEM_ElevFt     0.7495   1 76.2785 < 2.2e-16 ***
qmd            0.1058   1 10.7631  0.001131 ** 
Mat:qmd        0.0376   1  3.8262  0.051190 .  
Residuals      3.7238 379                      
```

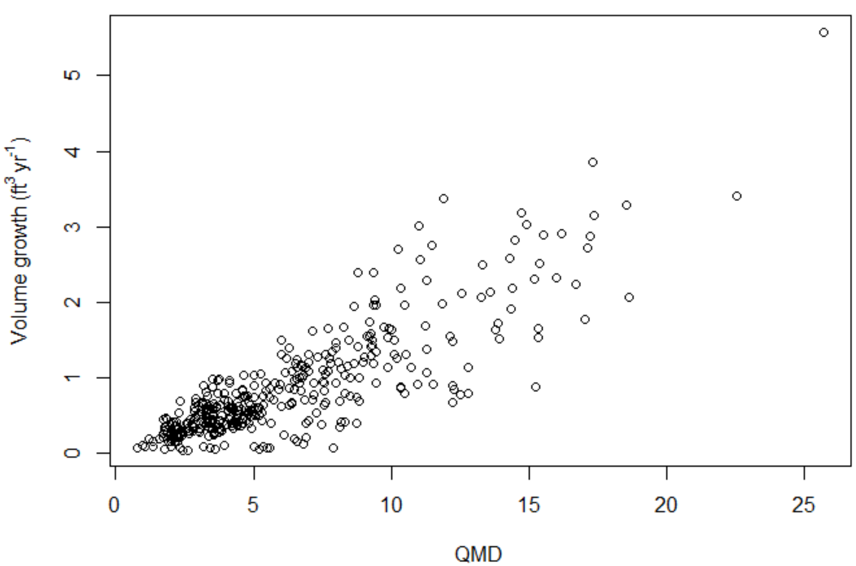


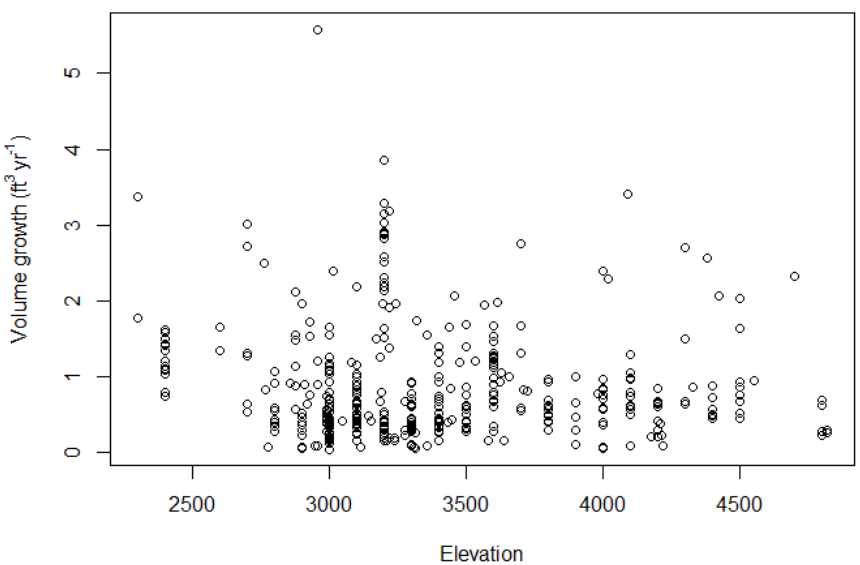


Other Measures of Site Productivity

- 10 Largest trees per plot
  - Volume growth $R^2 = 0.77$

```
Response: mean.dom.vg.yr
              Sum Sq  Df F value    Pr(>F)    
ElevFt         7.745   1  35.035 7.933e-09 ***
Mat            5.585   1  25.266 8.104e-07 ***
qmd          206.962   1 936.228 < 2.2e-16 ***
logit.shade    2.359   1  10.673  0.001199 ** 
Residuals     74.939 339                      
```

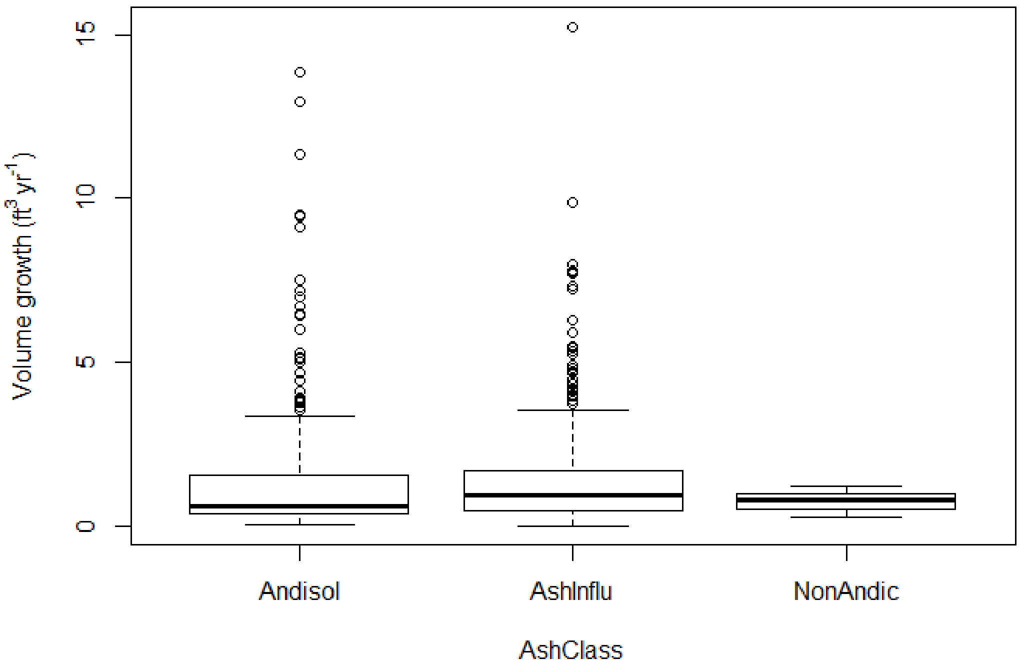




Other Measures of Site Productivity

- 10 fastest height growing trees
  - Volume growth $R^2 = 0.30$

```
Response: fast.mean.vg.yr
              Sum Sq  Df  F value    Pr(>F)    
(Intercept)    280.2   1  40.7850 3.042e-10 ***
qmd           2031.1   1 295.5997 < 2.2e-16 ***
qmd:AshClass   105.7   2   7.6893 0.0004962 ***
Residuals     4961.0 722
```



Conclusions

- Total stand site productivity is challenging to predict / model
  - Stand characteristics are important
- Looking at certain trees or classes in a stand could be representative of a site productivity
- Productivity of a class can be predicted fairly accurately with only a few site variables
  - Largest 10 trees
- Continue to look at volume growth as a measure of site productivity

Acknowledgements

- Mark Coleman, Mark Kimsey, Terry Shaw for guidance and funding
- Cooperators for providing the wealth of data
- Nick Crookston for the report and hard work

Questions?