Session 2 – Model Evaluation

Comparing Model Predictions With Data

Graphical first, then statistical

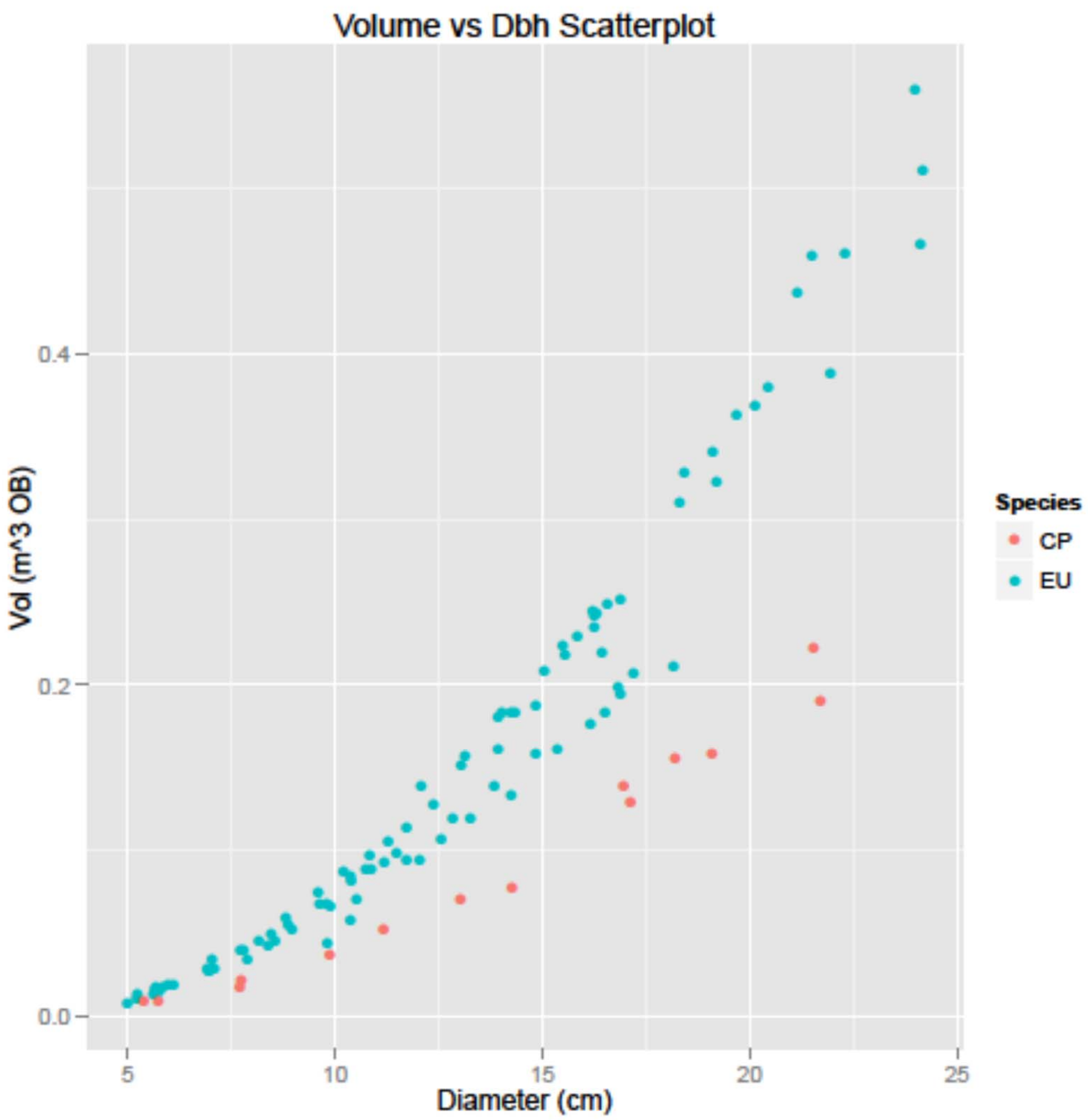
- **scatterplots can be quite revealing**
- **new data, old data, yield curves**

Re-measured plots

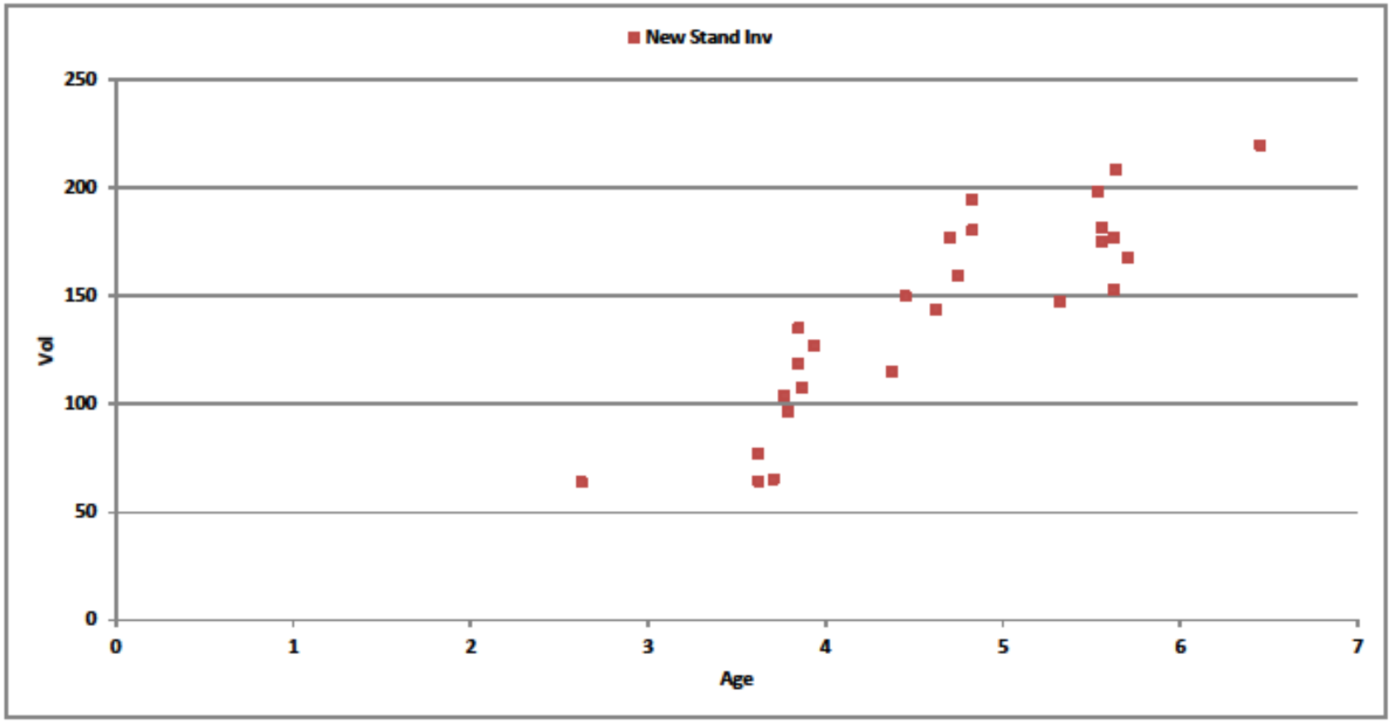
- **grow to cruise years option in FPS**
- **actual vs projected plots by species and component**

Statistics

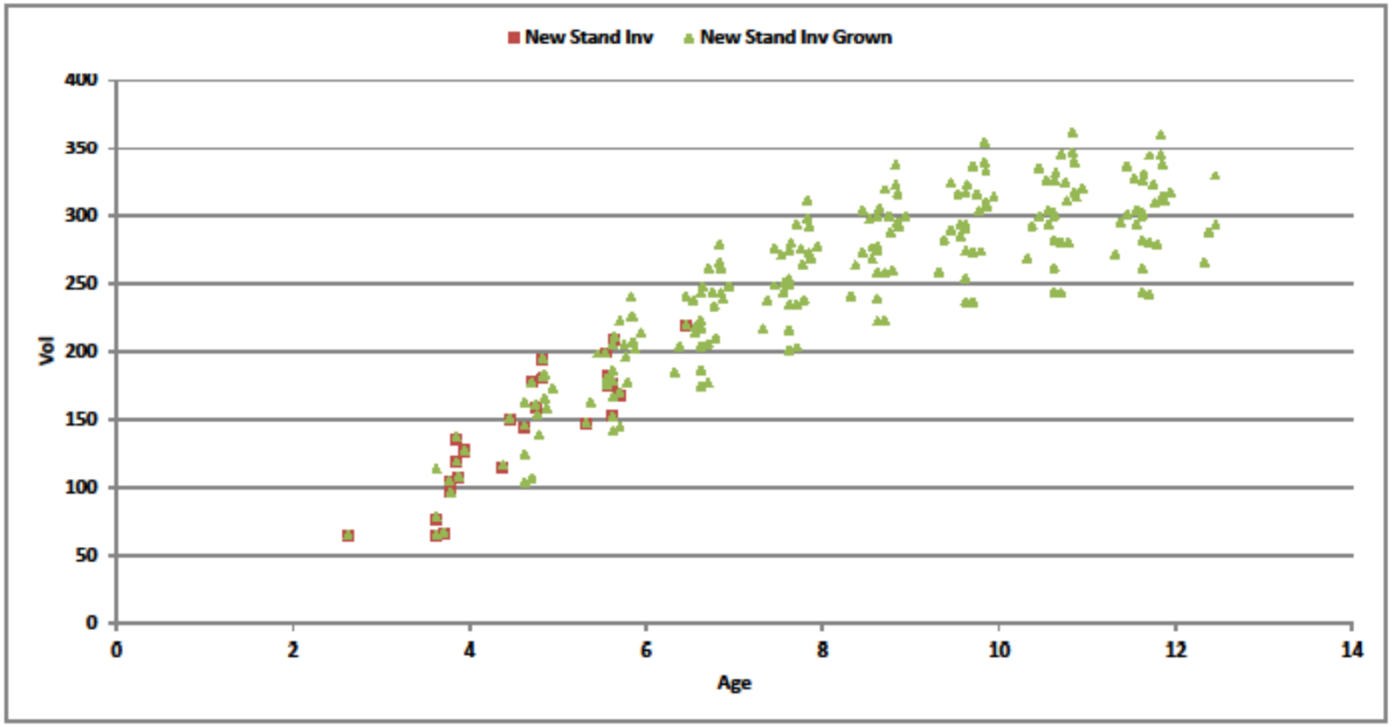
- **statistical significance? biological relevance?**
- **MAE & RMSE by species and component**
- **hypothesis tests by species and component**


Volume vs Dbh Scatterplot
Vol (m^3 OB)
0.0
0.2
0.4
5
10
15
20
25
Diameter (cm)
Species
CP
EU

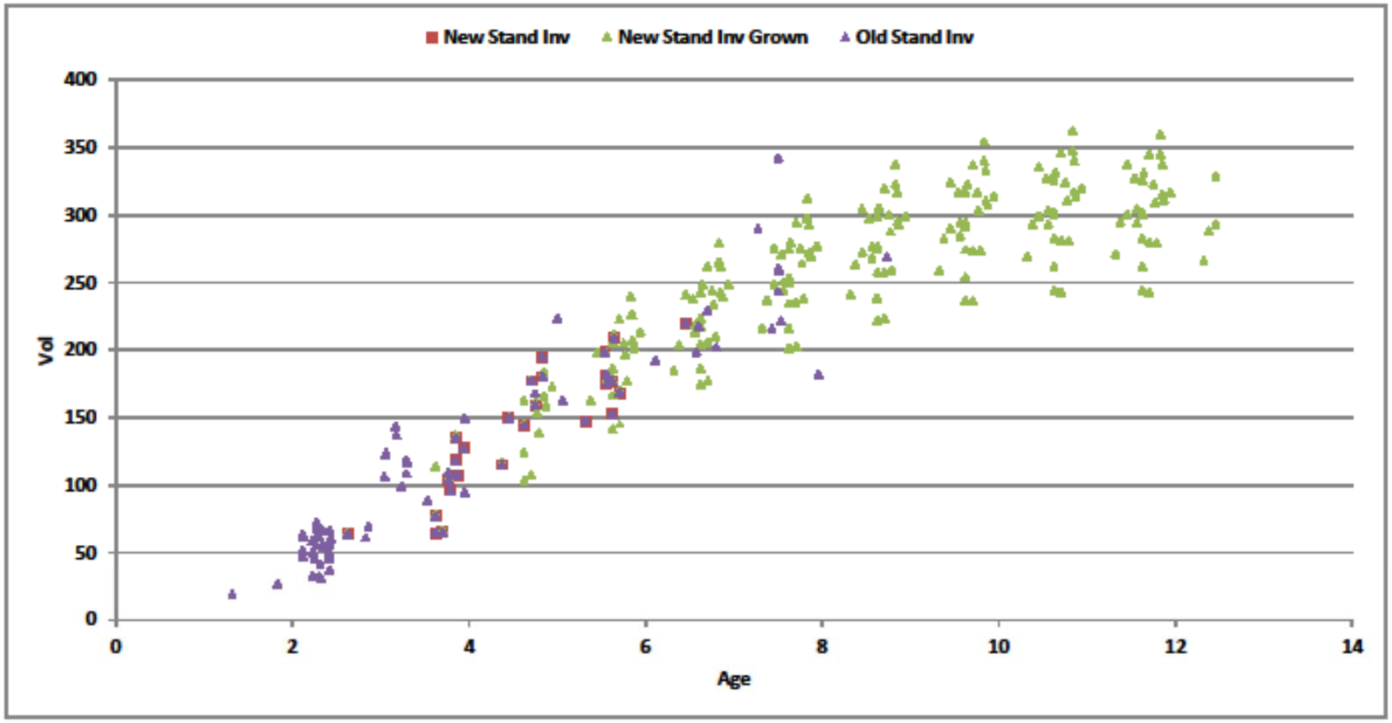
New Stand Inv Vol x Age



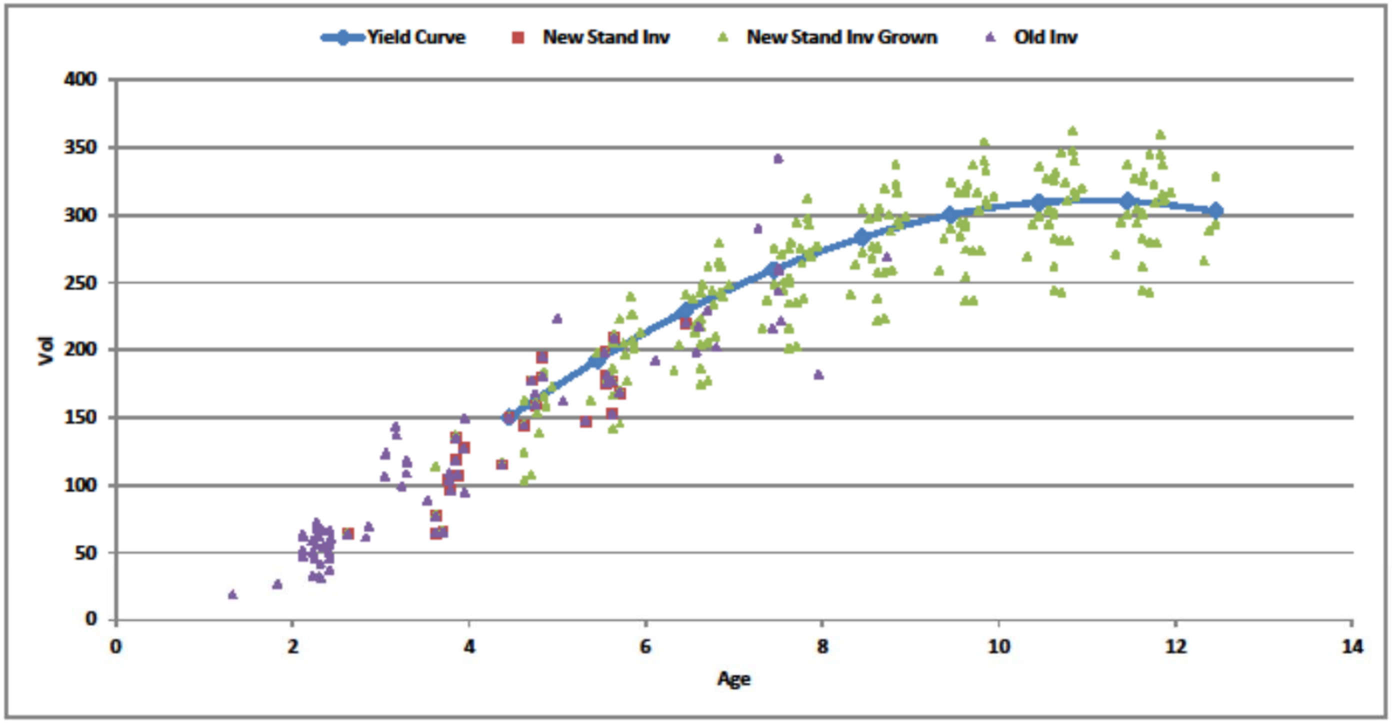
New Stand Inv & Grown Vol x Age



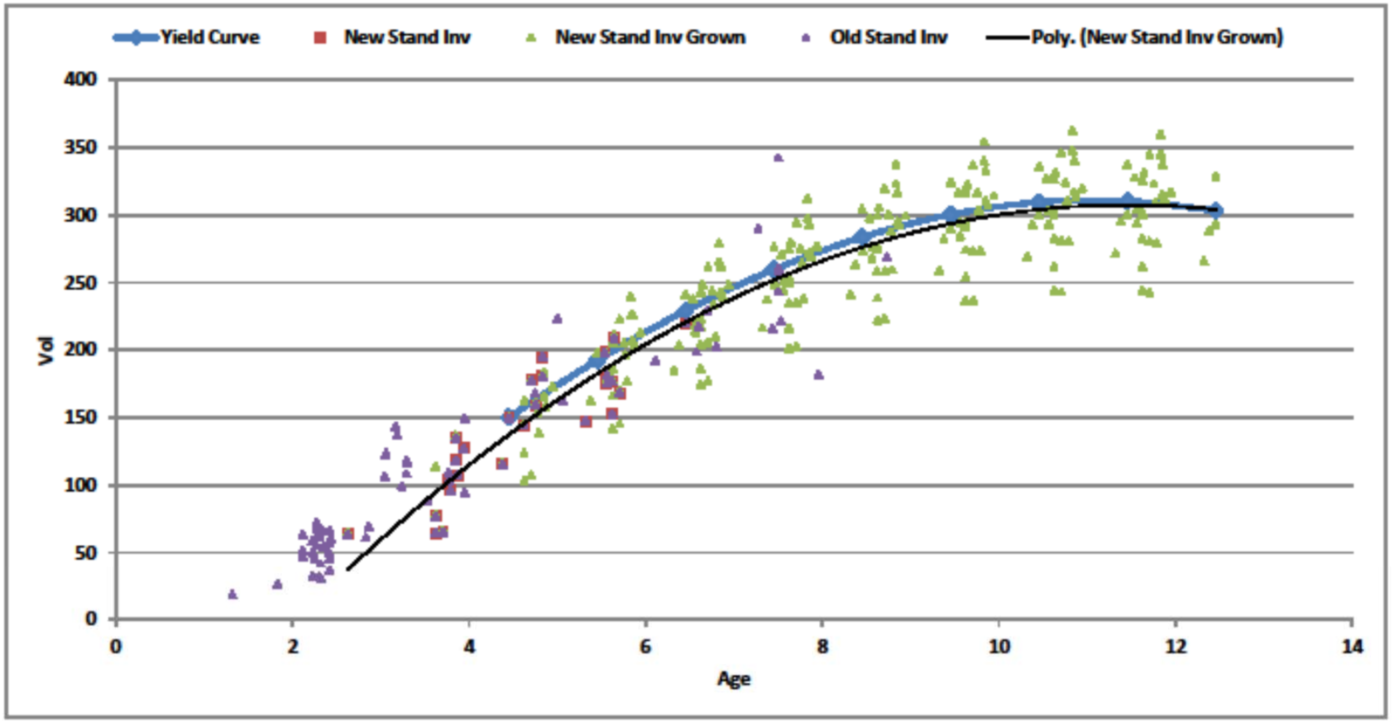
Old Inv Added



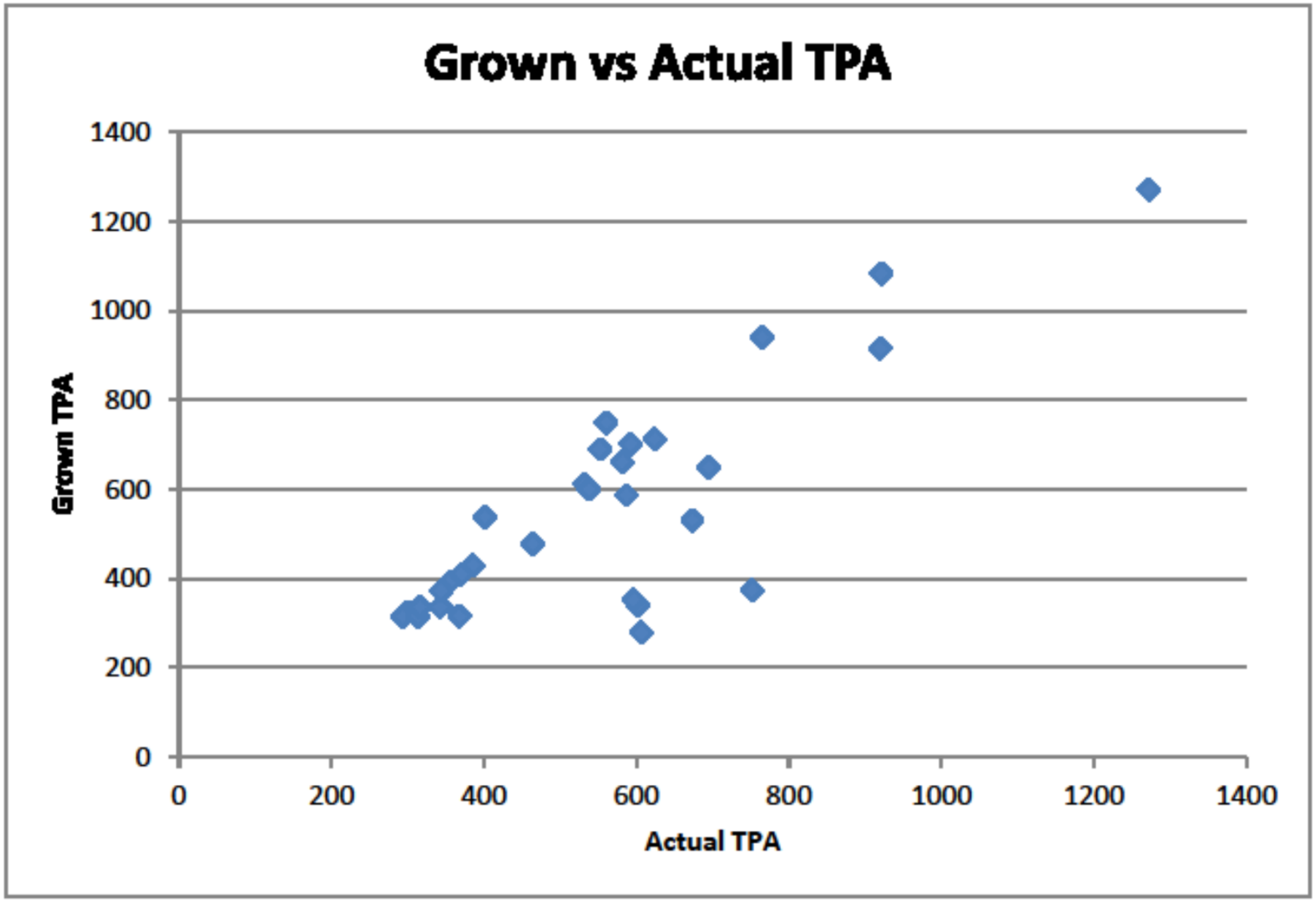
Yield Curve Added



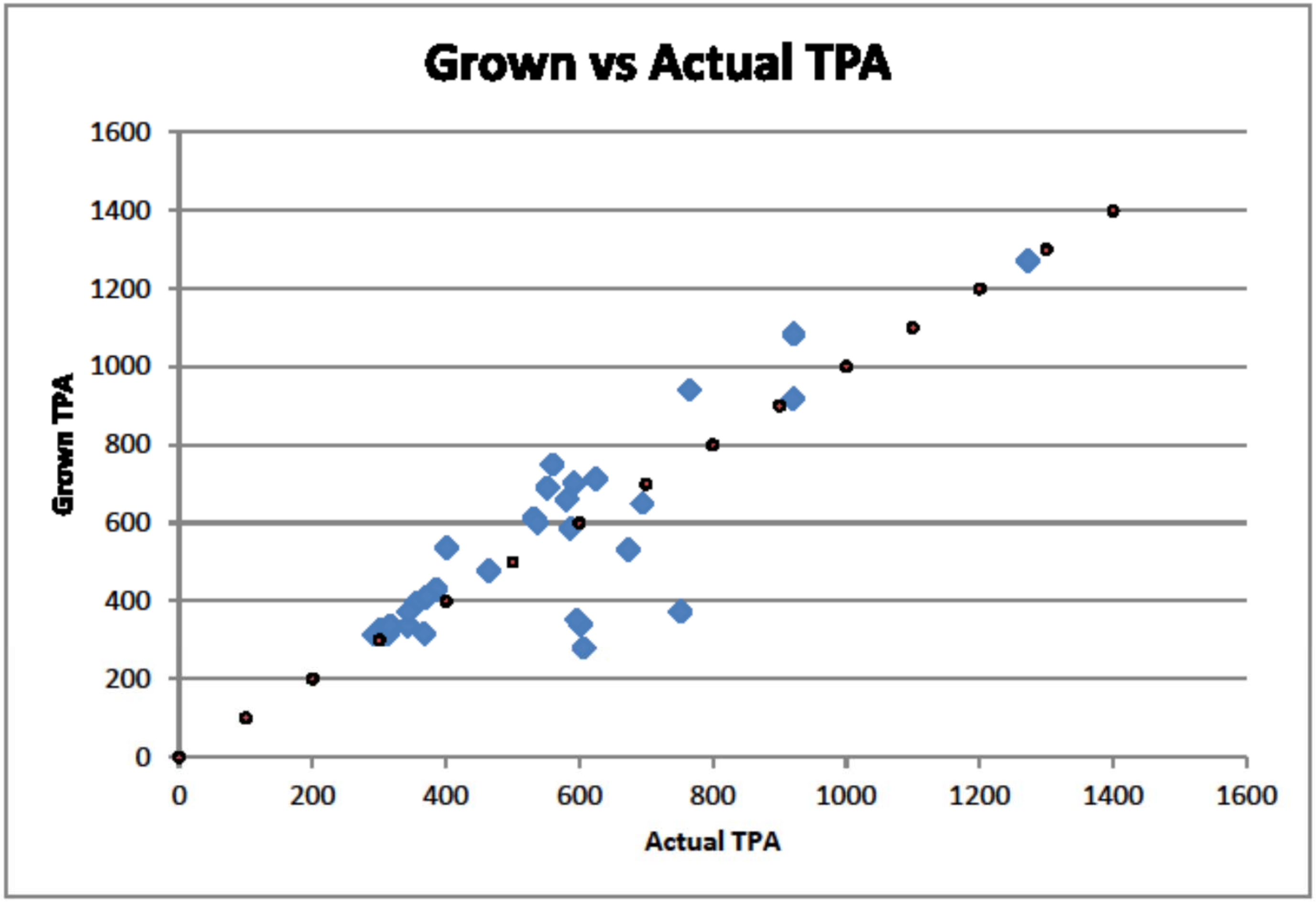
Trend Line of Grown Inv Added



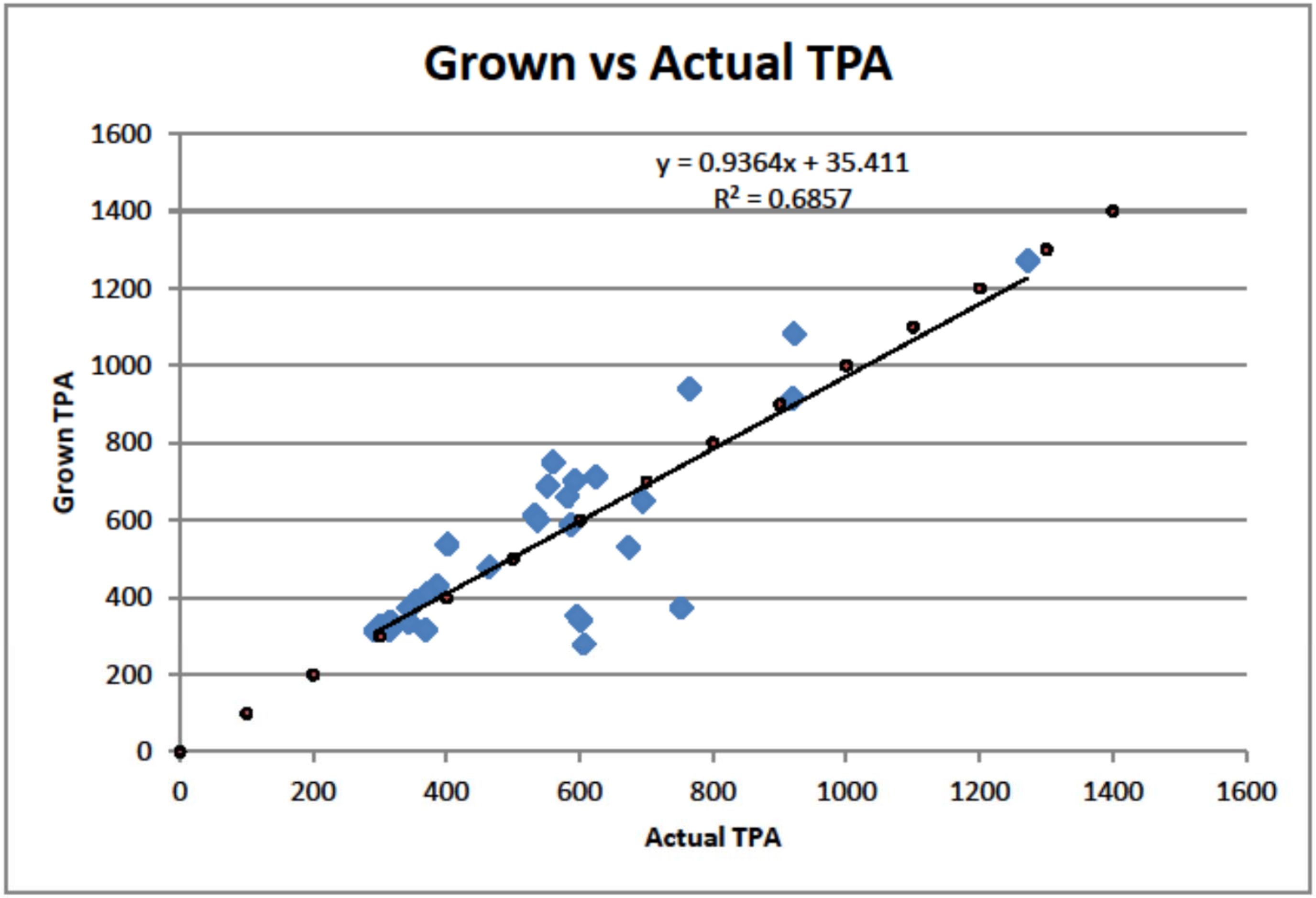
Proj v Act Series



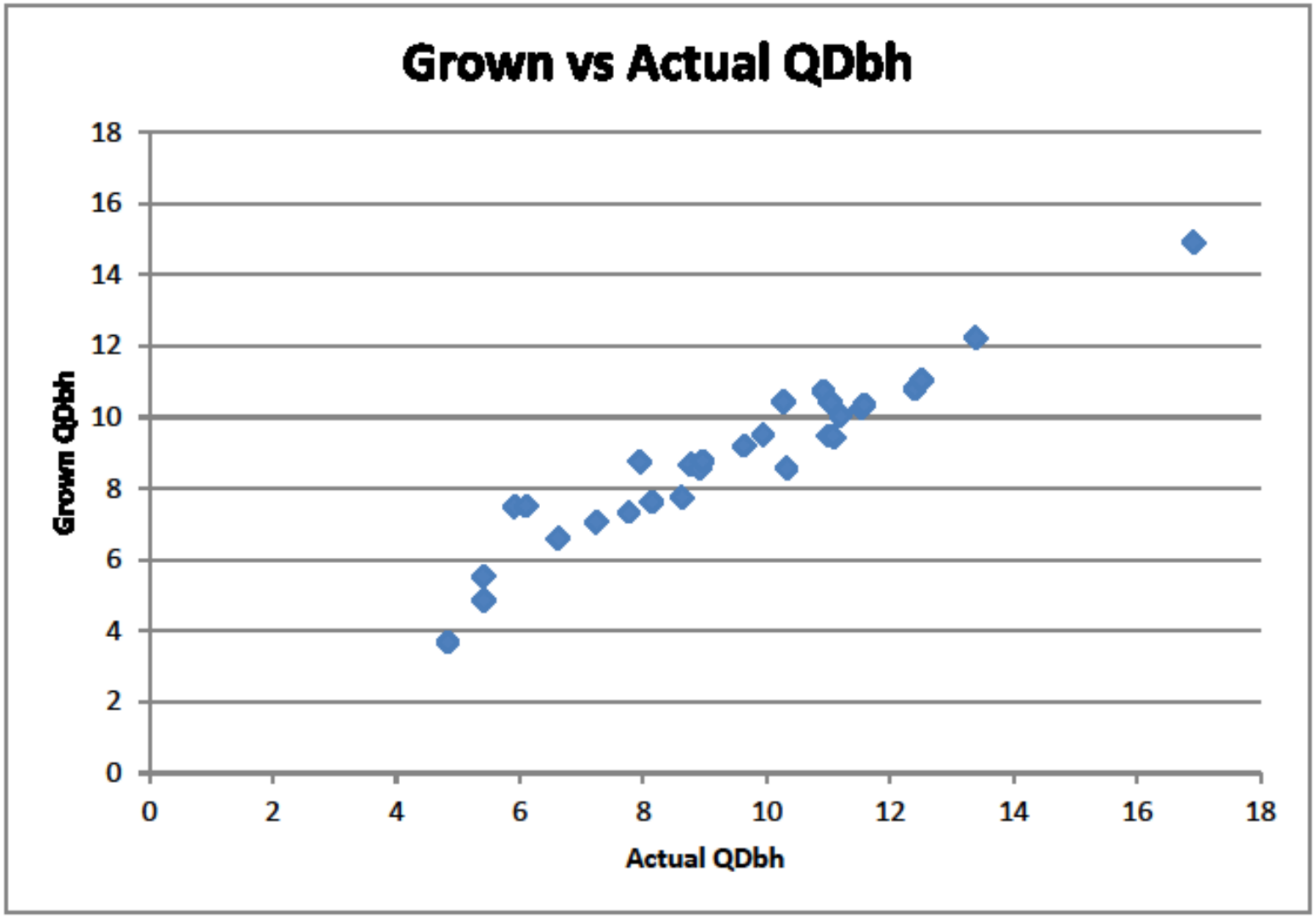
1:1 line added



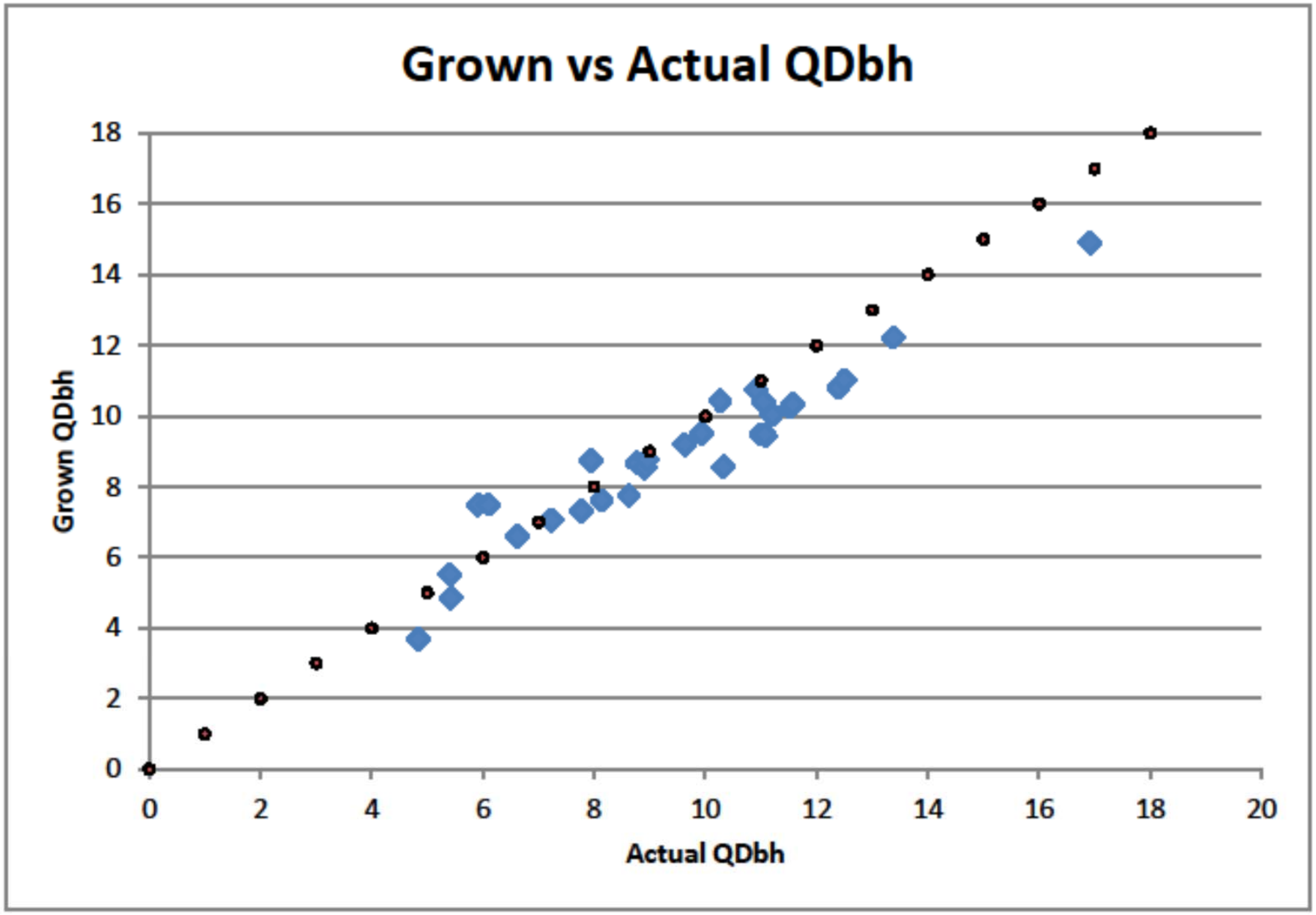
Trend line added



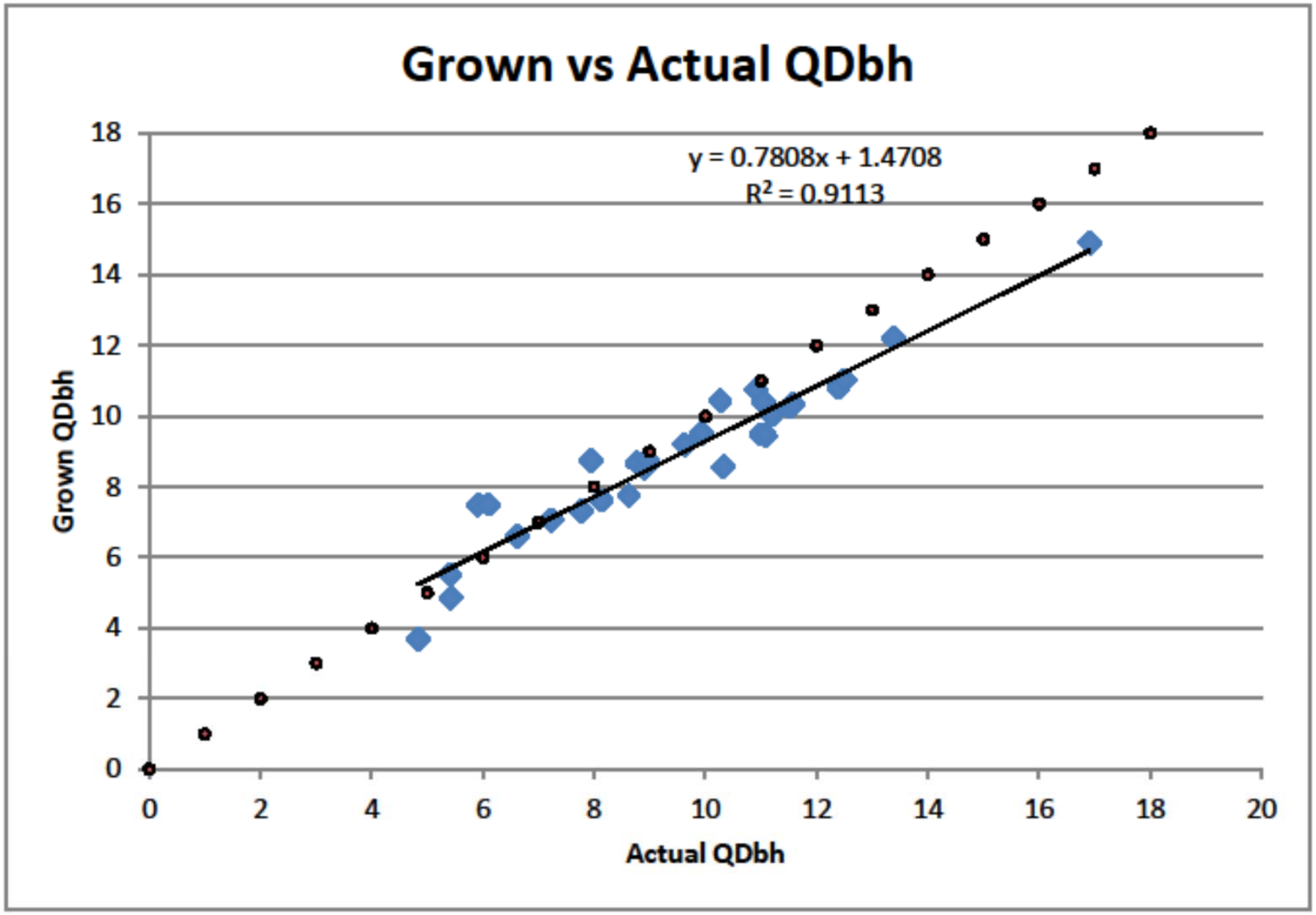
Proj v Act QDbh



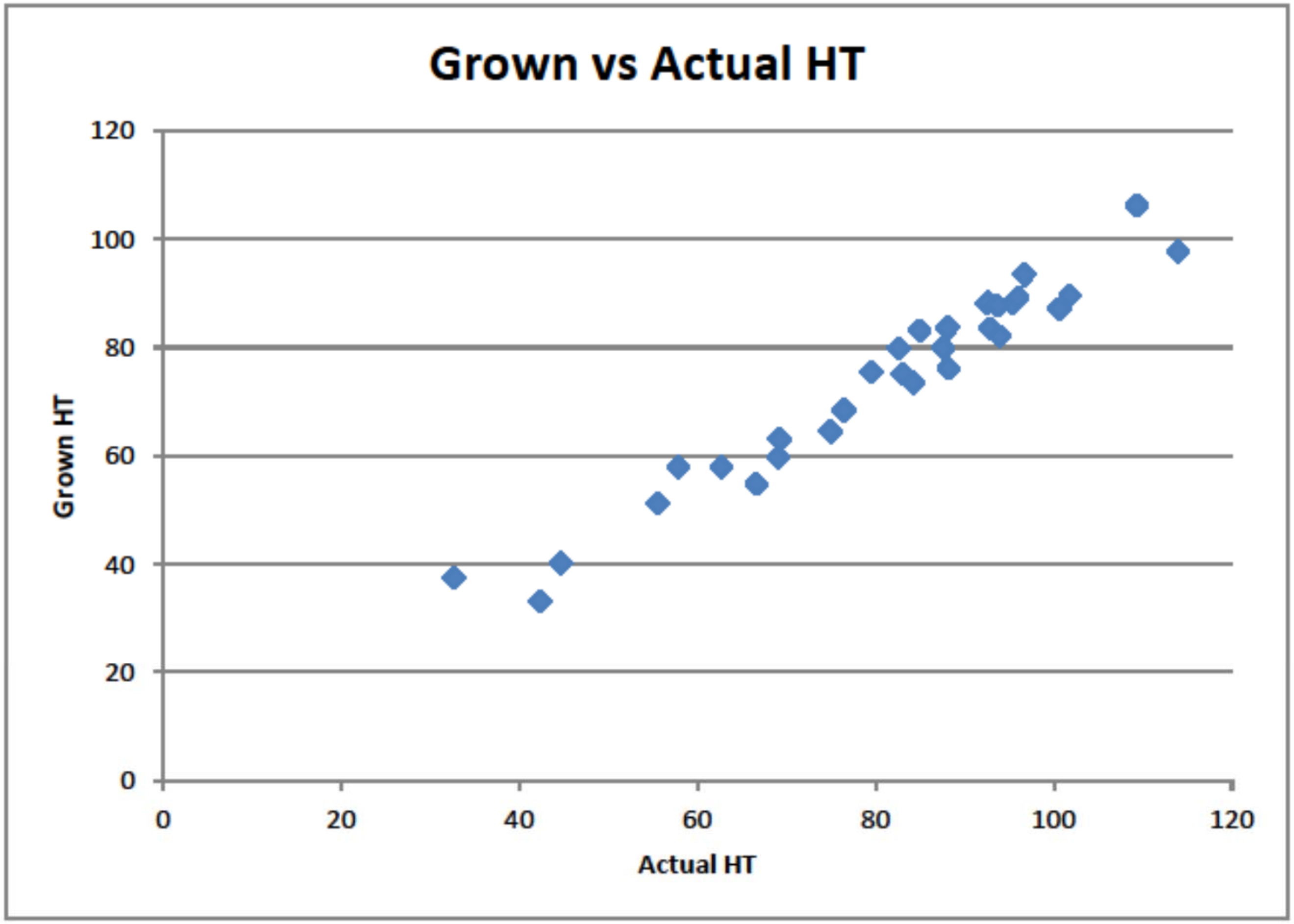
1:1 line added



Trend line added

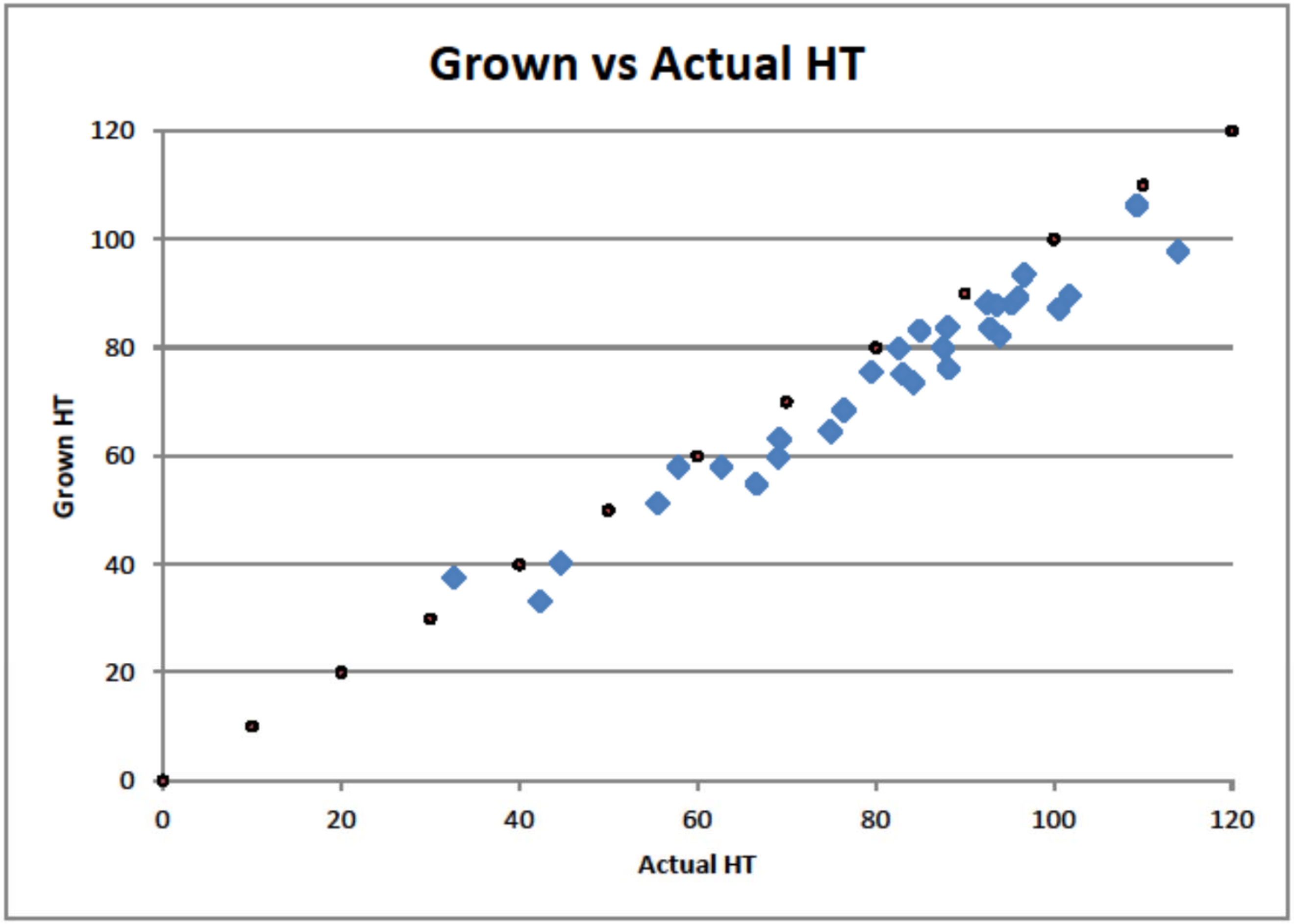


Proj v Act Ht

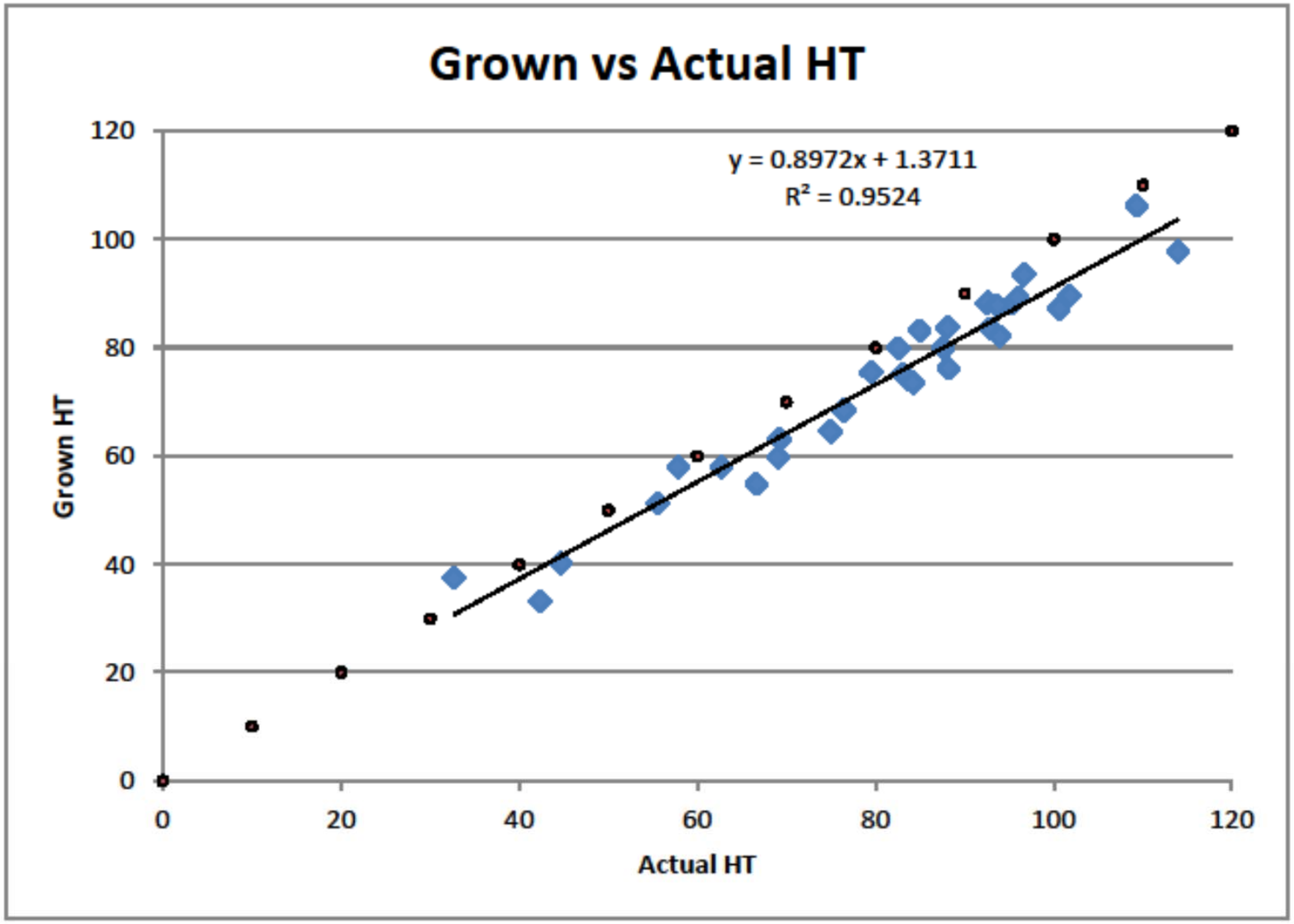


Grown vs Actual HT
120
100
80
60
40
20
0
Grown HT
0
20
40
60
80
100
120
Actual HT

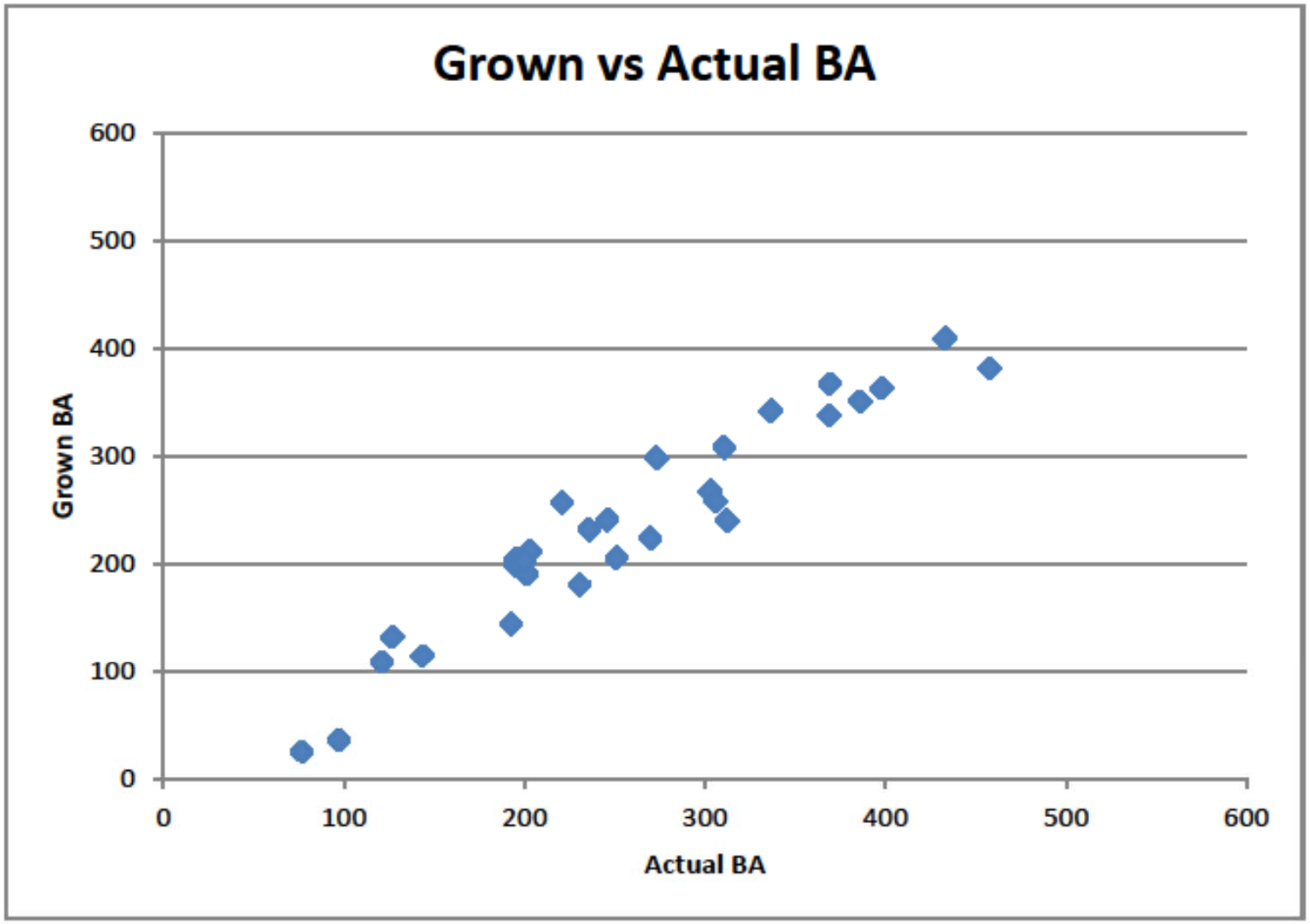
1:1 line added



Trend line added

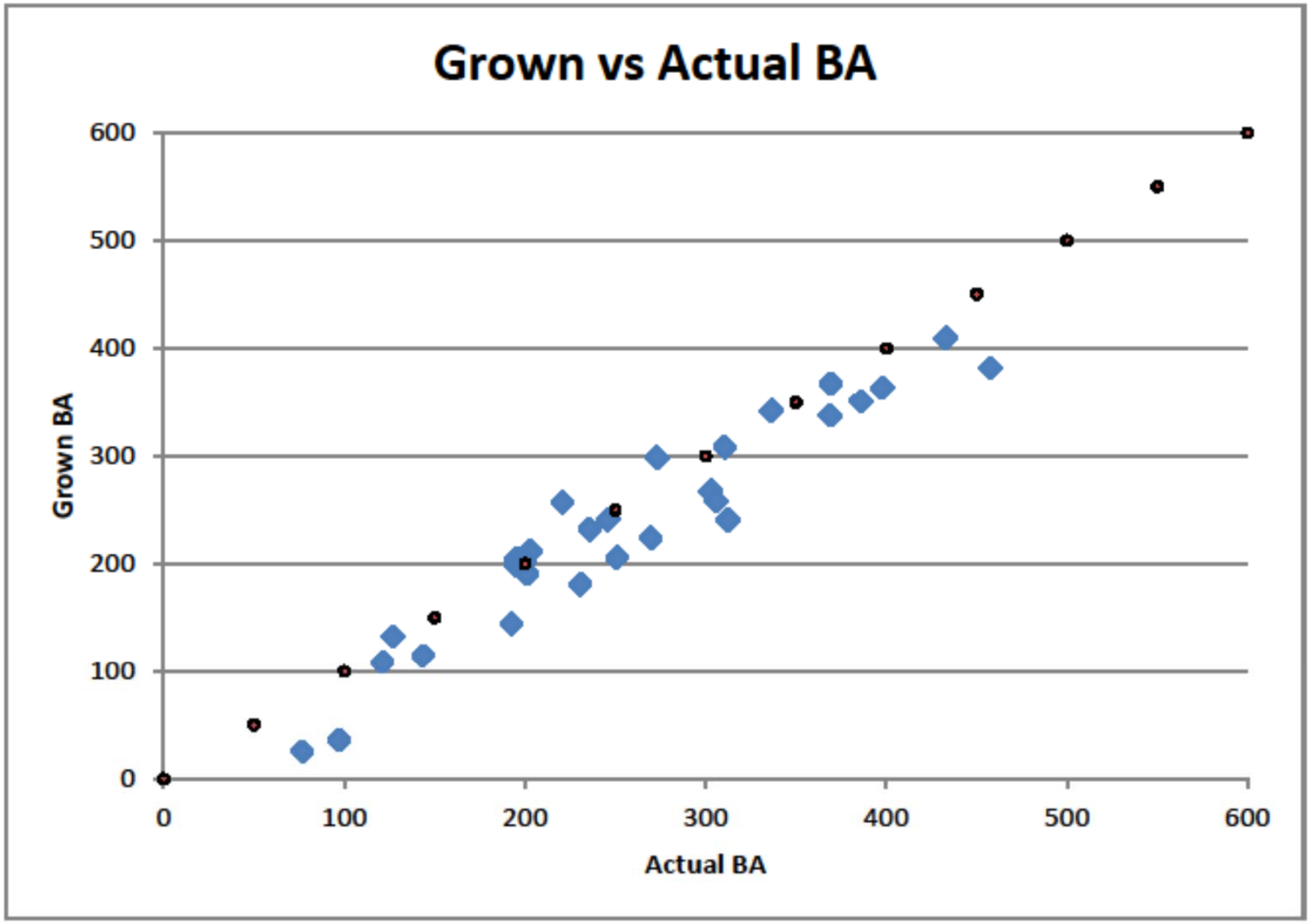


Proj v Act BA

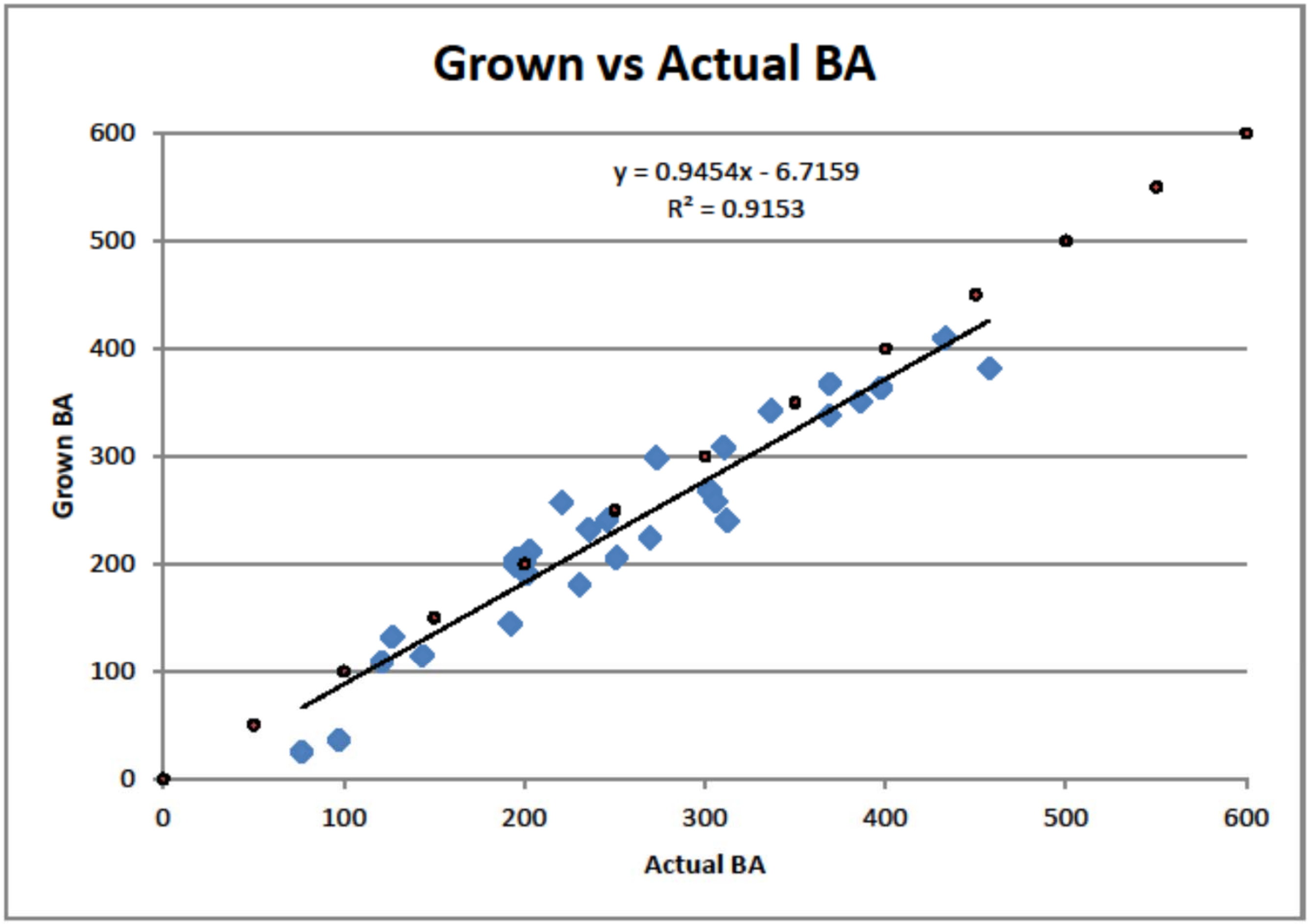


Grown vs Actual BA
600
500
400
300
200
100
0
Grown BA
0
100
200
300
400
500
600
Actual BA

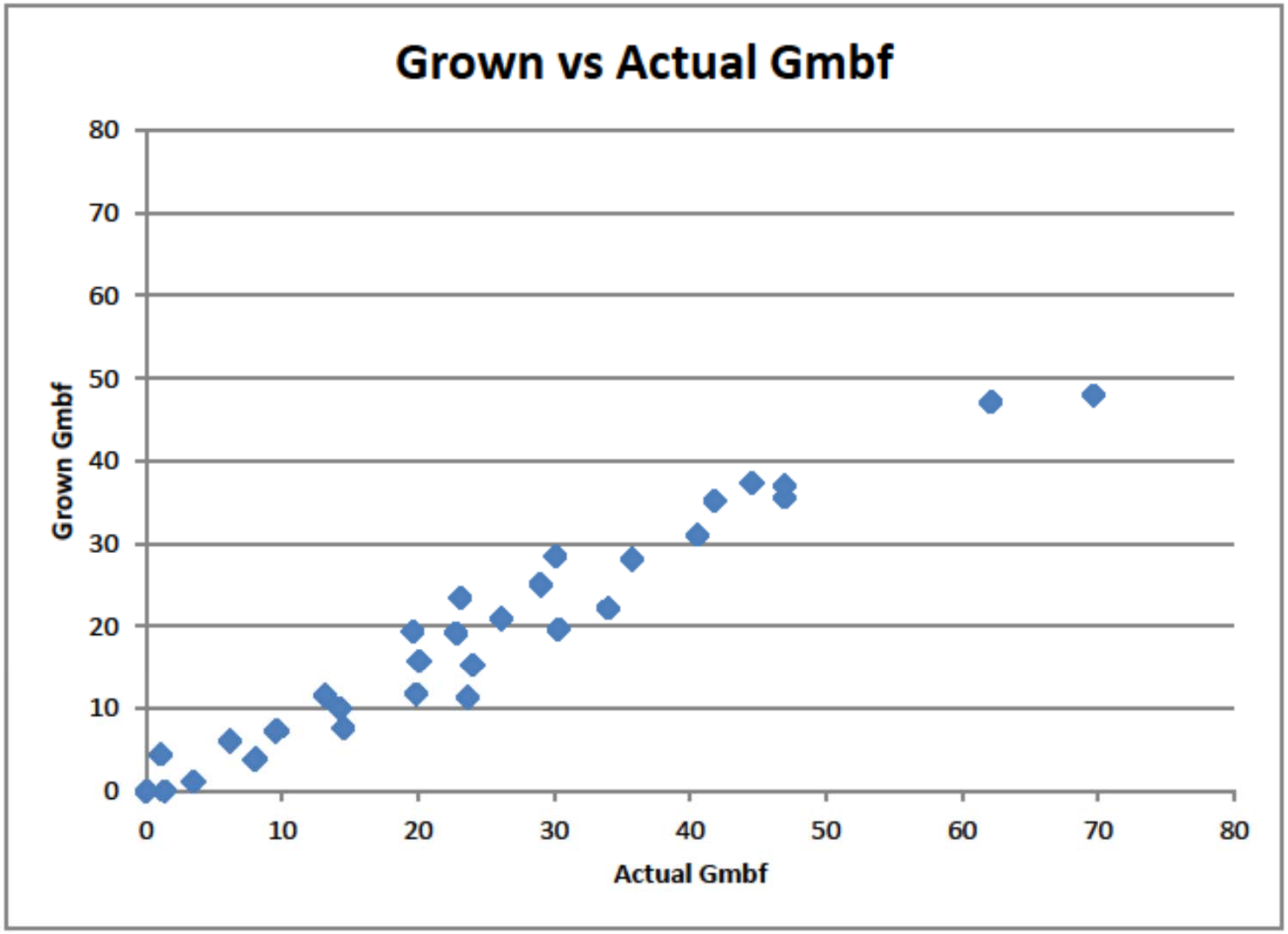
1:1 line added



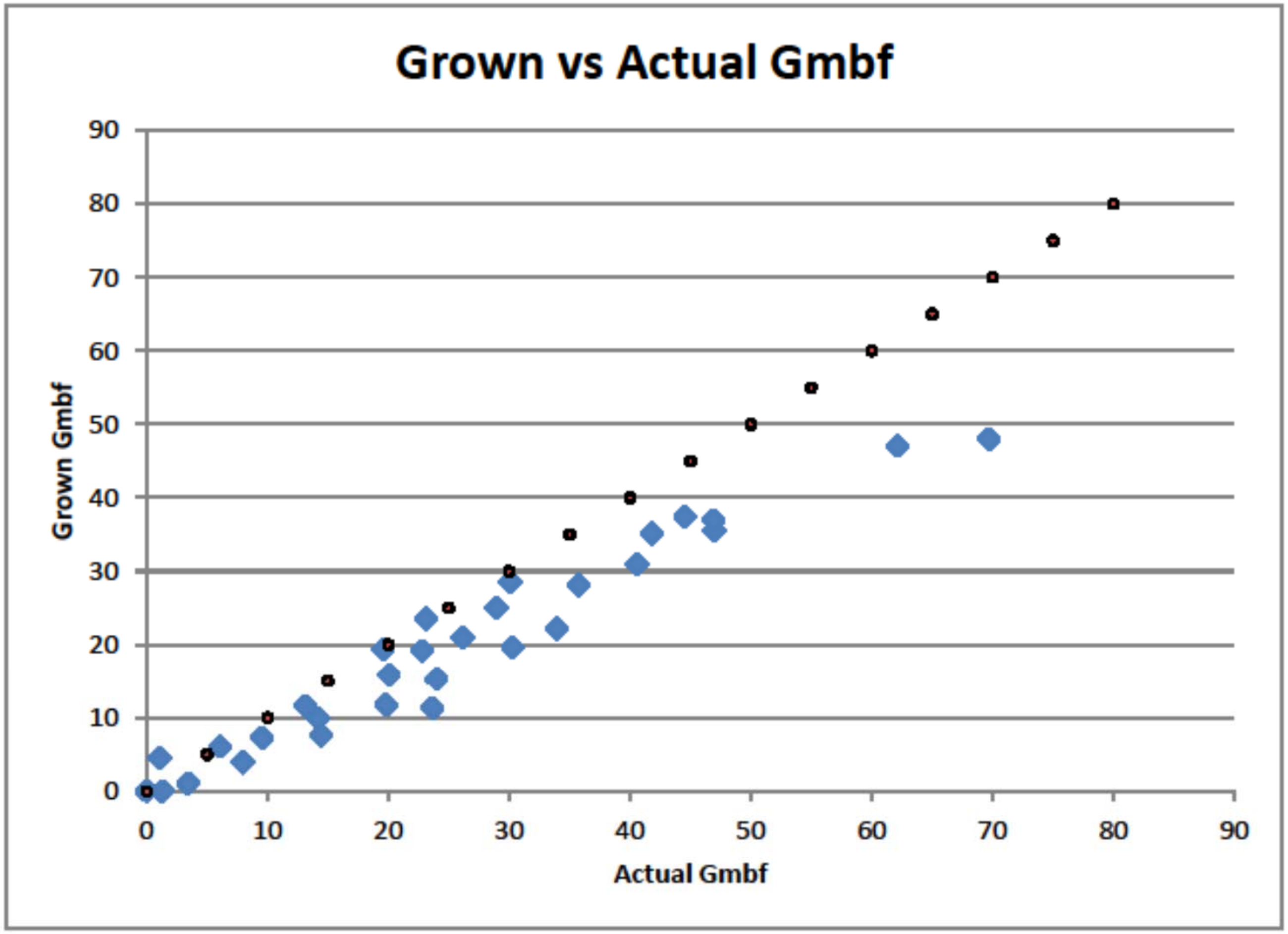
Trend line added



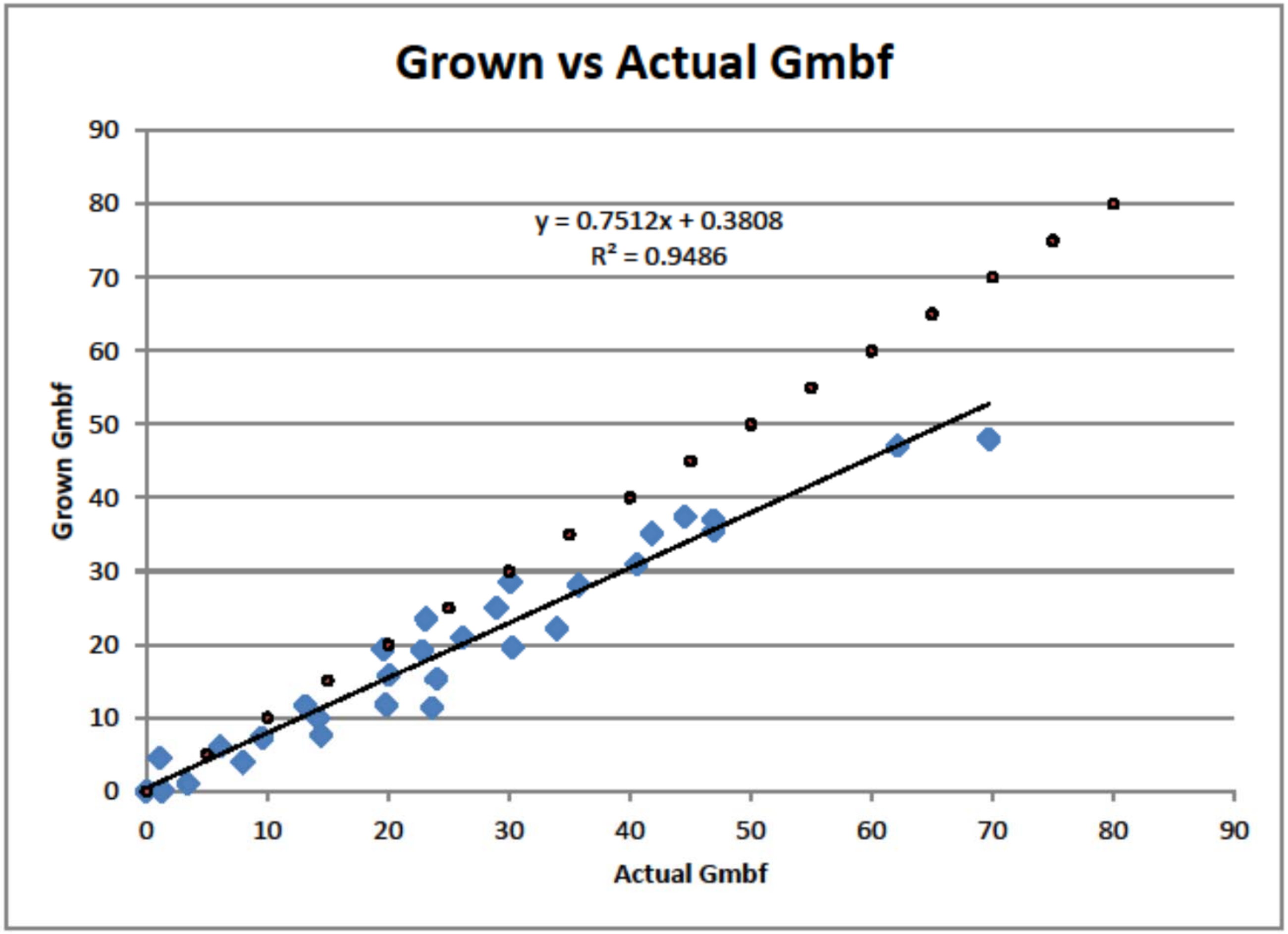
Grown v Actual Gmbf



1:1 line added



Trend line added



Grown v Act Regression Coefficient CIs

TPA	Coefficients	Lower 95%	Upper 95%
Intercept	35.41	-110.67	181.49
X Variable 1	0.94	0.69	1.18

QDbh	Coefficients	Lower 95%	Upper 95%
Intercept	1.47	0.54	2.40
X Variable 1	0.78	0.69	0.88

BA	Coefficients	Lower 95%	Upper 95%
Intercept	-6.72	-37.11	23.68
X Variable 1	0.95	0.83	1.06

HT	Coefficients	Lower 95%	Upper 95%
Intercept	1.37	-5.07	7.81
X Variable 1	0.90	0.82	0.97

Gmbf	Coefficients	Lower 95%	Upper 95%
Intercept	0.38	-1.70	2.46
X Variable 1	0.75	0.68	0.82

Accuracy Stats

Root Mean Square Error (RMSE)

sensitive to "outliers"

$$RMSErrors = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Mean Absolute Error (MAE)

RMSE >= MAE

"Smaller is better"

useful in comparing and refining models

before and after calibration

both estimate "absolute bias"

plots or means indicate bias direction

On Statistical Testing

Yang, Monserud & Huang, 2004:CJFR:34:619-629. An evaluation of diagnostic tests and their roles in validating forest biometric models

- *Compared 5 parametric and 5 non-parametric tests*

"It was shown that the usefulness of statistical tests in model validation is very limited. None of the tests seems to be generic enough to work well across a wide range of models and data. Each model passed one or more tests, but not all of them. Because of this, caution should be exercised when choosing a statistical test or several tests together to try to validate a model. It is important to reduce and remove any potential personal bias in selecting a favorite test, which can influence the outcome of the results."

Hypothesis Testing

Dbh & Height & BA

- tempting to use a standard paired-t test
- check distribution assumptions first
- log ratio of Grown over Actual is better
  - ln(G/A): H0: stat = 0, [+]: G>A; [-]:G<A

Tpa (survival and/or mortality)

- chi-square or K-S goodness-of-fit test

Equivalence testing

- set “practical or acceptable difference”
- TOST (two one-sided tests)
- Significance level to use?

Technique Summary

Graphical first, then statistical

- *proj vs actual scatterplots can be quite revealing*
  - *add 1:1 line (perfect agreement)*
  - *add linear trend line*
    - *slope (1)*
    - *intercept (0)*

Statistics

- *RMSE and MAE -->smaller is better*
- *distribution assumption for standard paired-t*
- Dbh, Ht, BA: ln(G/A):
  - H0: stat = 0; [+]: G>A; [-]:G<A
- Tpa: chi-squared or K-S goodness-of-fit test
- Equivalence tests --> incorporate “practical est”

Session 2 – Model Evaluation

Slide 2

Re-Cap Model Evaluation

- Evaluation/Validation is relative to Utility and Purpose
- Methods parallel the methods used to develop models
- Evaluation with Data
  - Predicted – Actual
- Evaluation without data
  - Patterns and principles
- Techniques
  - Bias, Precision