Nonlinear Relationship of Tree Features

Introduction

The project invests the relationship of the tree diameter at the breast height (DBH, in inches) to their total height (HT, in feet) for western white pine, there are 900 pairs of data.

The purpose of the study is to (1) develop the "best" model for predicting the tree total height using observed DBH in a long-term forest growth, (2) yield simulator among Five nonlinear asymptotic growth functions, (3) find the best model for large-sized trees.

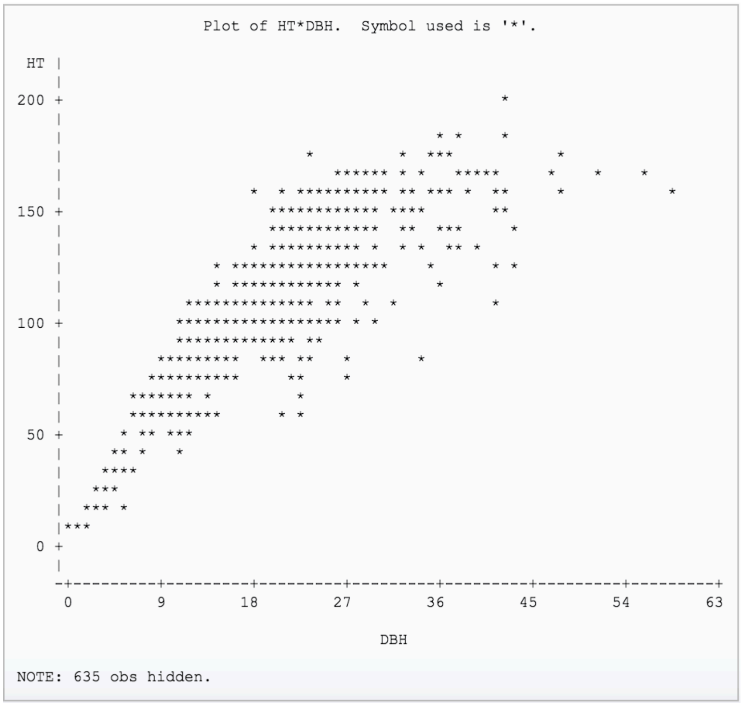
Data and Methods

There are 900 observations in total and Table 1 gives a brief summary of the variables. The maximum of HT is 200. Figure 1 is the scatter plot of tree total height (HT) against diameter (DBH). According to figure 1, the relationship between HT and DBH is not linear, thus we should apply nonlinear regression to analyze their relationship.

Table 1. *Describe Statistics of DBH and HT.*

Variable	N	Mean	Median	Std Dev	Minimum	Maximum	Corrected SS	Uncorrected SS
DBH	900	16.6719	18.05	10.857	0.1	58.1	105968.5388	356125.23
HT	900	90.0403	102	49.7959	4.5	200	2229192.606	9525728.07

Figure 1. Scatter Plot of Tree Total Height (HT) Against Diameter (DBH).



Five growth functions are applied and three iteration methods are used to converge the data (GAUSS, MARQUARDT, and NEWTON).

Results and Discussion

1. **Functions**

   1.1. **Exponential Function**

$$HT = a * e^{(\frac{-b}{DBH+c})}$$

Thus, $(-ln\frac{HT}{a})^{-1} = \frac{c}{b} + \frac{1}{b}DBH$

Table 2. Exponential Function Initial Coefficient Estimation

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.00734	0.05304	0.14	0.8900
DBH	1	0.10296	0.00267	38.53	<.0001

According to table 2, the initial model parameters are: a=max(HT)=200, $\frac{c}{b} = 0.00734, \frac{1}{b} = 0.10296$, thus, b=9.7125, c=0.07129.

Then we converge the parameters by GAUSS, MARQUARDT, and NEWTON.

Table 3. Exponential Function Parameter Iteration by GAUSS, MARQUARDT, and NEWTON

Method	**Parameters**	**Estimates**	**standard errors**	**asymptotic correlations**			**iterations**
				a	b	c	
NEWTON	a	229.3	5.4676	1	0.9585	0.7790	7
	b	16.8873	0.845	0.9585	1	0.9017	
	c	4.17	0.3849	0.7790	0.9017	1	
MARQUARDT	a	229.3	5.5063	1	0.9593	0.7825	4
	b	16.8873	0.8508	0.9593	1	0.9031	
	c	4.17	0.3861	0.7825	0.9031	1	
GAUSS	a	229.3	5.5063	1	0.9593	0.7825	4
	b	16.8873	0.8508	0.9593	1	0.9031	
	c	4.17	0.3861	0.7825	0.9031	1	

Table 3 gives a summary of the parameter estimates, asymptotic standard errors for the parameters, and asymptotic correlations among the parameters for the three methods. Newton took more steps to iterate, but it has smaller correlation between b and c, Marquardt and Gauss have exactly the same result, all three methods produce similar parameter estimations and standard error.

Exponential function: $\widehat{HT} = 229.3 * e^{-\frac{16.8873}{DBH+4.17}}$

1.2. Lundqvist Function

$$HT = a * e^{(-b*DBH^{-c})}$$

Thus, $ln\left(-ln\left(\frac{HT}{a}\right)\right) = \ln(b) - cln(DBH)$

Table 4. Lundqvist Function Initial Coefficient Estimation

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	0.98567	0.02207	44.67	<.0001
lnDBH	1	-0.53455	0.00822	-65.03	<.0001

Thus, the initial parameters are a=max(HT)=200, b=$e^{\beta 0} = 2.6796$, c=-$\beta 1$=0.53455.

Converge the parameters by three methods mentioned above.

Table 5. Lundqvist Function Parameter Iteration by GAUSS, MARQUARDT, and NEWTON

Method	Parameters	Estimates	standard errors	correlations			iterations
				a	b	c	
NEWTON	a	491	123.5	1	-0.5541	-0.9913	19
	b	4.6627	0.143	-0.5541	1	0.6576	
	c	0.3854	0.0641	-0.9913	0.6576	1	
MARQUARDT	a	491	94.0327	1	-0.2563	-0.9853	16
	b	4.6628	0.1173	-0.2563	1	0.4161	
	c	0.3854	0.0469	-0.9853	0.4161	1	
GAUSS	a	491	94.0548	1	-0.2559	-0.9853	24
	b	4.6627	0.1173	-0.2559	1	0.4157	
	c	0.3854	0.0469	-0.9853	0.4157	1	

Table 5 shows that all three methods produce the same parameter estimations, but Marquardt and Gauss have smaller correlation between b and c and smaller standard error than Newton, and Marquardt converge in fewer steps than Gauss. Thus, Marquardt overwhelming other two methods, we should use this method to do the simulation.

Lundqvist Function: $\widehat{HT} = 491 * e^{-4.6627*DBH^{-0.3854}}$

1.3. Modified Logistic Function

$$HT = \frac{a}{1 + \frac{1}{b} * DBH^{-c}}$$

Thus, $\ln\left(\frac{a}{HT} - 1\right) = \ln(b) - c * DBH$

Table 6. Modified Logistic Function Initial Coefficient Estimation

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	2.55308	0.04235	60.28	<.0001
DBH	1	-0.12587	0.00213	-59.00	<.0001

The initial a = 200, b=$e^{-\beta 0}$ =0.0778, c=-$\beta 1 = 0.1259$. Since the bad initial value, it's took more steps to converge to a stable number, we tried to converge with a=60 b=0.1 c=2.

Table 7. Modified Logistic Function Parameter Iteration by GAUSS, MARQUARDT, and NEWTON

Method	**Parameters**	**Estimates**	**standard errors**	**correlations**			**iterations**
				a	b	c	
NEWTON	a	114.4	1.0003	1	.	.	2
	b	0.0707	.	.	.	.	
	c	2.238	.	.	.	.	
MARQUARDT	a	226.6	13.3452	1	0.5651	-0.9193	14
	b	0.0311	0.00282	0.5651	1	-0.8410	
	c	1.1615	0.0636	-0.9193	-0.8410	1	
GAUSS	a	226.6	13.3459	1	0.5651	-0.9193	8
	b	0.0311	0.0028	0.5651	1	-0.8410	
	c	1.1614	0.0636	-0.9193	-0.8410	1	

According to SAS, Newton method is questionable, convergence criterion met but a note in the log indicates a possible problem with the model, the intercept was not specified for this model and the (approximate) Hessian is singular, thus, we are unable to obtain the estimated standard error and the corresponding correlations. Between Marquardt and Gauss, they have the same parameter estimations, standard error and correlations between parameters, but Gauss is preferred because it has fewer steps in converging.

Modified Logistic Function: $HT = \frac{226.6}{1 + \frac{1}{0.0311} * DBH^{-1.1614}}$

1.4. Richard Function

$$HT = a * (1 - e^{(-b*DBH)})^{c}$$

$$\ln\left(1 - \frac{HT}{a}\right) = -b * DBH$$

Table 8. Richard Function Initial Coefficient Estimation

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
DBH	1	-0.04194	0.00036561	-114.72	<.0001

The initial parameter estimations are: a=200, b=β and c=1.

Table 9. Richard Function Parameter Iteration by GAUSS, MARQUARDT, and NEWTON

Method	Parameters	Estimates	standard errors	correlations			iterations
				a	b	c	
NEWTON	a	179	5.7689	1	-0.9475	-0.7715	7
	b	0.0525	0.00517	-0.9475	1	0.9261	
	c	1.0499	0.0622	-0.7715	0.9261	1	
MARQUARDT	a	179	5.748	1	-0.9480	-0.7664	4
	b	0.0525	0.00508	-0.9480	1	0.9223	
	c	1.0499	0.0601	-0.7664	0.9223	1	
GAUSS	a	179	5.7479	1	-0.9480	-0.7664	5
	b	0.0525	0.00508	-0.9478	1	0.9223	
	c	1.0499	0.0601	-0.7664	0.9223	1	

From table 9, Marquardt is preferred because it has fewer steps to converge and smaller correlation between b and c compared to Newton method. All methods produced same parameter estimation whereas Marquardt and Gauss have smaller standard error.

Richard Function: $HT = 179 * (1 - e^{(-0.0525*DBH)})^{1.0499}$

1.5. Weibull Function

$$HT = a * (1 - e^{-b*DBH^{c}})$$

Thus, $\ln\left(-\ln\left(1 - \frac{HT}{a}\right)\right) = \ln(b) + c * \ln(DBH).$

Table 10. Weibull Function Initial Coefficient Estimation

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-2.81388	0.01675	-167.98	<.0001
lnDBH	1	0.87079	0.00624	139.54	<.0001

Initial parameter estimations are a=200, b=$e^{\beta 0}$=0.05997, c=$\beta 1 = 0.87079$.

With all the initial estimations, we can apply iteration methods to converge the estimated parameters.

Table 11. Weibull Function Parameter Iteration by GAUSS, MARQUARDT, and NEWTON

Method	**Parameters**	**Estimates**	**standard errors**	**correlations**			**iterations**
				a	b	c	
NEWTON	a	176.8	6.4967	1	0.4931	-0.8395	15
	b	0.0453	0.00364	0.4931	1	-0.8825	
	c	1.0442	0.0433	-0.8395	-0.8825	1	
MARQUARDT	a	176.8	6.4239	1	0.4661	-0.8338	10
	b	0.0453	0.00353	0.4661	1	-0.8726	
	c	1.0442	0.0419	-0.8338	-0.8726	1	
GAUSS	a	176.8	6.4241	1	0.4661	-0.8338	6
	b	0.0453	0.00353	0.4661	1	-0.8726	
	c	1.0442	0.0419	-0.8338	-0.8726	1	

Table 11 claims that Gauss is outperformed other two methods. All methods have same parameter estimations, whereas Gauss and Marquardt have smaller correlations between b and c, and smaller standard errors. Gauss is preferred than Marquardt due to its' fewer steps to convergence.

Weibull Function: $HT = 176.8 * (1 - e^{-0.0453*DBH^{1.0442}})$

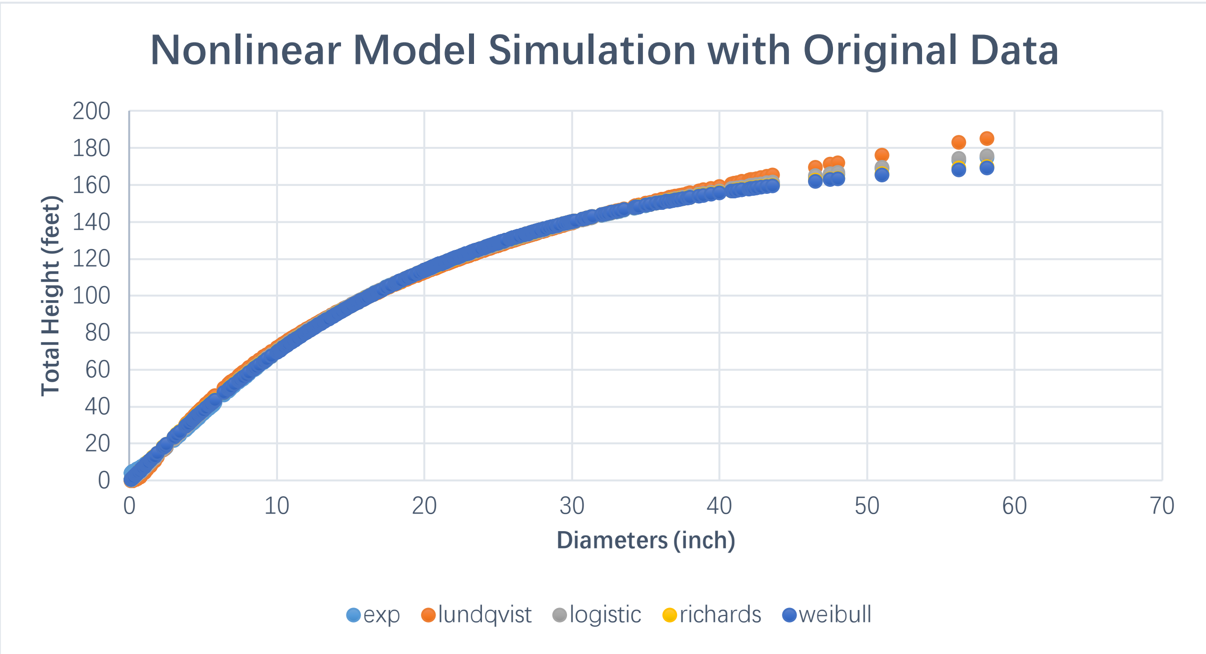
2. Comparison

To evaluate the fitness of the functions, we fit the five functions to the original data, by the analysis above, we choose Gauss as the iteration method. Figure 2 below is the plot of predicted value of HT using each function. Table 12 shows that Exponential function is the most efficient function and has the "best" statistical properties, it has the smallest RMSE, smallest standard errors of each parameter and largest R-square overall.

Table 12. Statistical Properties of the Five Functions.

Equation	Parameters	Estimates	SE	RMSE	R-square
Exponential	a	229.3	5.5063	14.96329	90.99%
	b	16.8873	0.8508		
	c	4.17	0.3861		
Lundqvist	a	491	94.0548	15.2217	90.68%
	b	4.6627	0.1173		
	c	0.3854	0.0469		
Modified Logistic	a	226.6	13.3459	15.02997	90.91%
	b	0.0311	0.00282		
	c	1.1614	0.0636		
Richards	a	179	5.7479	14.98666	90.96%
	b	0.0525	0.00508		
	c	1.0499	0.0601		
Weibull	a	176.8	6.4241	14.98332	90.97%
	b	0.0453	0.00353		
	c	1.0442	0.0419		

Figure 2. Nonlinear Model Simulation with Original Data



3. Simulation

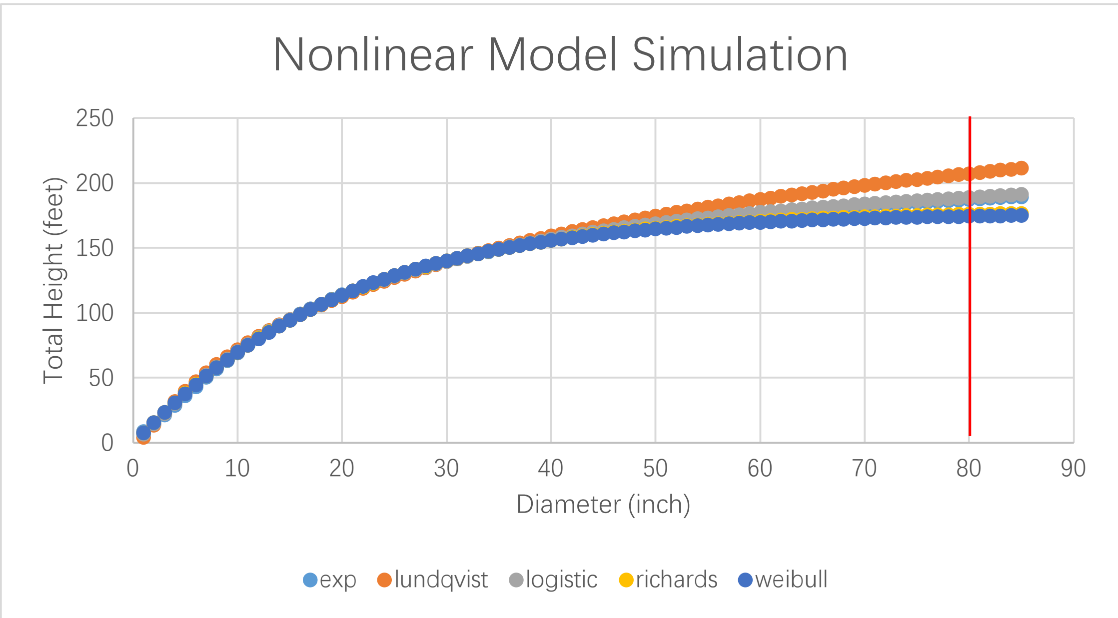
Table 12. Simulated HT (Partition)

DBH	exp	lundqvist	logistic	richards	weibull
1	8.746	4.636	6.835	7.893	7.830
……					

81.0	188.059	208.345	189.574	176.327	174.747
81.1	188.102	208.430	189.619	176.341	174.759
81.2	188.146	208.515	189.663	176.355	174.771
81.3	188.190	208.599	189.707	176.369	174.783
81.4	188.233	208.684	189.751	176.382	174.794
81.5	188.276	208.768	189.795	176.396	174.806
81.6	188.320	208.853	189.839	176.410	174.817
81.7	188.363	208.937	189.883	176.423	174.828
81.8	188.406	209.021	189.926	176.437	174.840
81.9	188.449	209.105	189.970	176.450	174.851
82.0	188.492	209.189	190.014	176.464	174.862

For estimating big trees, for example, DBH=81.8, lundqvist function gives the closest estimate: 209.021. Thus, lundqvist function has the "best" prediction performance for large-sized trees. Figure 3 shows lundqvist function gives the largest HT among all the functions whereas weibull function gives the smallest prediction on HT. when DBH is small, there is no large differences between these five functions.

Figure 3. *Nonlinear Model Simulation.*



Summary

The relationship between the tree's total height and their diameters is nonlinear, we can use lundqvist function to predict for large-sized trees and use exponential function to predict normal-sized trees since exponential function has more efficient parameters.