A Comparison of Regression Models for Developing Rating Curves

Cameron Bracken
Humboldt State University
ENGR 324

April 30, 2007

Abstract

Stream discharge and stage data was obtained for a gaging station near Tres Pinos, California. Two polynomial and two power curve regressions were carried out on groups of data prior to and including flows from large storm events. The goal of the analysis was to compare the commonly used power curve regression with other types of regression for obtaining rating curves. F-tests and r^2 values were calculated to evaluate the effectiveness of each regression relative to the power curve regression. The power curve model was statistically outperformed on every subset of data. Depending on the situation, a polynomial or quadratic power curve model may provide a fit that is statistically better than the power curve model. The power curve model tends to perform better when data points are closely spaced and no data points lie far from the main cluster. The traditional power curve method tends to overestimate high flows. Also using the r^2 value alone could be misleading in judging which model is statistically better. F-tests, runs tests and Ryan-Joiner tests are also important for evaluating the relative performance of each model. The choice of which model to use depends on the data. The conclusion of this analysis is that the power curve model does not provide the best fit for stage-discharge relationships but may be easier to carry out than the other models used.

Contents

List of Figures

List of Tables

Notation

h	=	gage height (ft)
Q	=	Discharge (cfs)
a_i	=	ith regression coefficient
n	=	Number of paired oservations (Q, h)
y_i	=	ith observed dedependent variable
r^2	=	Coefficient of (multiple) determination
r	=	Correlation coefficient
SSR	=	Sum of squared residuals
SST	=	Total sum of squares
R	=	Number of runs
μ_{R}	=	Expected number of runs
σ_{R}	=	Standard deviation of runs
Z_{R}	=	Standard normal Z score from runs test
Z_α	=	Standard normal value associated with significance level α
α	=	Significance level
$F_{\alpha,n-p_a,n-p_b}$	=	F statistic with signifigance level α and degrees of freedom $n-p_a$ and $n-p_b$
p_a	=	Degrees of freedom of model a
p_b	=	Degrees of freedom of model b
f	=	Observed F value for F test

1 Introduction

United States Geological Survey (USGS) site 11157500 is located near Tres Pinos, CA in San Benito County (Various 2007). The site measures the stream stage (ft) and stream discharge (ft^3/s) of Tres Pinos Creek (Figure 1).

(a) San Benito County (b) Tres Pinos, CA

Figure 1: Location of gaging station at Tres Pinos Creek (USGS 2007)

Daily discharge values were obtained from the USGS website (Figure 2) (USGS 2007). The period of record was examined for large storm events and data subsets were prepared to analyze leading up to and including the large events. This is useful because high flow measurements are difficult make and river geometry can change considerably during a large storm events (Dose et al. 2001). Six data sets were used corresponding to the storm events in 1996, 1997, 1998, 1999, and two in 2005.

A Rating curve is a relationship between stream height (stage) and stream flow (discharge) across a stream cross section. The goal in developing a rating curve is to predict the flow rate in a stream by measuring only the stage because flow measurements during storm condition are dangerous or impossible to make.

The goal of this report is to compare various interpolation methods to the most commonly used linearization method. In the literature review and methods section, this report will provide a brief review of literature related to developing rating curves and measuring stage and discharge. Additionally typical relationships for stage and discharge will be discussed. A review of the statistical methods used to compare models is provided. In the results section, the results of the regression and statistical analysis is provided. The discussion section describes and interprets the results and forms a conclusion.

2 Literature Review and Methods

2.1 Motivation

The motivation for developing stage-discharge relationships lies in the difficulty of making velocity (and therefore flow) measurements during storm conditions (Dose et al. 2001). Typically high flow measurements are inaccurate and few in number. Sometimes extrapolation techniques such as numerical modeling are used to predict stage discharge relationships where little data is available. Verifying the currently interpolation methods may help to better predict intermediate flows where little data is avalable.

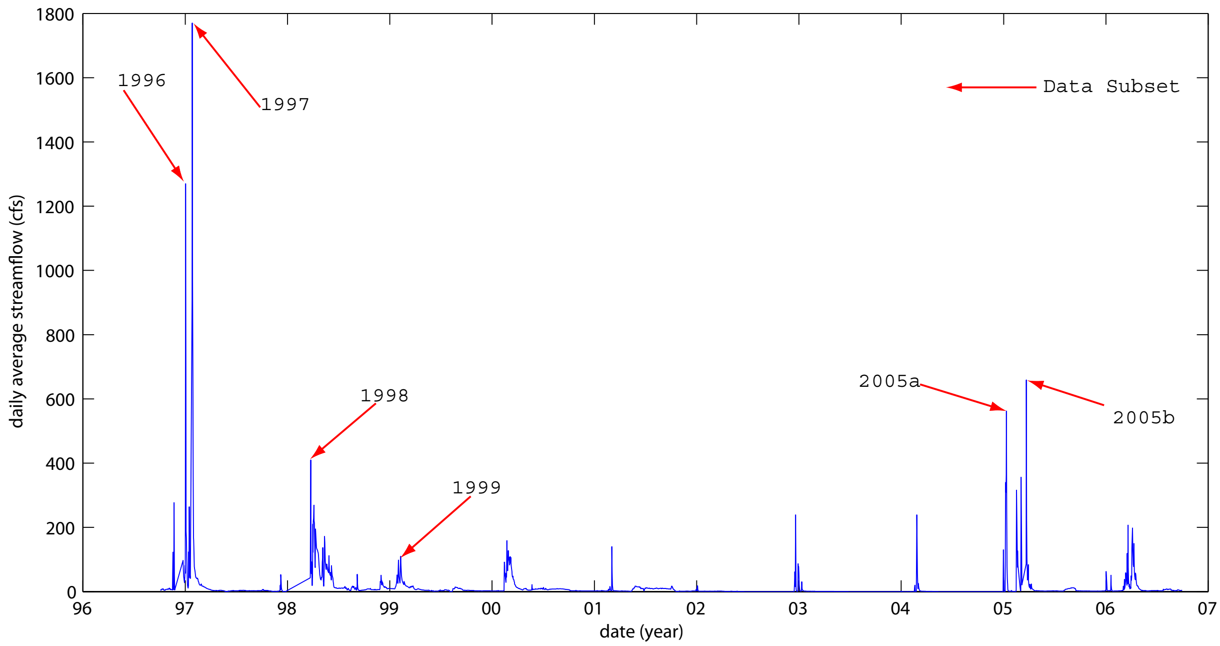


Figure 2: Daily Discharge values for the period of record (USGS 2007).

2.2 Measuring techniques

Many techniques are available for measuring both stage and discharge. The most common methods are discussed.

2.2.1 Discharge measurements

Point velocities are measured by breaking up a cross section of a stream into evenly spaced panels (Figure 3). The velocity is measured in each panel. Flow rate in each panel is calculated by multiplying the average velocity in each panel by the cross sectional area of the panel. The total flow rate is the sum of the flow rates across each panel. This method is analogous to integration of a 2 dimensional function (Department 2007). The average velocity in

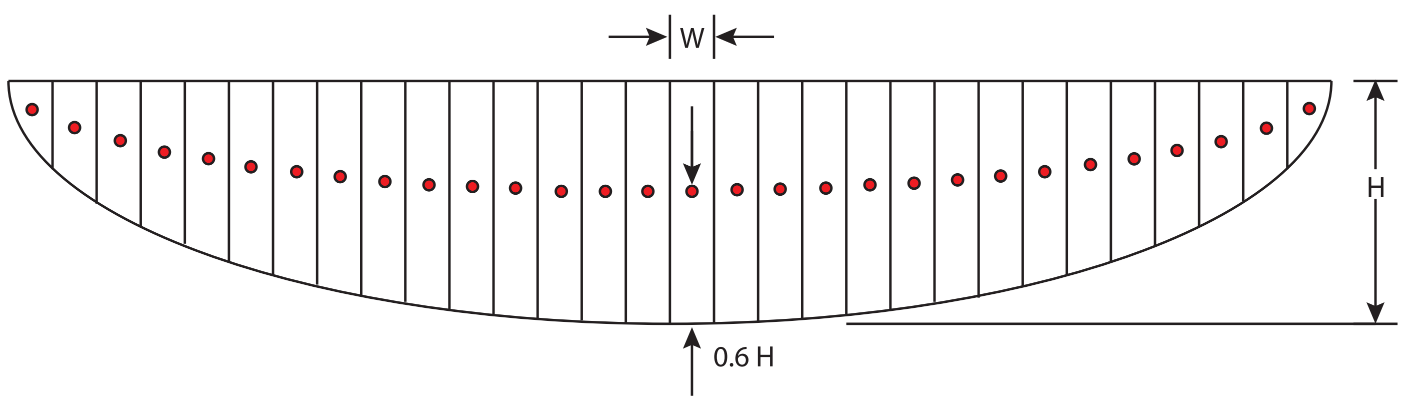


Figure 3: River cross section used to estimate flow rate (image courtesy of Jason Roberts).

each panel is estimated by measuring the velocity at one point in each panel which is located 60% of the total height from the bottom of the river (Pitlick 2002). Figure 3 shows the approximate location of the measuring points. The flow rate Q is calculated by multiplying the Area of the panel by the average velocity in that panel.

This "panel method" is the most widely used method for calculating flow rates. Other methods exist which use more or less sophisticated equipment and techniques.

2.2.2 Stage measurements

Stream stage is usually measured at a gaging station. Gaging Stations measure the stage of a stream by measuring the pressure head from a pipe that is installed in the side of the stream (Geology.com 2007). The measurements from a a gaging station are called gage height. The gage height measurements from most gaging station are sent by satellite to the USGS.

2.3 Extrapolation

Many techniques have been developed for extrapolating a full rating curve from only low flow measurements or very few high flow measurements. Dose et al. used second and third order polynomial curve fitting, least squares regression, linearization, and numerical modeling. Sivapragasam et al. (2005) implemented a method called Support Vector Machine in order to extrapolate results. Schmidt and Yen (2001) implemented an unsteady flow equation to extrapolate a stage-discharge relationship. The disadvantage to numerical modeling is that it relies on "the collection of a considerable amount of data" (Dose et al. 2001). This report will implement the same curve fitting methods as Dose et al. (2001) but interpolation is the focus of this report, not extrapolation. Interpolation is just as important as extrapolation because many times only a few high flow measurements are available and intermediate information is necessary.

2.4 Typical Relationships for Stage and Discharge

The traditional and most common model for developing rating curves is to plot the the stage-discharge points on a log-log scale (linearization) (Dose et al. 2001). Linear regression is then used to fit a trend line to the data by the method of least squares. A logarithmic relationship is back calculated using the coefficients a_1, a_2.

$$\begin{aligned}\ln Q_i &= a_1 \ln h_i + a_2 \\ Q_i &= e^{a_2 + a_1 \ln h_i} \\ Q_i &= e^{a_2} h_i^{a_1}\end{aligned}$$

where h is gage height and Q is flow rate. The curve generated from the linearization technique in called a power curve. This paper implements some lesser used but still common models for interpolating rating curves. The goal is to determine if some less commonly used models provide a better fit to the data than the power curve model.

This report also implements an uncommonly used method though. Higher order polynomial fits to linearized data have been suggested in the literature (Sivapragasam and Muttil 2005). A model will be implemented which will be called the quadratic power curve. This model involves fitting a second order polynomial to the linearized stage discharge data. A stage discharge relationship is back calculated using the coefficients a_1, a_2, a_3.

$$\begin{aligned}\ln Q_i &= a_1 (\ln h_i)^2 + a_2 \ln h_i + a_3 \\ Q_i &= e^{a_1 (\ln h_i)^2 + a_2 \ln h_i + a_3} \\ Q_i &= e^{a_1 (\ln h_i)^2} h_i^{a_2} e^{a_3}\end{aligned}$$

These two linearization methods method work well for lower flows. Because of the behavior of the logarithm function (reducing high numbers more than low numbers), the linearization method tends to work inaccurately for high flows (Sivapragasam and Muttil 2005).

Polynomial regression for interpolating rating curves involves polynomial fitting by the method of least squares. A second order polynomial relationship has the form

$$Q_i = a_1 h_i^2 + a_2 h_i + a_3$$

Similarly a third order polynomial relationship has the form

$$Q_i = a_1 h_i^3 + a_2 h_i^2 + a_3 h_i + a_4$$

Linearization is the most commonly used method for developing rating curves but is frequently inaccurate. The results section of this paper will compare the results of the power curve, quadratic power curve, second order polynomial and third order polynomial method of regression.

2.5 Statistical Comparisons of Models

The statistical methods used in this report are outlined below.

2.5.1 Gauss-Markov Theorem

The Gauss-Markov Theorem States that if $a_1, ..., a_n$ are linear estimators in terms of the dependent variable y and if:

1. The population mean of the residuals equals zero,
2. The variance of the residuals is constant,
3. And the residuals are statistically independent then

The estimators $a_1, ..., a_n$ are the best linear unbiased estimators (Chamberlain 2007a).

The first condition is guaranteed to be satisfied if least squares regression is used. The second condition is not considered quantitatively but residual plots are given in appendix Appendix B. The third condition can be tested using the runs test, explained below.

As for the linearity of the coefficients $a_1, ..., a_n$, the two linearized models can be transformed to models of a linear form. These are called intrinsically linear models (Devore 2000). The polynomial models are also linear with respect to the dependent variable.

2.5.2 Coefficient of Multiple Determination (r^2 value)

Another important quantity in regression is the total sum of squares (SST) given by

$$\text{SST} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$$

The total sum of squares is the deviation from the mean of the observed y values. The ratio SSR/SST is the proportion of variation that is not explained by the model (Devore 2000).

The coefficient of variation (r^2) is defined as the proportion of variation that is explained by the model. Therefore the coefficient of determination is given by

$$r^2 = 1 - \frac{\text{SSR}}{\text{SST}}$$

When this quantity is calculated for a regression other than a linear, it is called the coefficient of multiple determination. The value of r^2 is important for evaluating model performance.

2.5.3 Runs Test

Another desirable quality of regression is independence of residuals. The Runs test is used to evaluate this property. To carry out the Runs test:

1. Sort all n independent variables in ascending order.
2. Calculate residuals associated with each independent variable.
3. Count the number of times the residuals switch sign (+ to - or - to +), this is the number of Runs R.
4. Calculate $\mu_\text{R} = n/2 + 1$ and $\sigma_\text{R} = \sqrt{n-1}/2$
5. Calculate $Z_\text{R} = (\text{R} - \mu_\text{R})/\sigma_\text{R}$
6. If $Z_\text{R} > Z_\alpha$ do not reject the assumption that the residuals are independent (Chamberlain 2007b).

2.5.4 NSCORE plots and the Ryan-Joiner test

NSCORE plots and the associated Ryan-Joiner test can be used to test the normality of the residuals. To carry out the Ryan-Joiner test:

1. Assume normal distribution
2. Rank the residuals from most negative to most positive
3. Calculate the cumulative probability for each ith residual by $i/(n+1)$
4. Calculate the Z score associated with each cumulative probability.
5. Calculate the NSCORE associated with each cumulative probability by nscore=$4.91(cp^{.14} - (1-cp)^{.14})$
6. Plot the ranked residual vs. NSCORE.
7. Calculate the correlation coefficient (r) from a linear regression of ranked residual vs. NSCORE.
8. Compare the r value to the tabulated critical r value. If r>table value then do not reject the assumption of normality (Chamberlain 2007c).

A rejection of the assumption of normality of the residuals can indicate that some structure which is not explained by the model (NIST/SEMATECH 2007). The results of Ryan-Joiner test for each model and data set are given in the results section.

2.5.5 F-test for Model Comparison

The F test is a non-predictive test used to evaluate the relative performance of two models (Chamberlain 2007d). To carry out an F test:

1. Calculate $f = \text{SSR}_a/\text{SSR}_b$ such that f>1.
2. Look of $F_{\alpha,n-p_a,n-p_b}$, where p_a, p_b are the number of parameters being estimated in model a and b respectively.
3. If $f > F_{\alpha,n-p_a,n-p_b}$ then model b is better than model a.
4. If $f < F_{\alpha,n-p_a,n-p_b}$ then model a is about the same as b. The results of F tests relative to the power curve model are given in the results section.

3 Results

In the following tables, the "Pass Runs?" column is the answer to the question, do the residuals pass the runs test for independence? The column "NSCORE r" is the r value from the NSCORE plot for use with in the Ryan-Joiner test. The column "Reject Normality?" is the answer to the question of the Ryan-joiner test, are the residuals normally distributed? The column "Which Better?" is the conclusion from an F test performed relative to the power curve model. The power curve model is called PC, the quadratic power curve model is called Quad PC, the quadratic regression is called Quad, and the Cubic regression is called Cube. The units of SSR are cfs^2. For all tests $\alpha = 0.05$.

Table 1 gives regression statistics, runs test results and Ryan-Joiner results for the 1996 storm event. Table 2 gives the results of F tests relative to the power curve regression.

Table 1: Statistics 1996 subset with n=12.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	223481.406	0.921561	4	-0.603	yes	0.5789	yes
Quad	2812.976	0.998255	2	-1.809	no	0.8714	yes
Cube	19.156	0.999988	7	1.206	yes	0.8918	yes
Quad PC	3218.748	0.997957	3	-1.206	yes	0.7798	yes

Table 2: F test results 1996 subset with n=12.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	79.45	10	9	Quad
PC vs. Cube:	11666.63	10	8	Cube
PC vs. Quad PC:	69.43	10	8	Quad PC

Table 3 gives regression statistics, runs test results and Ryan-Joiner results for the 1997 storm event. Table 4 gives the results of F tests relative to the power curve regression.

Table 3: Statistics 1997 subset with n=11.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	650053.337	0.891435	2	-1.581	no	0.6335	yes
Quad	4267.045	0.998576	5	0.316	yes	0.9514	no
Cube	1843.745	0.999385	4	-0.316	yes	0.8845	yes
Quad PC	12275.057	0.995810	3	-0.949	yes	0.8734	yes

Table 4: F test results 1997 subset with n=11.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	152.34	9	8	Quad
PC vs. Cube:	352.57	9	7	Cube
PC vs. Quad PC:	52.96	9	7	Quad PC

Table 5 gives regression statistics, runs test results and Ryan-Joiner results for the 1998 storm event. Table 6 gives the results of F tests relative to the power curve regression.

Table 5: Statistics for 1998 subset with n=20.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	109.245	0.969468	6	-1.376	yes	0.7405	yes
Quad	6.501	0.997861	5	-1.835	no	0.9586	no
Cube	2.766	0.999091	7	-0.918	yes	0.9798	no
Quad PC	14.434	0.995380	5	-1.835	no	0.9141	yes

Table 6: F test results for 1998 subset with n=20.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	16.80	18	17	Quad
PC vs. Cube:	39.49	18	16	Cube
PC vs. Quad PC:	7.57	18	16	Quad PC

Table 7 gives regression statistics, runs test results and Ryan-Joiner results for the 1999 storm event. Table 8 gives the results of F tests relative to the power curve regression.

Table 7: Statistics for 1999 subset with n=21.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	2358.625	0.887895	3	-2.907	no	0.6583	yes
Quad	425.191	0.969380	8	-0.671	yes	0.7837	yes
Cube	327.872	0.976553	13	1.565	yes	0.8428	yes
Quad PC	319.266	0.976805	15	2.460	yes	0.8156	yes

Table 8: F test results for 1999 subset with n=21.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	5.55	19	18	Quad
PC vs. Cube:	7.19	19	17	Cube
PC vs. Quad PC:	7.39	19	17	Quad PC

Table 9 gives regression statistics, runs test results and Ryan-Joiner results for the 1996 storm event. Table 10 gives the results of F tests relative to the power curve regression.

Table 11 gives regression statistics, runs test results and Ryan-Joiner results for the 2005b storm event. Table 12 gives the results of F tests relative to the power curve regression.

Table 9: Statistics for 2005a with n=46.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	4.198	0.993897	6	-4.770	no	0.9868	no
Quad	1.784	0.997227	13	-2.683	no	0.9715	no
Cube	1.460	0.997733	15	-2.087	no	0.9699	yes
Quad PC	1.882	0.997059	13	-2.683	no	0.9719	no

Table 10: F test results for 2005a with n=46.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	2.35	44	43	Quad
PC vs. Cube:	2.88	44	42	Cube
PC vs. Quad PC:	2.23	44	42	Quad PC

Table 11: Statistics for 2005b with n=26.

Model	SSR	r^2	R	Z_R	Pass Runs?	NSCORE r	Reject Normality?
PC	6.069	0.998426	6	-2.400	no	0.5405	yes
Quad	1.028	0.999712	3	-3.600	no	0.9762	no
Cube	0.041	0.999988	12	0.000	yes	0.9624	no
Quad PC	0.072	0.999980	13	0.400	yes	0.9688	no

Table 12: F test results for 2005b with n=26.

Comparison	f	df 1	df 2	Which Better?
PC vs. Quad:	5.90	24	23	Quad
PC vs. Cube:	147.59	24	22	Cube
PC vs. Quad PC:	83.95	24	22	Quad PC

4 Discussion

The 1996 data subset contained 12 observations with 10 low flow measurements and 2 high flow measurements. The power curve (PC) model was outperformed by all other models in terms of SSR, runs test, r^2 value and the F test. The cubic model (Cube) was enormously better than the PC model in terms of the F test in particular (f=11666.63). None of the models had normally distributed residuals which suggests that some structure in the data is not explained by these models. The PC model is a highly lacking model compared to the others for this subset of data. Graphically the PC model deviated greatly at high flows where the other models match the data well (Figure 4). This is because of the property of the logarithm function in reducing larger values more than smaller values. None of the models provide a completely desirable fit to the 1996 subset of data.

The 1997 subset contained 11 observations with 9 medium-low flow measurements and two very high flow measurements. In the 1997 subset power curve (PC) model was outperformed by all other models in terms of SSR, runs test, r^2 value. All other models were statistically better than the PC model by the F test. The PC model did not pass the runs test where all other models did. The Cube model had the lowest SSR value but did not have normally distributed residuals. The Quad model provided the most desirable fit to the 1997 subset of data. Graphically the PC fits grossly overestimated the highest flow reading where all the other models fit relatively well (Figure 5).

The 1998 data subset contains 20 closely grouped observations with no very low or very high flow measurements. Once again, all models showed a much lower SSR value and a higher r^2 value than the PC model. The PC model did pass the runs test and failed to have normally distributed residuals. The PC model had a relatively high r^2 value but was still shown to be a insufficient model by the F test. The Cube model provided the most desirable fit to the 1998 subset. Graphically, all models fit relatively well but the PC model again overestimates the highest flow measurement (Figure 7). The PC model preformed better on the 1999 than on the previous two subsets where the data was more widely spaced.

The 1999 data subset contains 21 closely grouped observations with no high flow measurements. The PC model was again outperformed in all regards. None of the models had normally distributed residuals so none of the provided a completely desirable fit. Graphically, all the models provide a reasonably good fit to the data, but the PC model again over predicts the highest flow measurement (Figure 6). All of the models other than the PC model provided a competent fit to this subset.

The 2005a subset contains 46 observations with no high flow measurements. All models showed a low SSR and high r^2 value but PC model showed the highest SSR value. The PC model was outperformed by all other models on the F test. The PC model also did not have normally distributed residuals. Due to the high number of data points no model passed the runs test. The PC model preformed better on this subset which contained the most points. Graphically, all models fit very well but the PC model was still statistically insufficient (Figure 8). The Quad or Quad PC model provided the most competent fits to this subset.

The 2005b subset contains 26 observations with no high flow measurements but one that is relatively higher than all the others. All the models showed a very high r^2 value and the PC model again showed the highest SSR value. The PC model did not pass the runs test or have normally distributed residuals. The PC model was outperformed by all other models on the F test and also did not exhibit normally distributed residuals. Graphically, all models preformed similarly but again the PC model overestimated the highest flow measurement. The Quad PC and the Cube models provided desirable fits to the data.

Overall, the PC model was outperformed on every subset of data. Due to the logarithm function, the PC model tended to overestimate the highest flow measurement in a subset which was the main source of error for the model. The PC model also had the highest frequency of non-normal residuals. Despite its comparably poor fits, the PC model did follow the general trend of the data. The PC model is the only model used with only two parameters, which might be an advantage in some situations. Also it is relatively easy to carry out by hand. The other models were not compared to each other in terms of performance but one of the three could be chosen as the best fitting model for each subset. It would be wise to test the fit of all three models in order to asses the relative performance in any new set of data.

PC model tended to have a high r^2 that might be considered reasonable out of context of the other models. Using the r^2 value alone could be misleading in judging which model is better. F-tests, runs tests and Ryan-Joiner tests are important also for evaluating the relative performance of each model.

5 Conclusion

When developing rating curves based on existing data the traditional power curve model may not be the best choice. Depending on the situation, a polynomial or quadratic power curve model may provide a fit that statistically better than the power curve model. The power curve model tends to perform better when data points are closely spaced and no data points lie far form the main cluster. The traditional power curve method tends to overestimate high flows. Also using the r^2 value alone could be misleading in judging which model is better. F-tests, runs tests and Ryan-Joiner tests are important also for evaluating the relative performance of each model. The choice of which model to use depends on the situation of the data.

References

Charles Chamberlain. Gauss-markov theorem. Technical report, Humboldt State University, 2007a.

Charles Chamberlain. Runs test for independence. Technical report, Humboldt State University, 2007b.

Charles Chamberlain. Normal probability plots and the ryan-joiner test. Technical report, Humboldt State University, 2007c.

Charles Chamberlain. Model evaluation. Technical report, Humboldt State University, 2007d.

Agriculture Department. Field measurement of soil erosion and runoff. Technical report, Food and Agriculture Department of the United Nations, 2007. URL `http://www.fao.org/docrep/T0848E/t0848e-09.htm`.

Jay L. Devore. *Probability and Statistics for Engineering and the Sciences*. Duxbury Thomson Learning, 2000.

Torsten Dose, Gerd Morgenschweis, and Torsten Schlurmann. Extrapolating stage-discharge relationships by numerical modeling. Technical report, University of Wuppertal, Pauluskirchst, 2001.

Geology.com. What is a gaging station? stream discharge monitoring. Technical report, geology.com, 2007. URL `http://geology.com/articles/gaging-station.shtml`.

NIST/SEMATECH. Are the model residuals well-behaved? Technical report, Engineering Statistics Handbook, 2007. URL `http://www.itl.nist.gov/div898/handbook/pri/section2/pri24.htm`.

John Pitlick. Stream discharge measurement: Float and velocity-area methods. Technical report, University of Colorado, 2002.

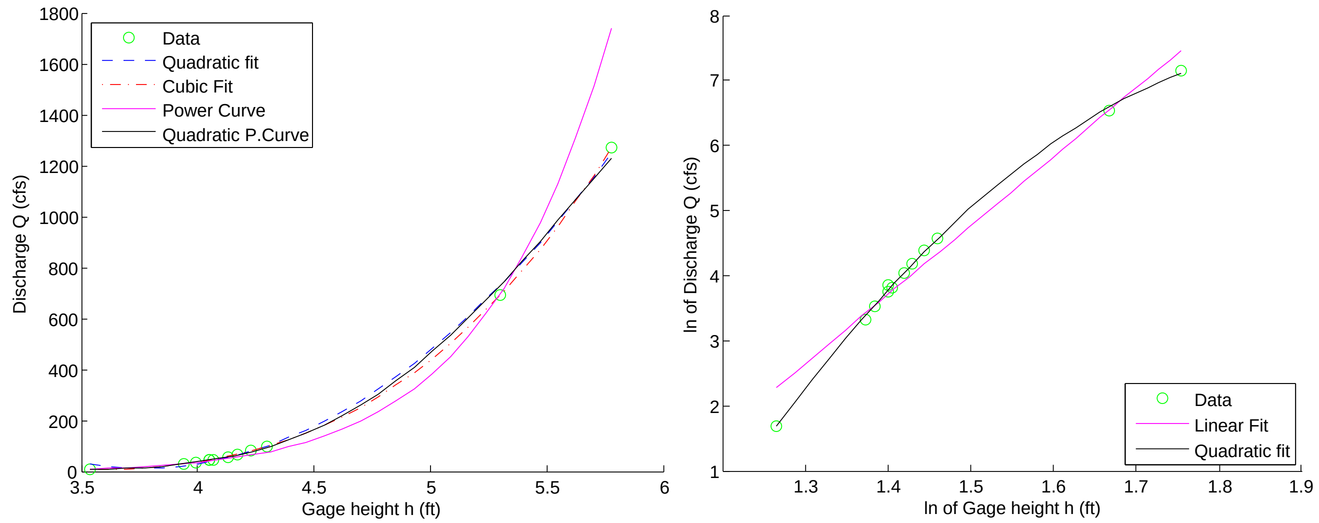
A. R. Schmidt and B. C. Yen. Stage-discharge relationship in open channels. *Proceedings of the 2001 International Symposium on Environmental Hydraulics*, 2001.

Chandrasekaran Sivapragasam and Nitin Muttil. Discharge rating curve extension - a new approach. *Water Resources Management*, pages 506–520, 2005.

USGS. Usgs 11157500 tres pinos c nr tres pinos ca. Technical report, National Water Information System, 2007.

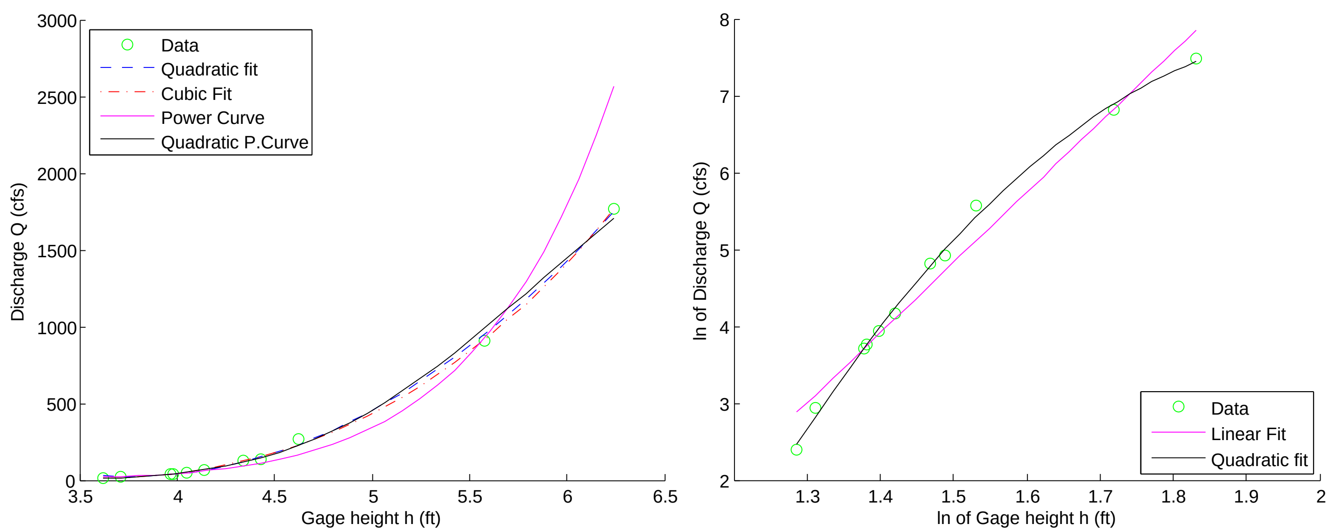
Various. Tres pinos, ca. Wikipedia, March 2007. URL `http://en.wikipedia.org/wiki/TresPinos2CCalifornia`.

Appendix A Regression Plots



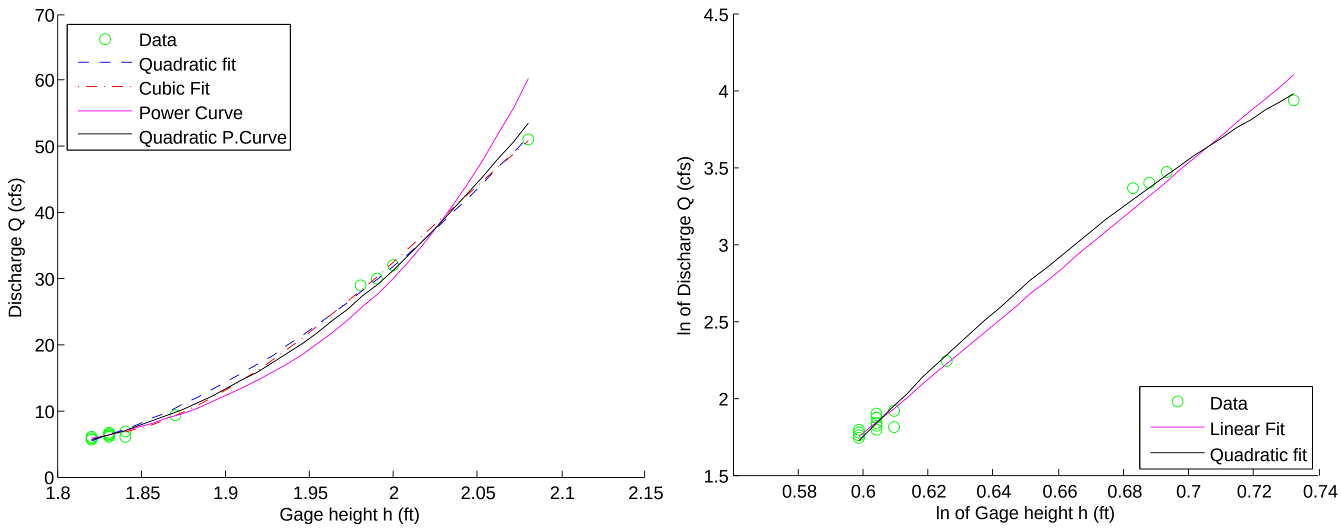
(a) Regressions (b) Linearized regressions

Figure 4: Graphs of 1996 Regressions.



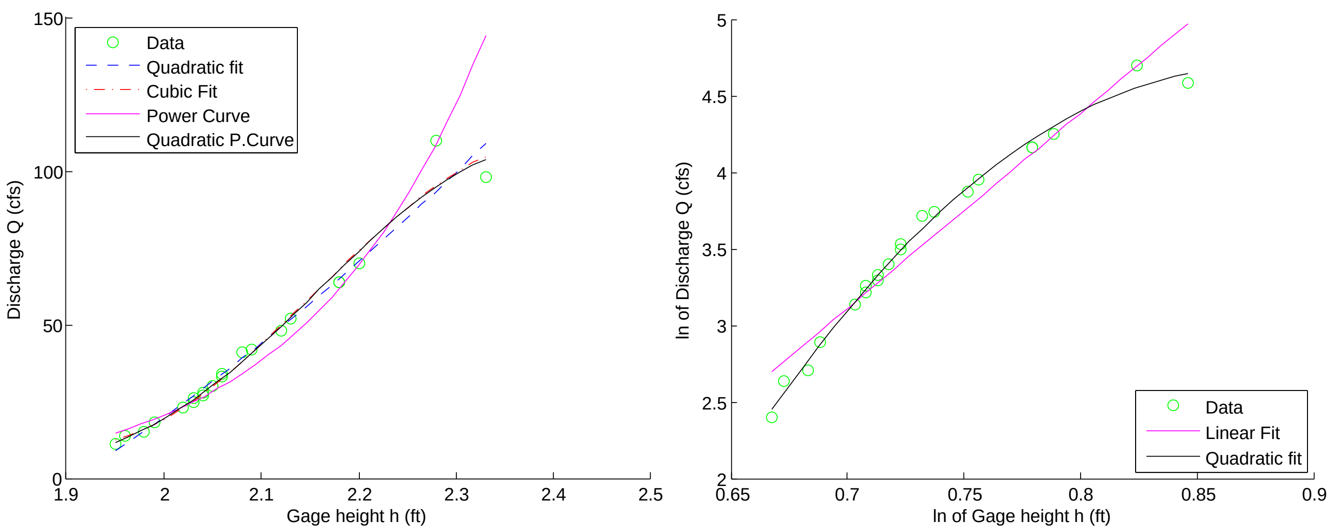
(a) Regressions (b) Linearized regressions

Figure 5: Graphs of 1997 Regressions.



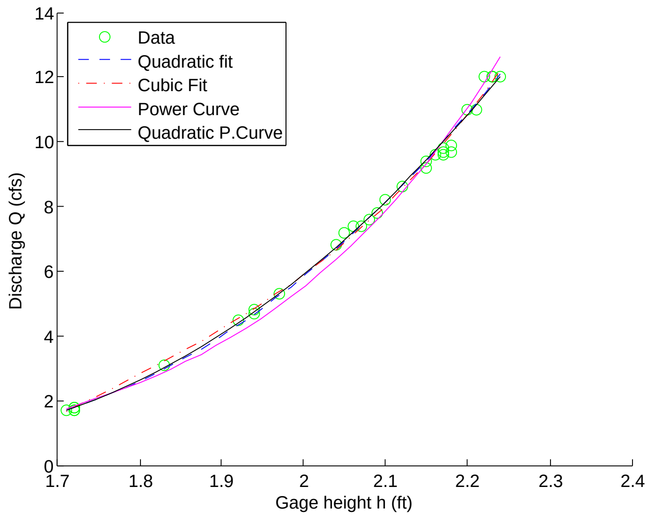
(a) Regressions (b) Linearized regressions

Figure 6: Graphs of 1998 Regressions.

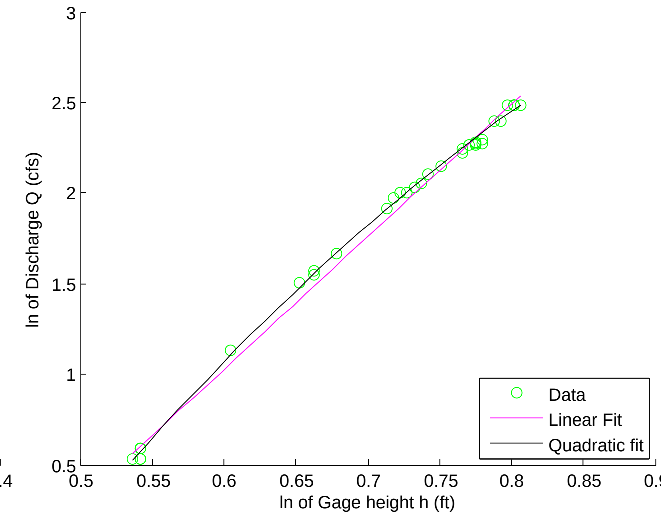


(a) Regressions (b) Linearized regressions

Figure 7: Graphs of 1999 Regressions.

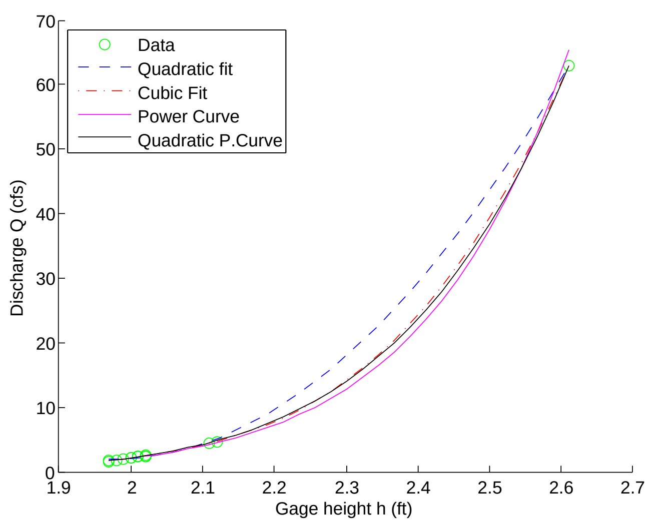


(a) Regressions

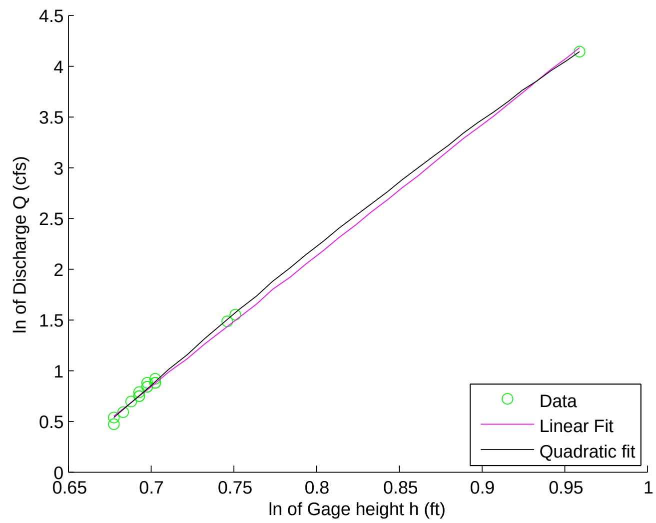


(b) Linearized regressions

Figure 8: Graphs of 2005a Regressions.



(a) Regressions



(b) Linearized regressions

Figure 9: Graphs of 2005b Regressions.

Appendix B Residual Plots

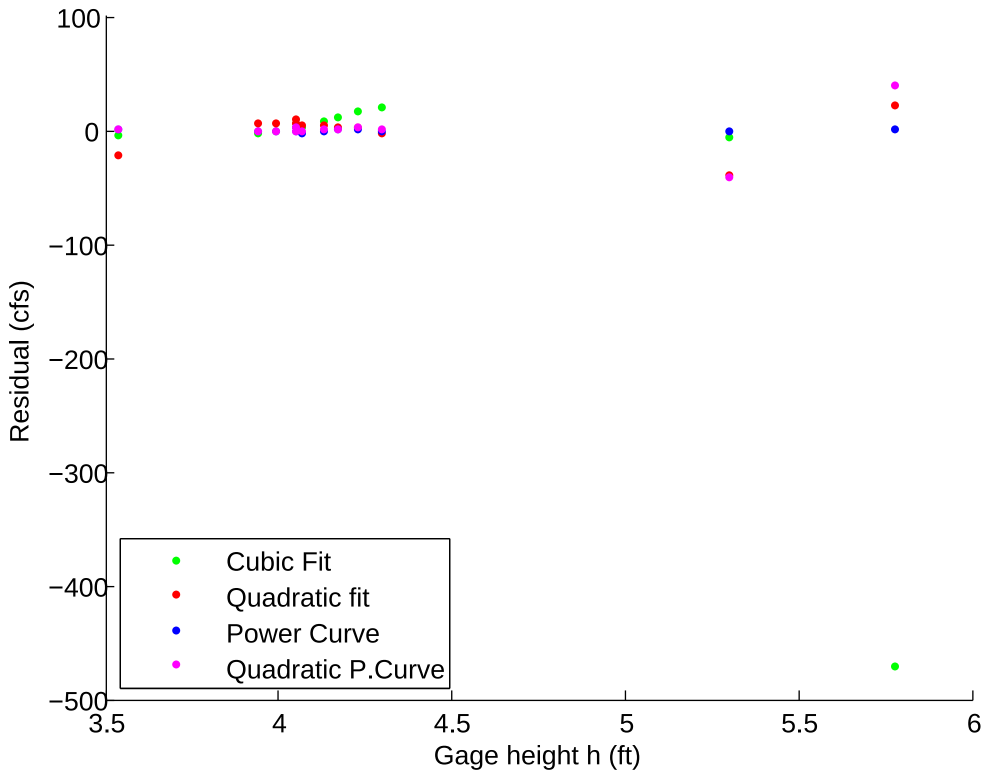


Figure 10: Plot of 1996 Residuals.

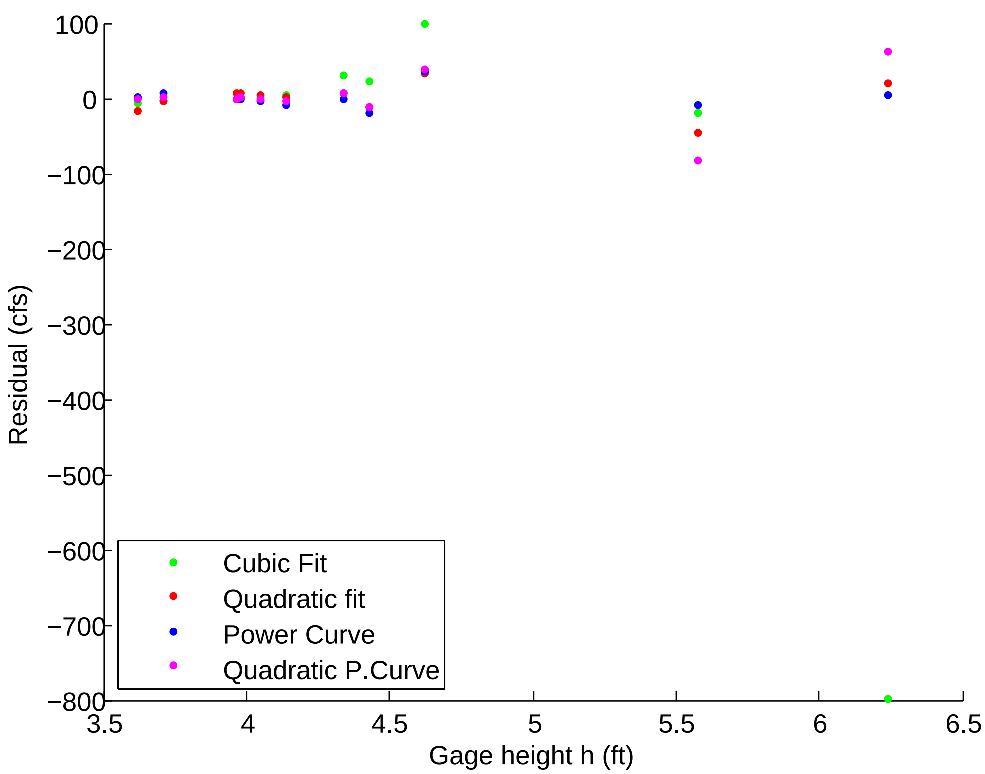


Figure 11: Plot of 1997 Residuals.

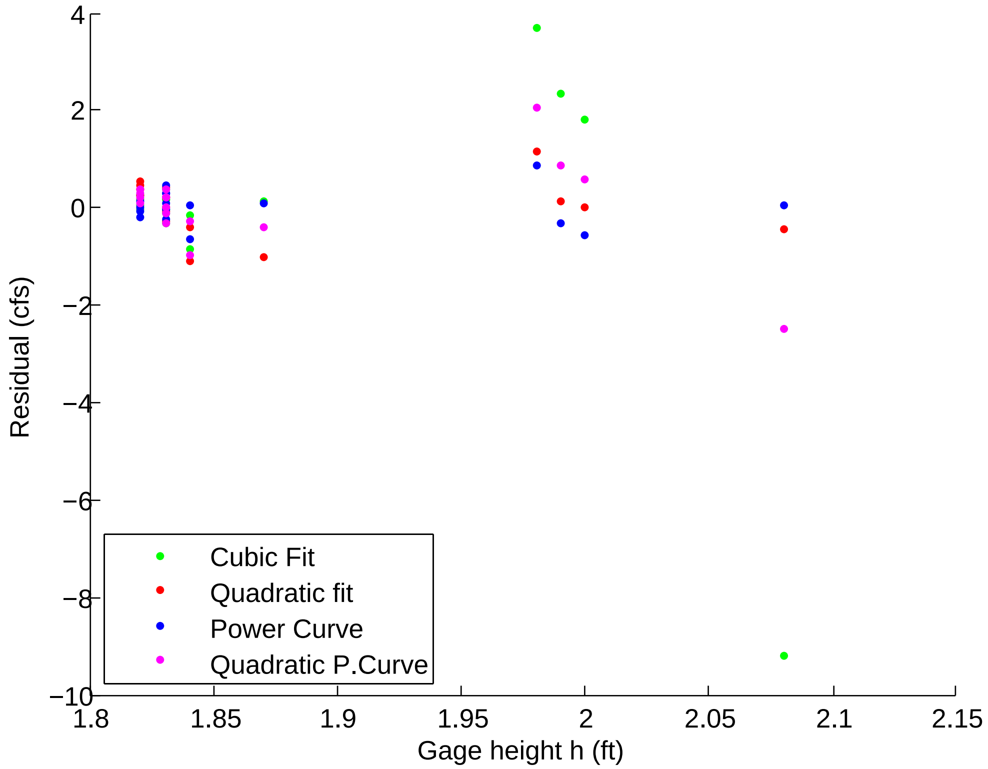


Figure 12: Plot of 1998 Residuals.

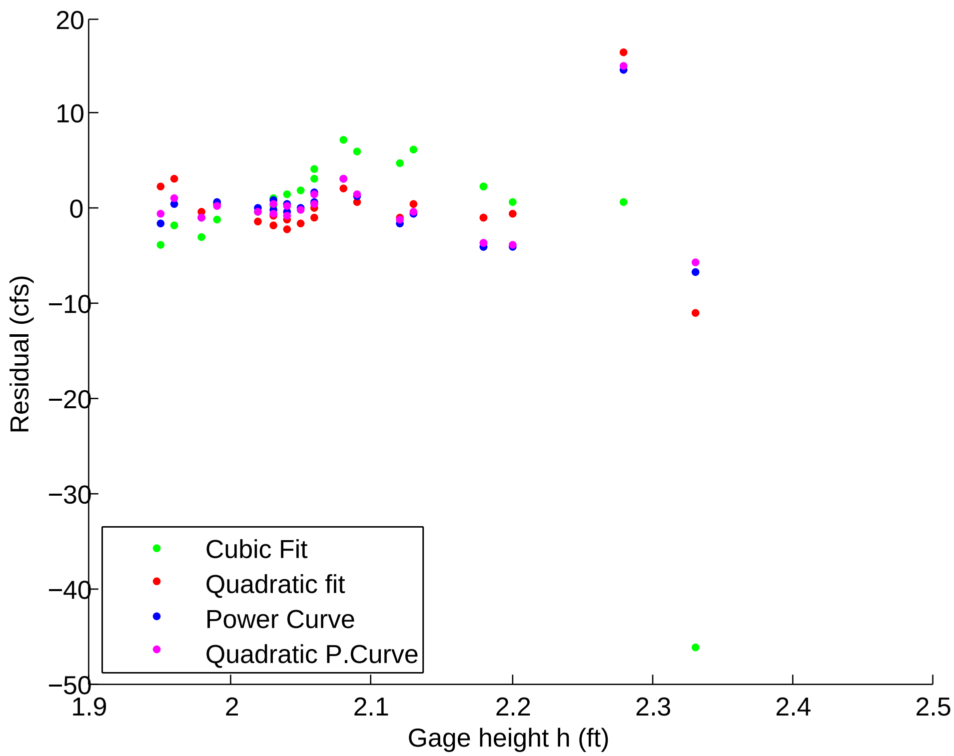


Figure 13: Plot of 1999 Residuals.

This was typeset with LaTeX

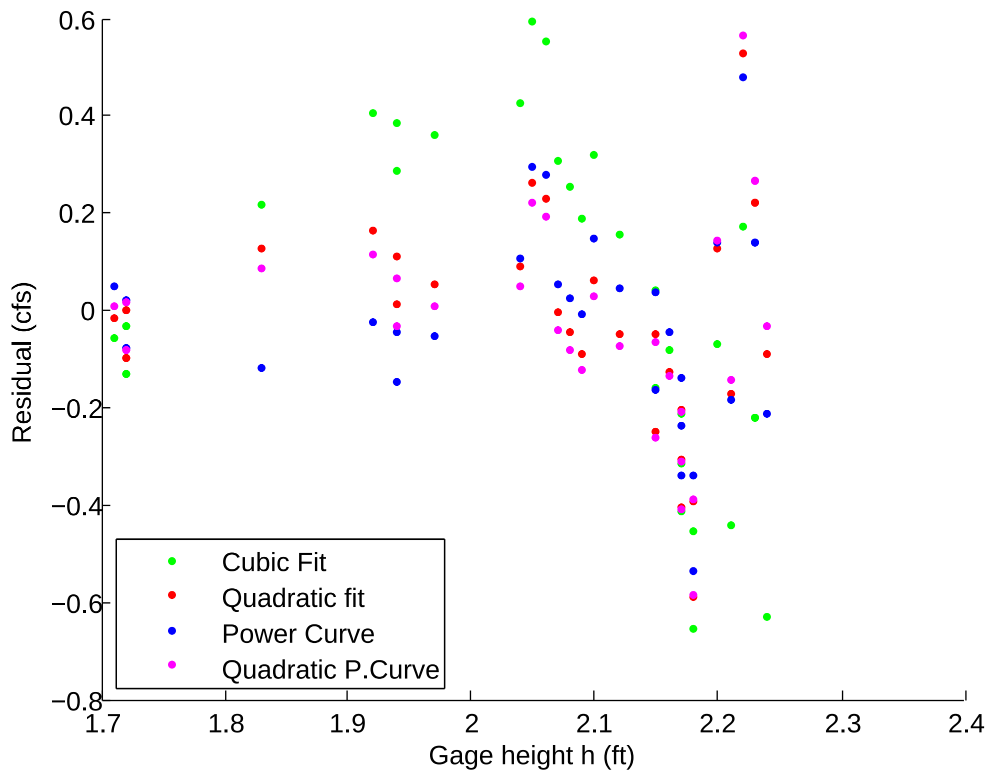


Figure 14: Plot of 2005a Residuals.

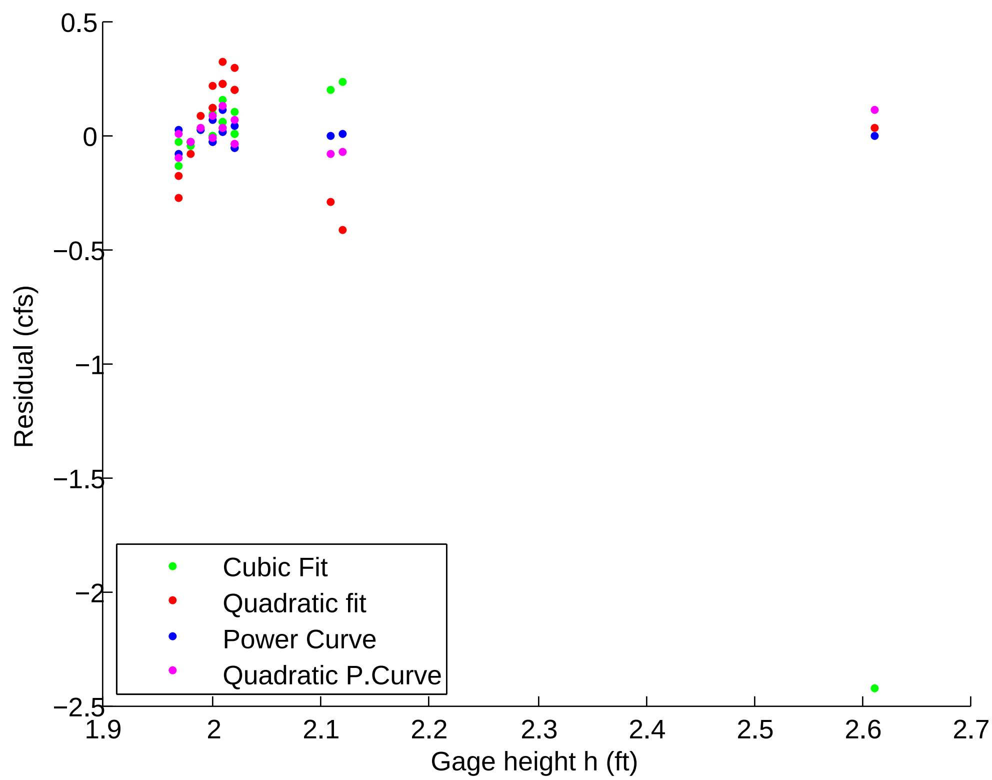


Figure 15: Plot of 2005b Residuals.