Why Do We Need Statistics to Detect Trends or Estimate Loads?

Statistical Techniques for Trend and Load Estimation
National Water Quality Monitoring Council
Chattanooga, TN, 17 May 2004

Tim Cohn

U.S. Geological Survey

Reston, Virginia

Some Questions

- Has water quality declined or improved in U.S. rivers?
- What is the magnitude of the trends?
- How large are nutrient loads entering Chesapeake Bay?
- Are the States in compliance with the Chesapeake Bay Agreement?

USGS

The Role of Statistics

- Uncertainty Abounds
  - Data
  - Models
  - Estimates

USGS

Role of Statistician

- Responses to Uncertainty
  - Eliminate uncertainty
  - Ignore uncertainty
  - Recognize and quantify uncertainty

USGS

The Problem

- **Estimate nutrient loads entering CB:**

$$L = \int_{t_a}^{t_b} L(t)dt$$

$$= \int_{t_a}^{t_b} KC(t)Q(t)dt$$

$$\cong \sum_i KC(t_i)Q(t_i)$$

USGS

The Problem (Cont'd)

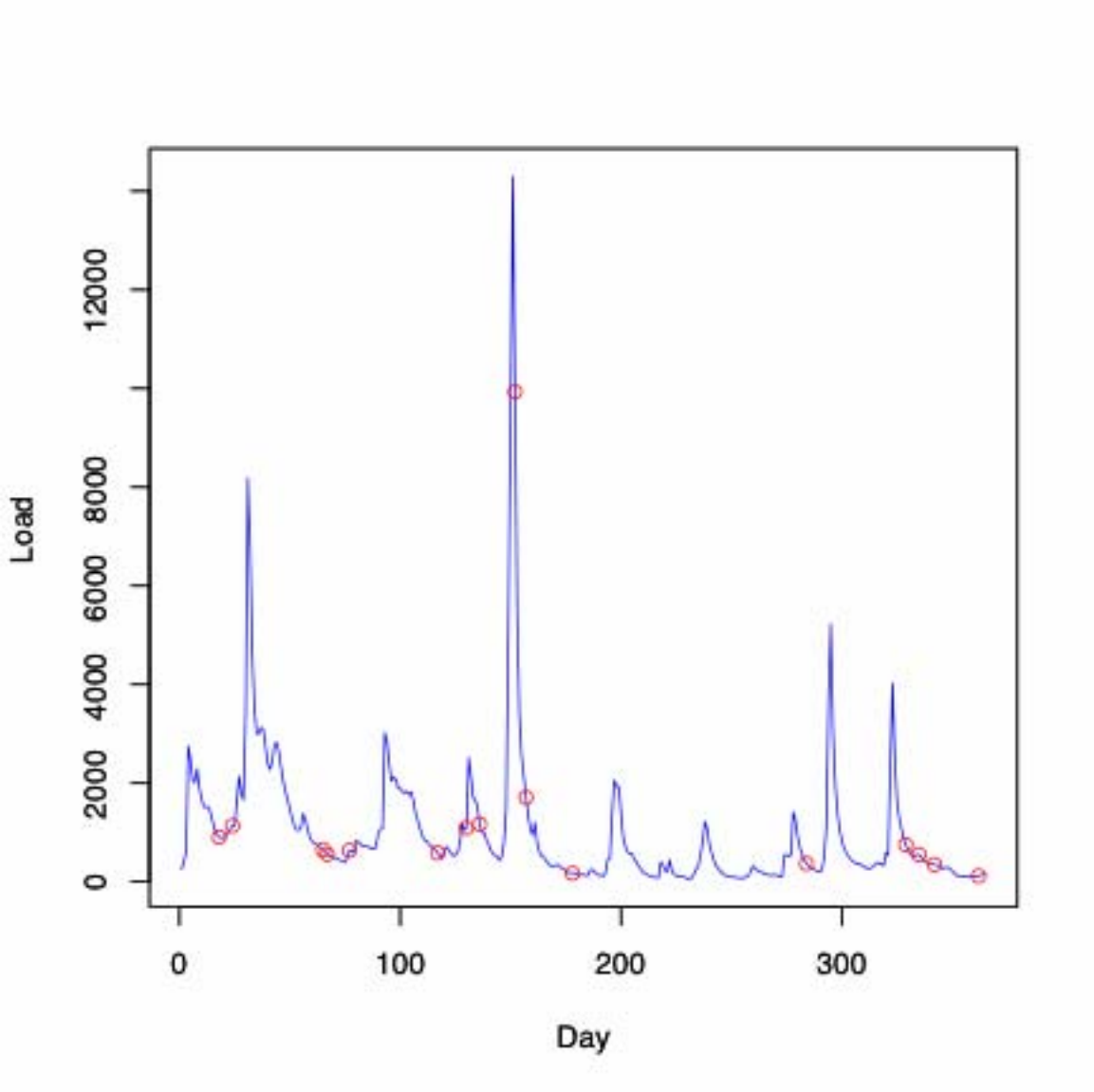
- Given:
  - A small sample of observed concentration values
  - A continuous record of discharge
  - Hydrological knowledge and judgment
  - Statistics

USGS

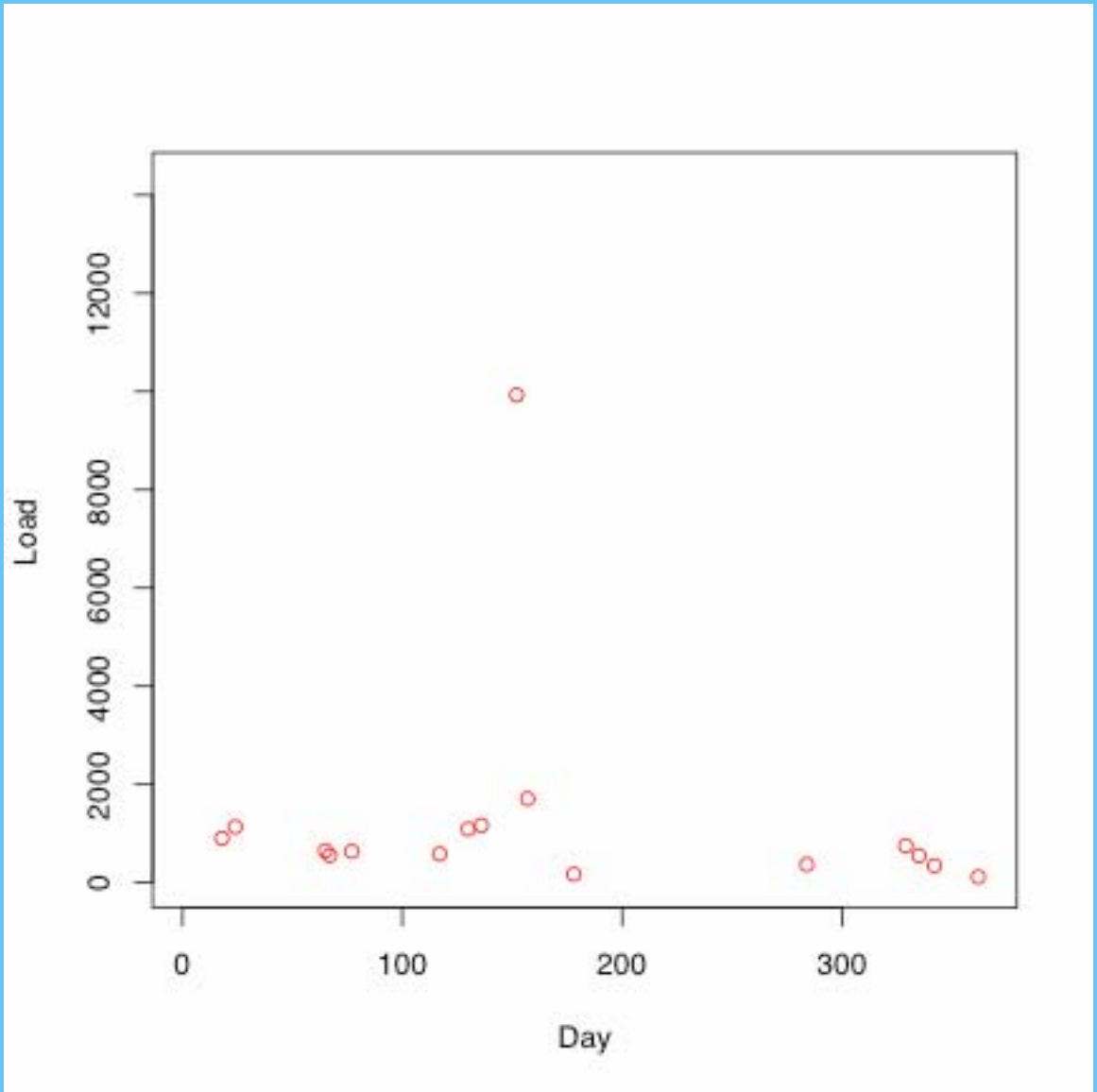
The Problem (Cont'd)

- Need to "fill gaps" in C record given
  - Continuous record of Q (and other cheaply measured surrogate variables)
  - A small sample of instantaneous observations of C
  - + … ???

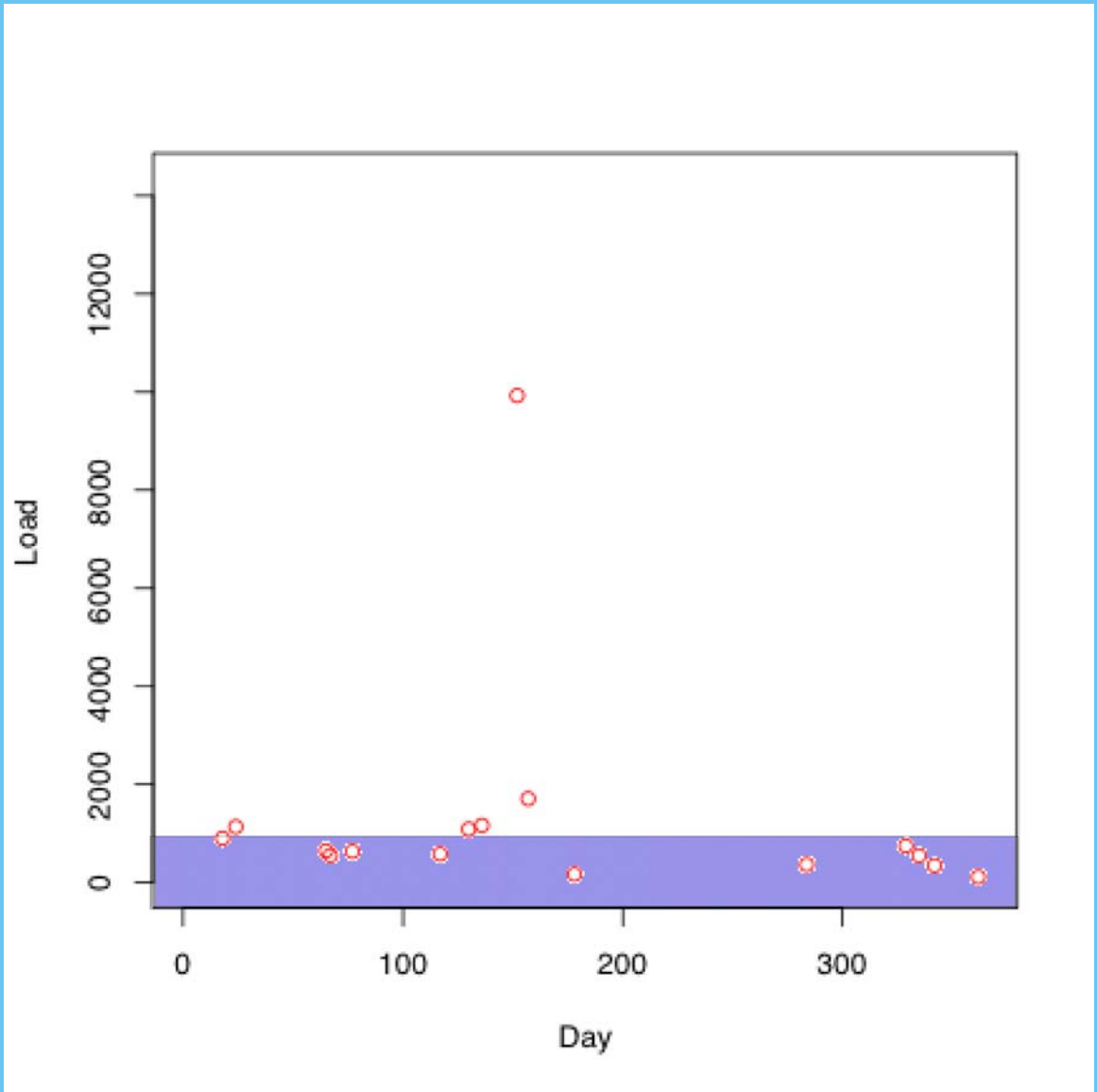
USGS



USGS



USGS



USGS

Why not just collect more data?

- Continuous measurement of C has been, until recently, costly or impossible
- Efficient estimates with known uncertainty may be obtained through modeling

USGS

What we need:

- An estimator…

USGS

…with desirable characteristics

- Statistical
  - Zero or low bias
  - Low (minimal) variance
  - Known uncertainty of estimates
- Practical
  - Low cost
  - Easy to understand, explain and justify
  - Permits design of sampling program

USGS

Load Estimation

Statistical Techniques for Trend and Load Estimation
National Water Quality Monitoring Council
Chattanooga, TN, 17 May 2004

Tim Cohn

U.S. Geological Survey

Reston, Virginia

Some Literature

- Cohn, T., "Recent Advances in Statistical Methods for the Estimation of Sediment and Nutrient Transport in Rivers," chapter 21 in *Contributions in Hydrology, US National Report to the IUGG*, pp. 1117-1124, 1995.
- Cohn, T., D. L. Caulder, E. J. Gilroy, L. D. Zynjuk, and R. M. Summers, "The Validity of a Simple Statistical Model for Estimating Fluvial Constituent Loads: An Empirical Study Involving Nutrient Loads Entering Chesapeake Bay, "*Water Resources Research* , 28(9), pp. 2353-2364, 1992.
- Cohn, T., E. J. Gilroy and W. G. Baier, "Estimating Fluvial Transport of Trace Constituents Using a Regression Model with Data Subject to Censoring," pp. 142-151, Proceedings of the Joint Statistical Meeting, Boston, August 9-13, 1992.
- Gilroy, E. J., R. M. Hirsch, and T. Cohn, "Mean Square Error of Regression-Based Constituent Transport Estimates," *Water Resources Research*, 26(9), pp. 2069-2077, 1990.
- Cohn, T., L. L. DeLong, E. J. Gilroy, R. M.. Hirsch, and D. K. Wells, "Estimating Constituent Loads," *Water Resources Research* , 25(5), pp. 937-942, 1989.

USGS

Estimators to Consider

- L_A: 7 [tons]
- L_B: L_{-1} (Last Year's Estimate)
- L_C: $365\ K\ C_1\ Q_1$
- L_D: $\Sigma\ K\ C_i\ Q_i$
- L_E: $\Sigma\ K\ C(Q_i,t_i)\ Q_i$

USGS

A Practical Approach to Load Estimation

1. Get to know your process and data
2. Develop a model relating concentrations to predictor variables
3. Use model to estimate loads (or log[loads]); being careful of bias
4. Characterize uncertainties

USGS

Step 1: Processes and Data

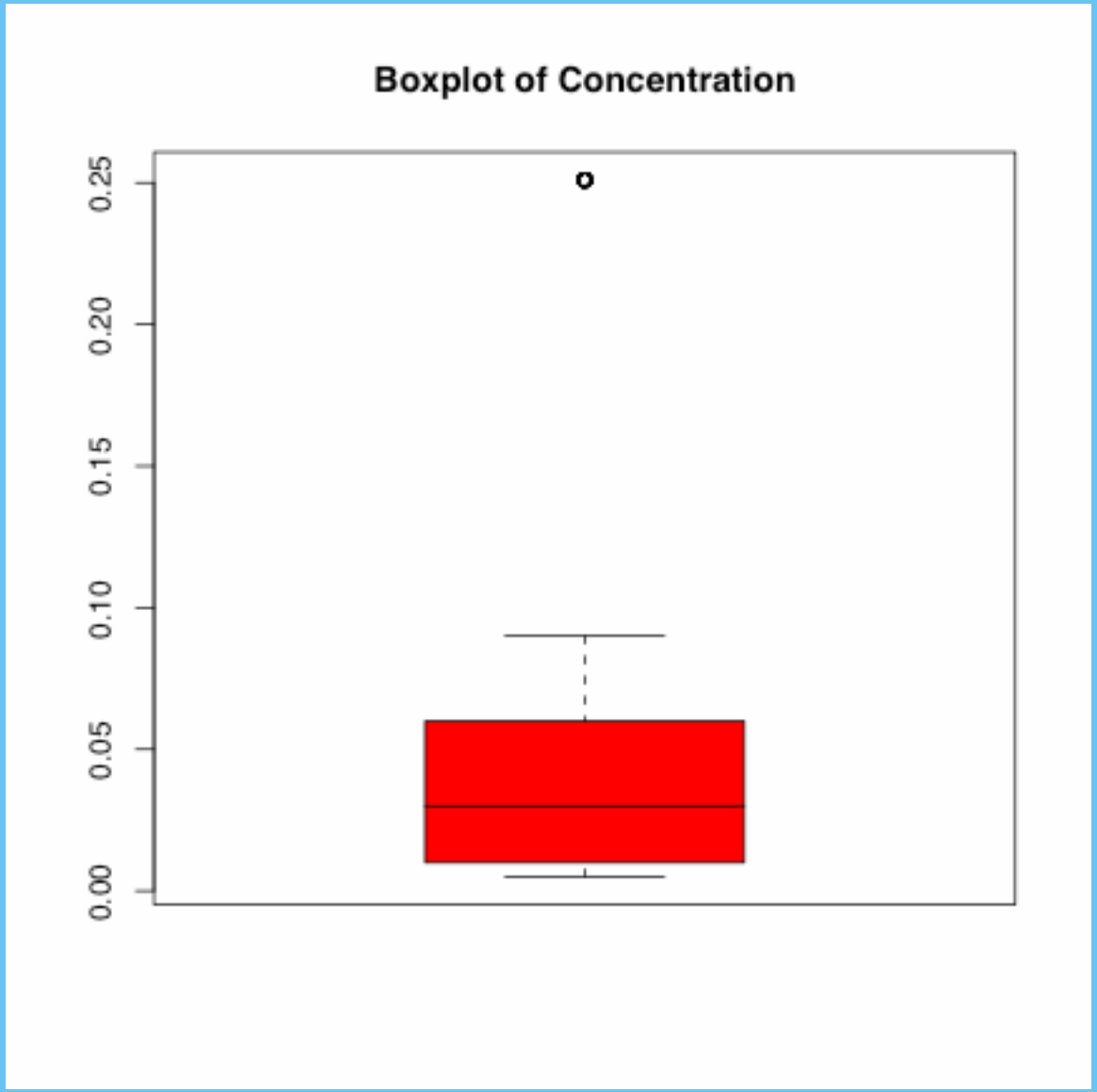
- Understand relevant hydrology
- Plot data!
  - Univariate Plots
    - Boxplots
    - Histograms
  - Bivariate Plots (Covariates)
    - Discharge: Q, Ln[Q]
    - Time
    - Boxplot by season, month
  - Whatever else you can think of

USGS

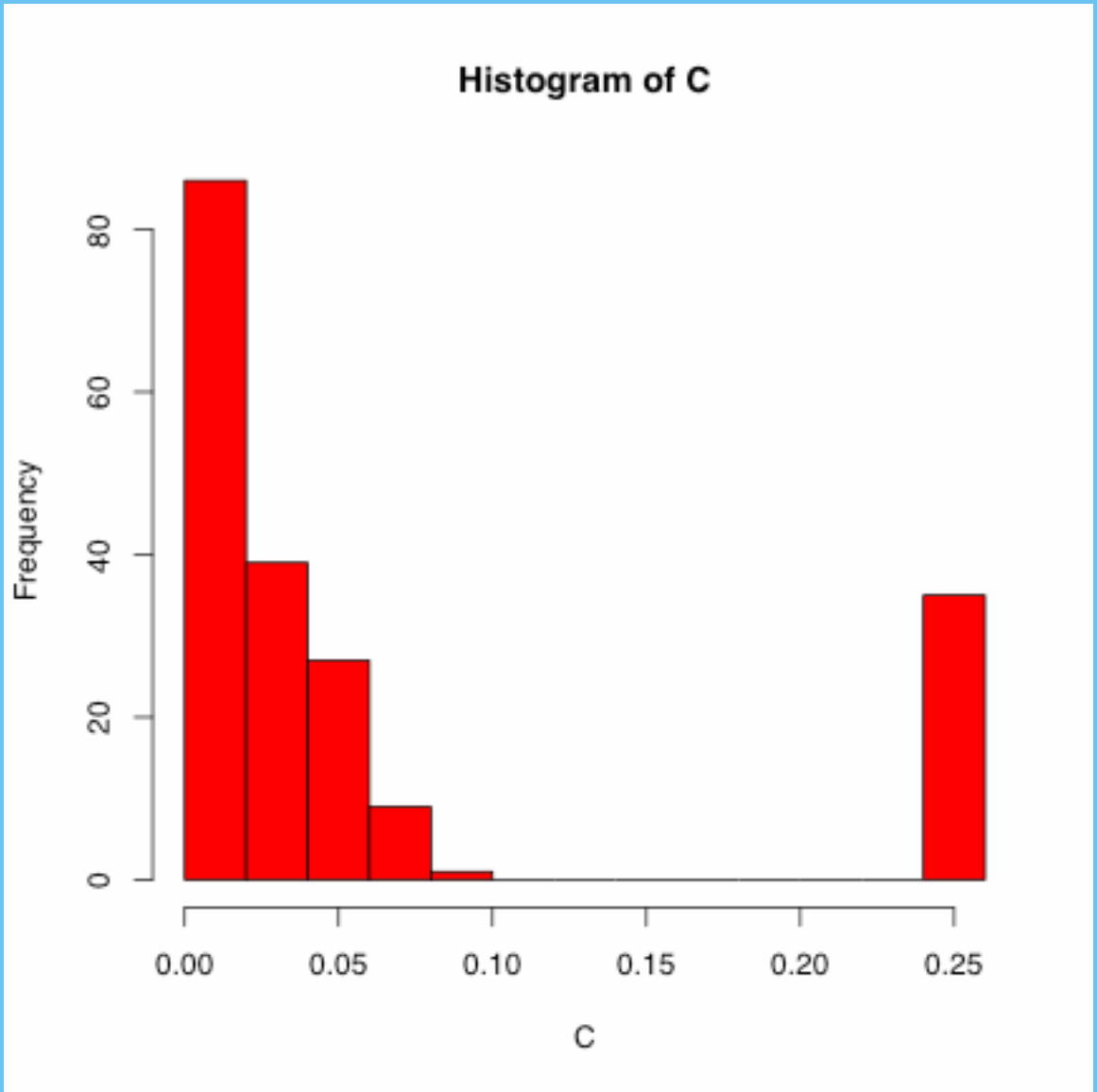
Some Commands

```
boxplot(C,main="Boxplot of Concentration",col="red")

hist(C,col="red")

boxplot(C,main="Boxplot of Concentration",col="red",log="y")

plot(time,C,main="C vs. Time",col="red",log="y")

plot(Q,C,main="C vs. Q",col="red",log="xy")

boxplot(split(C,MONTH),xlab="Month",ylab="C",main="C by
Month",col="red",log="y")
```

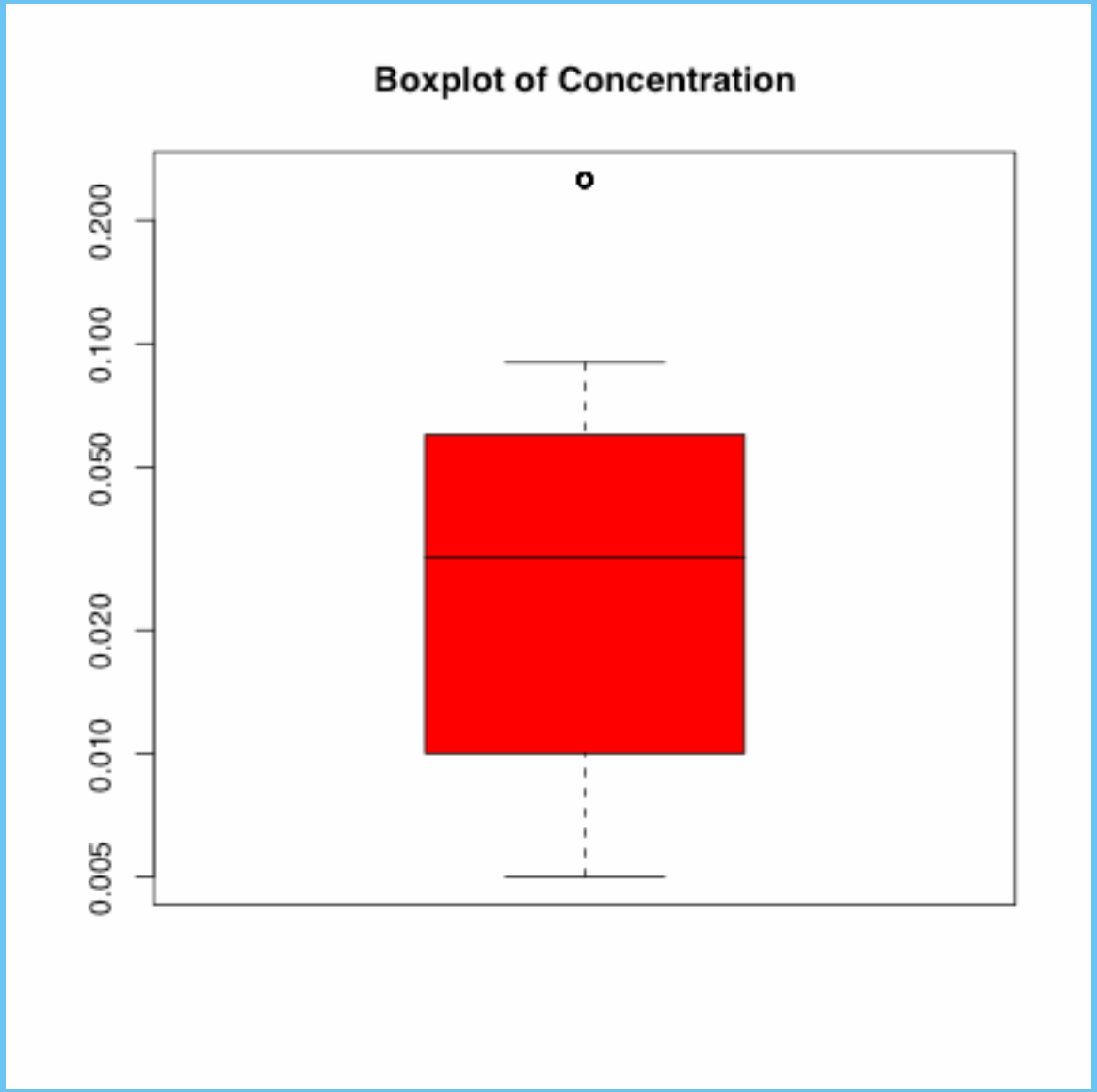
USGS



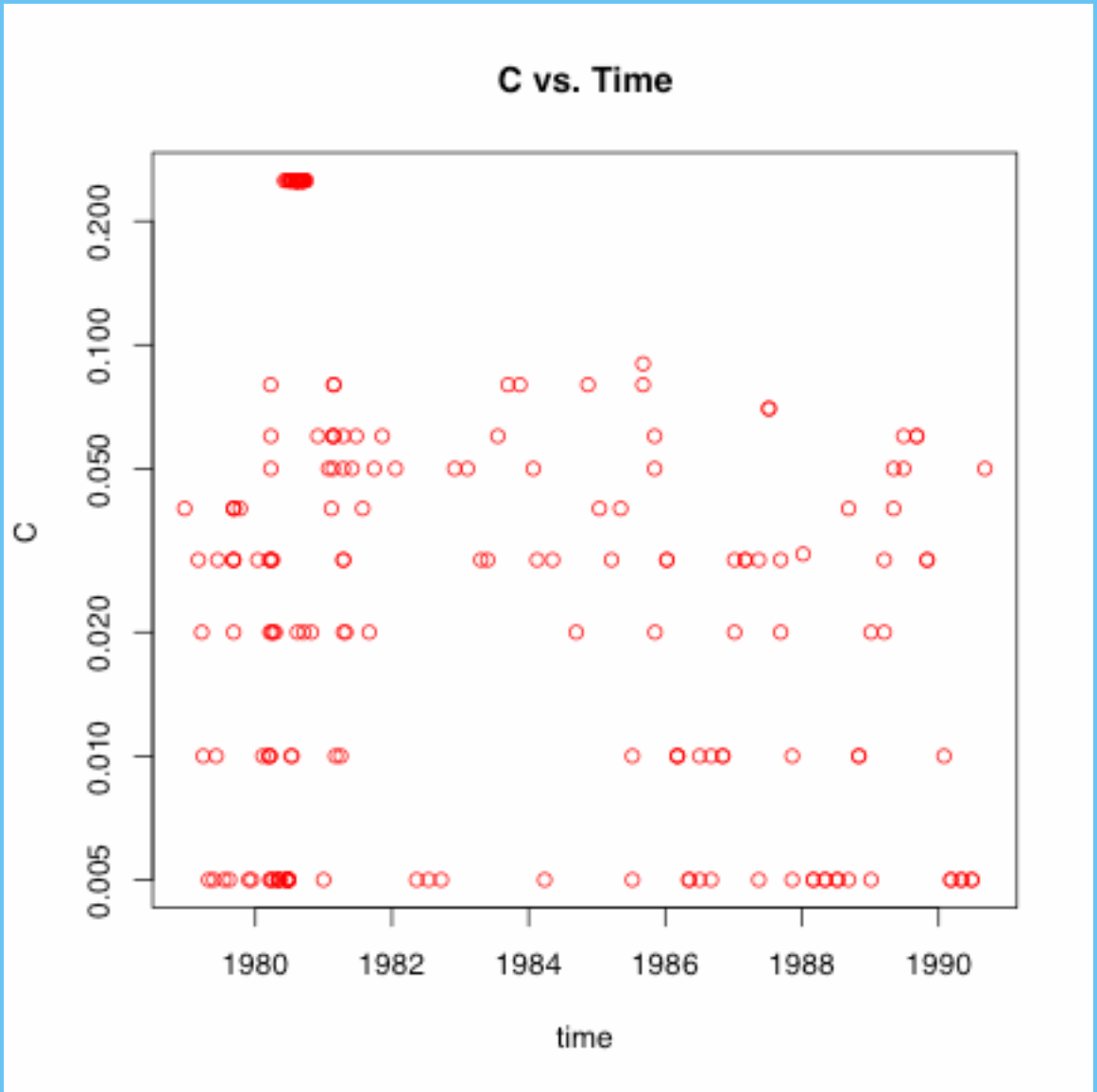
USGS



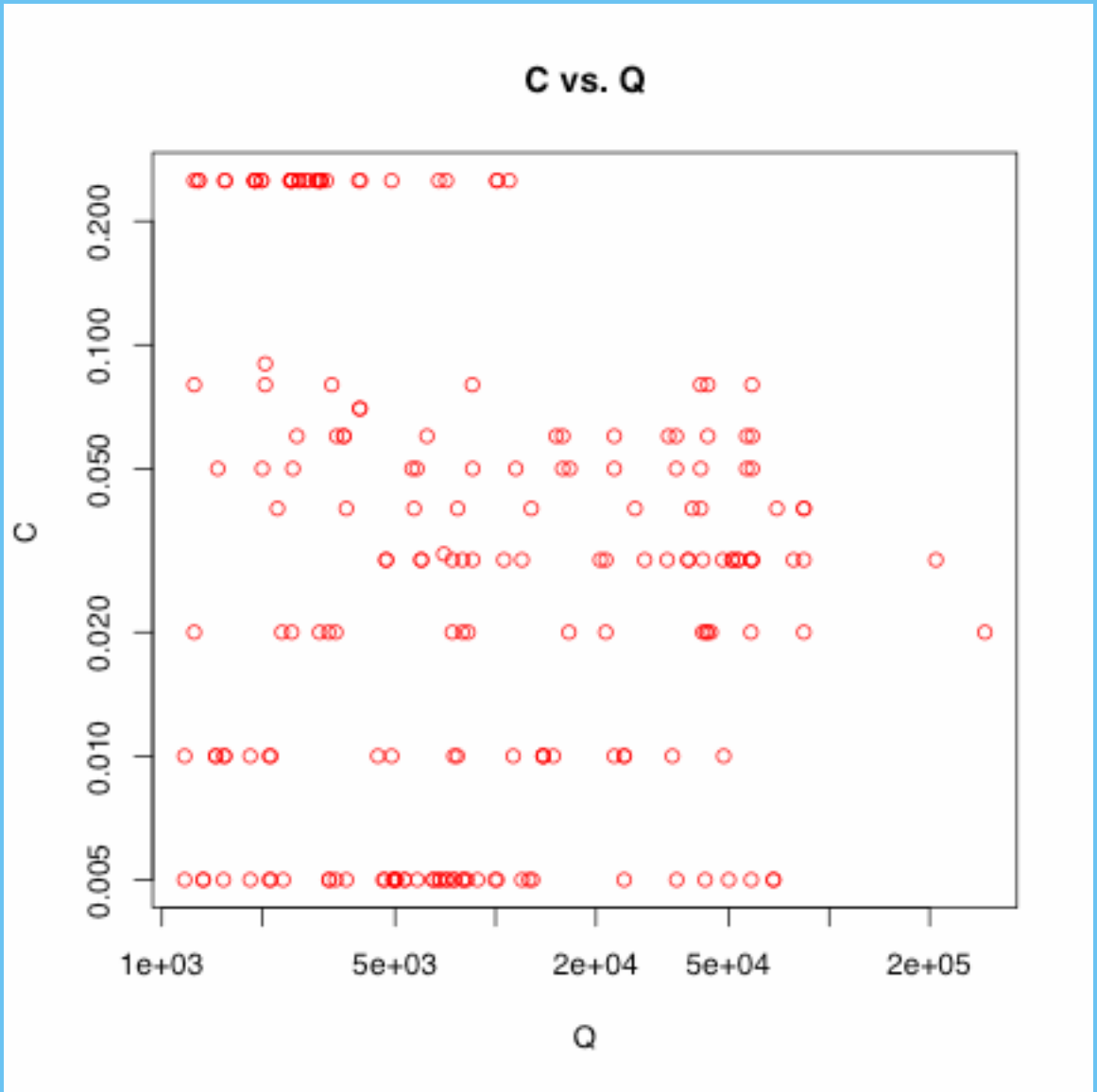
USGS



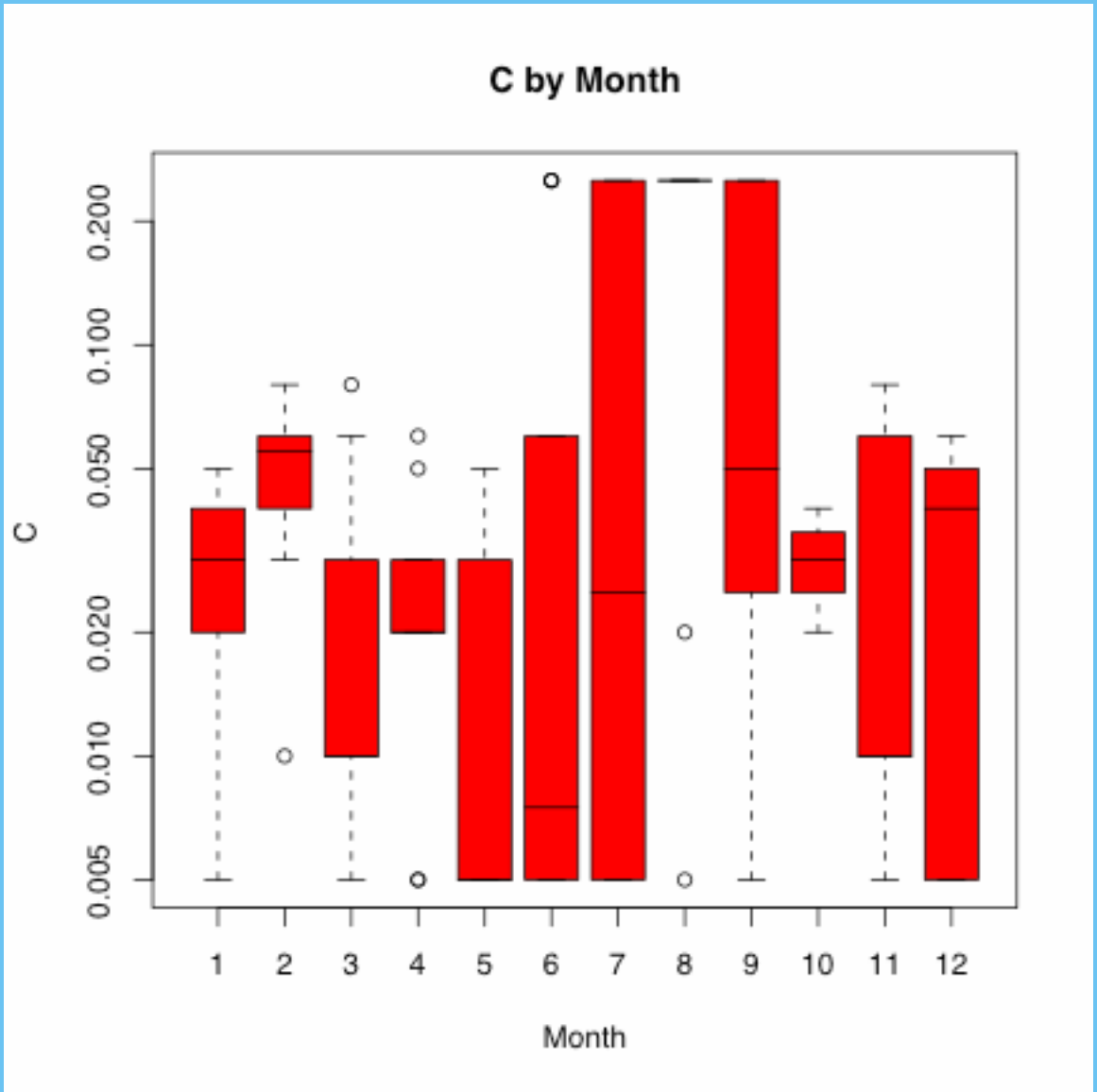
USGS


C vs. Time
C
0.005
0.010
0.020
0.050
0.100
0.200
1980
1982
1984
1986
1988
1990
time

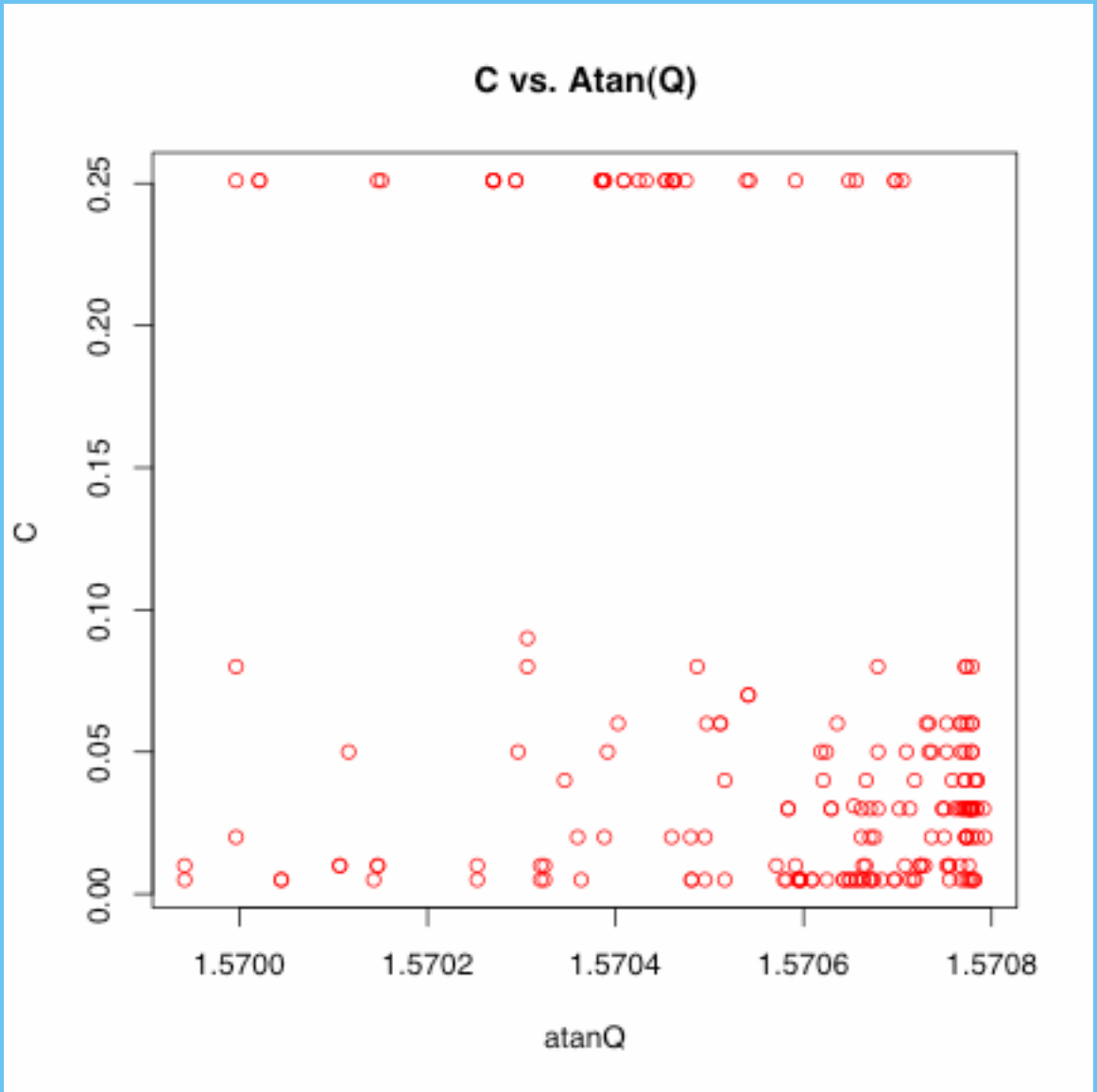
USGS


C vs. Q
C
0.005
0.010
0.020
0.050
0.100
0.200
1e+03
5e+03
2e+04
5e+04
2e+05
Q

USGS


C by Month
C
0.005
0.010
0.020
0.050
0.100
0.200
1
2
3
4
5
6
7
8
9
10
11
12
Month

USGS


C vs. Atan(Q)
C
0.00
0.05
0.10
0.15
0.20
0.25
1.5700
1.5702
1.5704
1.5706
1.5708
atanQ

USGS

How you *really* apply this…

- You do it yourself!

USGS

Step 2: Modeling

- Things to Consider
  - Linear model?
  - Fitting method?
    - Complete Data (OLS)
    - Censored data (Likelihood)
  - Transformation of response variable?
  - Predictor variables?

USGS

Some Guidelines for Model Building

1. Hydrological perspective
   1. Use an appropriate model
   2. Check all diagnostics, residuals, and fitted coefficient values
2. Statistical perspective:
   1. Explanatory variables: “All is fair” on the right-hand-side of the regression equation
   2. Response variable: Logarithmic transformation (or untransformed) is by far the easiest case to handle

USGS

The 7 Parameter Model

$$
\begin{aligned}
\ln[C] = \beta_0 & \\
& + \beta_1 \ln[Q/Q^*] + \beta_2 (\ln[Q/Q^*])^2 \\
& + \beta_3 (T - T^*) + \beta_4 (T - T^*)^2 \\
& + \beta_5 \sin(2\pi T) + \beta_6 \cos(2\pi T) \\
& + \varepsilon
\end{aligned}
$$

USGS

Flow Dependence

- C often depends on Q
- Relationship between ln[C] and ln[Q] may be approximated by polynomial:

$$\ln[C(Q)] = a_0 + a_1 \ln[Q] + a_2 \ln[Q]^2 + \ldots$$
$$= \beta_0 + \beta_1 (\ln[Q]-\ln[Q^*]) + \beta_2 (\ln[Q]-\ln[Q^*])^2$$
$$= \beta_0 + \beta_1 (\ln[Q/Q^*]) + \beta_2 (\ln[Q/Q^*])^2$$

- However, piecewise linear relationship may be necessary

USGS

Time Trends

- Many types
  - Step trends?
  - Gradual shifts?
  - Linear?
- Most can be approximated by polynomial

$$\ln[C(t)] = a_0 + a_1 t + a_2 t^2 + \ldots$$

$$= \beta_0 + \beta_3 (t-t^*) + \beta_4 (t-t^*)^2 \quad (\pm)$$

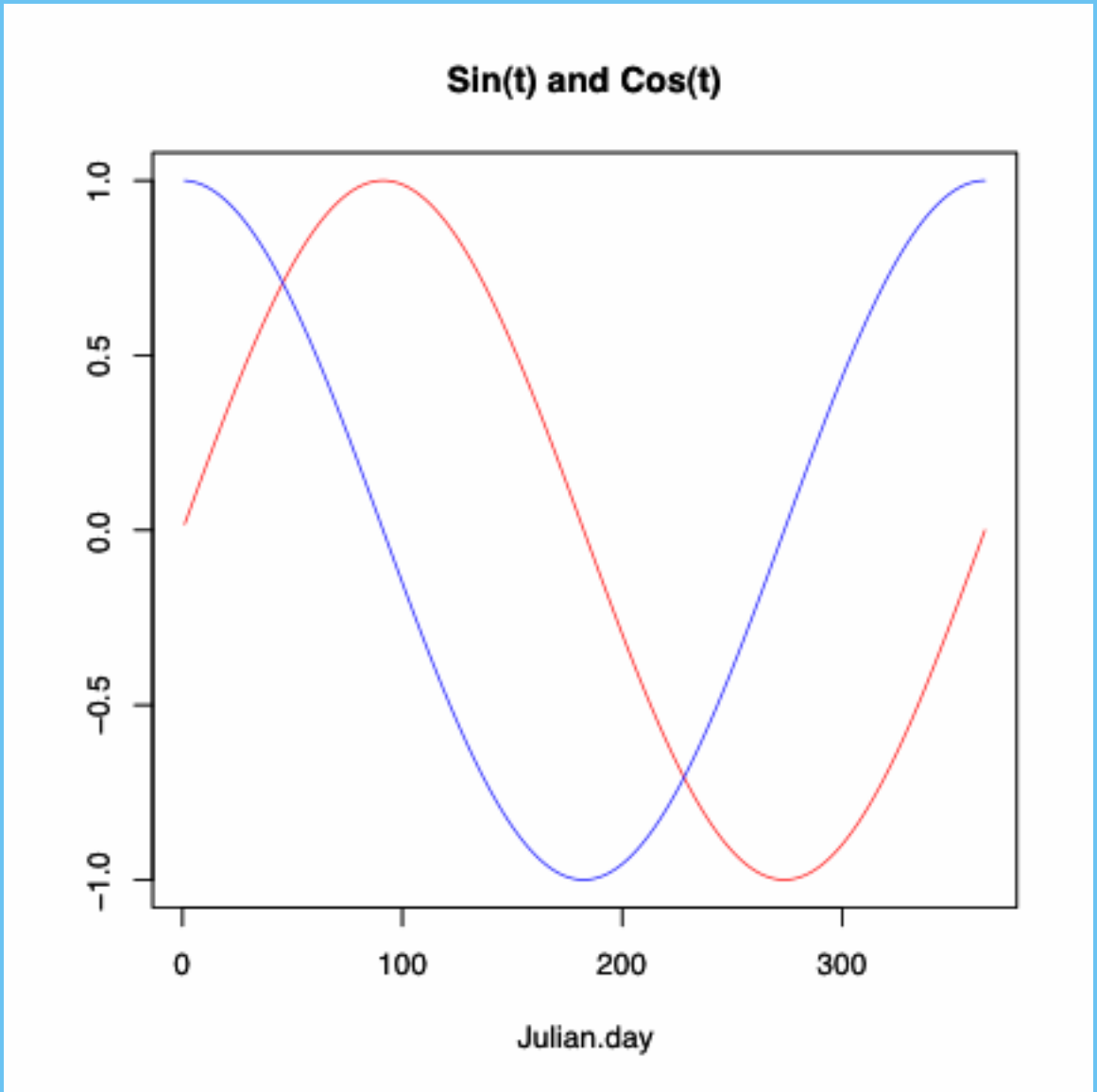
- "Moving window" may help

USGS

Seasonality by Fourier Series

- Linear model that describes annual cycle
- Easily fit by linear regression or tobit
- Result is easy to interpret

USGS



USGS

Seasonality by Fourier Series

- For continuous periodic function f(t)

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \sin(2\pi kt) + b_k \cos(2\pi kt)$$

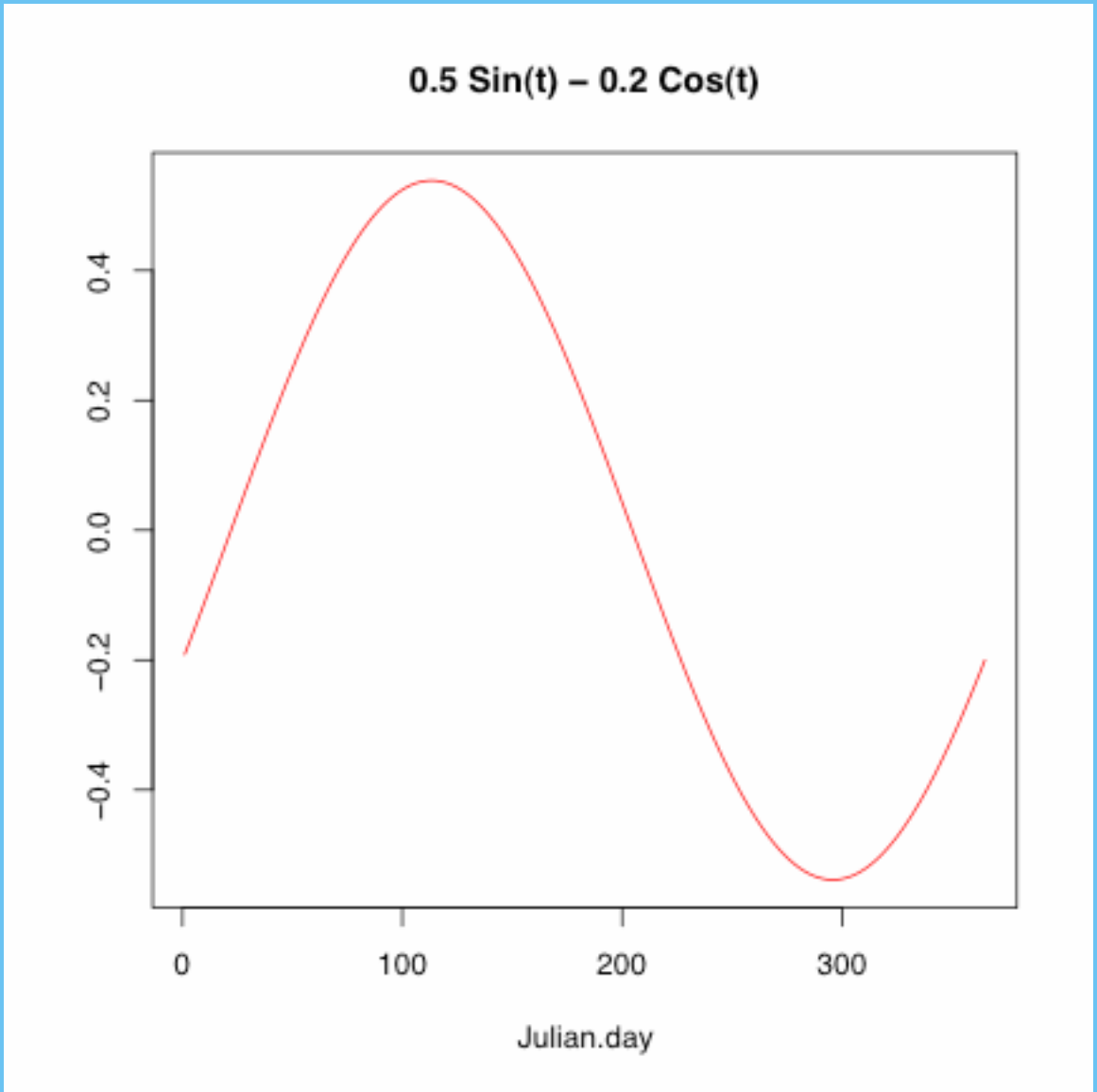
$$\cong a_0 + a_1 \cos(2\pi t) + b_1 \cos(2\pi t)$$

USGS

Example 1: (β_5 = 0.5, β_6 = -0.2)

```
> # Consider b5=0.5 and b6=-0.2
>   b5 <-  0.5
>   b6 <- -0.2
> # Compute the amplitude
>   sqrt(b5^2+b6^2)
[1] 0.5385165
> # Compute the phase (peak day)
>   365/(2*pi) * atan2(b5,b6)
[1] 113.3542
> # Plot the function
> f2 <- b5 * sin1 + b6 * cos1
> plot(Julian.day,f2,ylab=" ",main="0.5 Sin(t) - 0.2 Cos(t)",type="l",col="red")
```

USGS

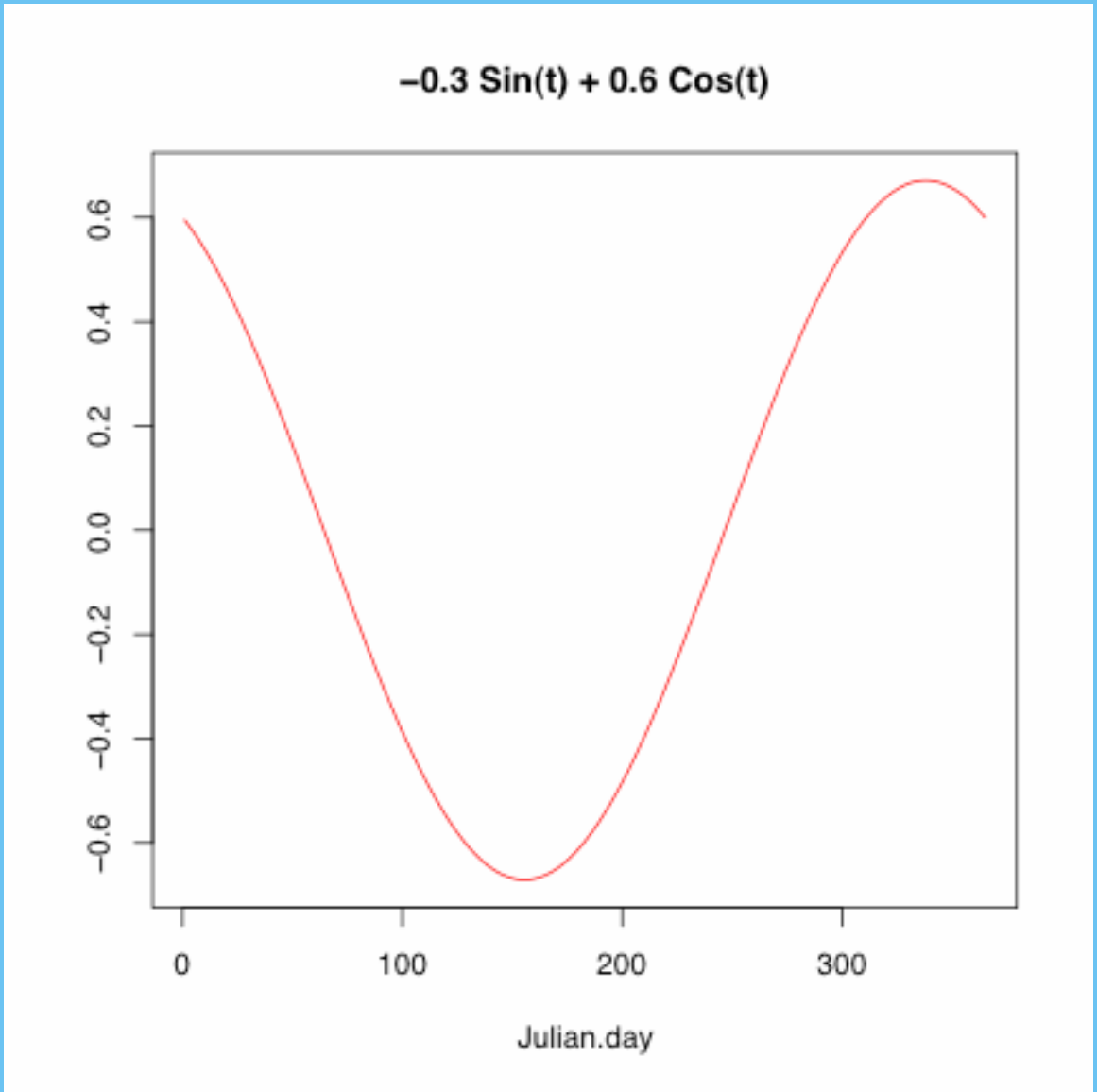


USGS

Example 2: (β_5 = -0.3, β_6 = 0.6)

```
> # Consider b5=0.5 and b6=-0.2
>   b5 <-  -0.3
>   b6 <-   0.6
> # Compute the amplitude
>   sqrt(b5^2+b6^2)
[1] 0.6708204
> # Compute the phase (peak day)
>   365/(2*pi) * atan2(b5,b6)
[1] -26.93401
> # Plot the function
> f2 <- b5 * sin1 + b6 * cos1
> plot(Julian.day,f2,ylab=" ",main="-0.3 Sin(t) + 0.6
    Cos(t)",type="l",col="red")
```

USGS



USGS

The Error (ε)

- "Explains" many things:
  - Natural variability of process
  - Measurement error
  - Model mis-specification
- Assumed to exhibit bell-shaped (possibly normal) distribution
  - $E[\varepsilon] = 0$
  - $E[\varepsilon^2] = \sigma^2$
- Independent and identically distributed ("iid")

USGS

How you *really* apply this…

- LoadEst software [Lorenz, 2004]
- Estimator
- “By Hand” on computer (only if you have to…)

USGS

Create Predictors

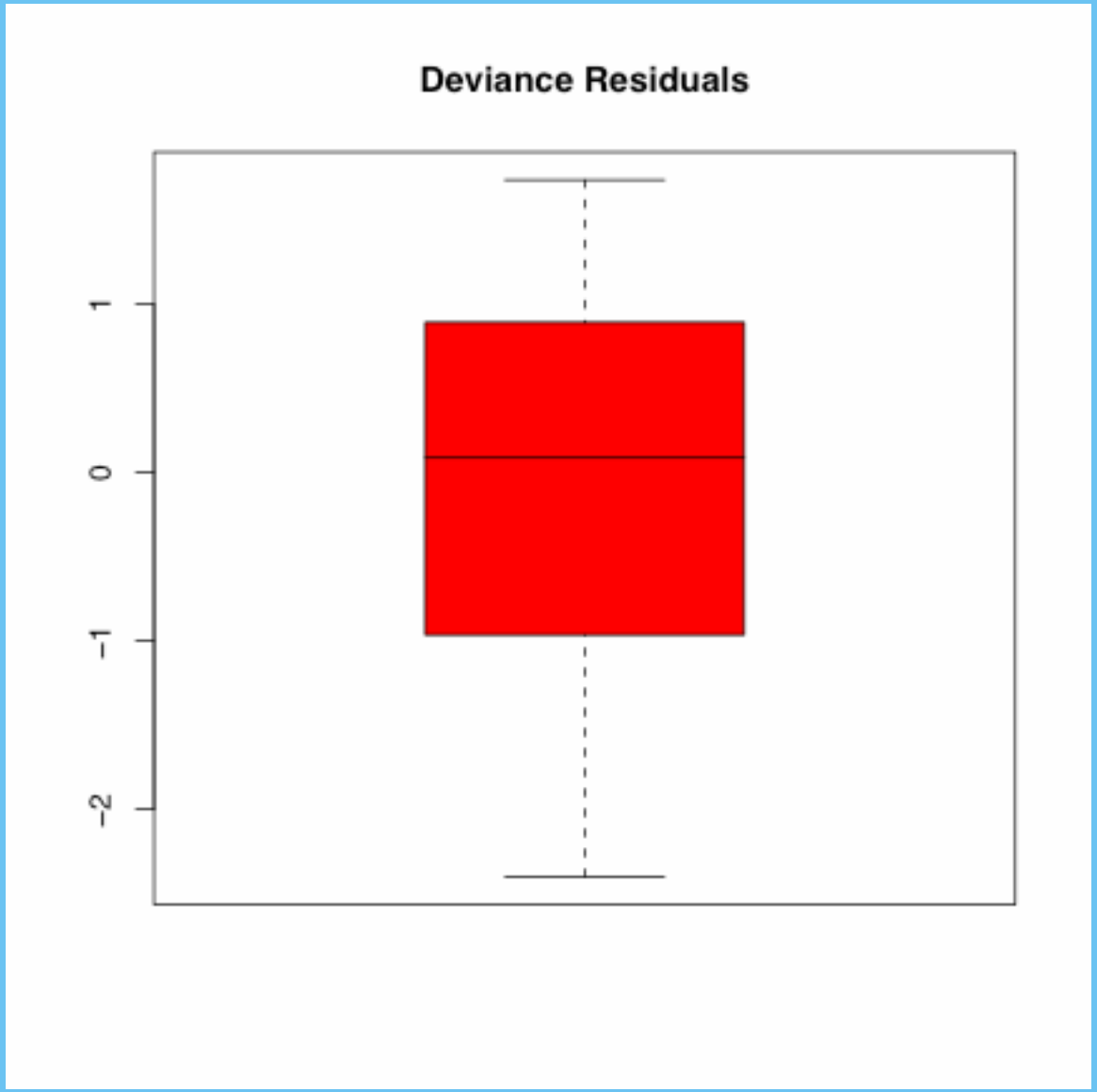
```
> ## Build the 7-parameter model and fit it
> y<-LN.LOAD.
> yd<-LN.DET..LIMIT.
> #
> lq<-log(Q)
>   lqm <- (lq-mean(lq))
>   lqs <-  mean(lq) + mean(lqm^3)/mean(2*lqm^2)
> lq1 <- lq-lqs
> lq2 <- lq1^2
> #
> dt<-DEC.TIME
>   dtm <- (dt-mean(dt))
>   dts <-  mean(dt) + mean(dtm^3)/mean(2*dtm^2)
> dt1 <-  dt-dts
> dt2 <- dt1^2
> sin1<-sin(2*3.14159*(dt-1983))
> cos1<-cos(2*3.14159*(dt-1983))
➢l1<-survreg(Surv(pmax(y,yd),y>=yd,type='left')~1+lq1+lq2+dt1+dt2+sin1+cos1,
➢    dist='gaussian')
> resdev<-residuals.survreg(l1,type="deviance")
> boxplot(resdev,main="Deviance Residuals",col="red")
> boxplot(split(resdev,MONTH),main="Deviance Residuals by Month",col="red")
> plot(time,resdev,main="Residuals vs. Time",col="red")
> plot(lq1,resdev,main="Residuals vs. Q",col="red")
```

USGS

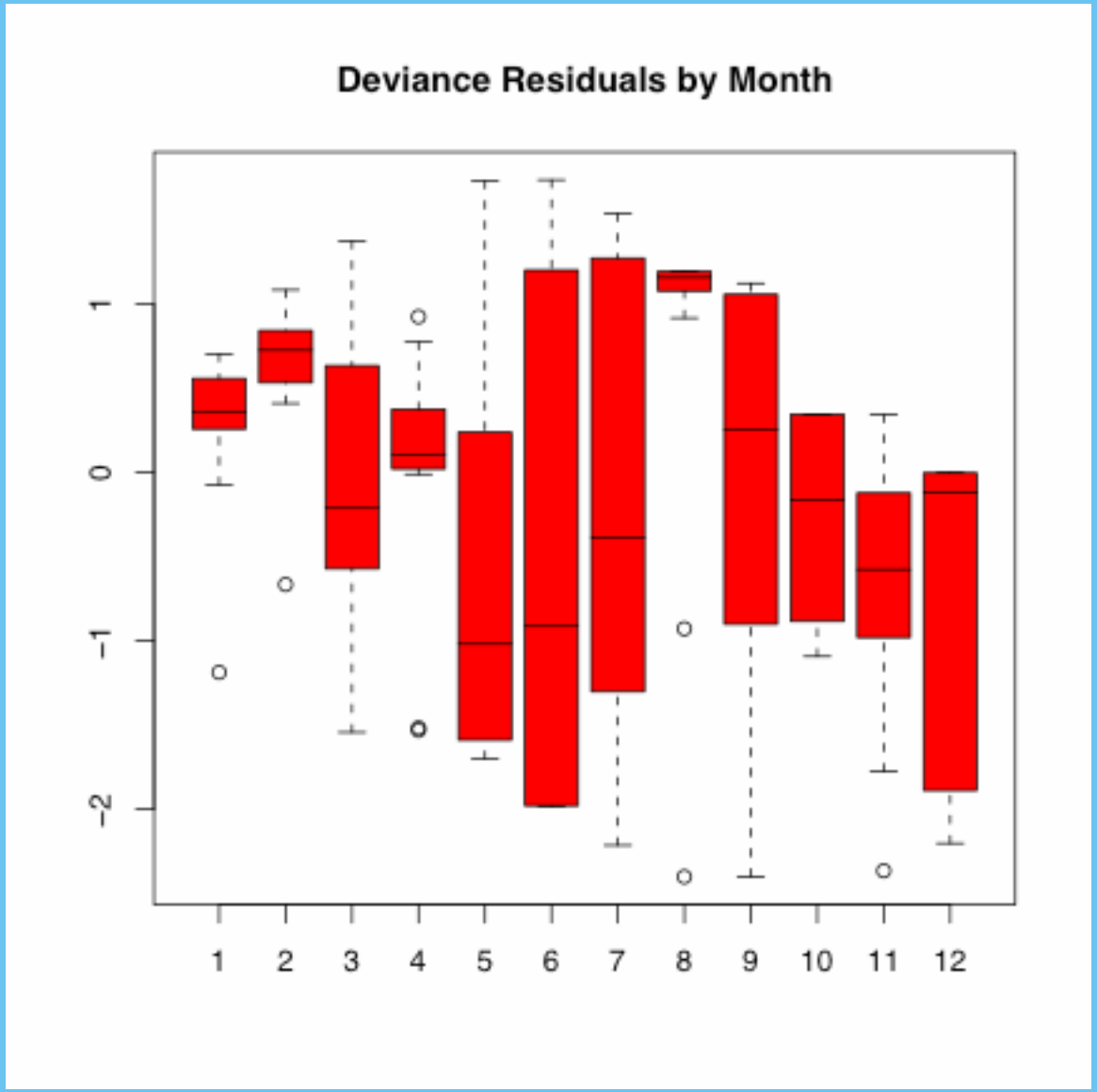
Regression Results

```
> summary(l1)

Call:
survreg(formula = Surv(pmax(y, yd), y >= yd, type = "left") ~
    1 + lq1 + lq2 + dt1 + dt2 + sin1 + cos1, dist = "gaussian")
              Value Std. Error      z         p
(Intercept)  6.8853     0.2334 29.497 3.13e-191
lq1          1.0829     0.0947 11.433  2.87e-30
lq2         -0.0569     0.0638 -0.892  3.73e-01
dt1         -0.1234     0.0272 -4.538  5.68e-06
dt2         -0.0214     0.0114 -1.886  5.93e-02
sin1        -0.8586     0.1704 -5.040  4.66e-07
cos1        -0.0483     0.1477 -0.327  7.44e-01
Log(scale)   0.1922     0.0605  3.176  1.49e-03

Scale= 1.21

Gaussian distribution
Loglik(model)= -284.2   Loglik(intercept only)= -351.6
                    Chisq= 134.63 on 6 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 4
n= 197
```

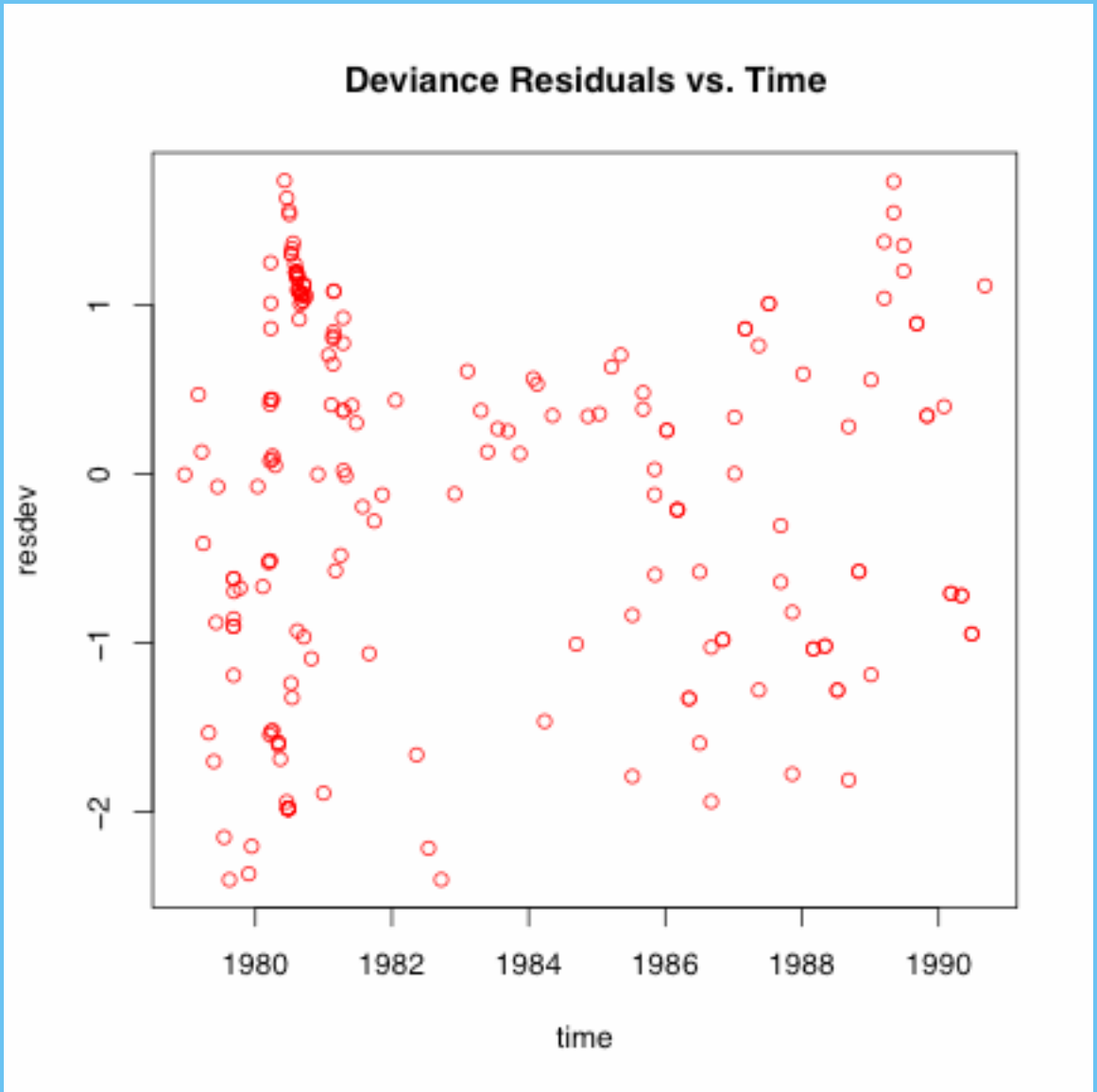
USGS



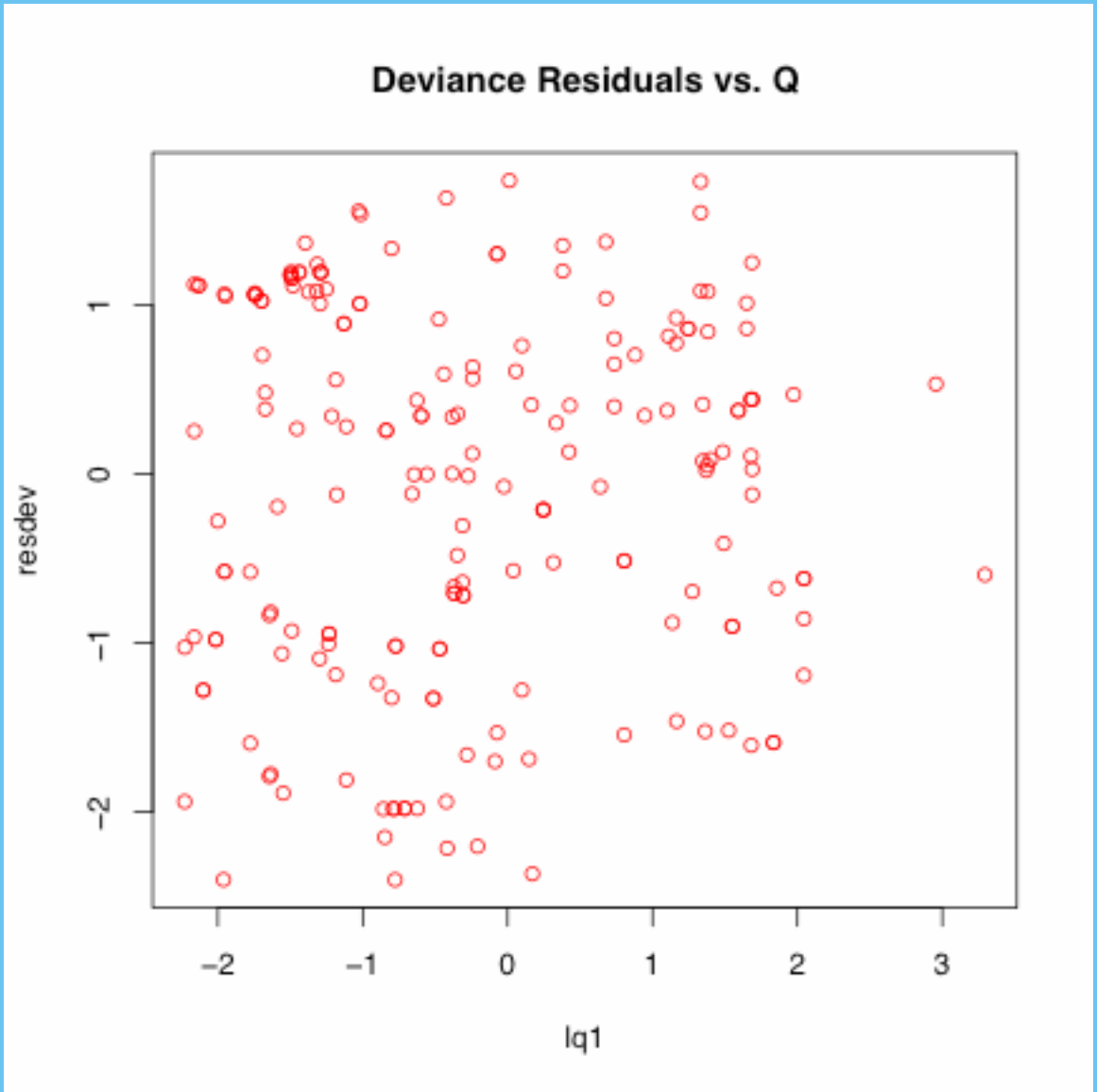
USGS



USGS


Deviance Residuals vs. Time
resdev
1
0
−1
−2
1980
1982
1984
1986
1988
1990
time

USGS



USGS

A Caution: Understanding "Lack of Fit"

- "Lack of fit" indicates a discrepancy between the observed data and the parametric model
- Rarely, if ever, can "lack of fit" be resolved by changing statistical fitting methods

USGS

Possible Solutions

- Alternative predictors
  - “Indicator variables”
  - Piecewise linear variables
- Stratify data by:
  - Season
  - Discharge range
  - Source
- Alternative models
  - Linear models
  - Non-linear models

USGS

Some Alternative Models

- **Complex *linear* right-hand sides**

$$\ln[C] = \beta_0 + \beta_1 \tan^{-1}\left(\frac{1}{X}\right) + \beta_2 \ln\left(\frac{X}{1+X}\right) + \varepsilon$$

- **Left-hand side untransformed**

$$C = \beta_0 + \beta_1 X + \varepsilon$$

USGS

Untransformed Response

$$C = \beta_1 (\frac{1}{Q}) + \varepsilon$$

$$C = \beta_0 + \beta_1 (\frac{1}{Q}) + \ldots + \varepsilon$$

USGS

Guidelines

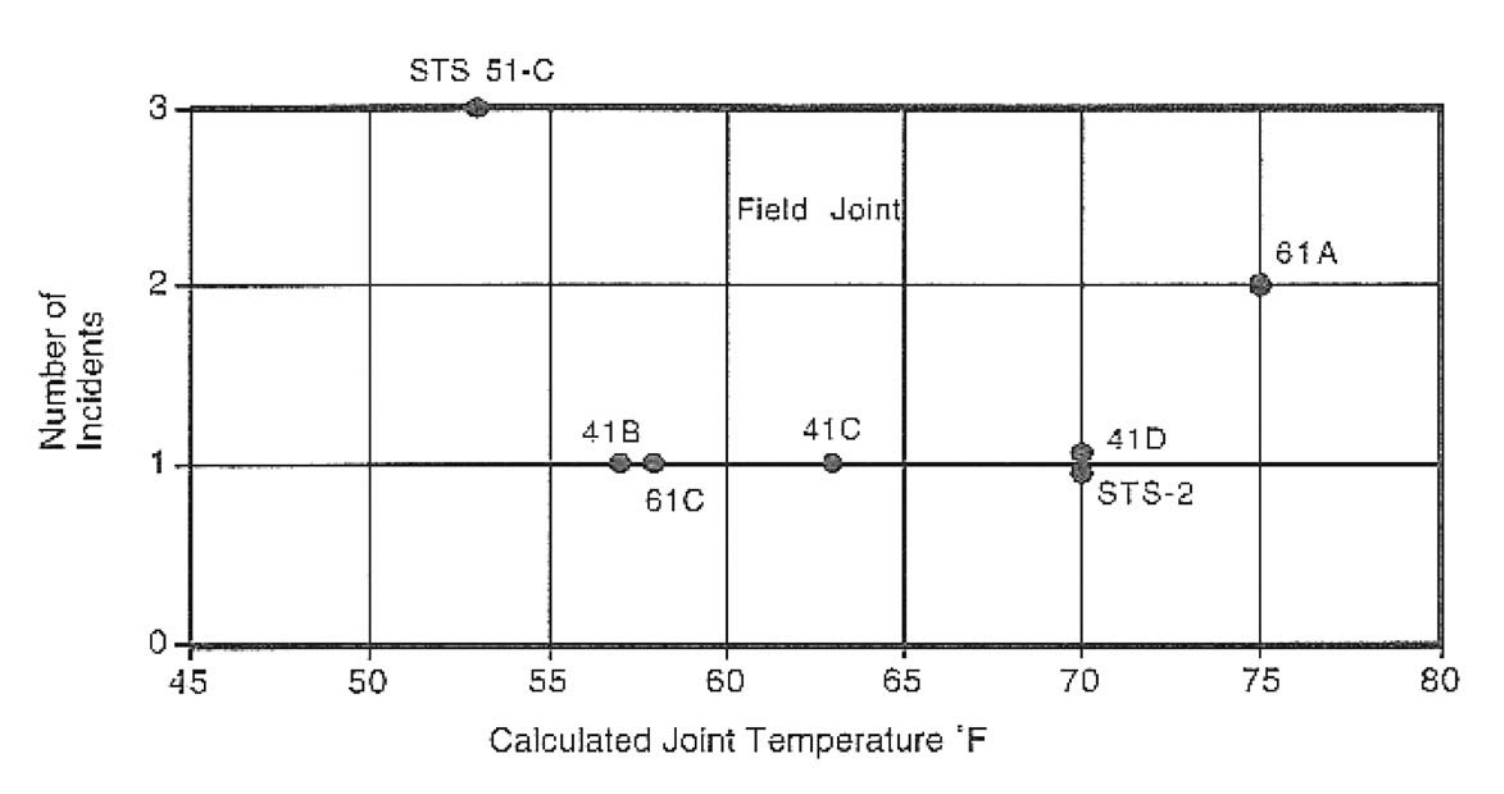
- Linear models are always easy to use
- Inherently non-linear regression models can be tricky
  - Uniqueness of results
  - Interpretation
  - Diagnostics
  - Uncertainty
- Advice: Avoid non-linear models whenever possible

USGS

Censored Data

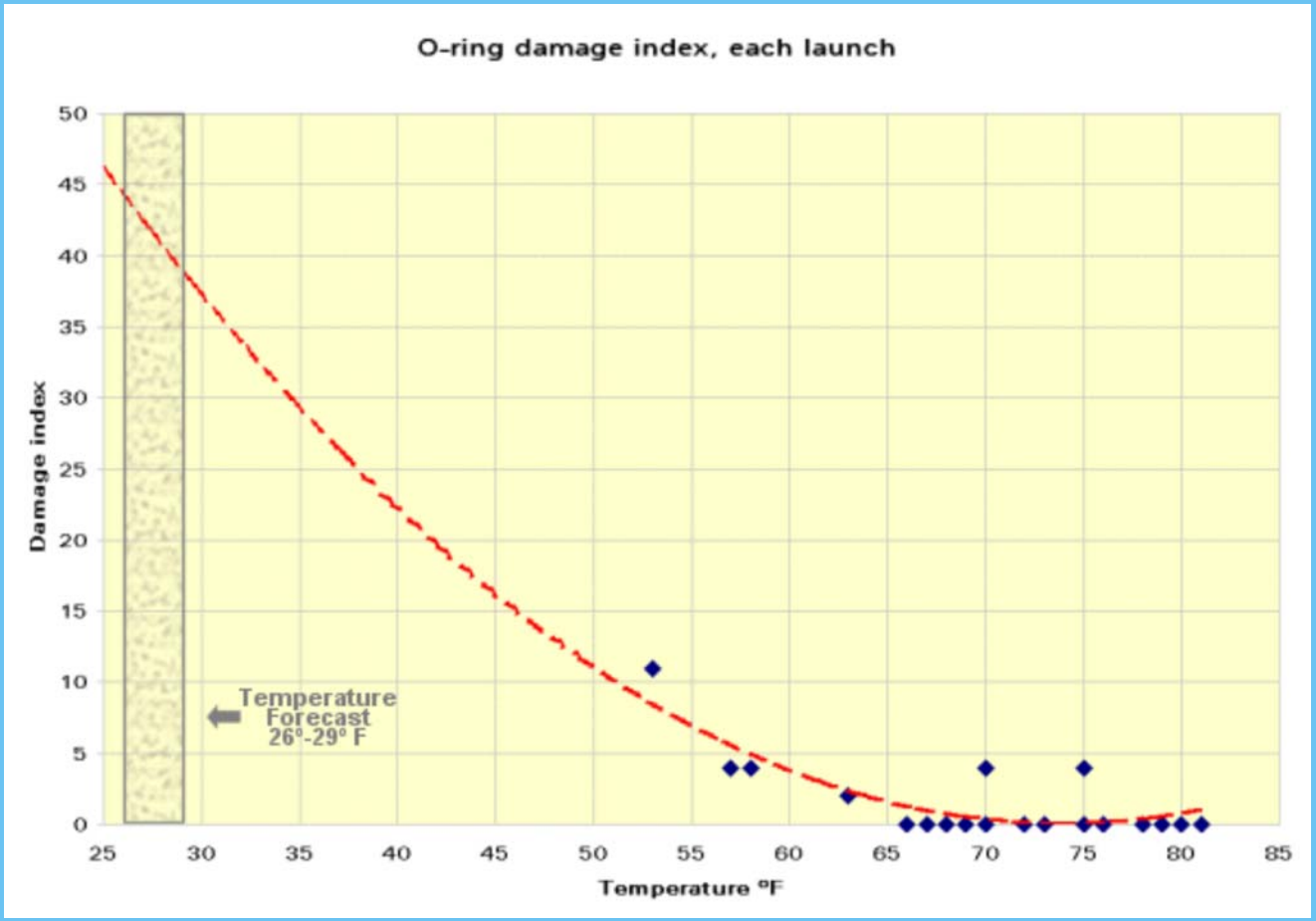
Challenger (1986)

Challenger Disaster (Tufte, 1997)



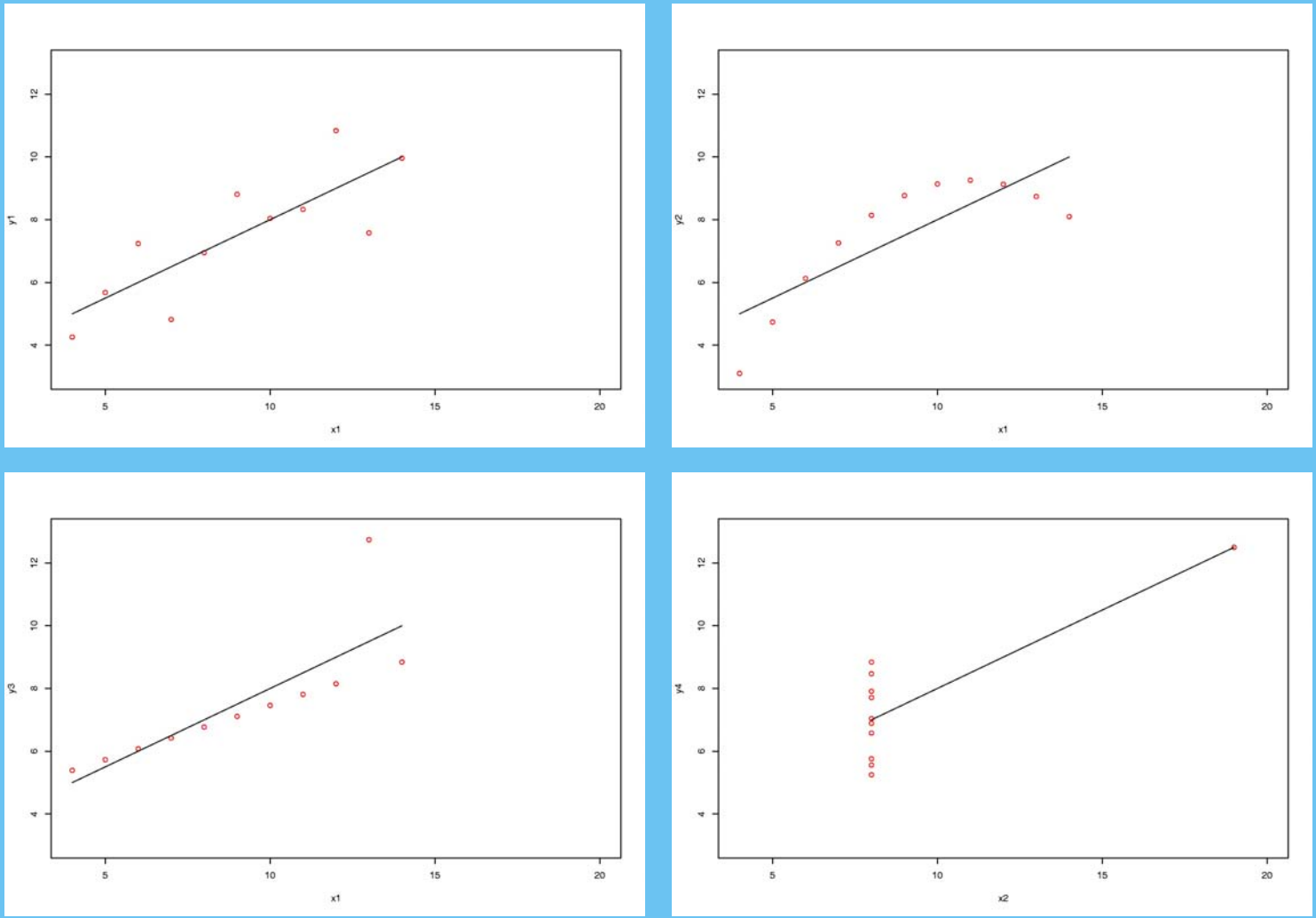
USGS

Challenger Disaster (Tufte, 1997)



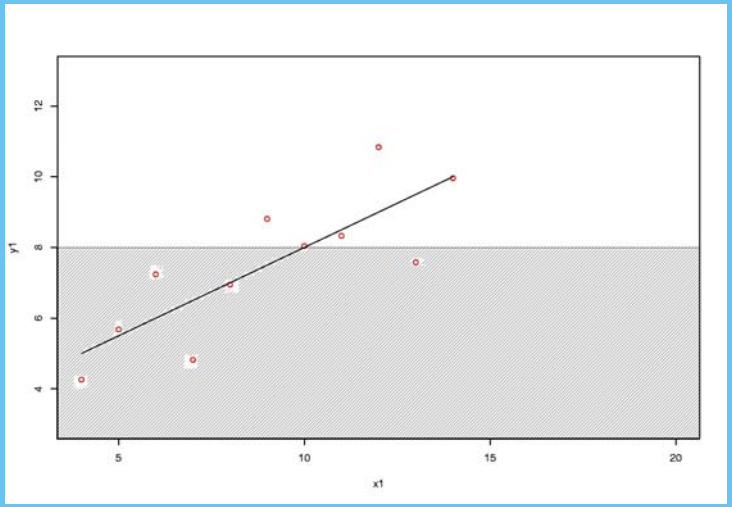
USGS

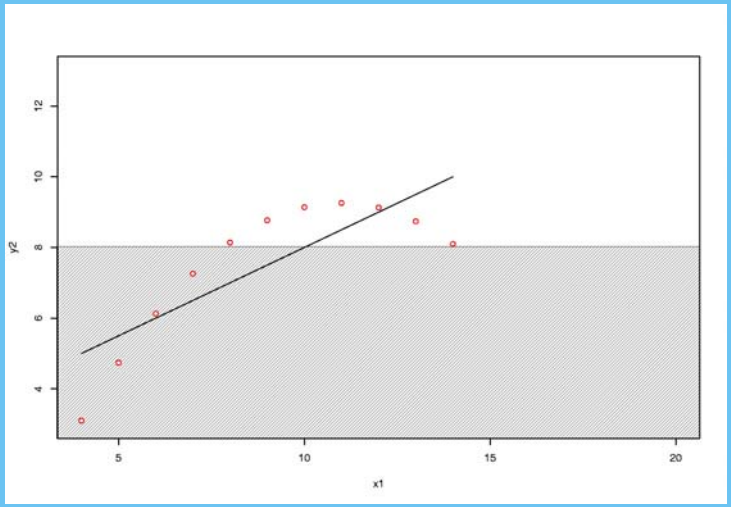
“Plot the Data!!!” (Anscombe, 1973)

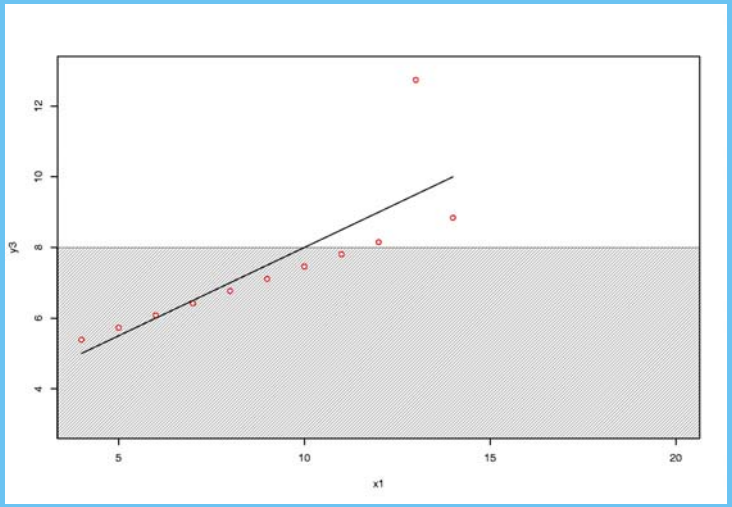


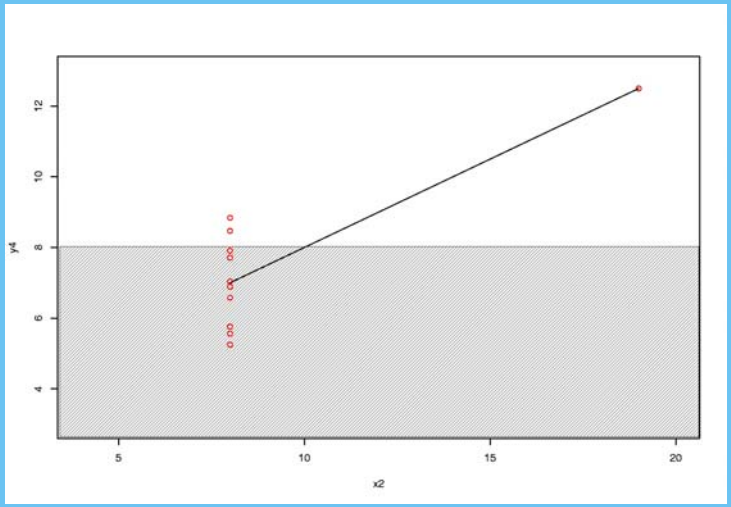
USGS

Censored Data????









USGS

Step 3: Load Estimation

- Given any value of Q and T, the calibrated model will give us an unbiased (nearly) estimate for ln[C].
- However, we want to estimate C, not ln[C].
- Retransformation can introduce bias.
- Solution:
  - MVUE for complete data
  - AMLE for censored data

USGS

Retransformation Bias

- For a constant value, X, 10^(log10(X)) = X
- However, the corresponding result is not generally true for random variables
- Consider Y=(1,10,100,1000,10000)

```
> y<-c(1,10,100,1000,10000)
> mean(y)
[1] 2222.2
> logy<-log10(y)
> logy
[1] 0 1 2 3 4
> mean(logy)
[1] 2
> 10^(mean(logy))
[1] 100
```

- Note: 10^(mean(log10(Y))) does not equal mean(Y)

USGS

Finney/Bradu/Mundlak Minimum Variance Unbiased Estimator (for complete samples)

$$\hat{L}_{MVUE} = \exp(X\hat{\beta}) \cdot g_m\left(\frac{m+1}{2m}[1-V]s^2\right)$$

$$E[\hat{L}_{MVUE}] = \exp\left(X\beta + \frac{\sigma^2}{2}\right)$$

$$= E[K \cdot Q \cdot C]$$

$$= E[L]$$

USGS

The Adjusted Maximum Likelihood Estimator (cases with censored data)

- Corrects for first-order bias with censored data
- Is not exact, but is nearly unbiased in small samples
- Details are in [Cohn et al., 1992]

USGS

How you *really* apply this…

- LoadEst software [Lorenz, 2004]
- Estimator
- “By Hand” on computer (only if you really, really, have to…)

USGS

Estimator Results (1)

MODEL FOR ESTIMATING CONSTITUENT LOADS
TIM COHN VERSION 2002.03

OUTPUT FILE:test_new.result
APPROXIMATE VARIANCES
INPUT FILE: q_01646580.csv
INPUT FILE: qw_01646580.csv

CALIBRATION FLOWS FROM DV FLOW FILE

USER SPECIFIED PERIOD: 1900.00-2100.00

DATE OF FIRST OBS. USED IN CALIBRATION: 12/19/1978
DATE OF LAST OBS. USED IN CALIBRATION: 09/05/1990
STATION NUMBER: 01646500

REGRESSION OF P00001
ON 7 REGRESSORS

			CONCENTRATION MODEL				LOAD MODEL			
NUMBER	NAME	CENTER	COEFF.	S.D.	T	P	COEFF.	S.D.	T	P
1	CONSTANT	0.000	-3.2981	0.2393	*****	0.000000	6.8904	0.2393	28.79	0.000000
2	LOG-FLOW	9.294	0.0835	0.0961	0.87	0.376162	1.0835	0.0961	11.27	0.000000
3	LOG-FLOW SQUARED	9.294	-0.0565	0.0650	-0.87	0.368438	-0.0565	0.0650	-0.87	0.368438
4	DEC_TIME	1984.477	-0.1223	0.0281	-4.35	0.000009	-0.1223	0.0281	-4.35	0.000009
5	DEC_TIME SQUARED	1984.477	-0.0214	0.0116	-1.85	0.055480	-0.0214	0.0116	-1.85	0.055480
6	SIN(2*PI*T)	0.000	-0.8575	0.1717	-4.99	0.000001	-0.8575	0.1717	-4.99	0.000001
7	COS(2*PI*T)	0.000	-0.0522	0.1504	-0.35	0.724568	-0.0522	0.1504	-0.35	0.724568
S				1.23287			1.23287			
R**2 (%)			27.1				53.1			

N 197
M 190
NCENS 46

SER. CORR. OF RES. 0.51659

USGS

Estimator Results (2)

	CONSTANT	LOG-FLOW	LOG-FLOW	DEC_TIME	DEC_TIME	SIN(2*PI*
LOG-FLOW	0.0000					
LOG-FLOW	0.0000	0.0000				
DEC_TIME	0.0000	-0.2007	-0.0834			
DEC_TIME	0.0000	0.1172	-0.1260	0.0000		
SIN(2*PI*	0.0000	0.5775	-0.3799	-0.0035	0.0341	
COS(2*PI*	0.0000	0.2247	-0.0211	0.1473	-0.1266	0.1764

PPCC CORRELATION = 0.98151

COMPUTED LOADS [KG/DAY] OR [G/DAY]

YEAR	MO.	LOAD:P00001	95% CONFIDENCE INTERVAL		S.E.	S.E. PRED.
1981 CALENDAR YEAR						
	1	195.6694	70.4654	437.3869	66.4320	96.3035
	2	1537.6285	534.7769	3511.1036	295.2624	782.7969
	3	721.3438	297.5790	1479.1419	150.1679	307.8625
	4	1299.9275	460.0853	2936.8720	229.7685	650.7298
	5	1265.6305	541.4102	2533.8004	245.9228	518.2232
	6	2359.8320	989.0216	4788.9295	505.7497	989.3241
	7	1144.8812	462.6759	2379.3361	224.9895	499.9095
	8	564.4426	236.3236	1146.2133	109.1965	236.9056
	9	1090.9093	453.7905	2224.7973	213.8104	461.2577
	10	942.6640	203.8472	2791.5722	211.4602	706.4105
	11	1083.5681	398.5133	2390.8171	253.6958	522.2743
	12	897.8394	357.4371	1884.1084	217.1514	398.4964
Q1	1981	794.2334	372.9771	1492.3641	155.1105	289.9610
Q2	1981	1637.6630	887.1445	2777.4265	306.9784	487.0748
Q3	1981	931.6991	513.2206	1560.6005	165.3769	269.7359
Q4	1981	973.5070	445.7005	1861.5685	219.9587	367.1508
CY	1981	1084.3488	720.9847	1568.0185	161.8535	217.0477
WY	1981	1057.9054	712.3826	1513.9721	153.0450	205.3474

USGS

Step 4: Uncertainty

- Errors to Consider
  - Sampling variability
  - Model Error
  - Other errors

USGS

“Real knowledge is to know the extent of one’s ignorance”

--Confucius

USGS

Estimator Results (1)

MODEL FOR ESTIMATING CONSTITUENT LOADS
TIM COHN VERSION 2002.03

OUTPUT FILE:test_new.result
APPROXIMATE VARIANCES
INPUT FILE: q_01646580.csv
INPUT FILE: qw_01646580.csv

CALIBRATION FLOWS FROM DV FLOW FILE

USER SPECIFIED PERIOD: 1900.00-2100.00

DATE OF FIRST OBS. USED IN CALIBRATION: 12/19/1978
DATE OF LAST OBS. USED IN CALIBRATION: 09/05/1990
STATION NUMBER: 01646500

REGRESSION OF P00001
ON 7 REGRESSORS

			CONCENTRATION MODEL				LOAD MODEL			
NUMBER	NAME	CENTER	COEFF.	S.D.	T	P	COEFF.	S.D.	T	P
1	CONSTANT	0.000	-3.2981	0.2393	*****	0.000000	6.8904	0.2393	28.79	0.000000
2	LOG-FLOW	9.294	0.0835	0.0961	0.87	0.376162	1.0835	0.0961	11.27	0.000000
3	LOG-FLOW SQUARED	9.294	-0.0565	0.0650	-0.87	0.368438	-0.0565	0.0650	-0.87	0.368438
4	DEC_TIME	1984.477	-0.1223	0.0281	-4.35	0.000009	-0.1223	0.0281	-4.35	0.000009
5	DEC_TIME SQUARED	1984.477	-0.0214	0.0116	-1.85	0.055480	-0.0214	0.0116	-1.85	0.055480
6	SIN(2*PI*T)	0.000	-0.8575	0.1717	-4.99	0.000001	-0.8575	0.1717	-4.99	0.000001
7	COS(2*PI*T)	0.000	-0.0522	0.1504	-0.35	0.724568	-0.0522	0.1504	-0.35	0.724568
S				1.23287			1.23287			
R**2 (%)			27.1				53.1			

N 197
M 190
NCENS 46

SER. CORR. OF RES. 0.51659

USGS

Estimator Results (2)

	CONSTANT	LOG-FLOW	LOG-FLOW	DEC_TIME	DEC_TIME	SIN(2*PI*
LOG-FLOW	0.0000					
LOG-FLOW	0.0000	0.0000				
DEC_TIME	0.0000	-0.2007	-0.0834			
DEC_TIME	0.0000	0.1172	-0.1260	0.0000		
SIN(2*PI*	0.0000	0.5775	-0.3799	-0.0035	0.0341	
COS(2*PI*	0.0000	0.2247	-0.0211	0.1473	-0.1266	0.1764

PPCC CORRELATION = 0.98151

COMPUTED LOADS [KG/DAY] OR [G/DAY]

YEAR	MO.	LOAD:P00001	95% CONFIDENCE INTERVAL		S.E.	S.E. PRED.
1981 CALENDAR YEAR						
	1	195.6694	70.4654	437.3869	66.4320	96.3035
	2	1537.6285	534.7769	3511.1036	295.2624	782.7969
	3	721.3438	297.5790	1479.1419	150.1679	307.8625
	4	1299.9275	460.0853	2936.8720	229.7685	650.7298
	5	1265.6305	541.4102	2533.8004	245.9228	518.2232
	6	2359.8320	989.0216	4788.9295	505.7497	989.3241
	7	1144.8812	462.6759	2379.3361	224.9895	499.9095
	8	564.4426	236.3236	1146.2133	109.1965	236.9056
	9	1090.9093	453.7905	2224.7973	213.8104	461.2577
	10	942.6640	203.8472	2791.5722	211.4602	706.4105
	11	1083.5681	398.5133	2390.8171	253.6958	522.2743
	12	897.8394	357.4371	1884.1084	217.1514	398.4964
Q1	1981	794.2334	372.9771	1492.3641	155.1105	289.9610
Q2	1981	1637.6630	887.1445	2777.4265	306.9784	487.0748
Q3	1981	931.6991	513.2206	1560.6005	165.3769	269.7359
Q4	1981	973.5070	445.7005	1861.5685	219.9587	367.1508
CY	1981	1084.3488	720.9847	1568.0185	161.8535	217.0477
WY	1981	1057.9054	712.3826	1513.9721	153.0450	205.3474

USGS

How you *really* apply this…

- LoadEst software [Lorenz, 2004]
- Estimator

USGS

Data Considerations for Load Estimation

National Water Quality Monitoring Council
Chattanooga, TN, 17 May 2004

Tim Cohn

U.S. Geological Survey

Reston, Virginia

Data Requirements for AMLE

- Observations above threshold
- “Less-than” values
- Censoring thresholds corresponding to *all* observations

USGS

Some Rules of Thumb

- Procedures have been used successfully to monitor nutrient loads from watersheds larger than 100 square miles.
- Typical sampling program for Loads involves:
  - At least 2 Years of data collections at 25 samples per year
- Half of samples should reflect high-flow events (if goal is long-term load estimate)

USGS

Complications

- Difficulties we can handle now:
  - Concentrations that vary dramatically
  - Transformed response variable(Retransformation bias)
  - Censored data
  - Nutrients and sediment
- Unresolved issues
  - If linear model cannot be made to fit
    - Small watersheds (<100 square miles)
    - Complicated situations
    - Toxics and metals
  - When errors are not independent
  - Etc.

USGS

S-LOADEST Demonstration

Statistical Techniques for Trend and Load Estimation
National Water Quality Monitoring Council
Chattanooga, TN, 17 May 2004

Tim Cohn

U.S. Geological Survey

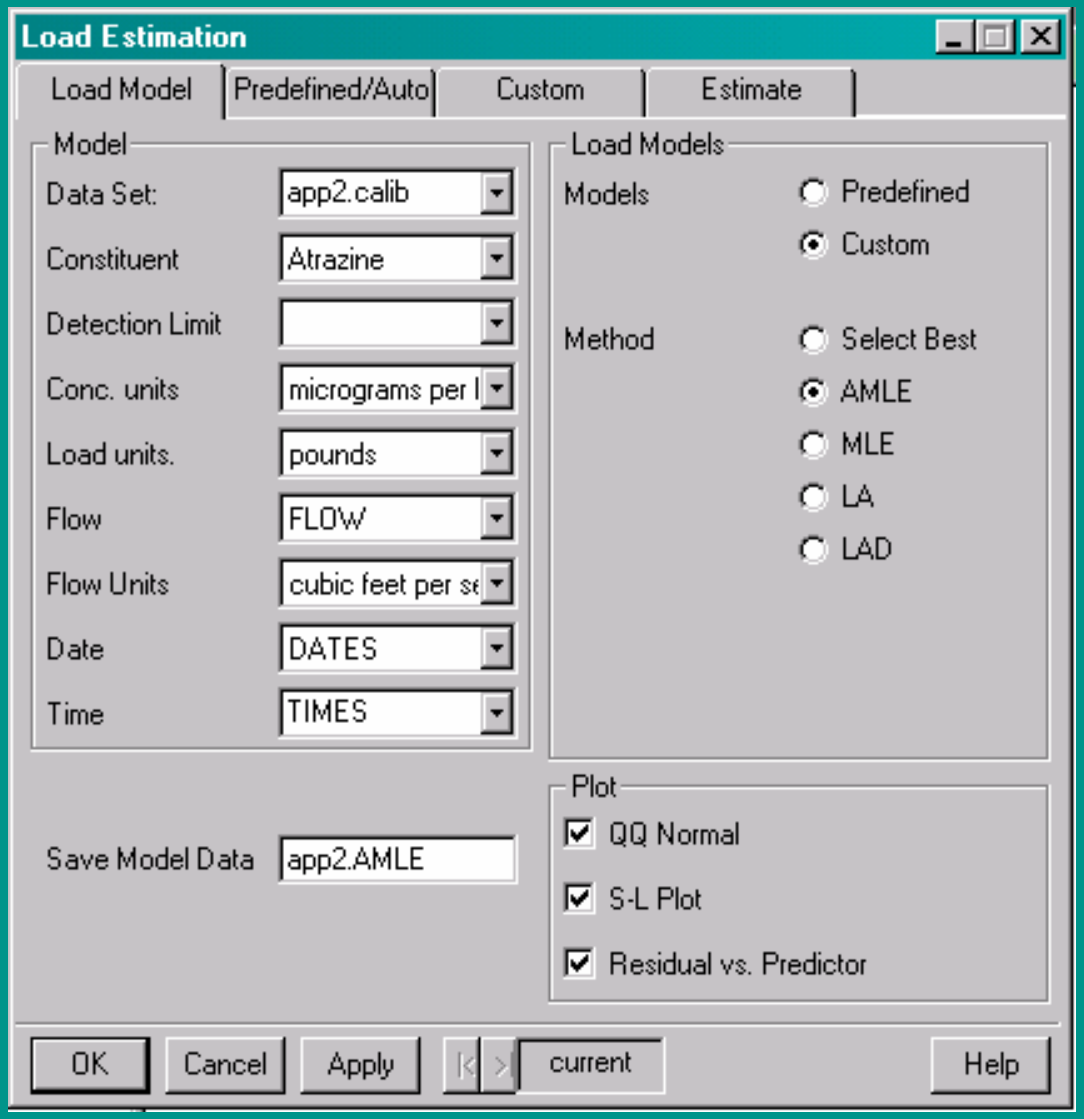
Reston, Virginia

Load Estimation

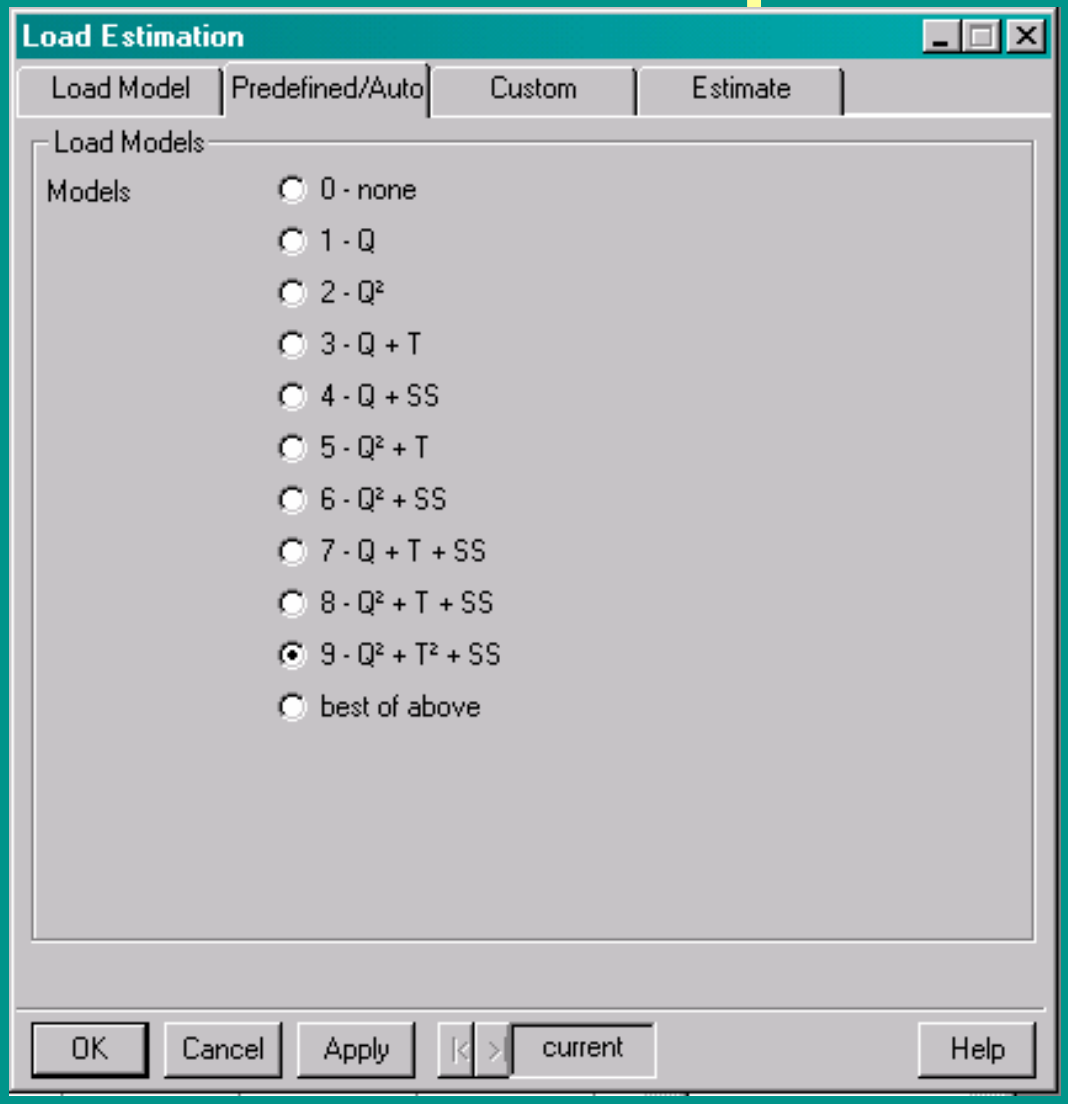
- Implemented in S-PLUS in two approaches.
  - A step-by-step menu system, similar to that of ESTREND useful for managing many stations and constituents.
  - A single dialog window for single-station, single-constituent analysis.
- Only the rating-curve method is available in LOADEST.

USGS

Calibration Data Information

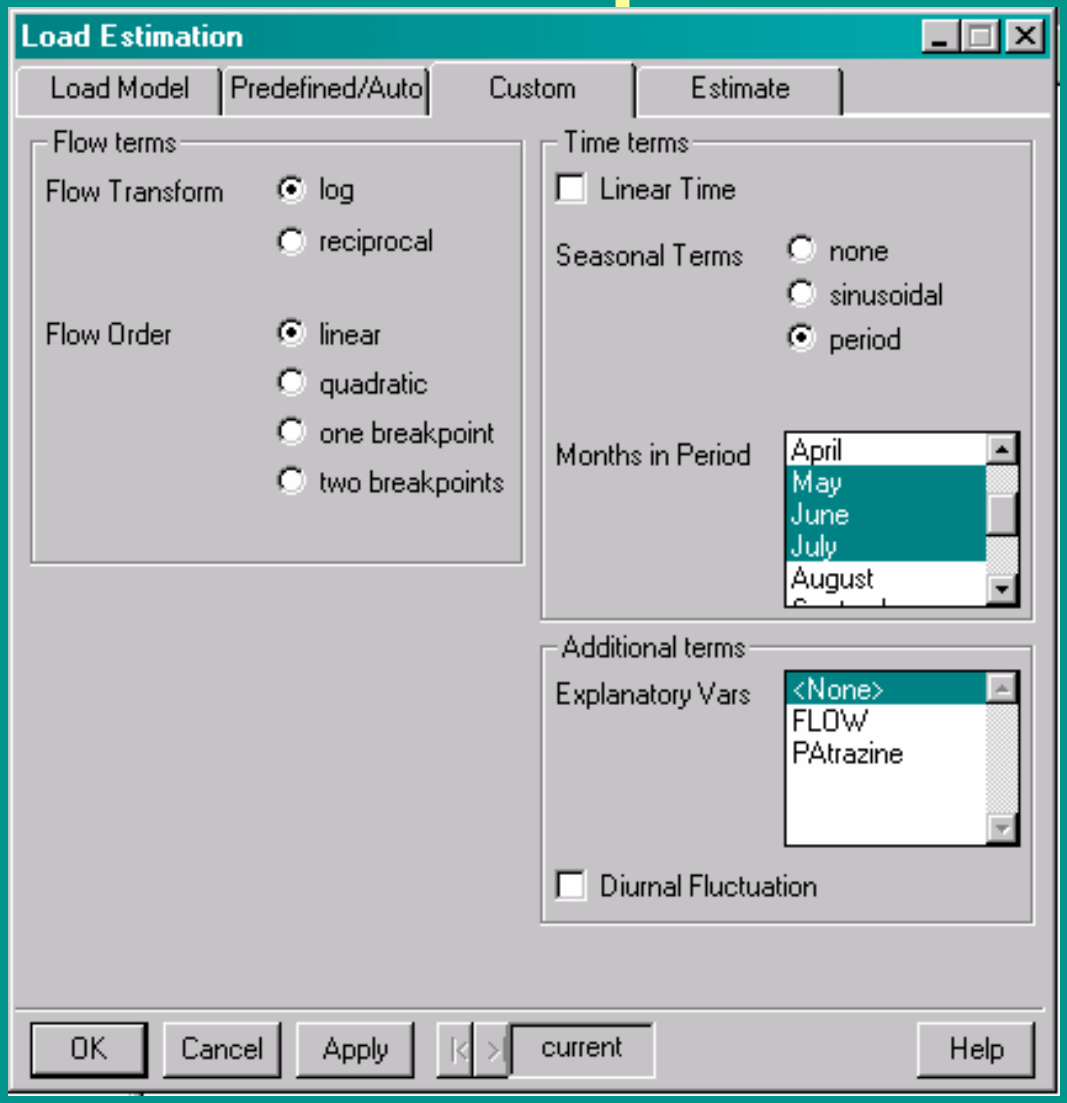


Predefined Model Options



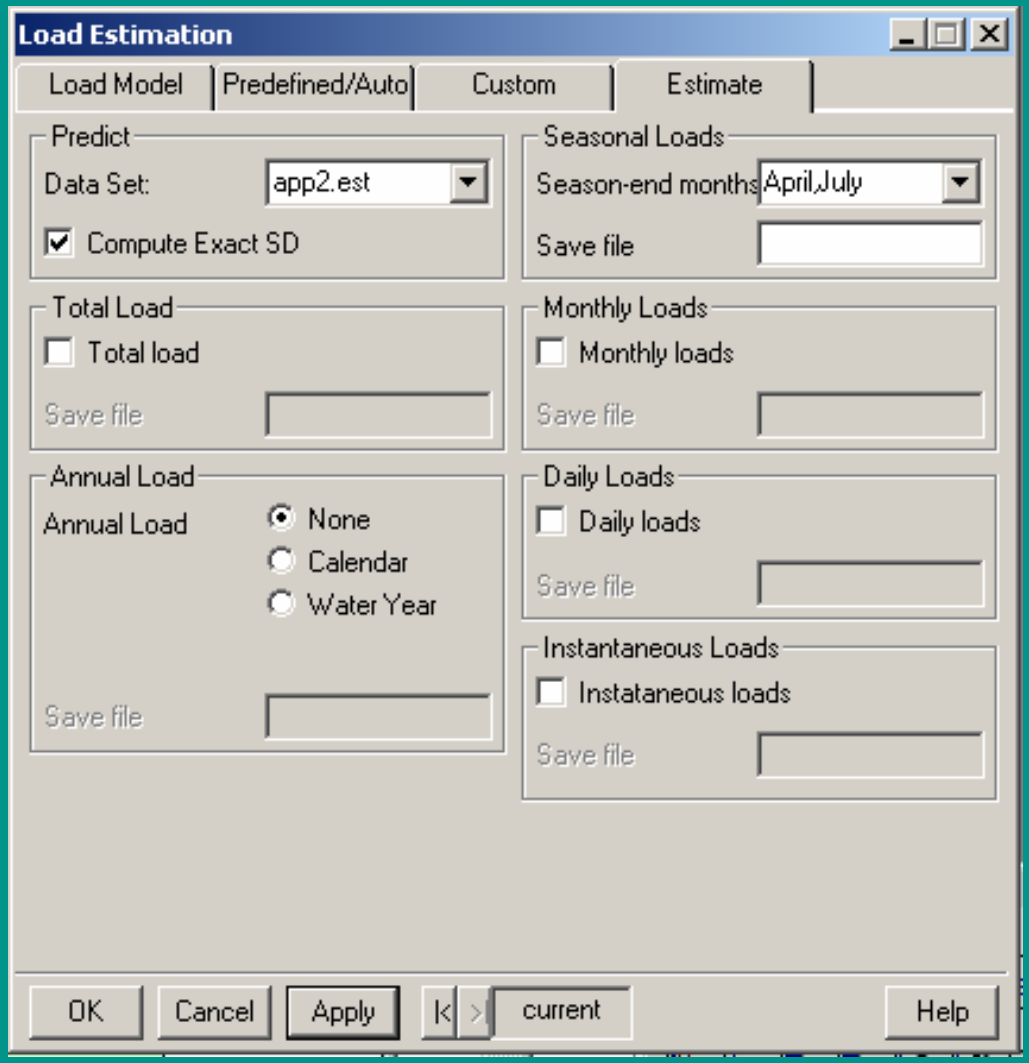
USGS

Custom Model Options



USGS

Estimation Options



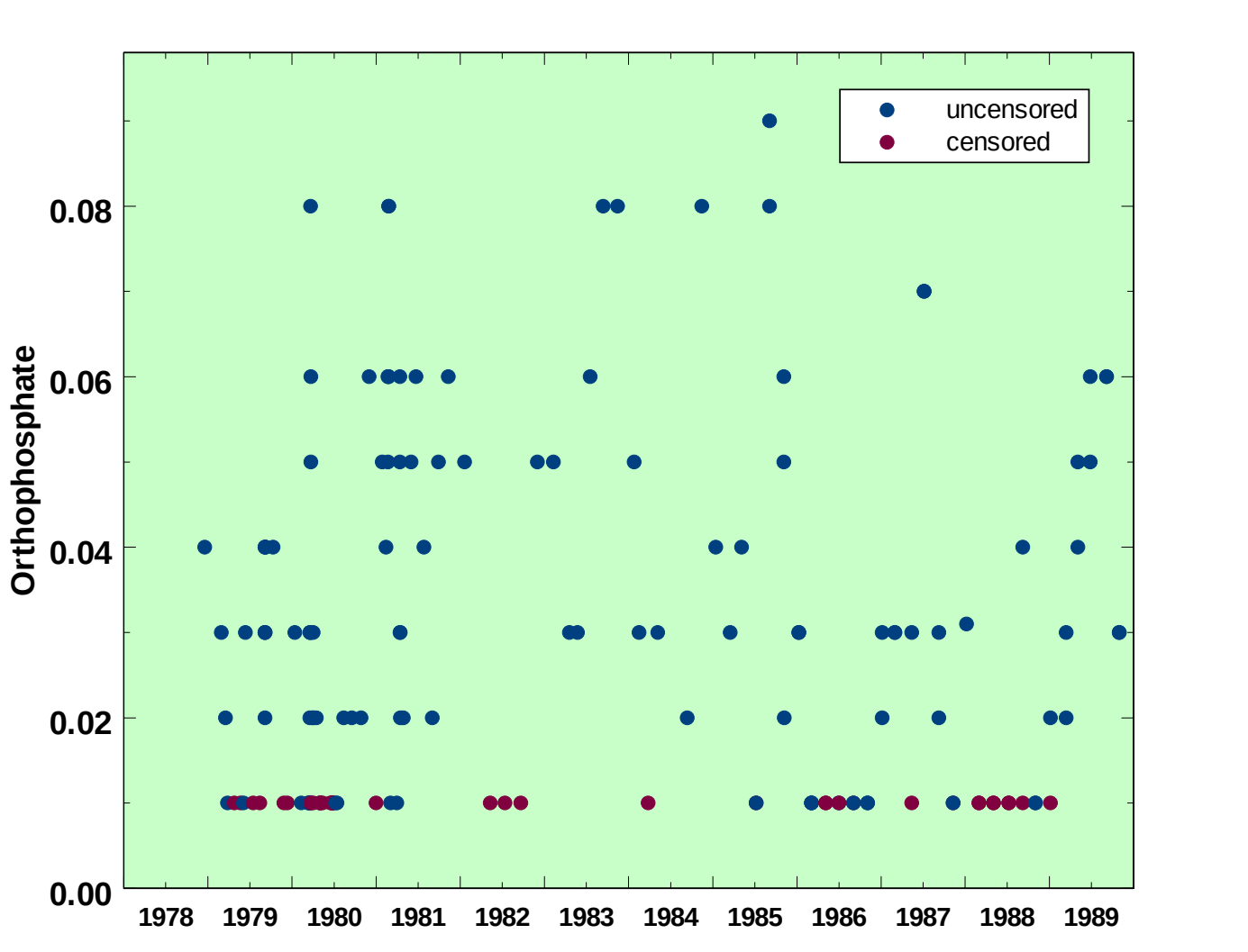
USGS

The Seven-Parameter Model

- The data are dissolved orthophosphate concentrations in the Potomac River collected from 1978 to 1989.
- Because the data are collected over a relatively long period of time and the advantages of using many terms in a predictive model, the 7-parameter model will be used for these data.

USGS

The Seven-Parameter Model



The Seven-Parameter Model

```
Total number of observations: 163, number of censored obs: 46.
Calibration data period of record: 12/19/1978 to 09/05/1990.
Calibration data streamflow summary statistics:
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
   1170   3920    8280  22964  36350  293000

 Adjusted Maximum Likelihood Estimate of the selected model.
  Coefficient   Std. Dev.       P-value
    Intercept  7.26431504 0.172606544 4.685199e-066
          lnQ  1.38892380 0.075990492 1.422784e-039
         lnQ2 -0.08468726 0.049363628 7.891491e-002
      DECTIME  0.04874243 0.024382206 4.276989e-002
     DECTIME2 -0.02430430 0.007853962 1.638463e-003
  sin.DECTIME -0.58992029 0.122709260 1.807739e-006
  cos.DECTIME  0.26803393 0.111101188 1.457480e-002
```

USGS

The Seven-Parameter Model

Estimated residual variance = 0.7380

R-squared of the regression = 77.6

Serial correlation of the residuals = 0.49

* NOTE: the following tests are merely informative, the user must evaluate the model more fully

Turnbull-Weiss normality test statistic = 84.9,

P-value = 2.27e-007, DF = 29

Grubbs & Beck 2-sided outlier test criteria at 5% significance

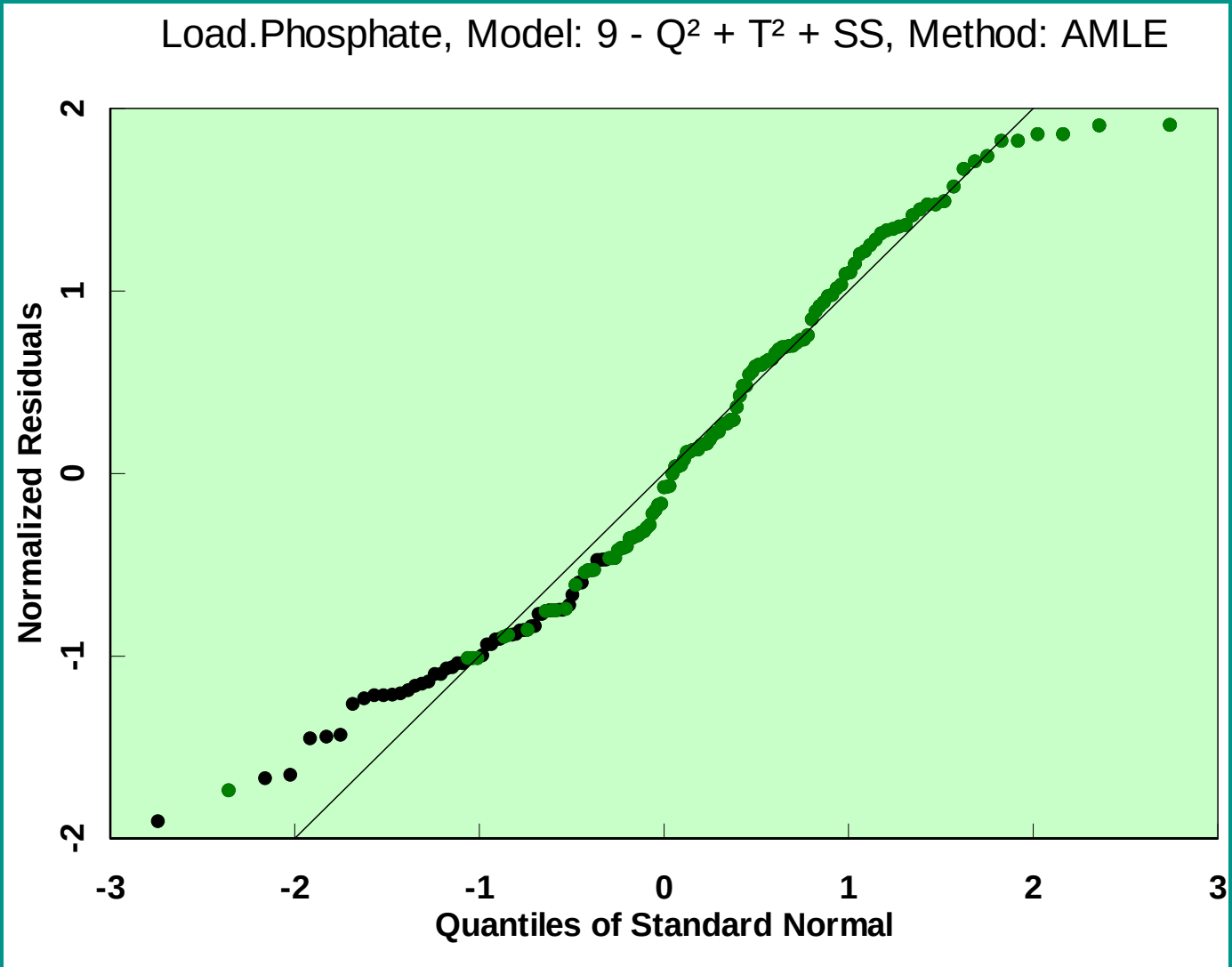
-3.54 3.54

Range of normalized residuals

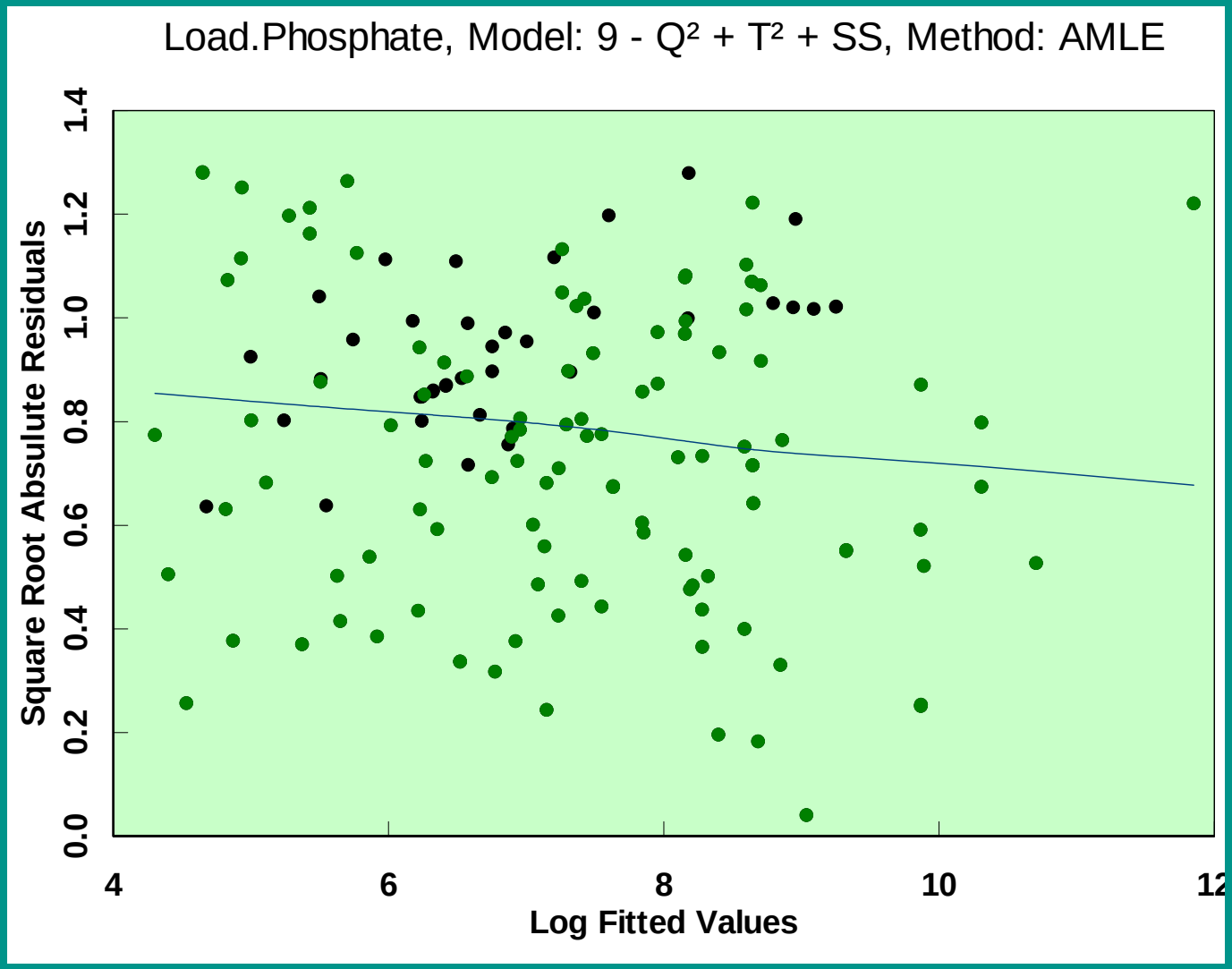
-1.905 1.91

USGS

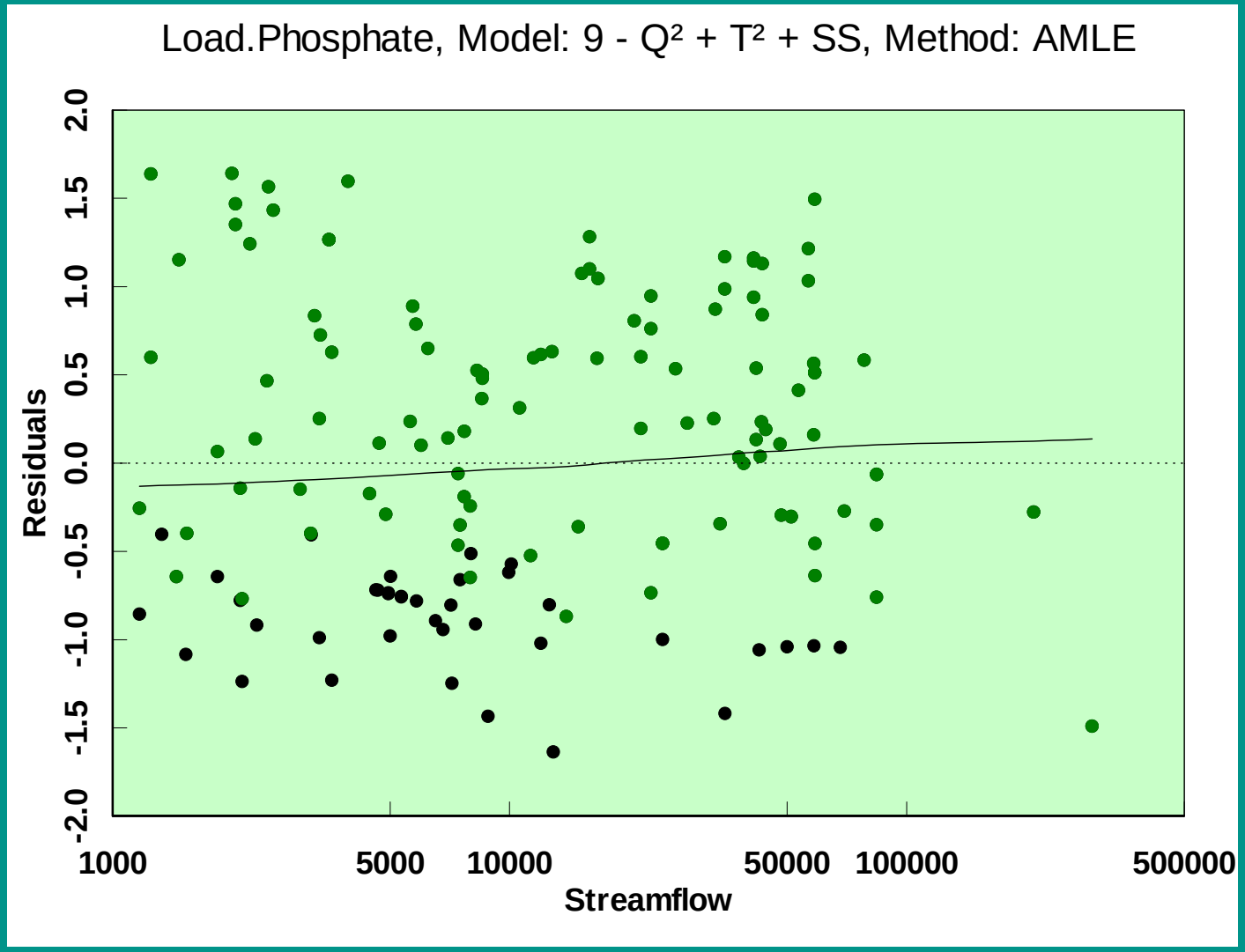
The Seven-Parameter Model



The Seven-Parameter Model



The Seven-Parameter Model



The Seven-Parameter Model

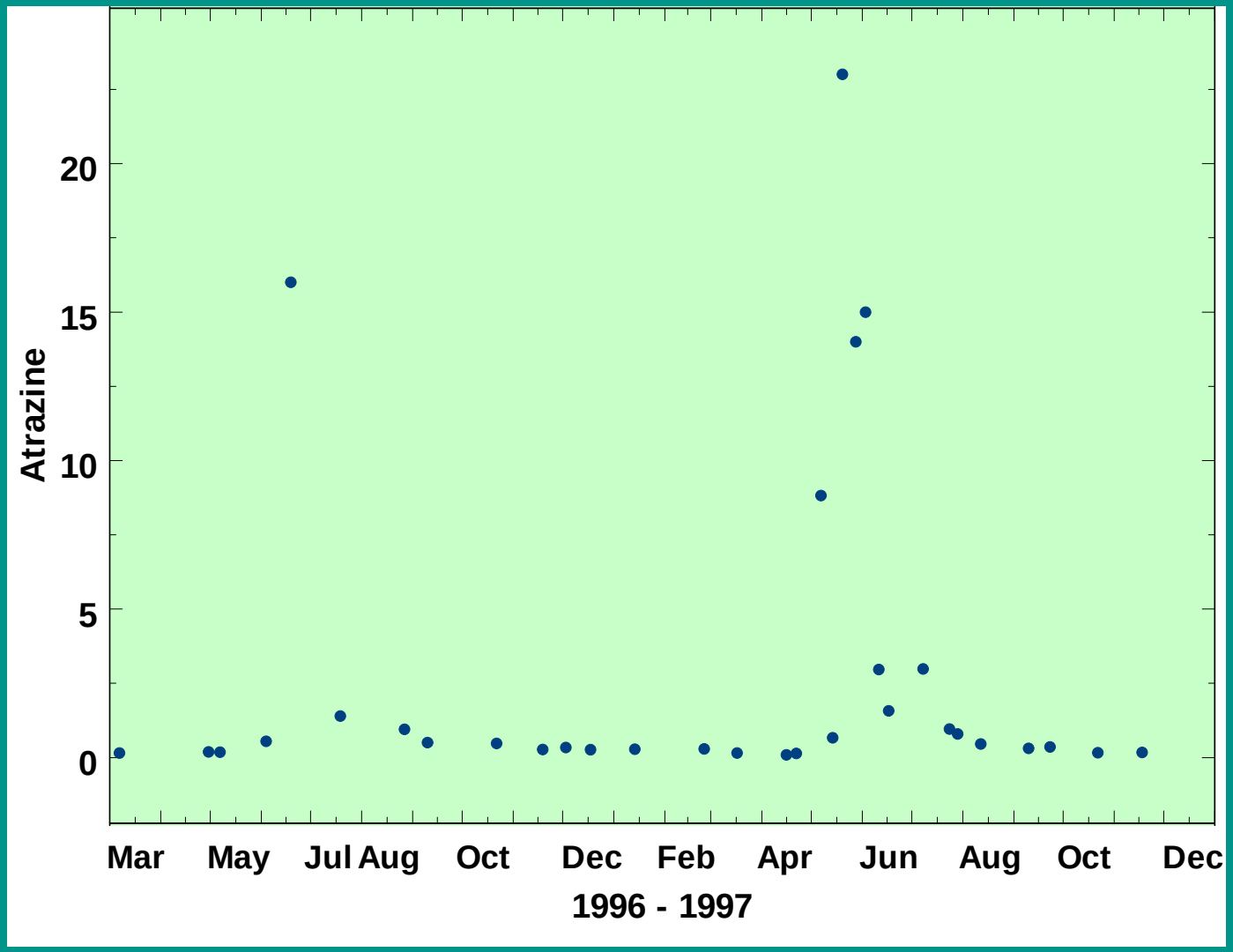
- Load estimates not done for this model.

Custom Seasonal Model

- The example data are atrazine concentrations collected from March 1996 to November 1997 in a river located in the “corn belt”.
- Atrazine exhibits a seasonal pattern, but not one that can be modeled using the Fourier transforms.
- Therefore a dummy variable corresponding to the period of high concentration is used.

USGS

Custom Seasonal Model



Custom Seasonal Model

- **For this model, The log transform is used with linear flow order term.**
- **Period is selected as the seasonal term and the Months are May, June, and July.**
- **The form of the model is**

$$\mathbf{Log(L) = \beta_0 + \beta_1 log(Q) + \beta_2 \mathit{Per} + \beta_3 log(Q)*\mathit{Per},}$$

where **L** is the estimated load, **Q** is the flow, and ***Per*** is the period (May-July).

USGS

Custom Seasonal Model

```
Total number of observations: 32, number of censored obs: 0.
Calibration data period of record: 03/07/1996 to 11/18/1997.
Calibration data streamflow summary statistics:
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
   68.0  230.0   469.0  747.1  951.3  3510.0

Adjusted Maximum Likelihood Estimate of the selected model.
           Coeffient Std. Dev.       P-value
Intercept -0.4196993 0.2154603 4.373650e-002
      lnQ  0.8366360 0.1958792 6.156590e-005
   period  2.3930913 0.3265727 4.808114e-009
 period.Q  1.2005271 0.3391639 5.816384e-004
```

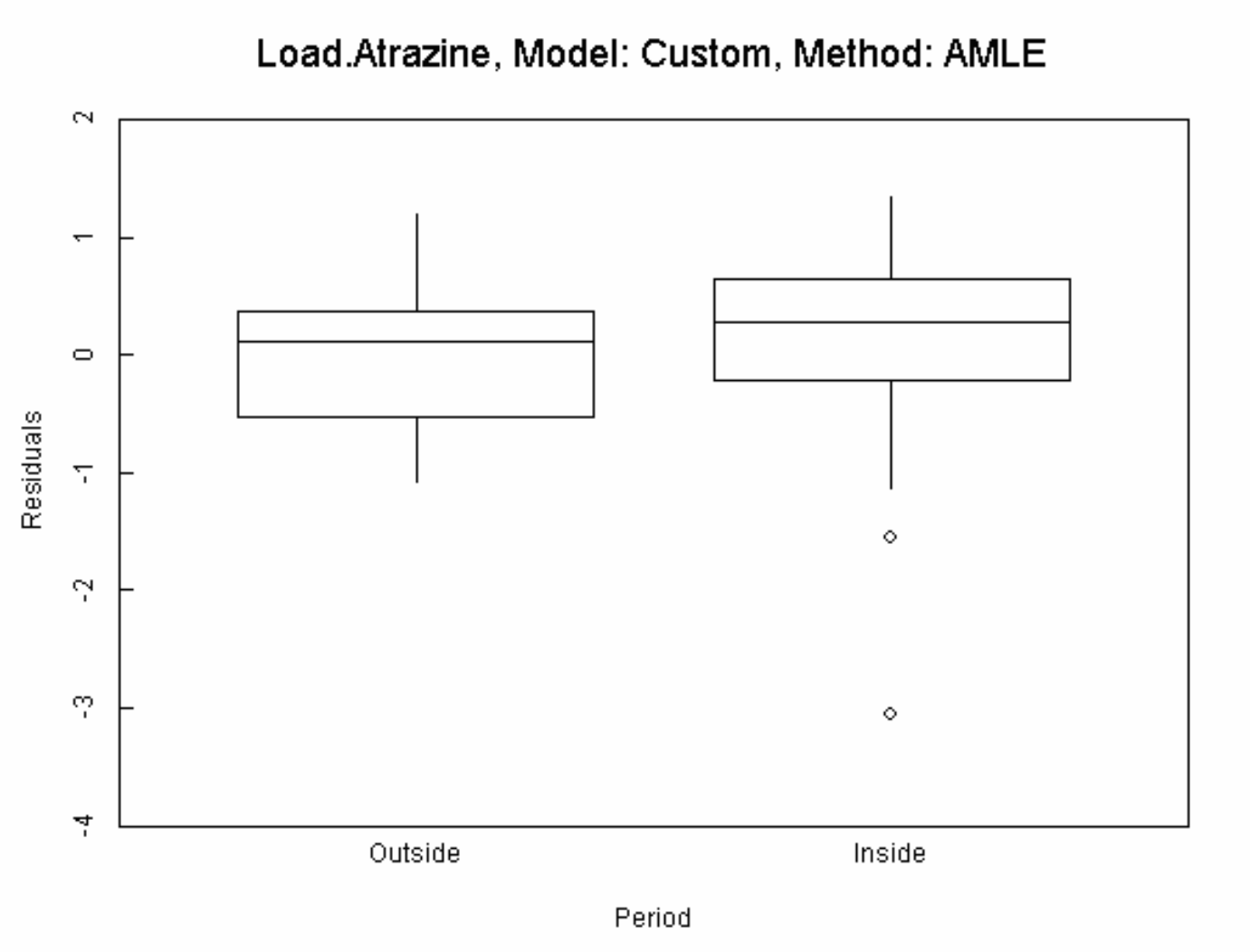
USGS

Custom Seasonal Model

```
Total number of observations: 32, number of censored obs: 0.
Calibration data period of record: 03/07/1996 to 11/18/1997.
Calibration data streamflow summary statistics:
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
   68.0  230.0   469.0  747.1  951.3  3510.0

Adjusted Maximum Likelihood Estimate of the selected model.
           Coeffient Std. Dev.       P-value
Intercept -0.4196993 0.2154603 4.373650e-002
      lnQ  0.8366360 0.1958792 6.156590e-005
   period  2.3930913 0.3265727 4.808114e-009
 period.Q  1.2005271 0.3391639 5.816384e-004
```

USGS

Custom Seasonal Model

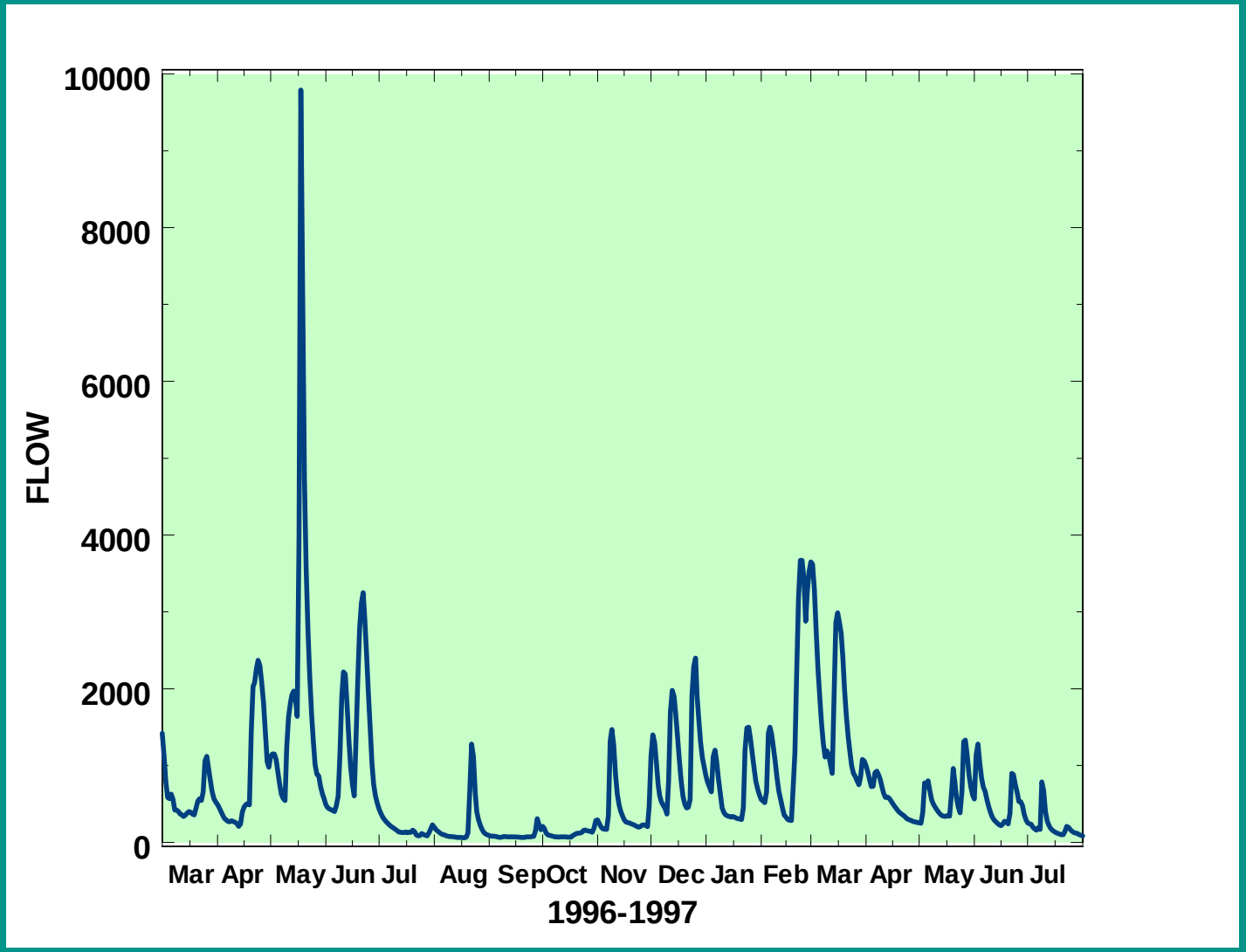


Custom Seasonal Model

- Load estimates for the high periods in 1996 and 1997 are of interest
- The estimation data are from March 1996 through July 1997.
- The only output requested on the Estimation page is the seasonal load for the periods ending in April and July.

USGS

Custom Seasonal Model



Custom Seasonal Model

```
Seasonal loads
                          Flux  Variance  L.95  U.95
SeasonEndingApril1996     1.35      .152   0.7   2.4
 SeasonEndingJuly1996   150.03 23730.777  17.9 574.9
SeasonEndingApril1997     1.24      .150   0.6   2.2
 SeasonEndingJuly1997    14.03    22.148   6.3  27.3
                           SEP NDAYS  Load
SeasonEndingApril1996      .45    61    83
 SeasonEndingJuly1996   163.66    92 13803
SeasonEndingApril1997      .40   273   341
 SeasonEndingJuly1997     5.46    92  1291
```

USGS

Custom Seasonal Model

- The load for the period from May-July 1996 is almost 14 thousand pounds!
- The drainage area of the river is 610 square miles.
- The yield is almost 23 pounds per square mile, which seems too large (application rates are less than 100 pounds per square mile).

USGS

Concluding Remarks

- Load estimation made too easy!
- As indicated by the last example, it is easy to collect the data, build a reasonable model and get unreasonable results.

USGS