Statistical Techniques for Trend and Load Estimation

National Monitoring Conference

Chattanooga, Tennessee

May 17, 2004

Instructors

Skip Vecchia, USGS, Bismarck, ND
(avecchia@usgs.gov)

Tim Cohn, USGS, Reston, VA
(tacohn@usgs.gov)

Dave Lorenz, USGS, Mounds View, MN
(course coordinator, lorenz@usgs.gov)

Course objectives

- ❑ Provide overview of basic statistical methods for detecting trends and estimating loads of chemical constituents in streamflow
- ❑ Introduce software packages (ESTREND, LOADEST, and QWTREND) developed by USGS for trend and load estimation
- ❑ Discuss general data requirements, guidelines, and philosophies for effectively monitoring trends and estimating loads

Course outline

Introduction: Why do we need statistics to detect trends or estimate loads? (Cohn)

Detecting trends in constituent concentrations (Vecchia)

- Overview of statistical methods
- Intro to S-ESTREND
- Data considerations for trend monitoring

Using time-series modeling to analyze trends (Vecchia)

- Overview of time series model and data requirements
- Intro to QWTREND

Lunch

Course outline (continued)

Estimating constituent loads (Cohn)

- Model development
- Unbiased load estimation
- Dealing with censored data
- Uncertainty in estimated loads
- General data requirements
- Intro to S-LOADEST

Discussion and wrap-up (Cohn and Vecchia)

Real-time demo of S-ESTREND and S-LOADEST (if time permits)

Informal poll

What do you want to gain from this course?

- **Basic understanding of statistical methods**
- **I understand the statistics – what about software**
- **Heck with statistics and software – show me how to set up a good monitoring program**

Detecting trends in concentration

What is a "trend" in concentration?

- ❑ Consistent increase or decrease through time
- ❑ Change in concentration resulting from a change in land use
- ❑ Shift in the seasonal timing of high or low concentrations
- ❑ Tendency for high or low concentrations to persist for several years before returning to "normal"
- ❑ A tendency for concentrations to follow a "predictable" pattern

Detecting trends in concentration (2)

It's much easier to define what we mean by "no trend"

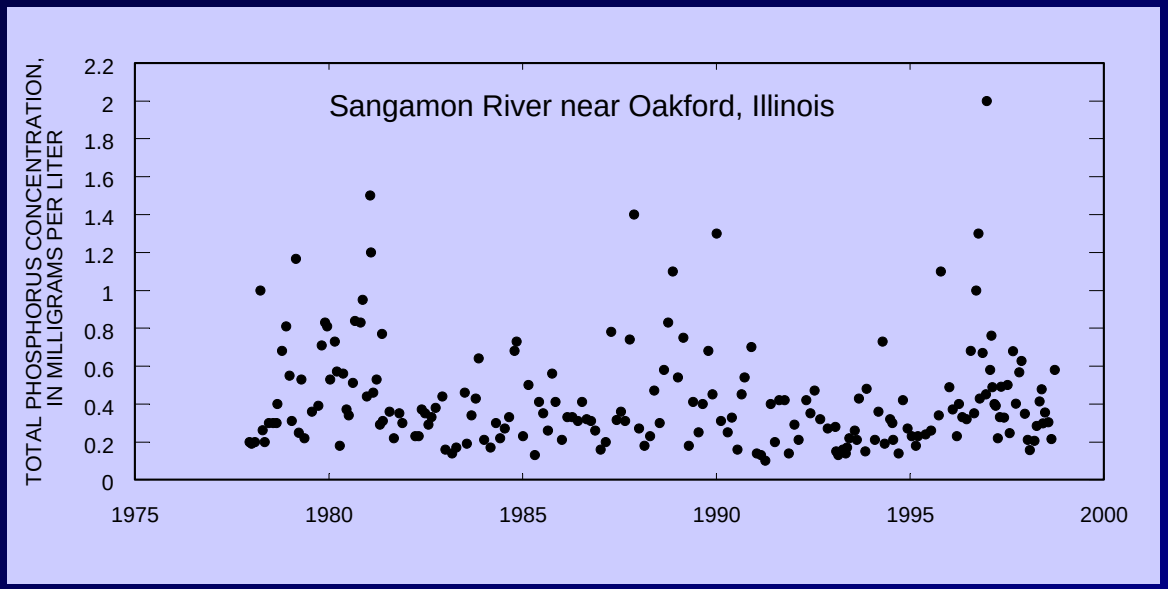
Let C(t) = L(t)/Q(t), where

- C(t) is concentration of a river cross-section, in milligrams per liter, at time t
- L(t) is the flux (milligrams per second) for the cross-section
- Q(t) is streamflow (liters per second) for the cross-section

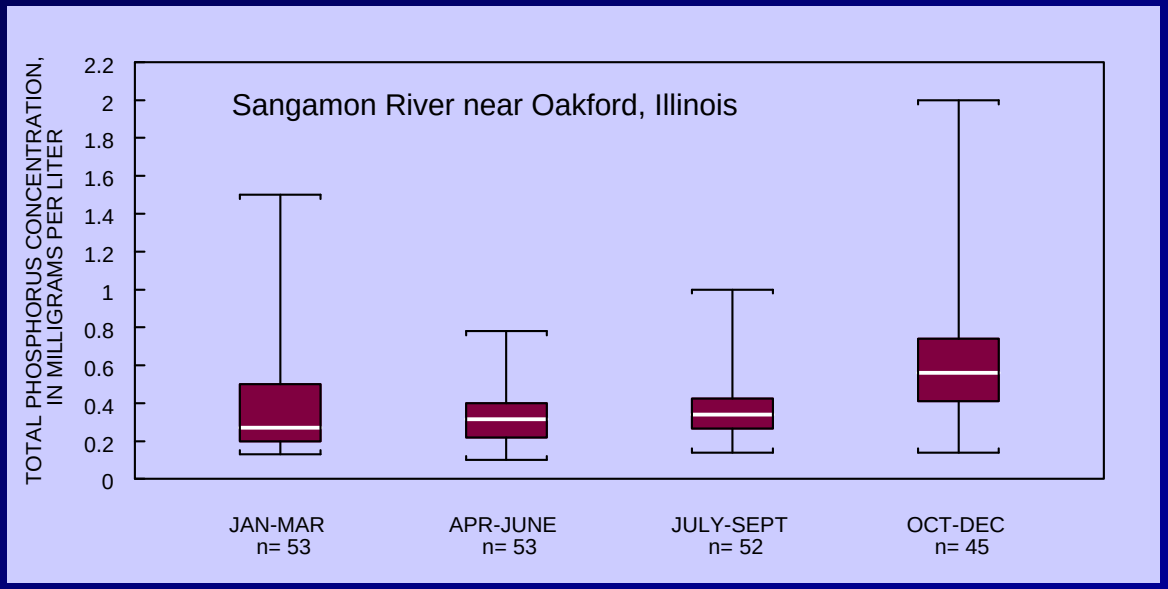
Definition 1: C(t) is trend-free if the probability distribution of C(t) is the same for all t [C(t1) ~ C(t2) for all t1,t2]

Problem: there are few, if any, data sets that are trend-free using definition 1 (Why?)

Example – total phosphorus



Time series plot of recorded data

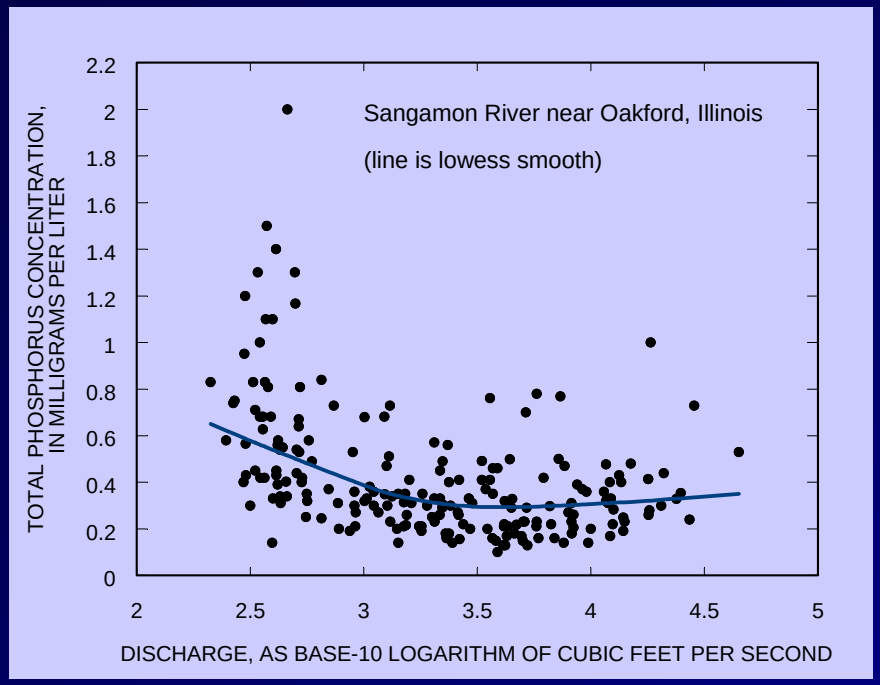


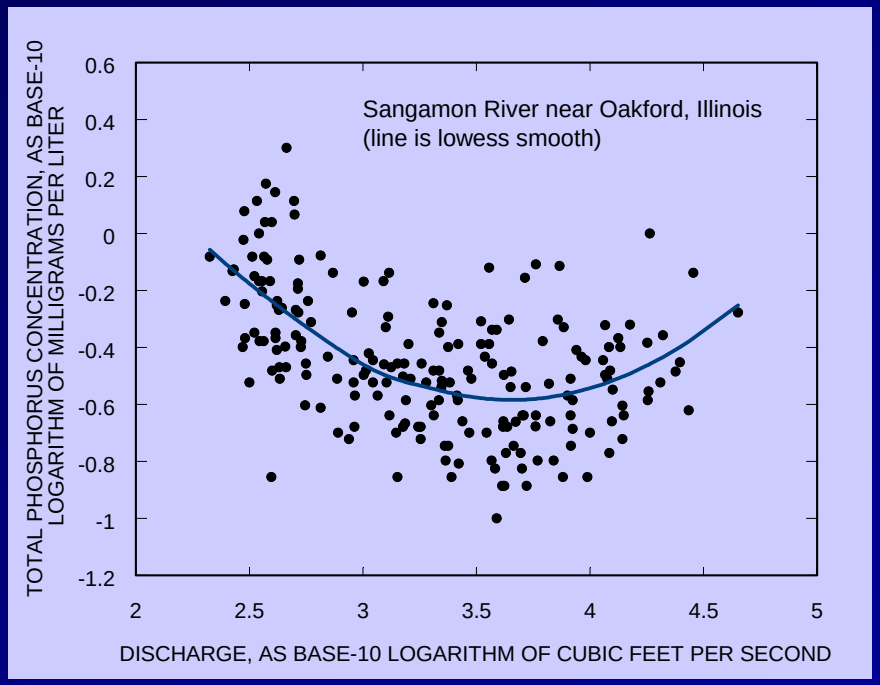
Boxplots of data by season, showing differences in probability distributions

Detecting trends in concentration (3)

- Largely due to the influence of streamflow, raw (unadjusted) concentrations usually
  - have seasonal patterns in their probability distribution
  - are highly variable and skewed (asymmetrical)
- Evaluating the relation between concentration and streamflow is an all-important consideration for interpreting trends.
- We will use a statistical model, called a flow-adjustment model, to allow us to adjust concentrations for variable streamflow conditions.

Example – total phosphorus (2)

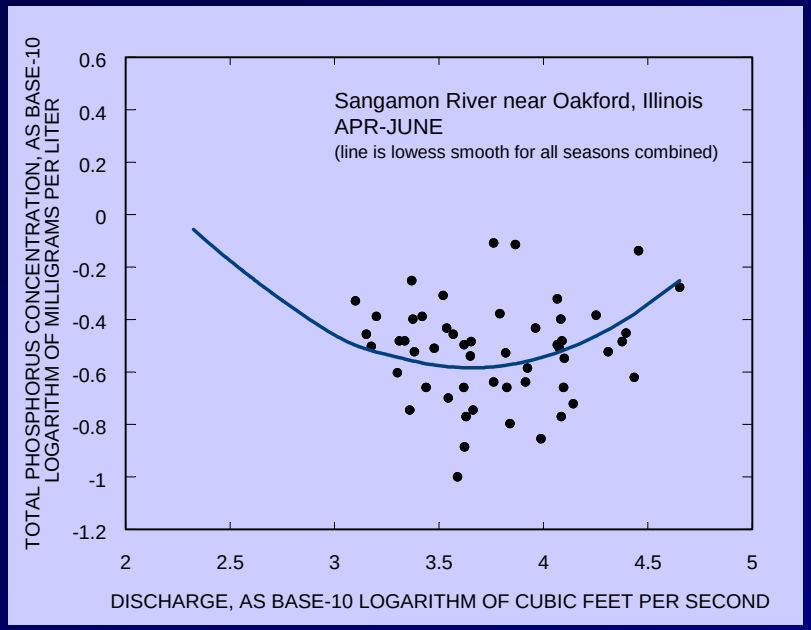


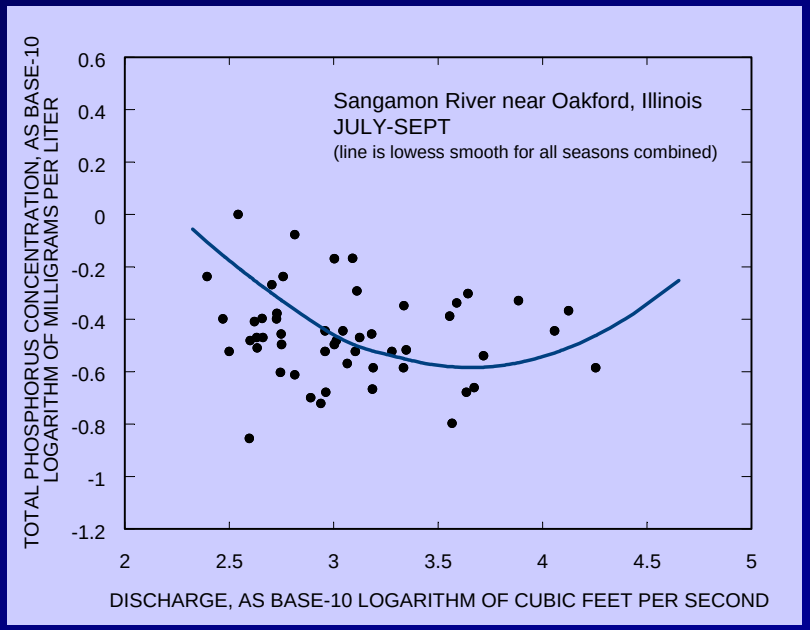


Conc. vs. log-discharge indicates skewed (asymmetrical) residuals and higher variance for low discharges

Log-conc. vs. log-discharge indicates more symmetric residuals and uniform variance, so use log-transformed conc.

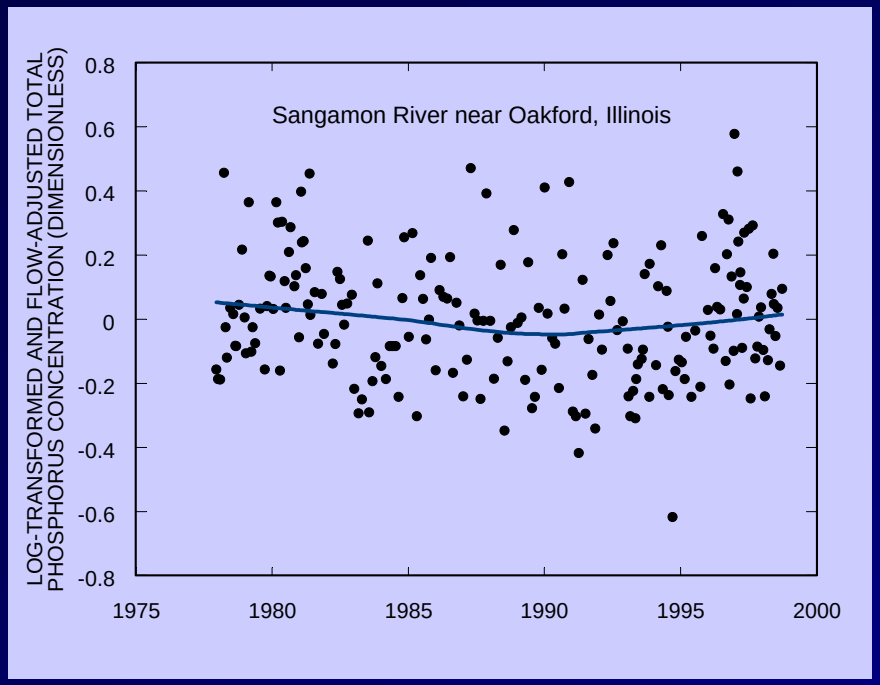
Example – total phosphorus (3)

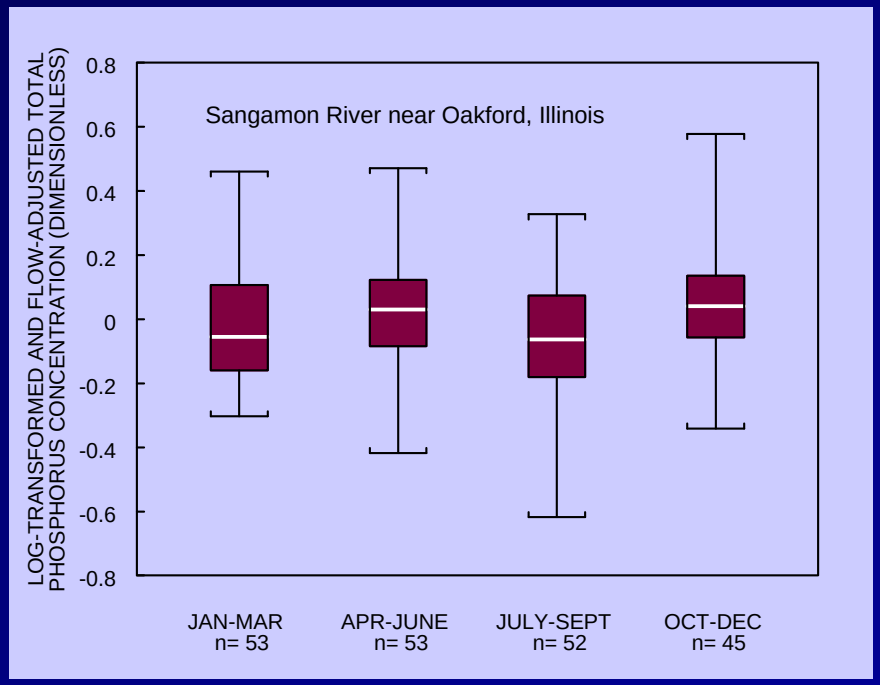




The same flow-adjustment model often works for all seasons. In this case, much of the seasonality in concentration can be explained by streamflow.

Example – total phosphorus (4)





"Flow-adjusted" concentrations (FAC's) are the residuals from the model relating concentration to streamflow. Because conc. was log-transformed, in this case the FAC's are dimensionless.

Flow-adjusment often removes much of the seasonality and results in more "normal" distributions

Detecting trends in concentration (4)

There may be some seasonality remaining in the flow-adjusted concentrations, because of seasonality in the sources or chemical processes that occur independently of streamflow. We can deal with the seasonality in a couple of ways:

(1) Remove the seasonality by including periodic functions (sines and cosines of various frequencies), along with streamflow, in the flow-adjustment model, or

(2) Look for trends that occur among samples collected during a particular season (trend in spring? fall?)

We will focus on (2) for now, and use an appropriate definition of “trend-free” that applies to seasonal data

Detecting trends in concentration (5)

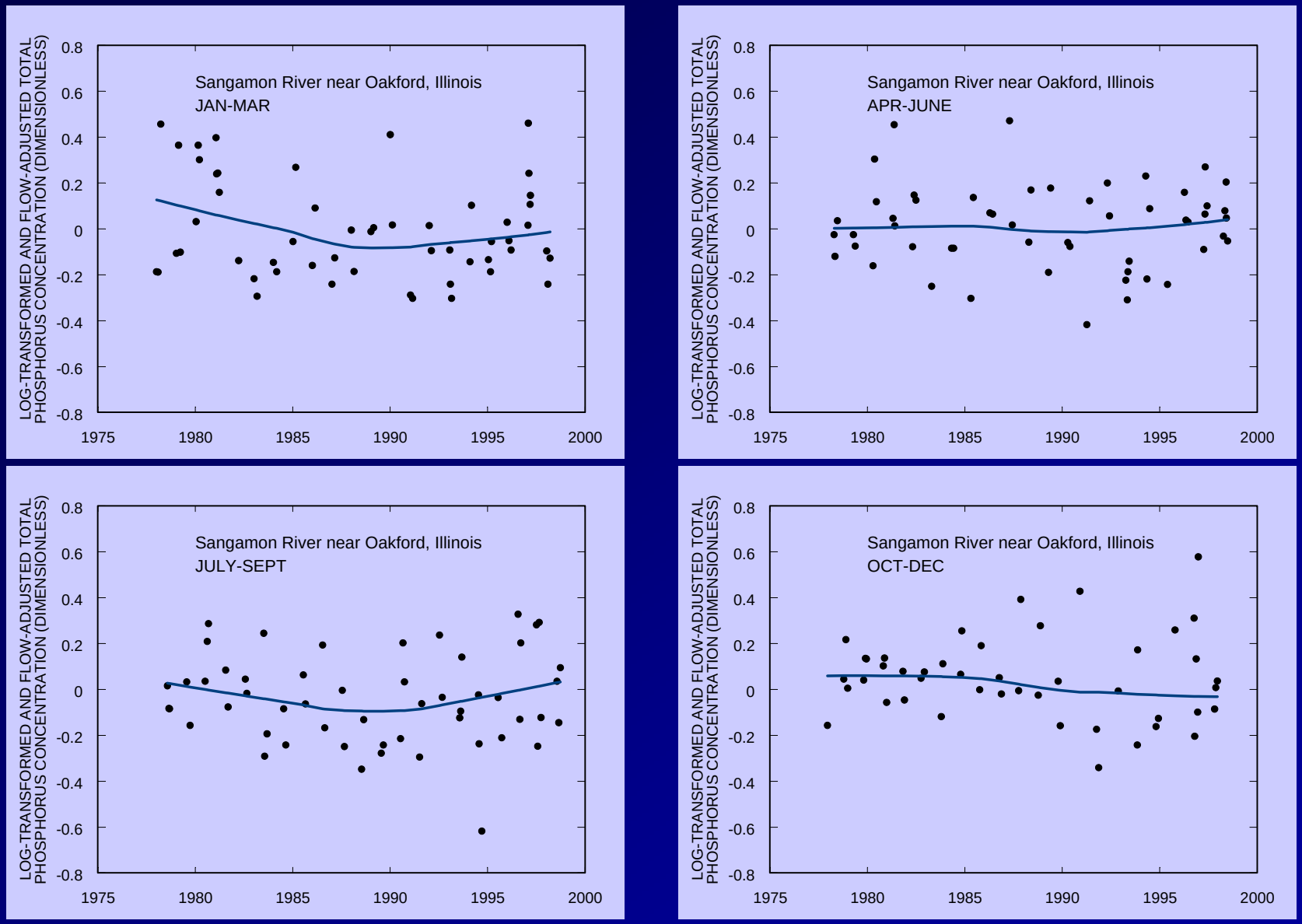
Null hypothesis for detecting seasonal trends [Let C(i, j) be concentration for year i and season j]

NH1 (without flow-adjustment): The probability distribution of C(i, j) depends on the season, but not the year, in which samples are collected [C(i1, j) ~ C(i2, j) for all years i1 and i2]

NH2 (with flow-adjustment): The probability distribution of FAC(i, j) depends on the season, but not the year, in which samples are collected [FAC(i1, j) ~ FAC(i2, j) for all i1 and i2]

It is recommended to use NH2 whenever possible. If the data indicate that the null hypothesis is not true (with a specified sig. level), a trend is indicated. Determining what the trend looks like or what caused the trend may take a good bit of ingenuity.

Example – total phosphorus (5)



Detecting trends in concentration (6)

An infinite variety of "trends" can occur

- A trend might occur uniformly throughout the year
- Trends may be occurring in some seasons but not others
- Conflicting trends (some up, some down) might occur among different seasons
- Trends may be monotonic (consistent increase or decrease through time) or non-monotonic (reverse direction)
- Monotonic trends may be linear (the same slope through time) or nonlinear

Detecting trends in concentration (7)

The ability to detect a trend depends on

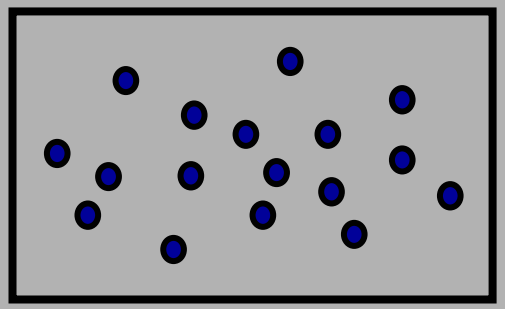
- The complexity and duration of the trend
- The number and timing of the water quality samples
- The quality of the data (below detection limit?, sampling or laboratory error?)
- The availability of good streamflow data
- The statistical method used to detect the trend

Seasonal-Kendall (SEAKEN) trend test for uncensored data

- ❑ Nonparametric test (based on ranks, so doesn't depend on an assumed distribution)
- ❑ Can be used for both raw and flow-adjusted concentrations
- ❑ Designed to detect monotonic trends in seasonal data
- ❑ Acceptable for records with at most 5 percent censored values and a single censoring threshold
- ❑ There are a number of requirements regarding sample size, definition of seasons, etc (considered later)
- ❑ Two versions of the test: one (uncorrected) that assumes no cross-correlation between seasons and another (corrected) that accounts for cross-correlation.

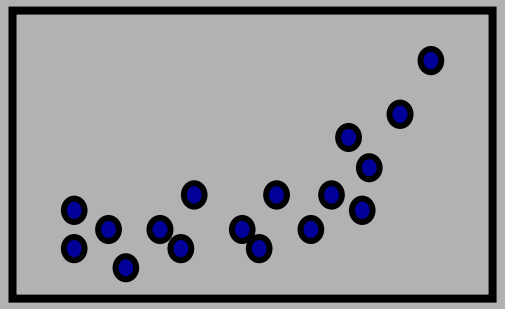
Kendall's τ

- Null hypothesis: $\tau = 0$
  - Y and X independent



Alternate hypothesis: τ not = 0

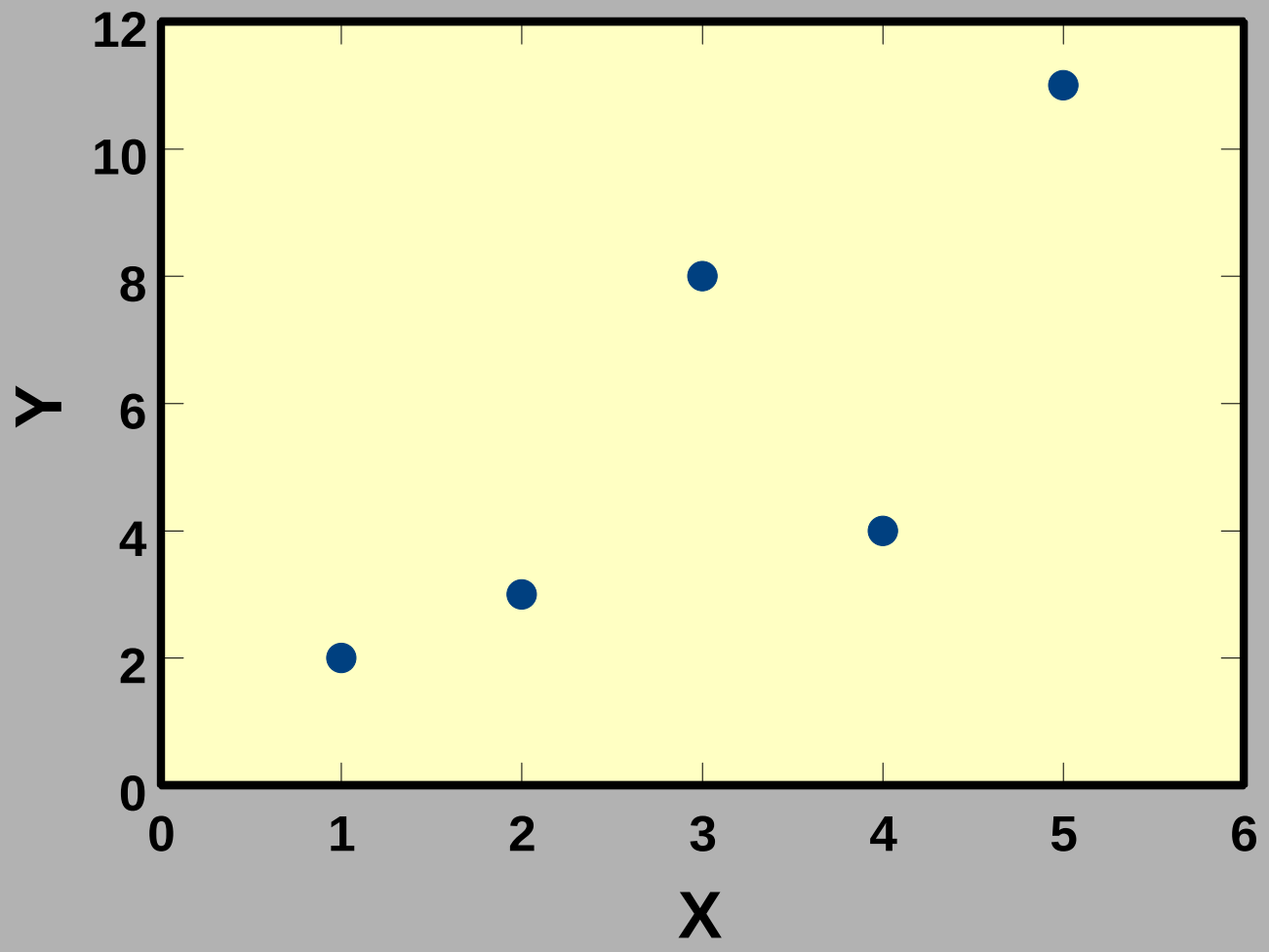
- **Y varies monotonically with X**



Computing Kendall's τ

Sort by X

X	Y
1	2
2	3
3	8
4	4
5	11



Computing Kendall's τ

S = total pluses – total minuses

$$\tau = \frac{S}{\text{total pairwise comparisons}}$$

$$= \frac{S}{n\,(n-1)\,/\,2}$$

For the example:

$$\tau = \frac{8}{5\,(5-1)\,/\,2}$$

$$= 0.80$$

Test for significance

How likely is it to get a τ of 0.80 if the null hypothesis is true?

$$\sigma_S = \sqrt{\left(\frac{n}{18}\right)(n-1)(2n+5)}$$

$$Z_\tau = \frac{|S| - 1}{\sigma_S}$$

Compare Z_τ to the standard normal distribution

For the example:

$$\sigma_S = \sqrt{\left(\frac{5}{18}\right)(5-1)(2(5)+5)}$$

$$= 4.08$$

$$Z_\tau = \frac{8-1}{4.08}$$

$$= 1.75$$

$$p = 0.080$$

Therefore, the trend (correlation between X and Y) is significant

Correction for ties

If any X,Y pairs have a common X or Y value, the pairwise comparison is a "tie".

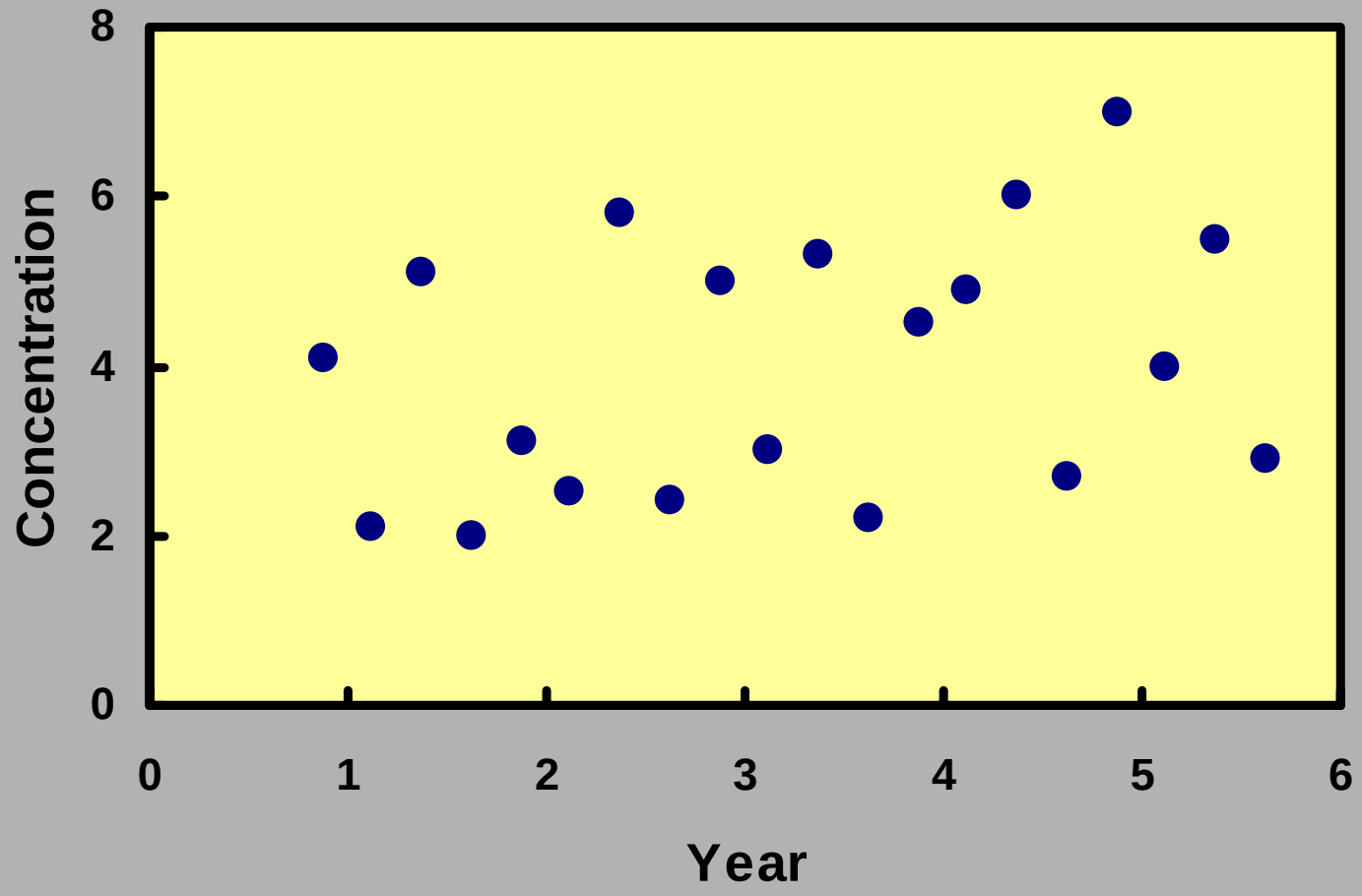
Ties do not contribute to S, but they do contribute to σ_S :

$$\sigma_S = \sqrt{\frac{\left\{ n(n-1)(2n+5) - \sum t_i (i)(i-1)(2i+5) \right\}}{18}}$$

Where: i = the number of points in a tied group

t_i = the number of groups with i ties

Seasonal Kendall Test



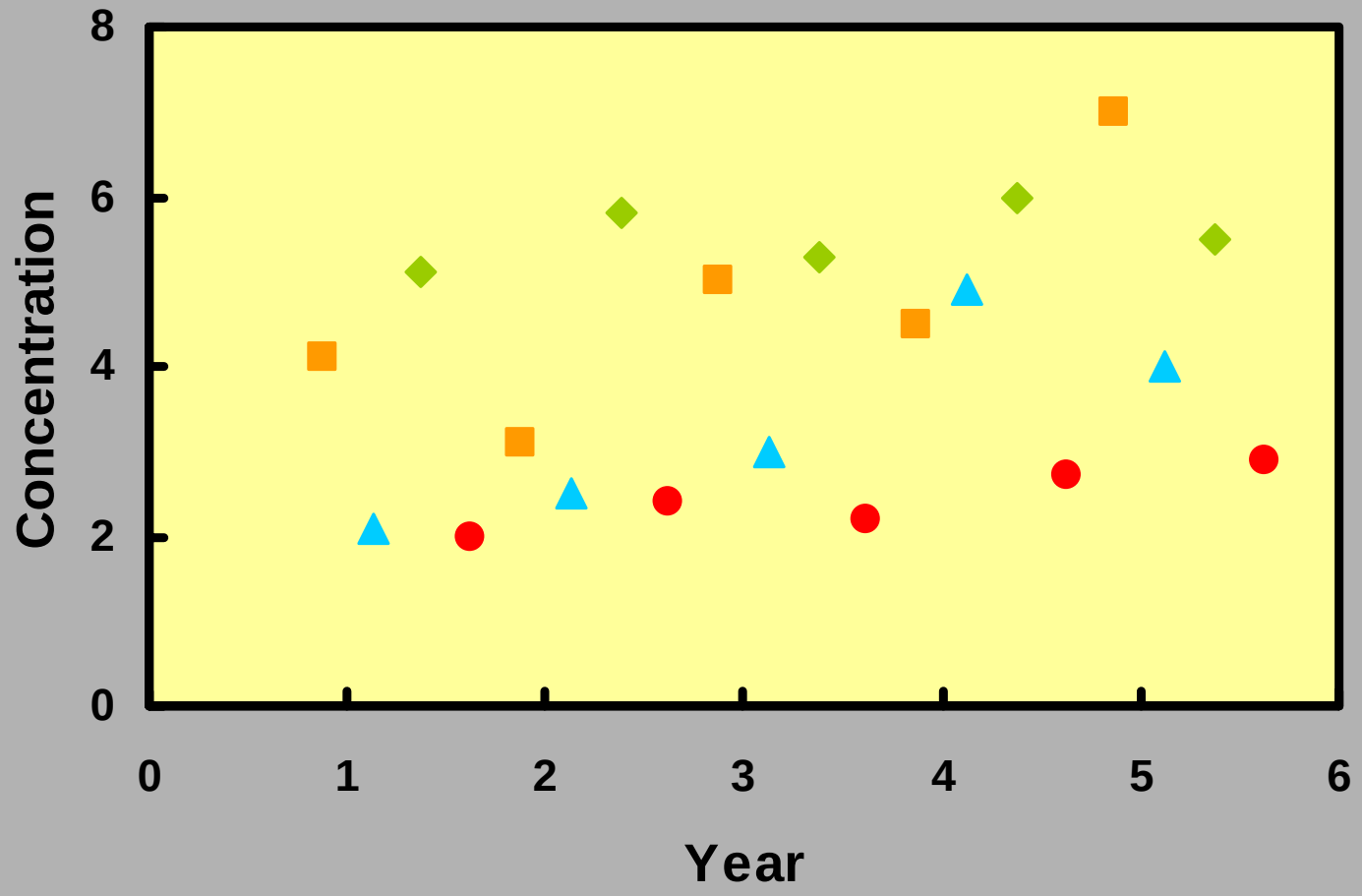
116 pluses
74 minuses

S = 42

τ = 0.22

P = 0.183

Seasonal Kendall Test



33 pluses
7 minuses

S = 26

τ = 0.65

P = 0.002

Slope = 0.375

Typical "seasons" for water-chemistry trends are monthly and quarterly

Seasonal Kendall Test

S is computed using pluses and minuses from pairwise comparisons within seasons.

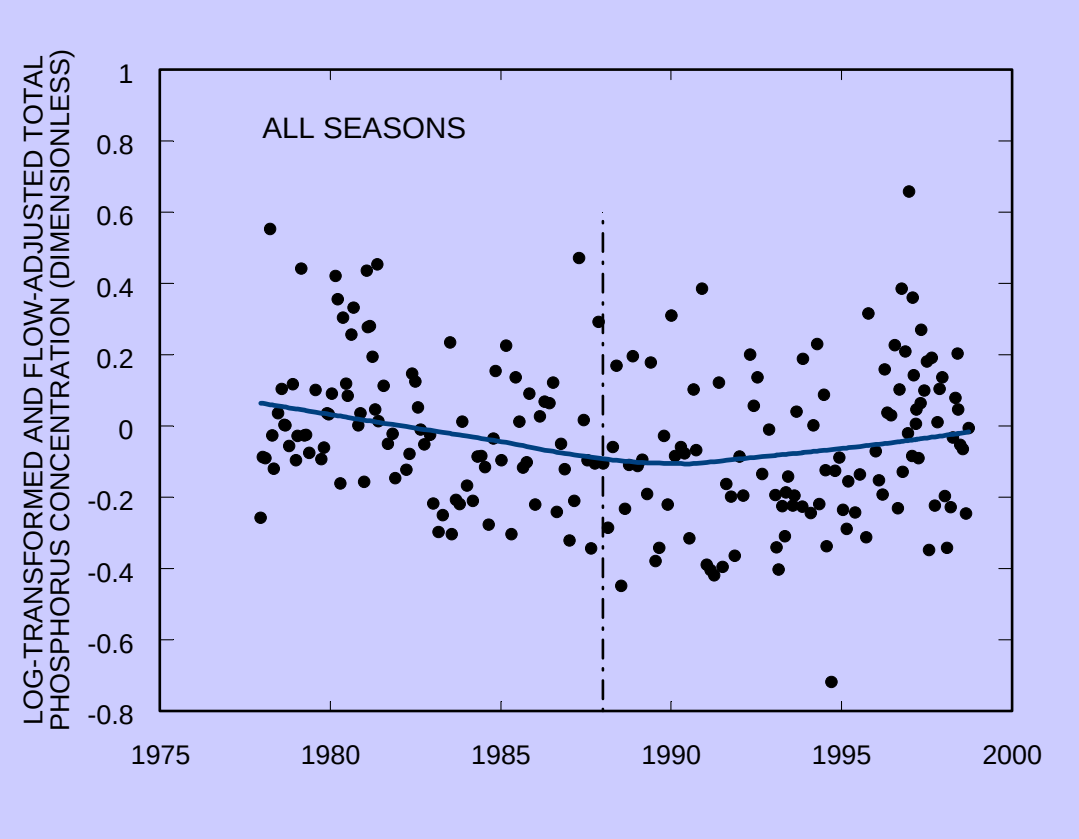
σ_S is computed using the total number of seasonal pairwise comparisons.

For more than 10 years of data, σ_S must be adjusted to account for serial correlation within seasons.

Slope is the median of the pairwise slopes within seasons.

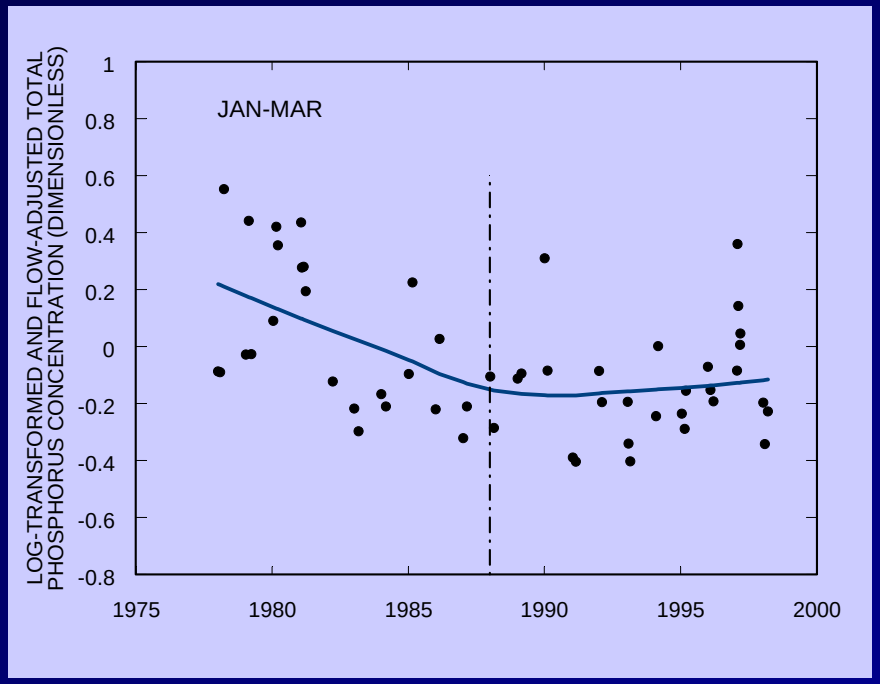
Example – SEAKEN TEST for (slightly modified) total phosphorus concentrations

Tau = - 0.08 (p = 0.11) for 1975-2000
Tau = - 0.15 (p = 0.043) for 1975-1987
Tau = 0.17 (p = 0.012) for 1988-2000

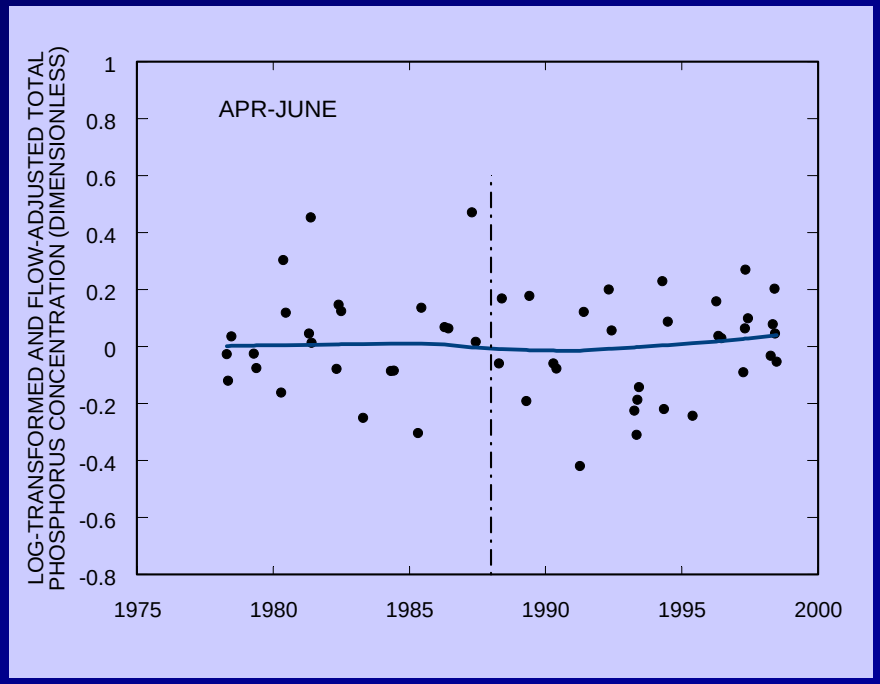


Example – SEAKEN TEST (continued)

Tau = - 0.24 (p = 0.01) for 1975-2000
Tau = - 0.39 (p = 0.009) for 1975-1987
Tau = 0.11 (p = 0.398) for 1988-2000



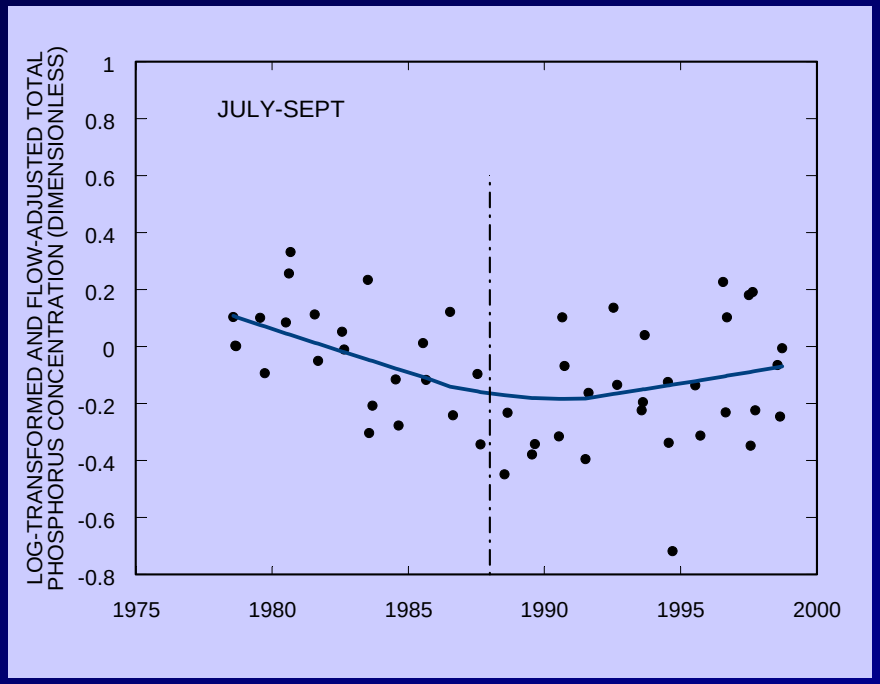
Tau = 0.03 (p = 0.73) for 1975-2000
Tau = 0.10 (p = 0.526) for 1975-1987
Tau = 0.13 (p = 0.335) for 1988-2000

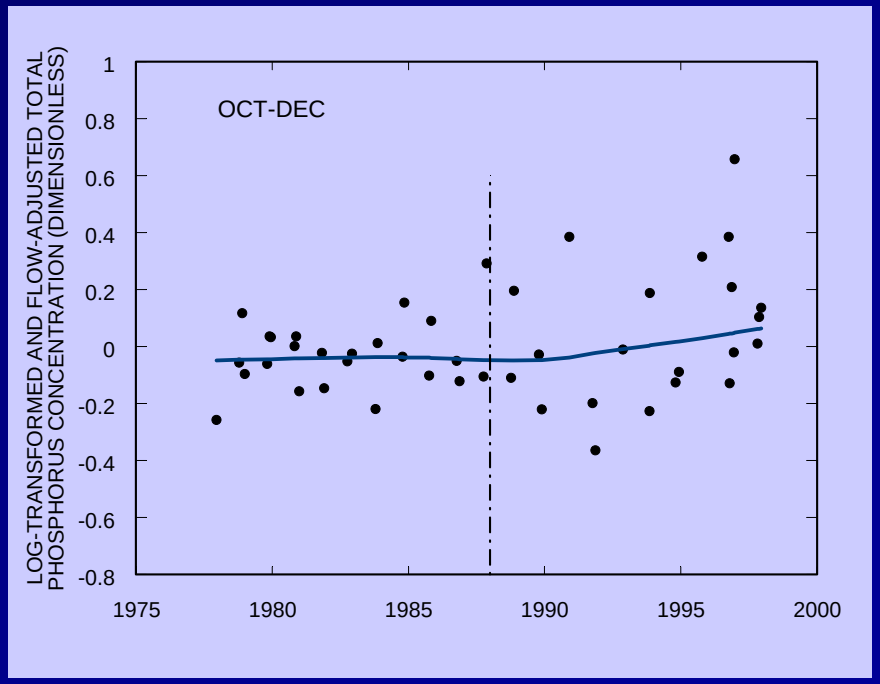


Example – SEAKEN TEST (continued)

Tau = - 0.19 (p = 0.048) for 1975-2000
Tau = - 0.36 (p = 0.017) for 1975-1987
Tau = 0.24 (p = 0.075) for 1988-2000

Tau = 0.15 (p = 0.153) for 1975-2000
Tau = 0.05 (p = 0.747) for 1975-1987
Tau = 0.25 (p = 0.116) for 1988-2000





Example – SEAKEN test (continued)

An insignificant result could mean

- The trend is not monotonic
- Conflicting trends occur among seasons
- Too few data were collected to detect the trend
- There is, in fact, no trend

A significant result indicates that a trend did occur. However, the trend may not be monotonic and it may not be the same for each season. There's no substitute for graphical analysis and common sense when interpreting trend results!

SEAKEN trend test for censored data

- Acceptable for records with more than 5 percent censored values (must set a single threshold value)
- Can be used only on raw concentrations (can't flow-adjust if too many concentration values are censored)
- Just like regular SEAKEN test, but there are lots of "ties" because every censored value is assigned the same rank
- Generally not recommended if more than 50 percent of values are censored

Minimum data requirements for applying SEAKEN test (censored or uncensored)

- At least 5 years of record, with at most one observation per season (multiple observations need to be culled)
- Number of observations > max(10, 3*number of seasons) [more data required as degree of censoring increases]
- If missing data for some seasons, the number and length of seasons should be defined so that, for each season, at least 50 percent of the possible pairwise comparisons can be made among both the middle 3/5 of the record and among the beginning and ending 1/5's of the record

Defining seasons for SEAKEN test (example) [X denotes an observation]

Season		BEG 1/5		MIDDLE 3/5						END 1/5	
		1	2	3	4	5	6	7	8	9	10
1	Jan	X	X	X	X			X	X		
	Feb	X				X	X			X	X
	Mar							X	X		
2	Apr	X		X	X	X	X				X
	May	X						X	X		X
	Jun	X				X	X			X	
3	Jul	X	X	X	X			X	X	X	X
	Aug		X			X	X				X
	Sep	X						X	X		X
4	Oct	X	X	X	X	X	X			X	
	Nov	X							X	X	
	Dec	X		X		X	X				X

Advantages and disadvantages of nonparametric methods

Advantages (+) and disadvantages (-) of nonparametric methods such as Kendall's tau:

+ Easy to compute
+ Requires few assumptions
+ Robust to outliers
+ Handles censoring (single threshold, no flow-adjustment)

- Assumes monotonic trend
- No multiple detection limits
- Must define seasons, perhaps combine or throw out data
- Difficult to interpret trends in "ranks"

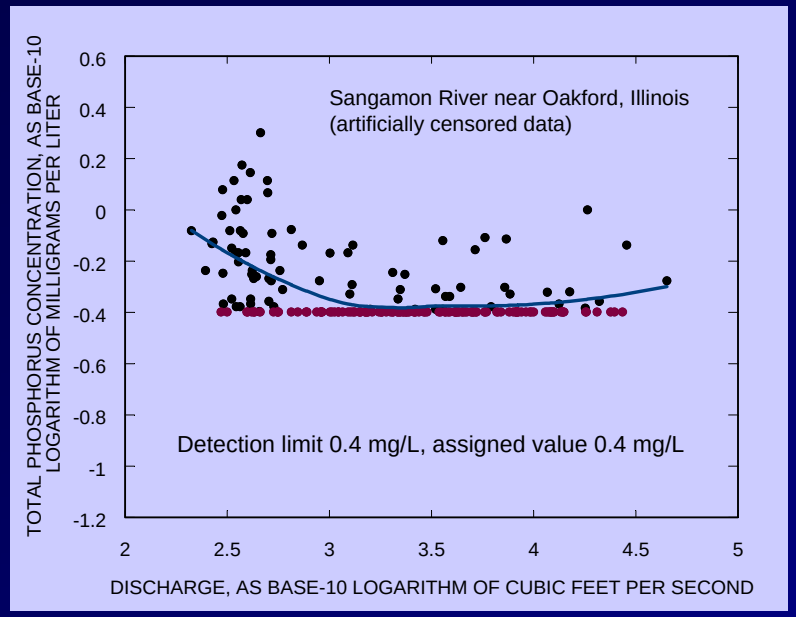
Parametric regression for censored data (Tobit regression)

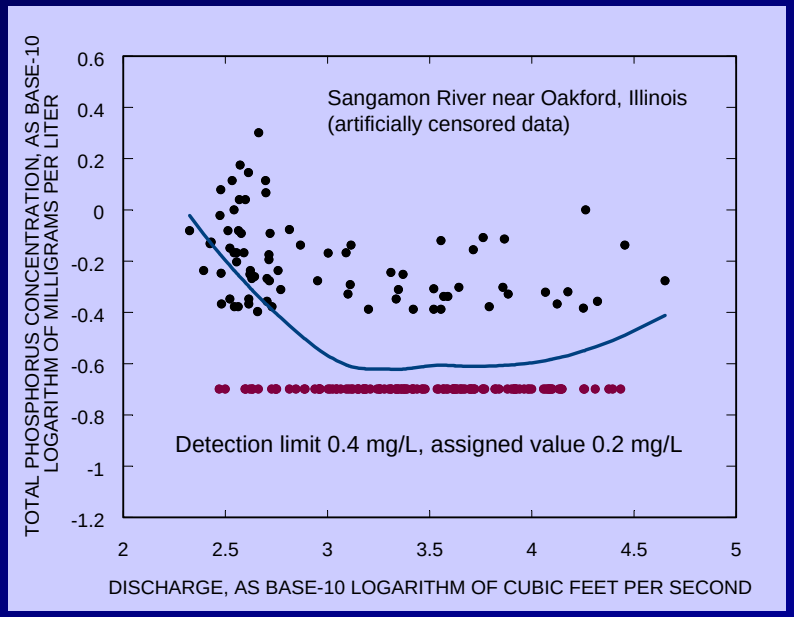
Tobit regression uses maximum likelihood estimation to fit a parametric model with censored data. Like non-parametric methods, there are both advantages and disadvantages:

- Requires specification of a parametric model
- Requires care fitting the model and verifying assumptions
- May require more data than non-parametric methods

+ Can be used to model complex (non-monotonic) trends
+ Multiple detection limits, lots of censored data OK
+ Good for explanatory, not just exploratory, analysis
+ Uses full power and flexibility of maximum likelihood theory
+ Can be used to estimate constituent loads

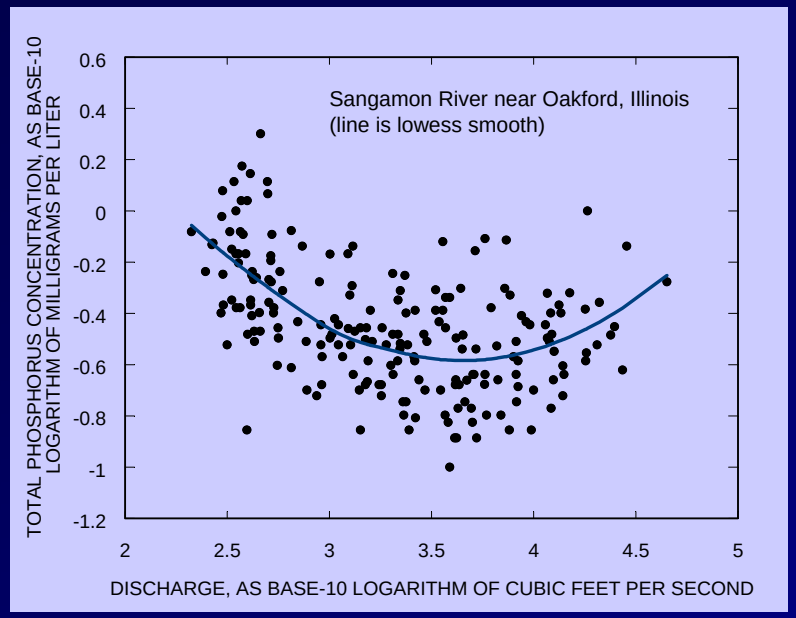
Tobit regression example

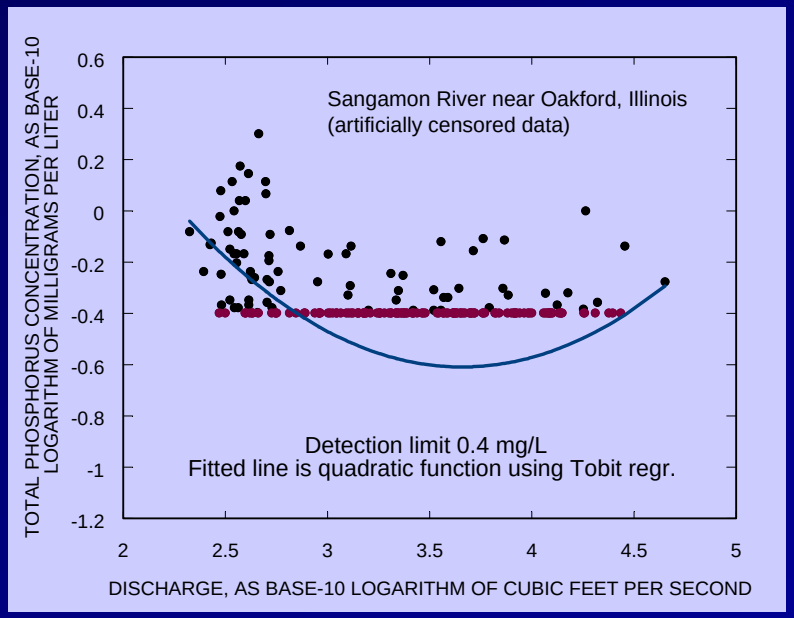




Treating censored data as “known” values can radically alter the model, depending on what values are assigned to censored observations

Tobit regression example - continued





Tobit regression fits a parametric model without assigning values to censored data (only “less thans”). The line on the right is the fitted equation,

$$\log(\text{Conc}) = a0 + a1 \log(\text{Discharge}) + a2 [\log(\text{Discharge})]^2$$

Overview of S-ESTREND

- Original software package (ESTREND) developed by Schertz, Alexander, and Ohe (USGS WRIR 91-4040)
- New version is an S-Plus 6.1 for Windows application developed by Dave Lorenz of the USGS
- Well-suited for trend analysis of seasonal data with many stations and constituents (regional or national investigations)
- Many choices for flow-adjustment models + good model diagnostics (residual plots, etc)
- Performs seasonal Kendall test (uncensored or censored) and Tobit analysis

ESTREND Basics

A system for estimating monotonic trends in water quality

U.S. Department of the Interior
U.S. Geological Survey

Requirements

- S-PLUS 6.1 (should also work on 6.2).
- The USGS library is available at: **http://water.usgs.gov/software/library.html.**

USGS

Step by step process

1. Initial screening of data.
2. Graphically explore the data.
3. Define seasons.
4. Select flow-adjustment model.
5. Run trend tests.

USGS

Initial screening of data

- Designed to screen large retrievals of data (many stations and constituents).
- Printed report shows summary statistics for each station and constituent.
- The printed output is sorted by constituent.
- A summary data set is also created.
- Stations and constituents that do not meet selection criteria can be eliminated from subsequent analysis.

USGS

Initial screening ouput

```
Constituent:  Total.P

   Station:  07300000
106  observations from  10/25/1974  to  08/16/1989
28  less than values and  0 zero values ( 26.4  %)
Statistics of the raw data
    0%  10%  25%  50%  75%  90%  100%
 0.003 0.01 0.01 0.02 0.05 0.09 0.951

Constituent:  Chloride

   Station:  07300000
275  observations from  10/01/1968  to  08/16/1989
0  less than values and  0 zero values ( 0  %)
Statistics of the raw data
   0%   10%  25%  50%  75%  90% 100%
   65 150.4  210  260  300  330  530
```

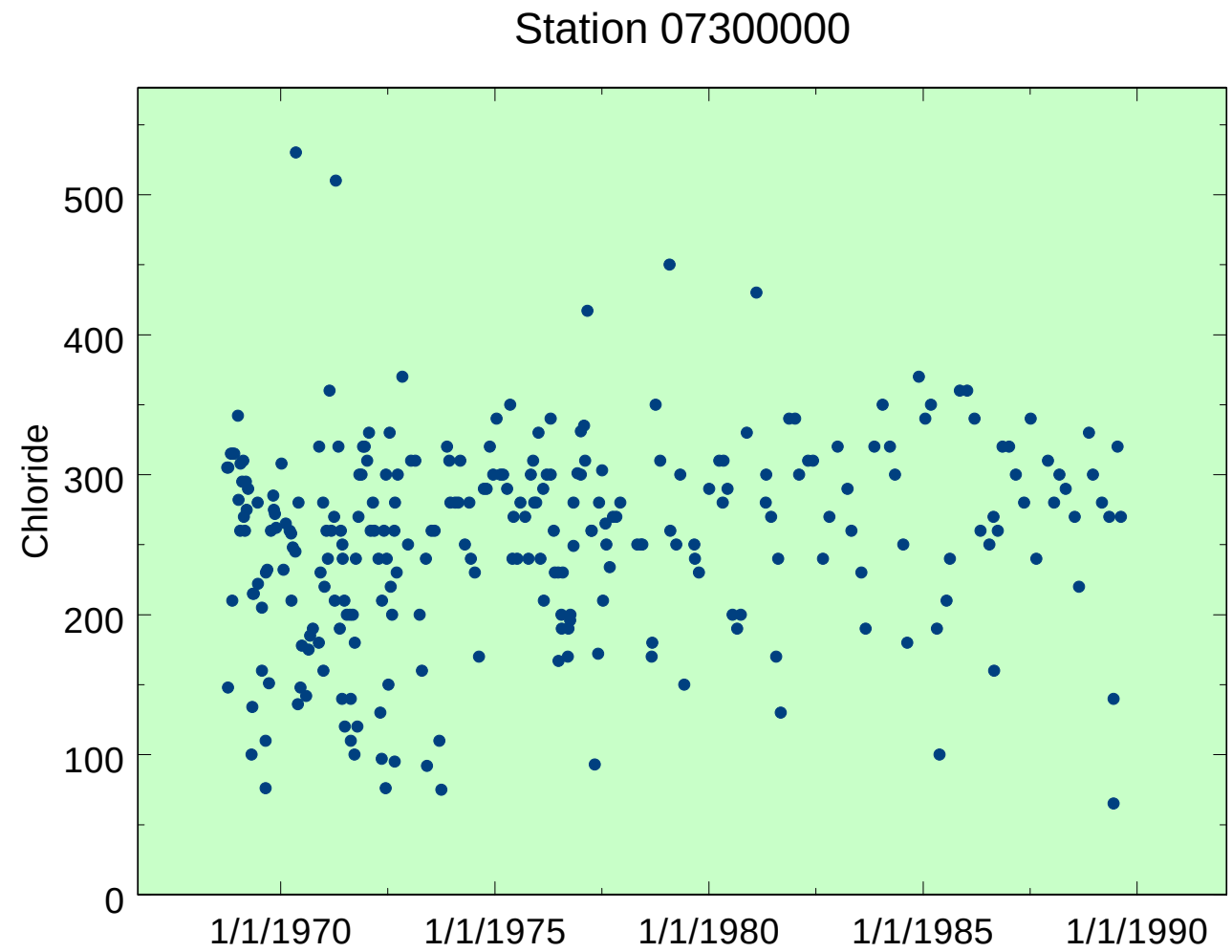
•

Graphically explore the data

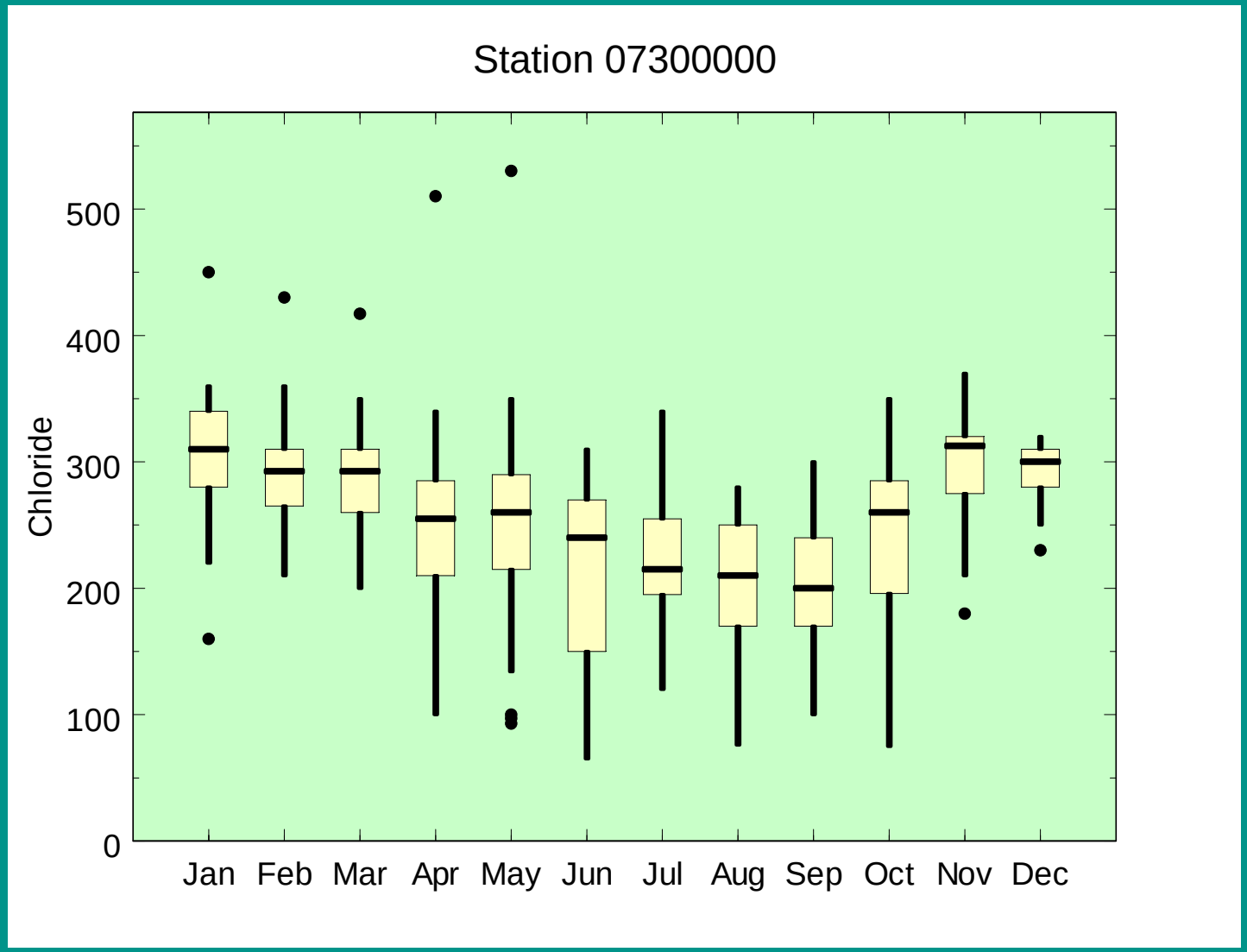
- Three plots are available, shown on next 3 slides.
- Time-Series plot shows the distribution of concentration over time.
- Monthly Boxplots show the range of concentration by month.
- Barcharts of annual samples show the number of samples per year.

USGS

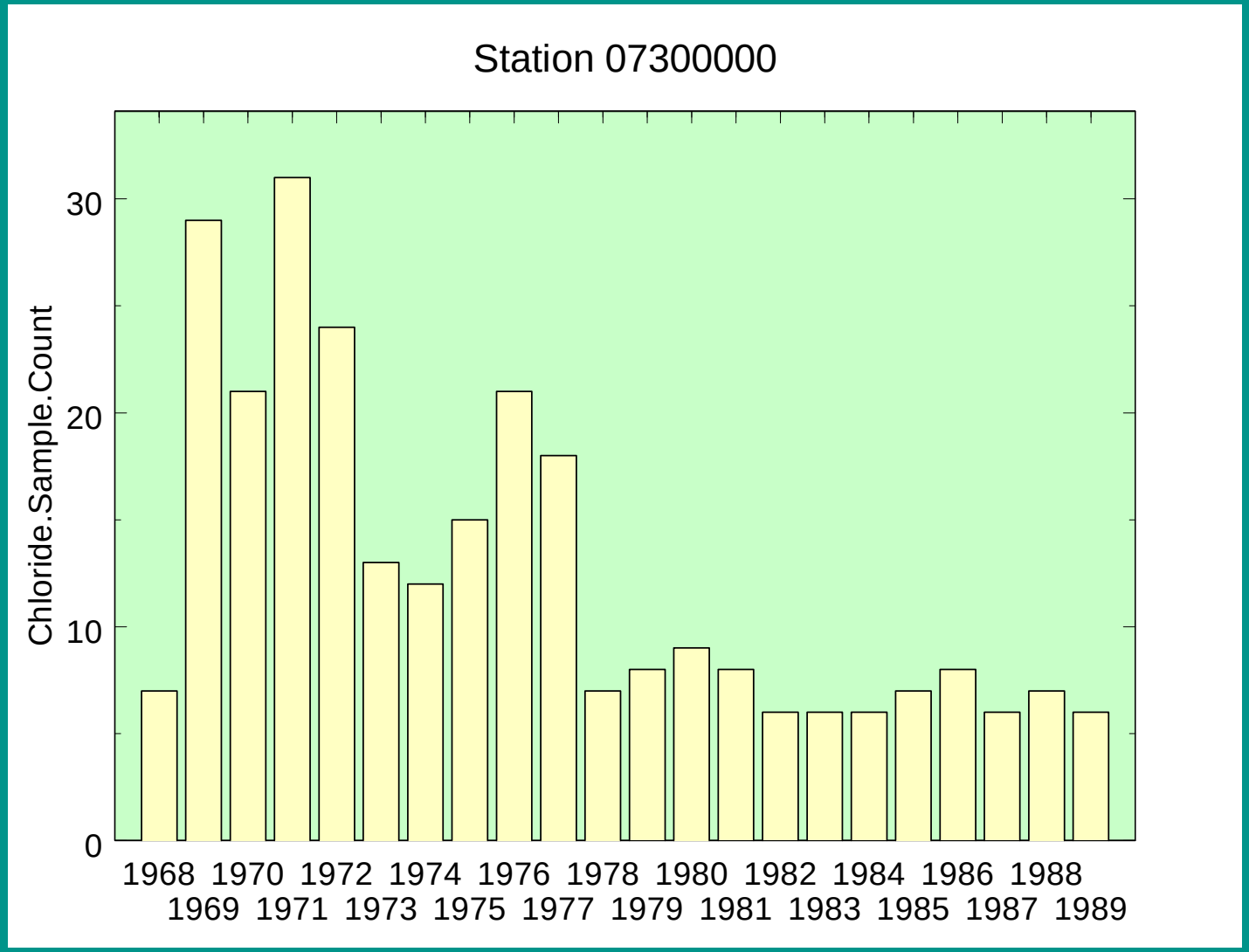
Time-Series Plot



Monthly Boxplots



Barchart of annual samples

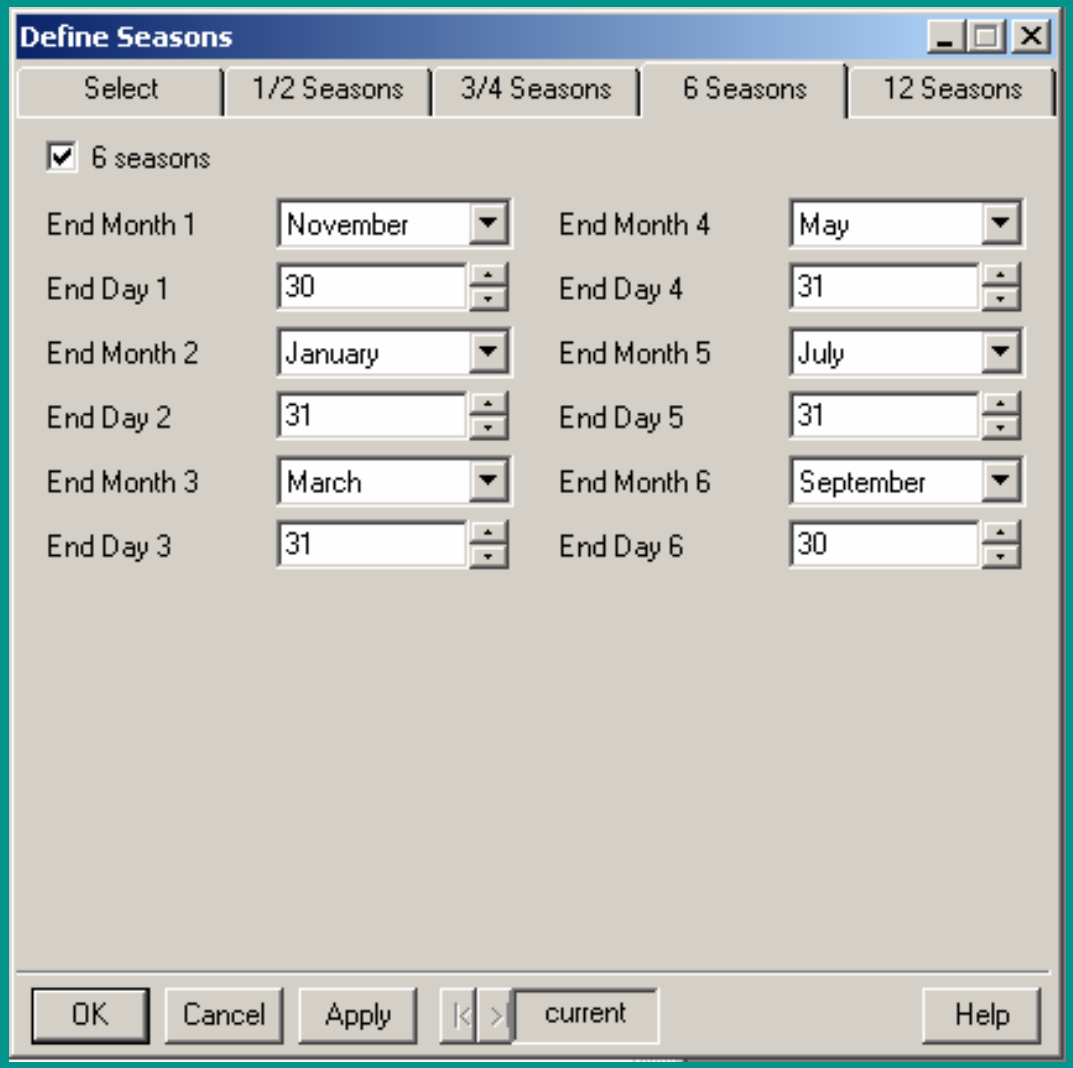


Define Seasons

- Several pre-defined seasonal periods are available in ESTREND, all based on the water year.
- 2 periods of six months.
- 3 periods of four months.
- 4 periods of three months.
- 6 periods of two months.
- 12 periods of one month.

USGS

Define Seasons



Define Seasons

- This analysis divides the data into two sections: the beginning and ending (BE) and the middle (MI).
- The BE section is composed of approximately the first and last 1/5th of the period of record and the middle is the remainder.
- This analysis calculates how many pairwise comparisons can be made within each section for each period.

USGS

Define Seasons

```
Constituent:  Chloride

    Station: 07300000 Record length -- 20.9  years,
             samples 10/01/1968 to 08/16/1989
       Period of seasonal analysis: 1969 to 1989
       Beginning-ending record number of years = 8
       Middle record number of years = 13
  Percent of possible within-season comparisons for the
  6-season definition.
       1      2      3   4   5   6
BE  87.5 100.00 100.00 100 100 100
MI 100.0  92.31  92.31 100 100 100
```

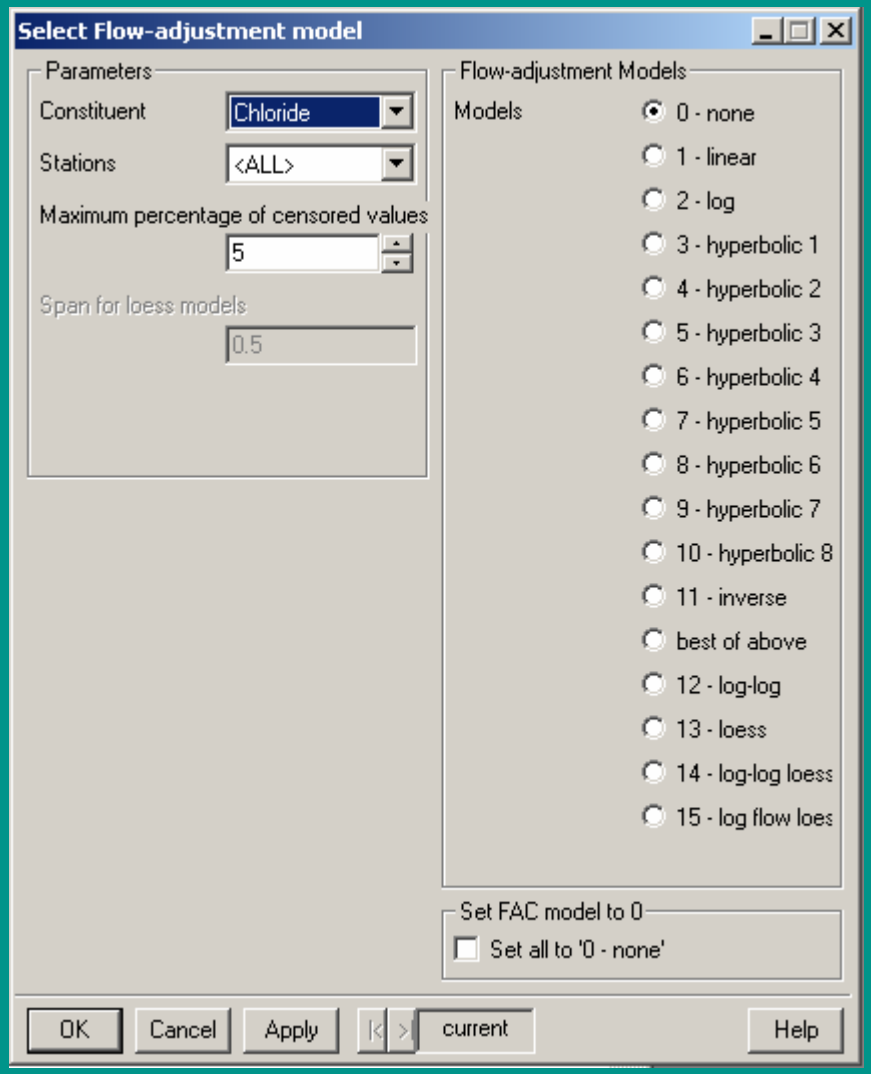
USGS

Select Flow Adjustment Model

- Several pre-defined models that relate concentration to flow are available.
- There are 11 linear regression models that include linear-, log-, reciprocal- and hyperbolic-transforms of flow.
- Several log concentration models are also available, including LOWESS models.

USGS

Select Flow Adjustment Model



Select Flow Adjustment Model

```
Constituent:  Chloride
 Station: 07300000
  Selecting 'best' model for station 07300000
 Press for model 1 1394358.72953868
 Press for model 2 1343088.2247578
 Press for model 3 1130297.45447716
 Press for model 4 1214408.18589433
 Press for model 5 1344245.66778992
 Press for model 6 1423339.03263617
 Press for model 7 1432994.54395171
 Press for model 8 1421928.12045204
 Press for model 9 1436116.67238252
 Press for model 10 1519723.34889941
 Press for model 11 1836065.81687706
 Model number 3 selected as the 'best' model
```

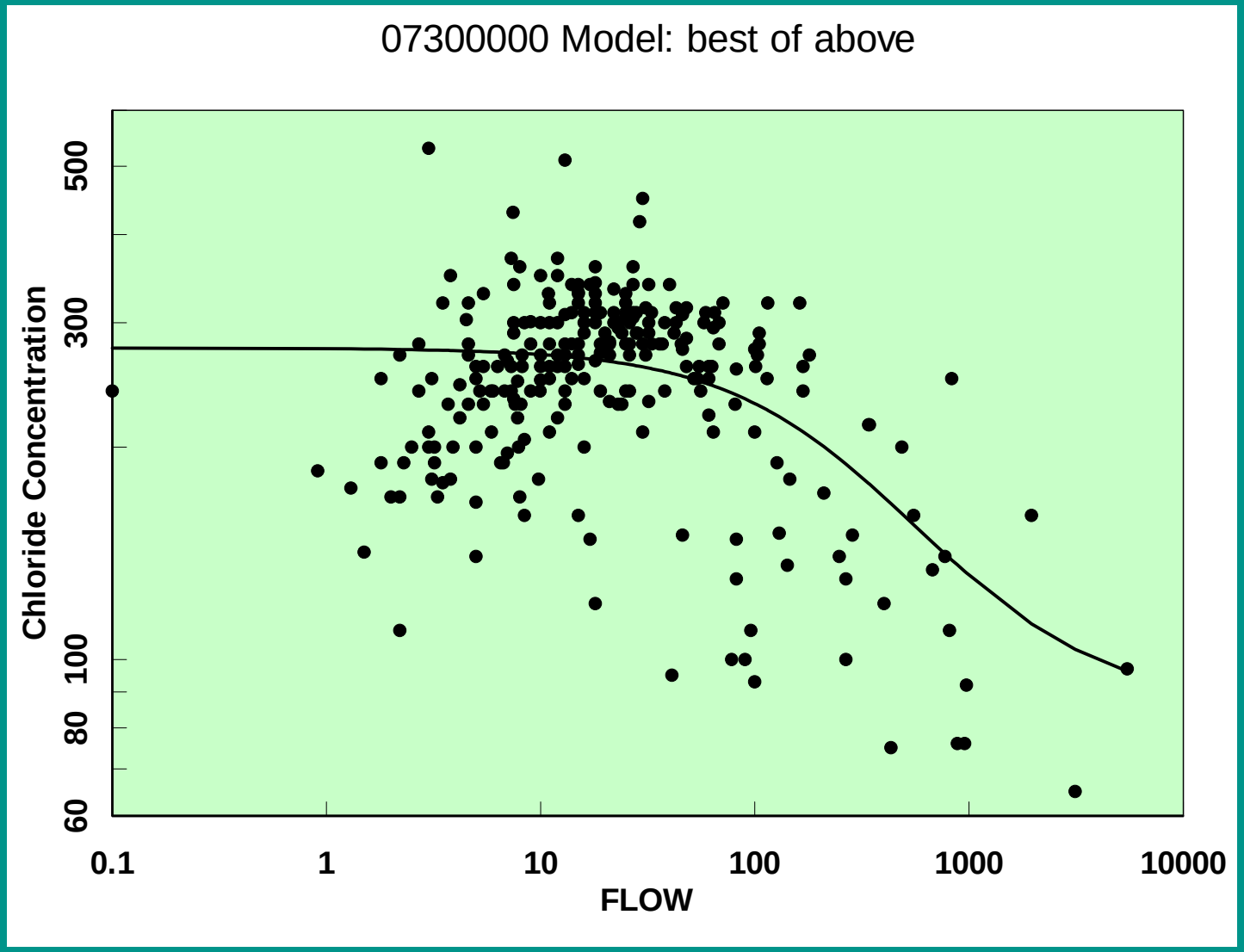
USGS

Select Flow Adjustment Model

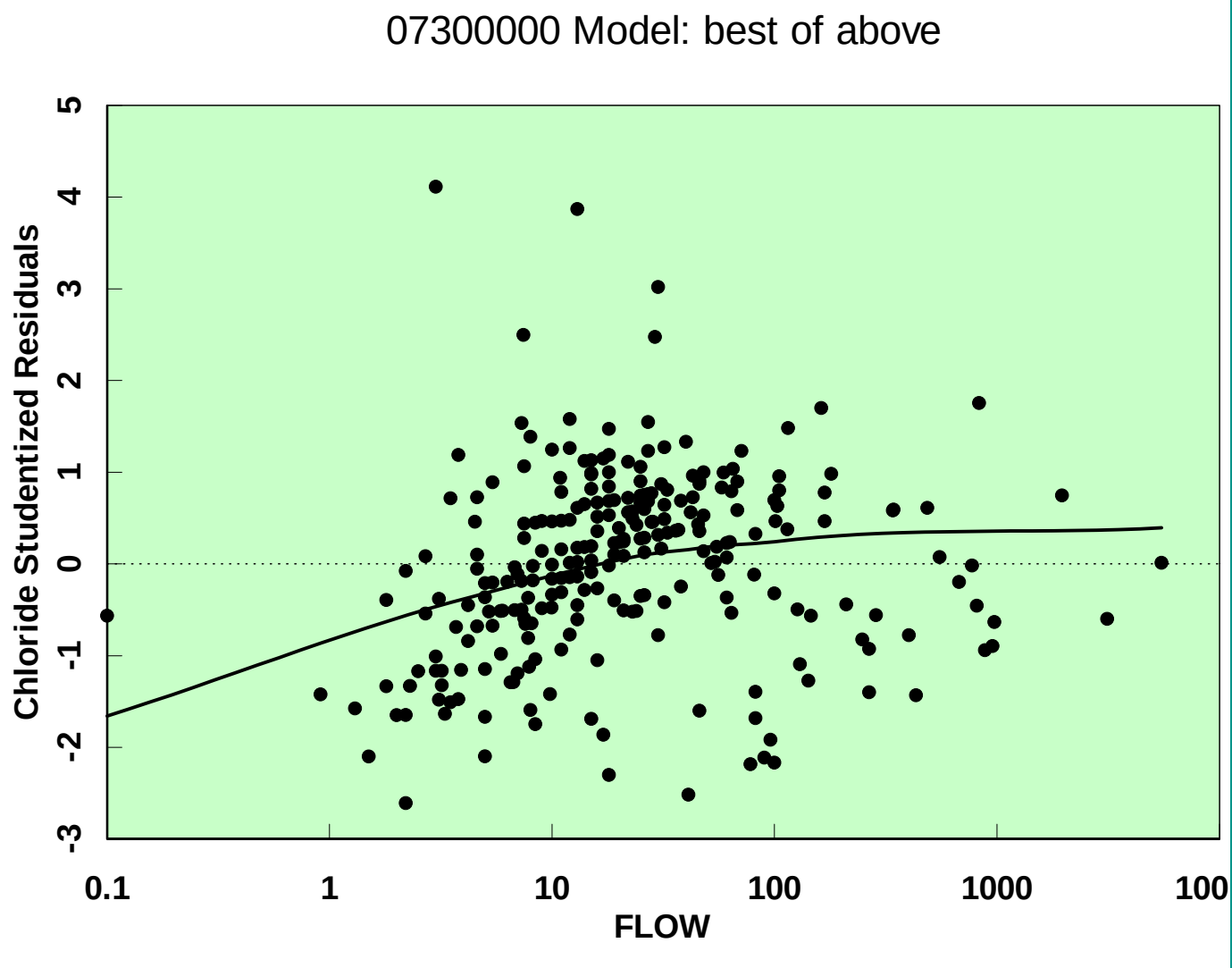
```
Residuals:
  Min      1Q Median     3Q   Max
 -165 -36.19  4.777 44.05 255.5

Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)  85.8663  20.1393     4.2636   0.0000
      Indep 190.4248  22.4035     8.4998   0.0000

Residual standard error: 63.93 on 273 degrees of
   freedom
Multiple R-Squared: 0.2093
F-statistic: 72.25 on 1 and 273 degrees of freedom,
   the p-value is 1.221e-015
```

USGS

Select Flow Adjustment Model



Select Flow Adjustment Model

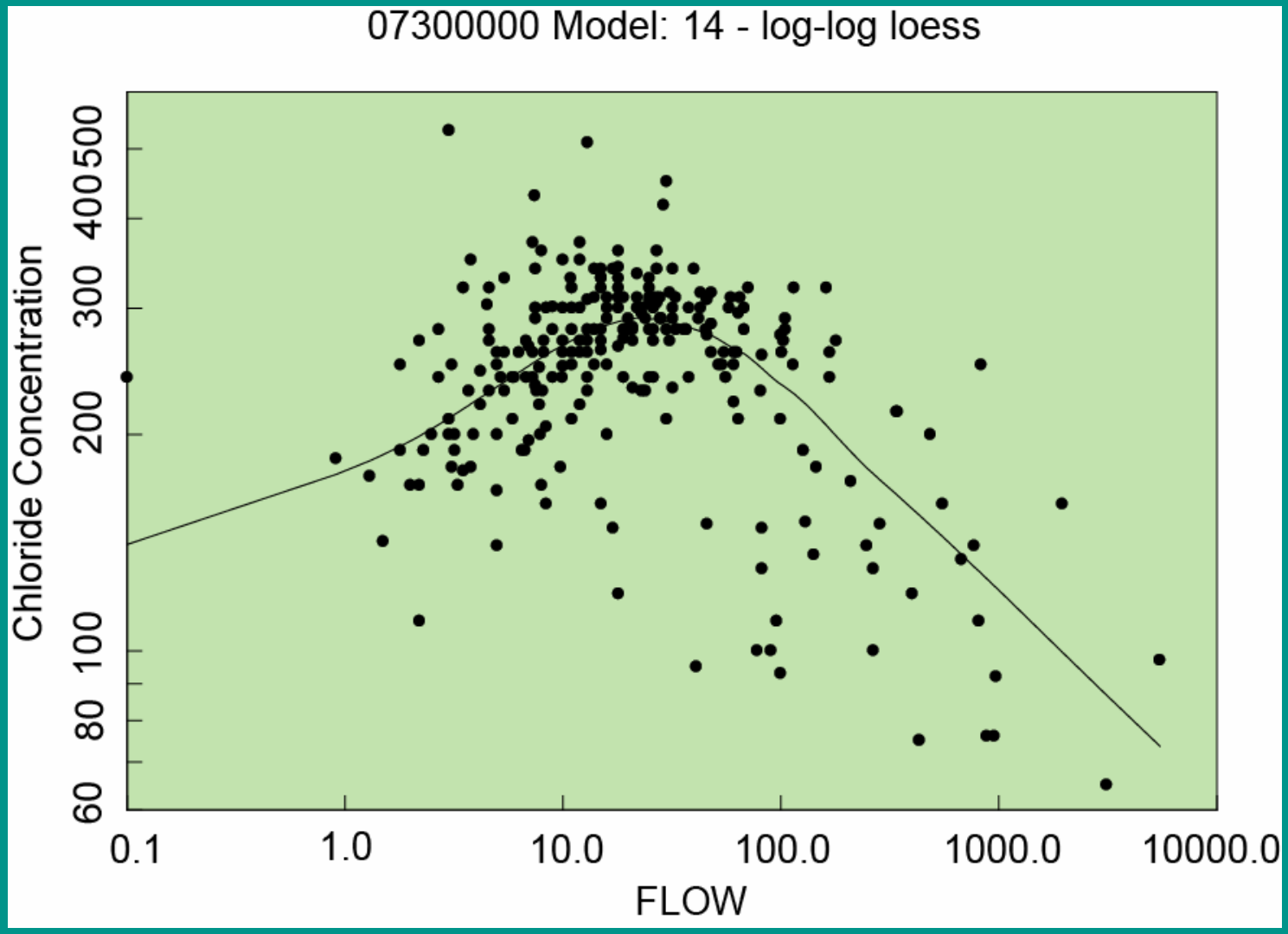


Select Flow Adjustment Model

```
Constituent:  Chloride
 Station: 07300000
Call:
loess(...)

 Number of Observations:          275
 Equivalent Number of Parameters: 4.5
 Residual Scale Estimate:         0.1983
 Residuals:
   min   1st Q   median  3rd Q    max
 -1.08 -0.1125 0.002432 0.1227 0.9215
```

USGS

Flow Adjustment model for Chloride



Run Trend Tests

- The tests are run on selected stations for each constituent.
- Two choices for the test: seasonal Kendall and Tobit regression.
- Each requires the starting year and number of years of analysis.
- Each has additional requirements for running the trend test.

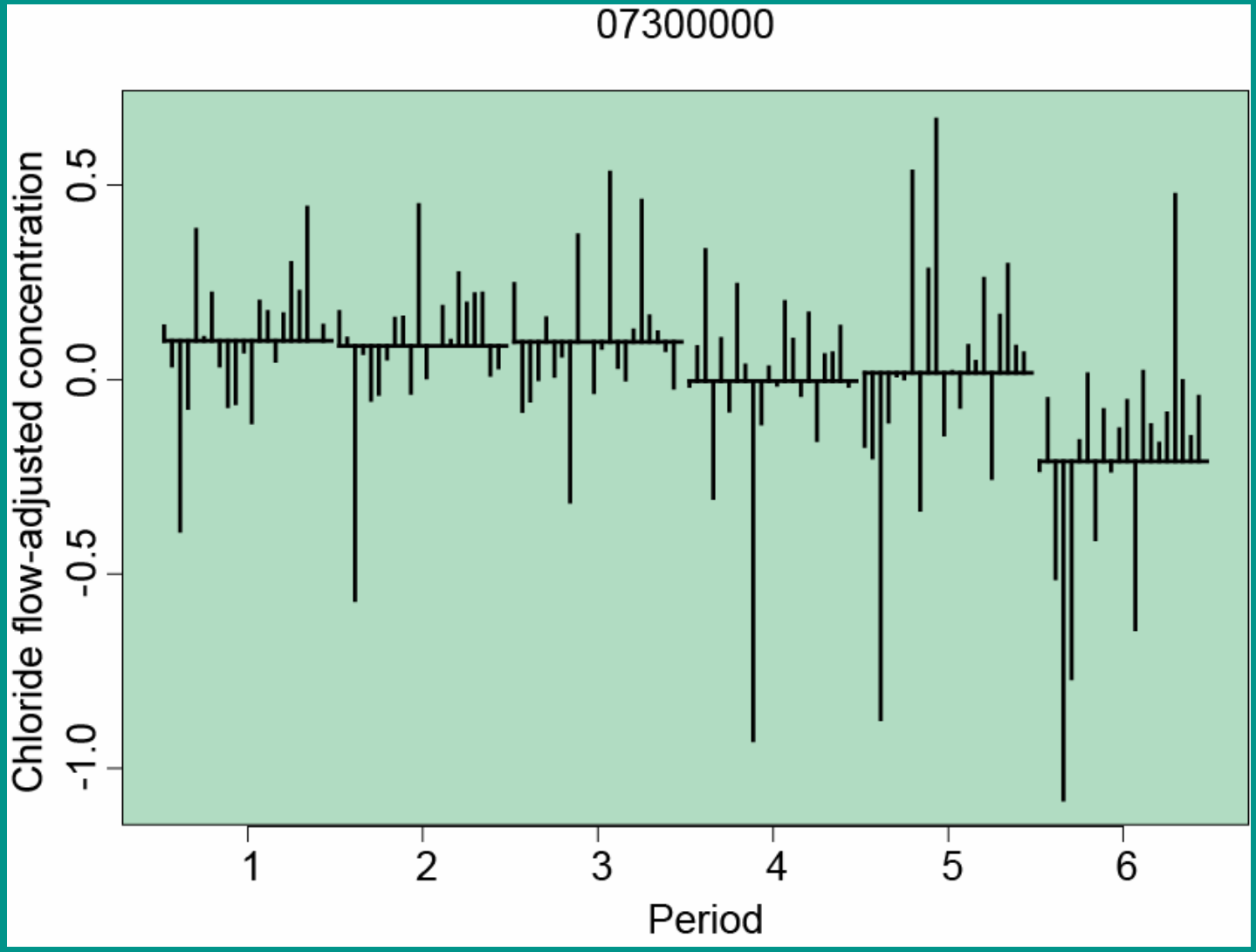
USGS

Chloride seasonal Kendall test

```
Constituent:  Chloride

  Station: 07300000
uncensored Seasonal Kendall test.
Begin year: 1968, end year: 1989, number of seasons: 6.
                                  Corrected
         Nobs   Tau  p.value p.value  Slope   Median
Raw data  123 0.241 0.000222 0.00375 2.5000 270.0000
FAC data  123 0.235 0.000336 0.01181 0.0084   0.0323
```

USGS

Chloride Subseries plot

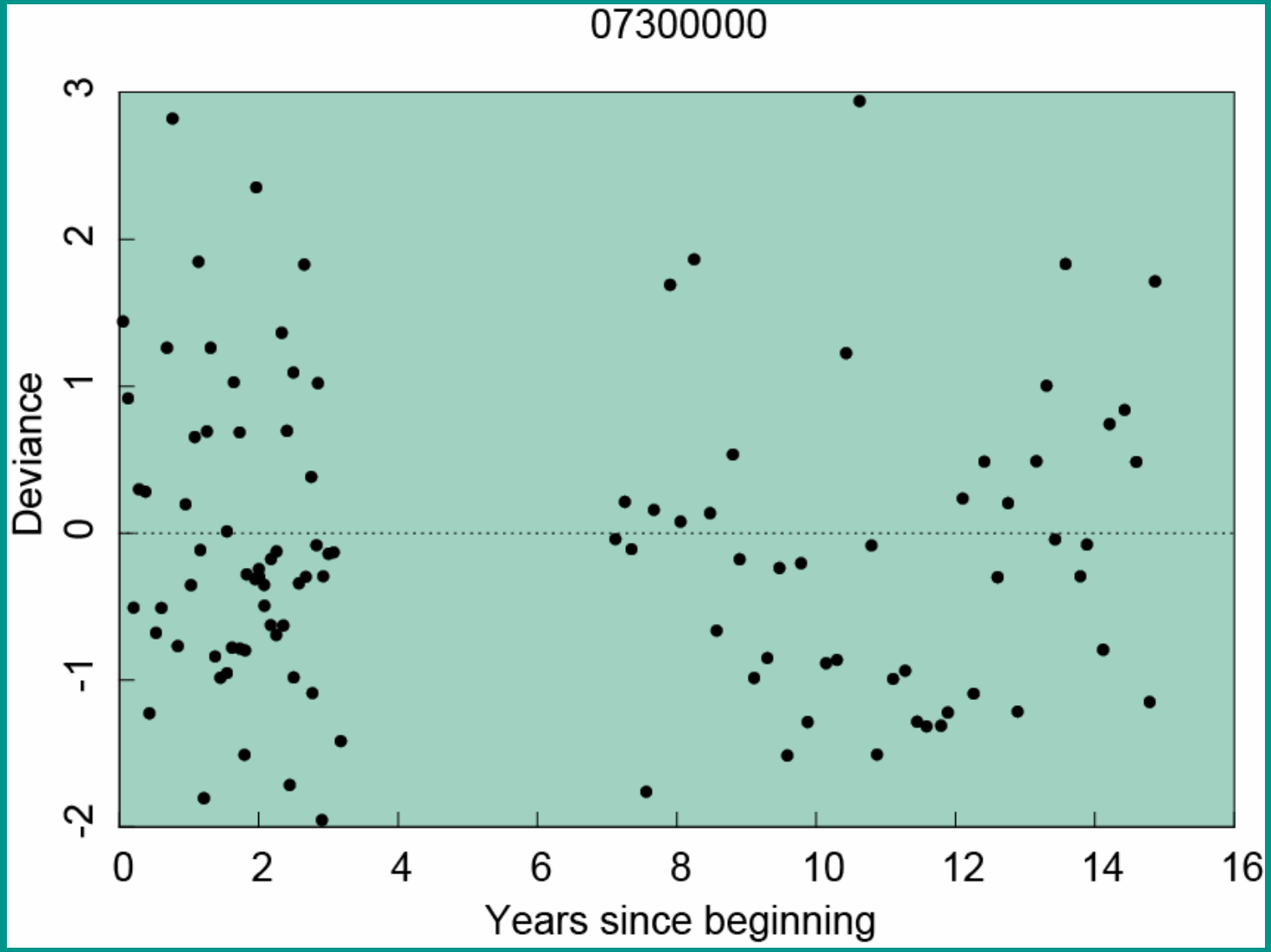


Total Phosphorus Tobit test

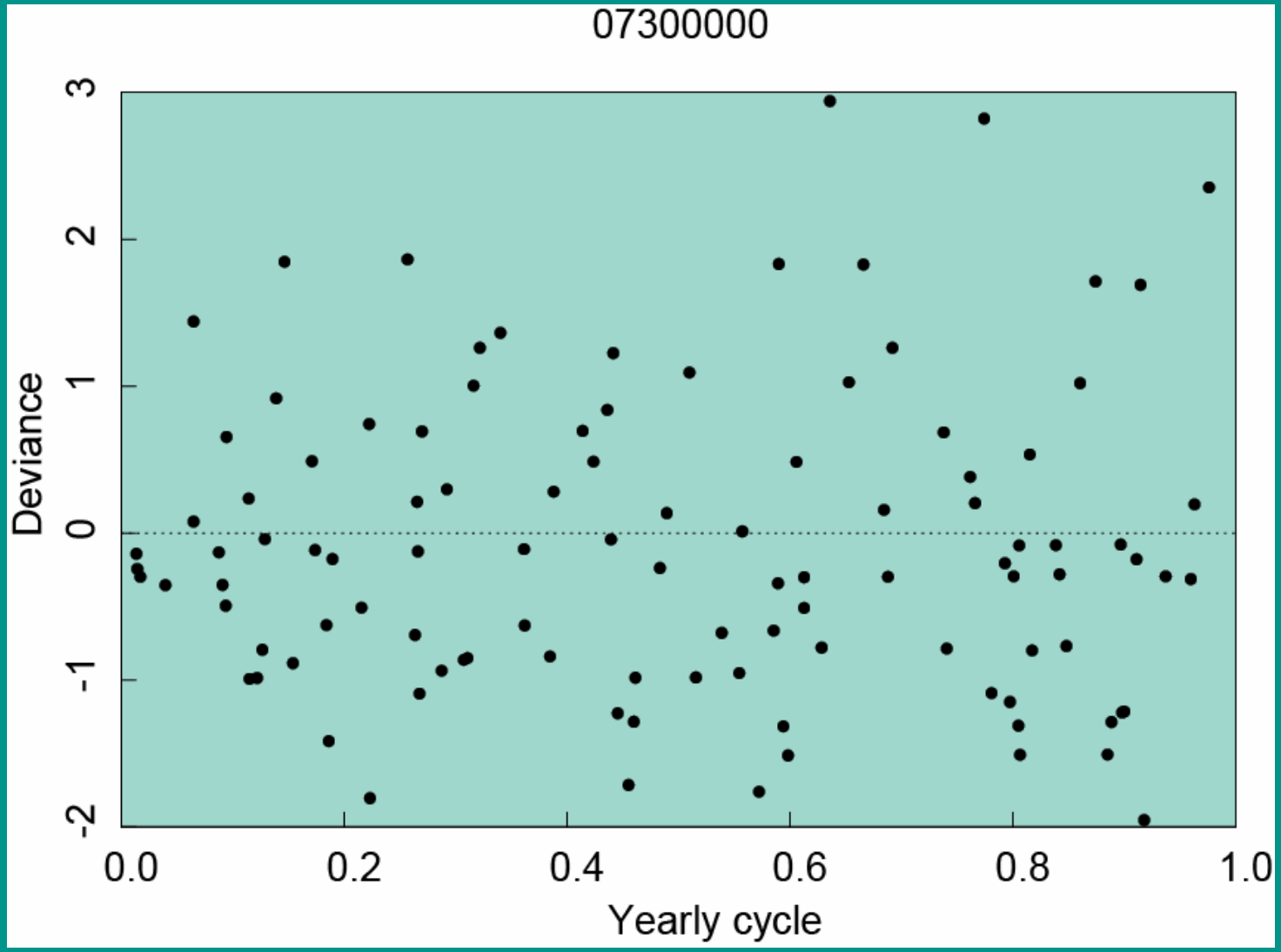
```
Constituent:  Total.P

  Station: 07300000

                      Value Std. Error        z         p
       (Intercept) -4.36441     0.3454 -12.6363 1.33e-036
              Year -0.10032     0.0219  -4.5876 4.48e-006
         log(FLOW)  0.38100     0.1127   3.3812 7.22e-004
cos(2 * pi * Year) -0.00454     0.1578  -0.0288 9.77e-001
sin(2 * pi * Year) -0.76359     0.1630  -4.6838 2.82e-006

Scale= 1.03

n= 106
```

USGS

Total P trend diagnostic plot



Total P trend diagnostic plot



Concluding remarks

- The results of the ESTREND trend tests can be extracted and processed by other software.
- The results need to be put into the context of regional patterns, or human activities.

USGS

Concluding remarks

- This short section is not designed to equip anyone with the knowledge needed to do trend studies or use the ESTREND software.
- More complete documentation for ESTREND is available in the USGS library.
- The complete, one-week class on trend and load estimation that will be open to the public is not yet scheduled.

USGS

Data considerations for trend monitoring

Monitoring objectives can be targeted toward specific constituents and specific questions, such as:

- "Did a sewage plant upgrade at location x reduce total phosphorus concentrations downstream?"
- "Do livestock grazing practices increase erosion in small watersheds?"
- "Has dissolved sulfate concentration in streamflow decreased as a result of reduced sulfur emissions?"
- "Has increased water use lowered dissolved oxygen concentrations during low-flow conditions?"

Data considerations for trend monitoring (2)

Designing a sampling program for targeted objectives requires a case-by-case analysis. We will focus instead on some general design concepts that apply for monitoring "generic" trends in a wide variety of constituents

A "generic" trend model, as defined here, is

$$\log C(t) = \alpha(t) + \beta_1 F[Q(t)] + \beta_2 \{F[Q(t)]\}^2 + \beta_3 \sin(2\pi t) + \beta_4 \cos(2\pi t) + \varepsilon(t)$$

- $C(t)$ is concentration at time t, where t is in decimal-years;
- $\alpha(t)$ is the trend;
- $F[Q(t)] = \log Q(t) - E\{\log Q(t)\}$, where $Q(t)$ is discharge; and
- $\varepsilon(t)$ is the error

Data considerations for trend monitoring (3)

General observations regarding the generic trend model:

- For samples separated by 2 weeks or more, serial correlation in the errors can usually (not always!) be ignored

- Even if there is no trend [$\alpha(t)$ is constant] and the errors are well-behaved (no outliers), to estimate the 5 generic model coefficients and verify the model, at least 20 samples are recommended

- The samples should represent different flow conditions (from low to high flow) and be reasonably balanced among seasons

Data considerations for trend monitoring (4)

Some "good" designs for establishing a baseline for evaluating trends:

- Bi-weekly sampling for 1 year
- Monthly sampling for 2 years
- Bi-monthly sampling for 4 years
- Quarterly sampling for 5 years

Some "not-so-good" designs for establishing a baseline:

- Weekly sampling for 6 months
- Two samples per year for 10 years
- Highly unbalanced designs (most samples collected during high-flow conditions or during a certain time of year)

Data considerations for trend monitoring (5)

- ❑ Once a baseline is established, sampling frequency can be reduced to monitor trends in subsequent years (from monthly to quarterly, for example)
- ❑ "Over-sample" at the beginning of the monitoring, not at the end!
- ❑ Censoring generally reduces the amount of information in the samples, but collecting more samples may not be the answer (are 40 "less-thans" really better than 20 "less-thans"?)
- ❑ Don't confuse trend monitoring with load estimation – load estimation may require more samples, especially during high-flow conditions (more on load estimation later)