## Updates of Action Items from Previous Meetings

Presented by Bob Larson to the DMAS subcommittee, 25 March 2003

## Animated Maps

- $SO_4$, $NO_3$ and $NH_4$
- Flash, PowerPoint

## Site Classification

- Analysis repeated with 2000 census, 1999 emission data
- 13 urban sites
  - NTN: NJ99, NC41, MO43,MA13,IL19,VA10
  - MDN: TX50, GA22, CA72, IN26, FL97, WI22, OR01
- Useful in recent site variance petitions
- Web application developed that will determine population and emission values for any location.

## Site Classification

- Partially integrated into sites database, not on web yet
- Documentation
  - Site Classification and Characterization
  - Paper developed for AWMA conference
  - Proposal developed for determining regionally-representative sites

## Mercury Data Management

- Double data entry system developed for HAL
  - Partitions data into data entry, data validation, final data
  - Used by HAL for 2002 Q4 data
  - Does not yet incorporate methyl Hg

## Hysplit Trajectories Back on the Web

- Re-programmed web interface to Hysplit
- Multiple (3) trajectories can be run at different elevations
- Using our own mapping routines

## Site Selection Maps

- Now dynamic
- Working on enhanced interface using Flash

## MDN Maps

- 1998-2001 on the web

## Isopleth Grids Available

- 1994 – 2001
- All analytes
- Need to create metadata

# Data Audit Topics

Presented by Bob Larson to the DMAS subcommittee, 25 March 2003

## What is a Data Systems Audit?

- Traditional “data audit”
  - check accuracy of a random selection of samples
  - Analogous to a QA sample program
  - Too time intensive for a 2 ½ day visit

## What should a data system audit include

- Review of sample validation procedures
- Review of data management practices

## Data validation methods

- Is the approved data validation scheme appropriate?
- Is the approved data validation scheme being accurately followed?

## Data entry/validation

- Is data entry accurate and complete?
- Does data entry take place in a timely manner?
- Is data entry verified?
- Are validation rules accurately followed?
- Are validation rules appropriate?

## Data Management practices

- Is the DBMS appropriate?
- Is use of the DBMS appropriate?
- Is security adequate?
- Is the program code adequately documented?
- Is the program code maintainable?
- Is the backup/recovery plan adequate?
- Are data management staff qualified and adequately trained?

## End User Issues

- Are data available in formats useful to end-users?
- Is data retrieval easy/flexible enough?
- Are data well documented?

# The NAtChem Data Exchange Standard for Precipitation Chemistry

C. Ro, J. Narayan, B. Sukloff, and R. Vet

Meteorological Service of Canada

## Outline

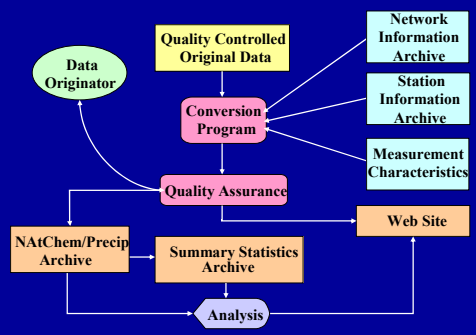
1. NAtChem/Precip Database status
2. Database management system
3. Data Exchange Standard (DES)
4. Structure of DES file
5. Summary

## NAtChem/Precip Database Status

**Monitoring Networks and Sites:**
**Total Number and (in 2001)**

| | | | | |
|---|---|---|---|---|
| 9 | (6) | Provincial Networks | 331 | (78) Sites |
| 4 | (1) | Federal Networks | 136 | (24) Sites |
| 10 | (2) | US Networks | 367 | (249) Sites |
| 23 | (9) | Networks | 834 | (351) Sites |

## NAtChem/Precip System




## Data Exchange Standard (DES)

- The data formatting and metadata protocols provide information about the measurements that users need.
- The protocols include the following features:
  - Validity flag
  - The CAS Registry number and the CAS -9CI name
  - Data type and format
  - Files are archived in ASCII CSV

## DES: Quality Assurance Report (QAR)

- A QAR is required for DES, which is completed by each network.
- It includes:
  1. Operating Information
  2. Measurement details
  3. Quality Assurance/Quality Control
  4. Data Management and Quality Control
  5. Measurement Data Quality
  6. References

# DES: QAR and Metadata

- The QAR provides detailed metadata for the instrumentation and measurements.
- The key characteristics of the measurements are input to the DES format along with the actual measurement data.
- The entry is facilitated by the DES template, an Excel® spreadsheet that contains drop-down pick lists.
- The completed DES file in ASCII CSV format and the QAR are archived by NAtChem.

# Structure of DES File

The basic structure of a DES file consists of four major sections:

1. Heather
2. Standard flags
3. Site information
4. Data

# DES: The Header Section

It describes the file contents and ownership.

| | |
|---|---|
| *DATA EXCHANGE STANDARD VERSION | NATCHEM PRECIP 2003/01/17 (1.01) |
| *COMMENT | NAtChem/Precipitation Chemistry Template |
| *QUALITY CONTROL LEVEL | 1 (a complete data set of specified quality that consists of research pro |
| *DATE THIS FILE GENERATED/ARCHIVE VERSION NUMBER | 2003-03-23 1 |
| *ORGANIZATION ACRONYM | NADP |
| *ORGANIZATION NAME | National Acid Deposition Program |
| *STUDY OR NETWORK ACRONYM | NADP-NTN |
| *STUDY OR NETWORK NAME | National Acid Deposition Program-National Trends Network |
| *FILE CONTENTS DESCRIPTION--SHORT/LONG | NPRDF_NADP_2001 NAtChem PRecipitation Data File of NADP for 2001 |
| *PRINCIPAL INVESTIGATOR NAME--LAST/FIRST | Bowersox, Van C. |
| *PRINCIPAL INVESTIGATOR AFFILIATION | Illinois State Water Survey |
| *CO-INVESTIGATOR NAME--LAST/FIRST | Larson, Bob |
| *CO-INVESTIGATOR AFFILIATION | Illinois State Water Survey |
| *COUNTRY CODE | US (UNITED STATES) |
| *STATE OR PROVINCE CODE | IL |
| *SAMPLING FREQUENCY OF DATA IN MAIN TABLE | Weekly |
| *PRINCIPAL INVESTIGATOR CONTACT INFORMATION | Bowersox, Van, NADP Program Office Illinois State Water Survey 2204 |
| *DATA USAGE ACKNOWLEDGEMENT | NADP-NTN, NADP Program Office Illinois State Water Survey 2204 Griff |
| *NAME AND AFFILIATION OF PERSON WHO GENERATED THIS FILE | Chul-Un Ro, NAtChem Database and Analysis Facility, Environmet Cana |
| *DATE OF LAST MODIFICATION TO DATA IN MAIN TABLE | 2003-03-23 |
| *FILE CHANGE HISTORY--VERSION NUMBER/DESCRIPTION | N/A |
| *NAME AND VERSION OF SOFTWARE USED TO CREATE THIS FILE | MS Excel/2000 |
| *STANDARD CHARACTERS | !#$%&'()*,-+=./0123456789:;<>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ |
| *COMPANION FILE NAME/FORMAT AND VERSION | N/A |

The sampling frequency of the data in the main data table in this file. It may help to think of this as the pattern of reporting of data. Select from the drop-down list.

This is a mandatory key phrase

# DES: The Standard Flags Section

It describes the validity flag codes assigned to every measurement value.

| | | |
|---|---|---|
| *TABLE NAME | NAtChem standard flags | |
| *TABLE FOCUS | Metadata | |
| *TABLE COLUMN NAME | Flag: NAtChem | Description |
| *TABLE COLUMN NAME TYPE | Variable | Variable |
| *TABLE COLUMN UNITS | None | None |
| *TABLE COLUMN FORMAT TYPE | Char | Char |
| *TABLE COLUMN FORMAT FOR DISPLAY | 2 | 120 |
| *TABLE BEGINS | | |
| | V0 | Valid value |
| | V1 | Valid value but set equal to the detection limit (DL) because the measured value was below the DL |
| | V2 | Valid estimated value |
| | V3 | Valid value but qualified because of possible contamination (e.g., pollution source, laboratory contamination source) |
| | V4 | Valid value but qualified due to non-standard sampling conditions (e.g., instrument malfunction, sample handling) |
| | M1 | Missing value because no value is available |
| | M2 | Missing value because invalidated by data originator |
| | M3 | Missing value because value failed NAtChem/Precipitation criteria |
| | H1 | Historical data that have not been assessed or validated |

# DES: The Site Information Section

It describes the location, site characteristics and instruments.

| TABLE NAME | Site information | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TABLE FOCUS | Metadata | | | | | | | |
| TABLE COLUMN NAME | Site ID: NAtChem | Co-location ID | Site ID: networ | Description | Country | State | Latitude: deci | Longitude: decir |
| TABLE COLUMN UNITS | None | None | None | None | None | None | Decimal degr | Decimal degrees |
| TABLE COLUMN FORMAT TYPE | Char | Char | Char | Char | Char | Char | Decimal | Decimal |
| TABLE COLUMN FORMAT FOR DI | 12 | 2 | 3 | 50 | 50 | 20 | 10.4 | 10.4 |
| TABLE COLUMN MISSING CODE | None | None | None | None | None | None | -999.99999 | -999.99999 |
| TABLE USER NOTE | The final character of site ID identifies single collector at the site (_), co-located collector at the site (1,2) or | | | | | | | |
| TABLE BEGINS | | | | | | | | |
| | NADPCAAB1 | P | NADP80CA | Lethbridge | CA (CA | AB | 49.64389 | -112.78778 |
| | NADPCANS1 | P | NADP86CA | Kejimkujik | CA (CA | NS | 44.42361 | -65.20556 |
| | NADPCAON1 | P | NADP87CA | Mount For | CA (CA | ON | 43.99972 | -80.74611 |
| | NADPCAPQ1 | P | NADP89CA | Frelighsbu | CA (CA | PQ | 45.05028 | -72.86167 |
| | NADPCAPQ1 | P | NADP89CA | Sutton | CA (CA | PQ | 45.06778 | -72.67556 |
| | NADPUSAK1 | P | NADP02AK | Denali Nat | US (UNI | AK | 63.72861 | -148.96389 |
| | NADPUSAK1 | P | NADP02AK | Poker Cre | US (UNI | AK | 65.15250 | -147.48611 |
| | NADPUSAL1 | P | NADP01AL1 | Black Belt | US (UNI | AL | 32.45750 | -87.24222 |
| | NADPUSAL1 | P | NADP01AL9 | Sand Mou | US (UNI | AL | 34.28806 | -85.96889 |
| | NADPUSAL1 | P | NADP01AL0 | Delta Eler | US (UNI | AL | 30.79639 | -87.84972 |
| | NADPUSAL1 | N | NADP01AL2 | Bay Road | US (UNI | AL | 30.47444 | -88.14111 |
| | NADPUSAL1 | S | NADP0199A | Sand Mou | US (UNI | AL | 34.28806 | -85.96889 |
| | NADPUSAR1 | P | NADP05AR1 | Buffalo Na | US (UNI | AR | 36.08472 | -92.58694 |
| | NADPUSAR1 | P | NADP05AR0 | Caddo Va | US (UNI | AR | 34.16944 | -93.09861 |
| | NADPUSAR1 | P | NADP05AR2 | Fayettevill | US (UNI | AR | 36.10167 | -94.17333 |

# DES: The Data Section

It contains 13 rows of metadata fields for each measurement species.

| | | | | |
|---|---|---|---|---|
| *TABLE FOCUS | Surface--fixed | | | |
| *TABLE EXPLANATION OF ZERO OR NEGATIVE VALUES | | | | |
| *TABLE USER NOTE | | | | |
| *TABLE KEY FIELD NAMES | Site ID: standard | | | |
| *TABLE COLUMN NAME | Site ID: standard | Sulphate | Sulphate | Sulphate |
| *TABLE COLUMN NAME TYPE | Variable | Variable | Flag--study | Flag--NAtChem |
| *TABLE COLUMN CAS IDENTIFIER | None | C14808-79-8 | C14808-79-8 | C14808-79-8 |
| *TABLE COLUMN USER NOTE | None | SO4= | SO4= | SO4= |
| *TABLE COLUMN USER NOTE2 | None | None | None | None |
| *TABLE COLUMN UNITS | None | mg/L (milligram per liter) | None | None |
| *TABLE COLUMN FORMAT TYPE | Char | Decimal | Char | Char |
| *TABLE COLUMN FORMAT FOR DISPLAY | 12 | 7.3 | 2 | 2 |
| *TABLE COLUMN MISSING CODE | None | -99.999 | None | None |
| *TABLE COLUMN LOOKUP TABLE NAME | Site information | None | None | None |
| *TABLE COLUMN OBSERVATION TYPE | Supplementary data | Precipitation chemistry | Precipitation chemistry | Precipitation chemistry |
| *TABLE COLUMN LABORATORY ANALYTICAL METHOD | Not applicable | IC (Ion chromatograph) | Not applicable | Not applicable |
| *TABLE COLUMN DETECTION LIMIT | Not applicable | 0.01 | Not applicable | Not applicable |

## DES: NAtChem Data Analysis Centre

- The NAtChem Data Analysis Centre works with Networks iteratively to ensure that their data files conform to the DES protocols.
- This is done through “read and verify” computer programs that identify formatting and data integrity errors and produce time series plots.

## Summary

- The completeness of the data archive for both measurements and metadata is very important.
- NAtChem developed a DES to meet the needs of both data originators and users.
- It is the result of an international effort addressing the data archiving needs for atmospheric measurement data.

**Performance and Acceptance Criteria (PAC)**
/
**Data Quality Objectives (DQOs)**

for the

**NADP/NTN**
**USGS EXTERNAL QUALITY ASSURANCE PROJECT**

Greg Wetherbee: wetherbe@usgs.gov

**Objectives of the USGS External QA Project**

1. Estimate total error associated with NADP chemical measurements?
2. Determine portion of total error attributed to each step in the data-collection process?
3. Determine whether known and measurable sources of error are controlled to acceptable levels?
4. Determine what unmeasured sources of error can be identified, measured, and controlled?

**USGS QUALITY ASSESSMENTS**

1. Document past performance of laboratories, site operators, and field equipment in terms of absolute and relative error.
2. Document "trends" in performance from one year to next.
   Improving? No change?
3. Never state whether performance meets expectations. ~~Good?~~ ~~Bad?~~

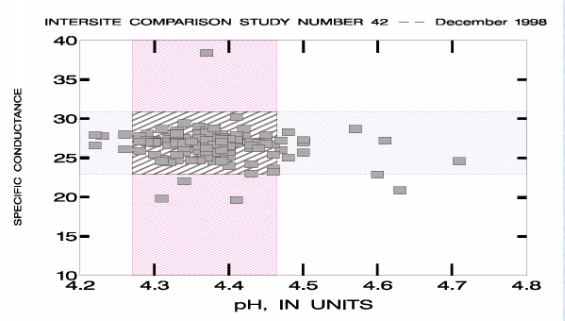
**The Performance and Acceptance Criteria Process (PAC)**

1. State the Problem
2. Identify the Study Questions
3. Identify Types of Information Needed
4. Establish Study Design Constraints
5. **Specify Information Quality**
6. Develop a Strategy for Information Synthesis
7. Optimize the Design for Collecting Information

*USEPA, October 2002, EPA QA/G-4A, Peer Review Draft*

**Example PAC for Intersite Program**

| Data Quality Indicator | Measurement Quality Objectives for Performance Criteria | Acceptance Criteria |
|---|---|---|
| Precision | 100% within .02 pH Units, 1 µS/cm | >90% within .05 pH Units, 2 µS/cm |
| Bias | Less Than +/- 5% | Less Than +/- 10% |
| Representativeness | 4.0<pH<6.0<br>Sc < 50 | 3.5<pH<6.5<br>Sc < 100 |
| Comparability | 0.02 pH Units<br>2 µS/cm | 0.04 pH Units<br>4 µS/cm |
| Completeness | 100% Sites Respond | 95% Sites Respond |
| Sensitivity | 0.02 pH Units, 1 µS/cm | 0.04 pH Units, 2 µS/cm |



## Example PAC for Interlaboratory Program

| Data Quality Indicator | Measurement Quality Objectives for Performance Criteria | Acceptance Criteria |
|---|---|---|
| Precision | 2 f-psuedosigma | 3 f-psuedosigma |
| Bias | 0% | +/- 5%, No Trends |
| Representativeness | 25th-75th NTN Percentile | 10th-90th NTN Percentile |
| Comparability | Median Values 95% Accurate Compared to Target Values. | Median Values 90% Accurate Compared to Target Values. |
| Completeness | 100% Lab Analyses | 95% Lab Analyses |
| Sensitivity | No ultrapure D.I. detections | ≤ 2 ultrapure D.I. detections |

## Example PAC for SHE and Field Audit Programs

| Data Quality Indicator | Measurement Quality Objectives for Performance Criteria | Acceptance Criteria |
|---|---|---|
| Precision | 5% Absolute Error | <10% Absolute Error |
| Bias | 0% | Less than +/-5% |
| Representativeness | Protocol performed correctly by all site operators. | Greater than 90 percent site operators perform protocol correctly. |
| Comparability | 2 f-pseudosigma of median concentration | 3 f-pseudosigma of median concentration |
| Completeness | 100% Samples Processed | >90% Samples Processed |
| Sensitivity | 0.02 mg/L Absolute Difference | 0.05 mg/L Absolute Difference |

## Example PAC for Collocated-Sampler Program

| Data Quality Indicator | Measurement Quality Objectives for Performance Criteria | Acceptance Criteria |
|---|---|---|
| Precision | <10% Absolute Error | <25% Absolute Error |
| Bias | 0% | Less than +/-10% |
| Representativeness | Less than 5 percent difference in sample volumes. | Less than 10 percent difference in sample volumes. |
| Comparability | Data for 2 samplers correlated & within historic site data range . | Data within range of historic data for site. |
| Completeness | 100% | 75% - Less than 13 weeks missed |
| Sensitivity | Precipitation Depth: 0.02 inches, Concentrations: 0.02 mg/L Absolute Difference | Precipitation Depth: 0.05 inches, Concentrations: 0.05 mg/L Absolute Difference |

## The Data Quality Objectives (DQO) Process

**7 Steps** for DQO Planning Team

1. State the Problem
2. Identify the Decision
3. Identify the Inputs to the Decision
4. Define the Boundaries of the Study
5. Develop a Decision Rule
6. Specify Tolerable Limits on Decision Errors
   (e.g. $\alpha$= 0.05, $\beta$=0.20)
7. Optimize the Design for Obtaining Data
   (e.g. cost effectiveness)

***How are DQOs different from PAC?***

…specify tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions. (USEPA, 2002)

***How are DQOs and PAC related?***

DQOs define the performance and acceptance criteria that limit the probabilities of making decision errors by considering the purpose of collecting the data; defining the appropriate type of data needed; and specifying tolerable probabilities of making decision errors. (USEPA, 2002)

## Step 2: Identify the Decision(s)

Potential Decisions:

A) Constituent concentrations in precipitation are decreasing [or increasing].

B) NTN data quality is "acceptable."

C) Others?

## Step 5: Develop Decision Rule(s)

*…if, then statements*

Potential Decision Rules:

**A)** If a Seasonal Kendall Test detects a negative [or positive] slope, then constituent concentrations in precipitation are decreasing [or increasing].

**B)** If median collocated-sampler [or substitute other program] absolute error is less than or equal to *X%* percent, then data quality is "acceptable."

## Step 6: Specify Tolerable Limits on Decision Errors

Step 6 determines:

A) How many samples need to be collected (N)

…generally, N becomes larger as $\alpha$ and $\beta$ get smaller

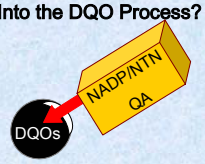
B) Spatial distribution of samples (e.g. grid spacing)

…generally, grid spacing tighter as $\alpha$ and $\beta$ get smaller

C) Temporal distribution of samples (e.g. seasonality)

## Does the NADP/NTN Fit Into the DQO Process?

**No**:




**DQOs:**

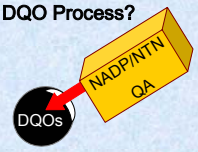
DQOs define number (N), quality, and spatial/temporal distribution of samples required to make decisions with a pre-specified level of statistical confidence.

**NADP/NTN:**

Natural environment and funding control the number and spatial distribution of NTN samples. Therefore, $\alpha$ and $\beta$ would have to vary geographically. This complexity would limit spatial interpretation of the data.

## Does the NADP/NTN Fit Into the DQO Process?

**No**:




**DQOs:**

DQOs are for making decisions about two clear alternatives (e.g. whether action levels are exceeded or not; clean precipitation vs dirty; etc.).

**NADP/NTN:**

Data analysis not always conducive to making yes/no decisions. Lots of "gray areas." Probability of Type II error ($\beta$) would likely be high.

# Data Quality Objectives: What do the Trends Show?

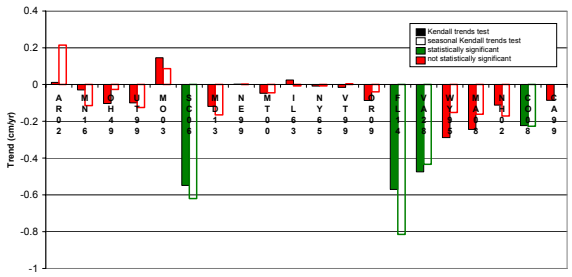
C. Lehmann

NATIONAL ATMOSPHERIC DEPOSITION PROGRAM

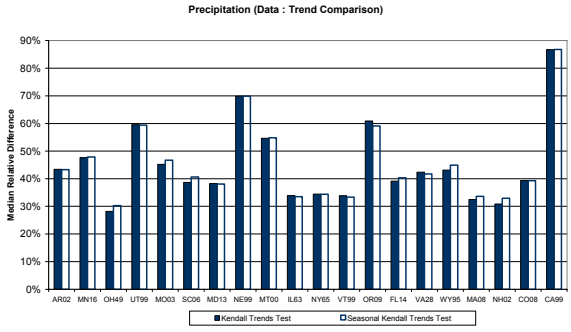
## Objective

- Compare variability in data trends with measurement variability.
- Indicator of measurement system performance

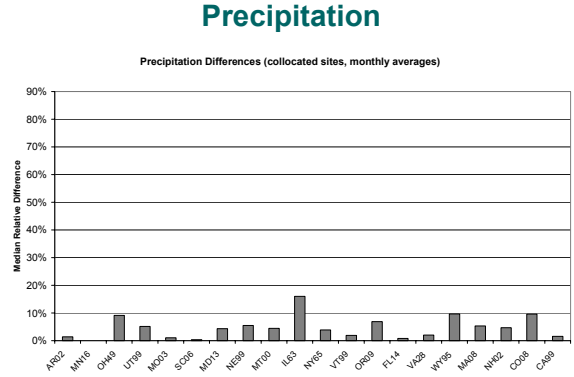
## Precipitation



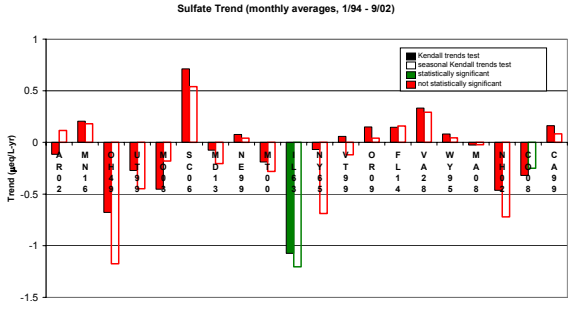

## Precipitation

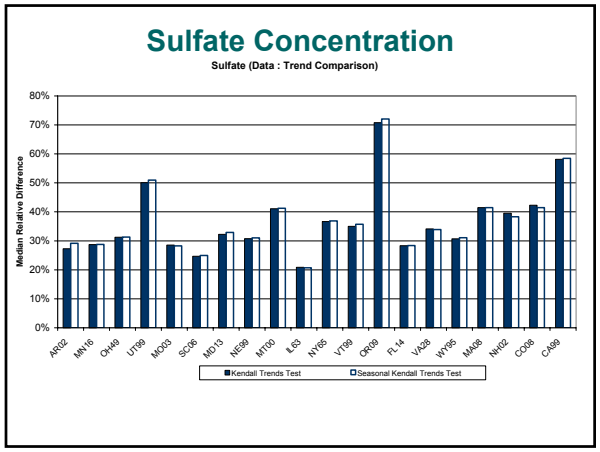


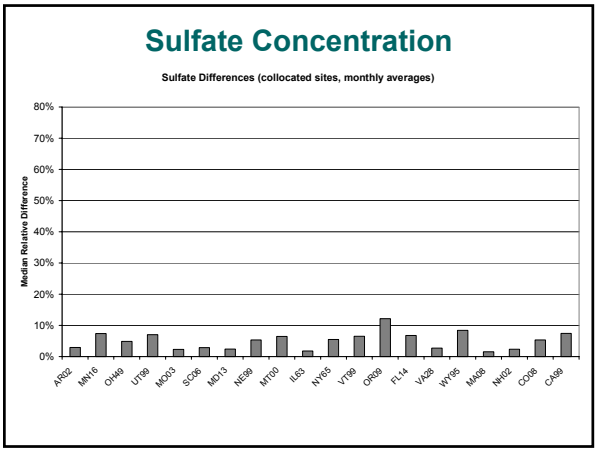

## Precipitation

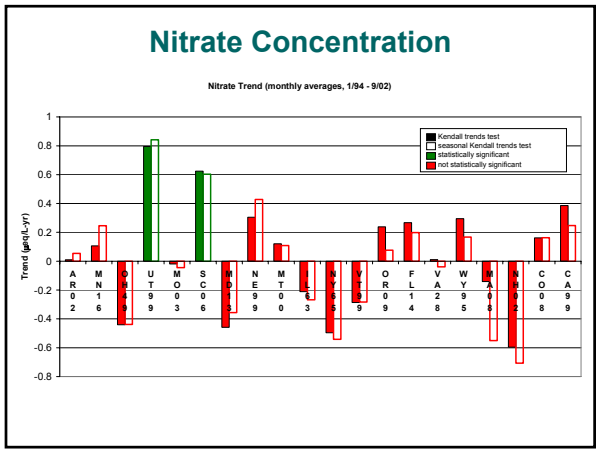


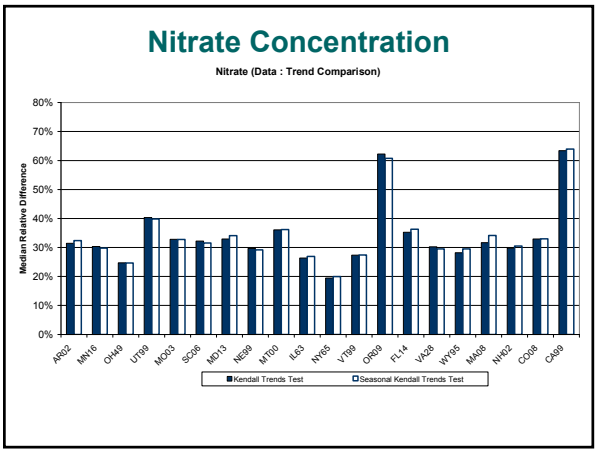

## Sulfate Concentration

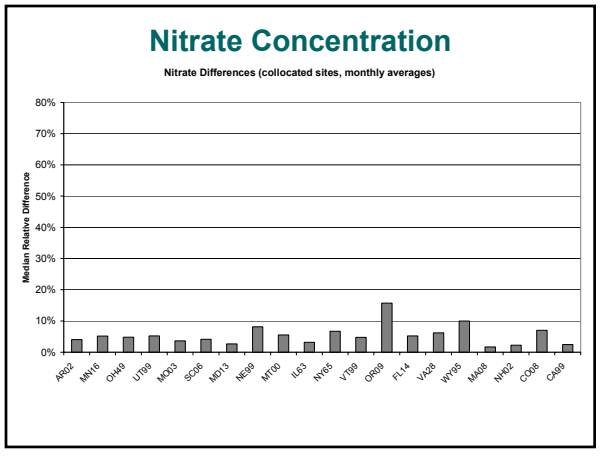


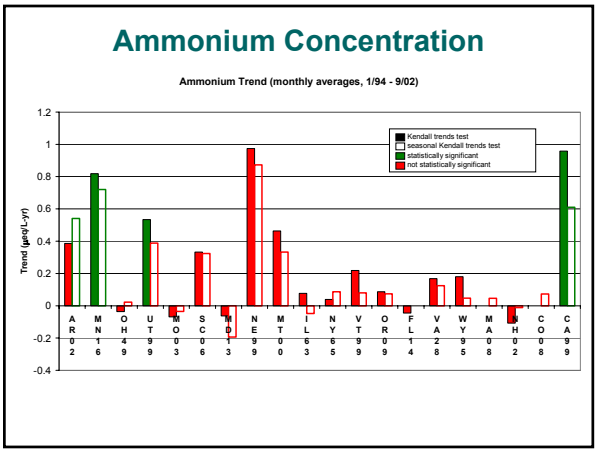

Sulfate Concentration
Sulfate (Data : Trend Comparison)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99
Kendall Trends Test
Seasonal Kendall Trends Test

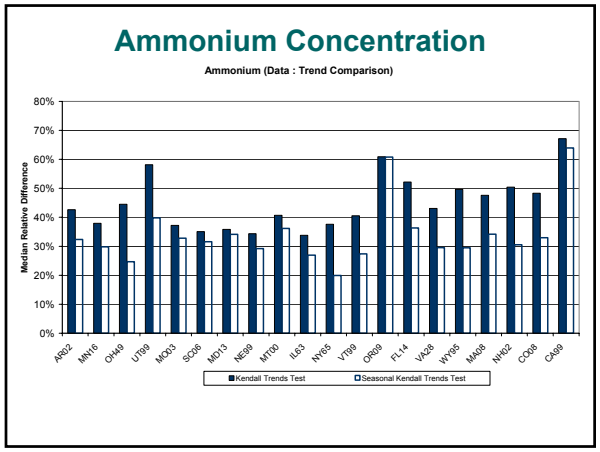

Sulfate Concentration
Sulfate Differences (collocated sites, monthly averages)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99

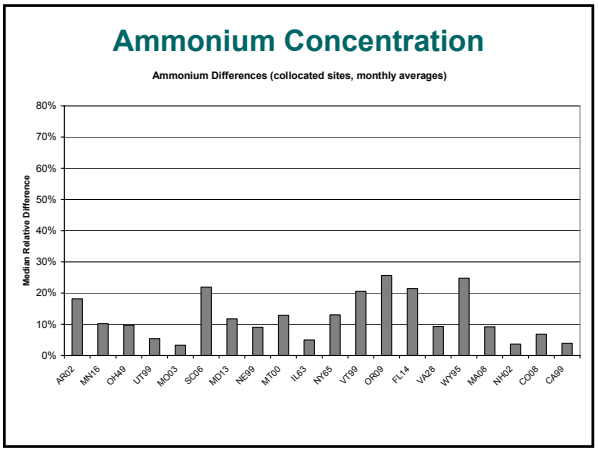

Nitrate Concentration
Nitrate Trend (monthly averages, 1/94 - 9/02)
Trend (µeq/L-yr)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Kendall trends test
seasonal Kendall trends test
statistically significant
not statistically significant
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99

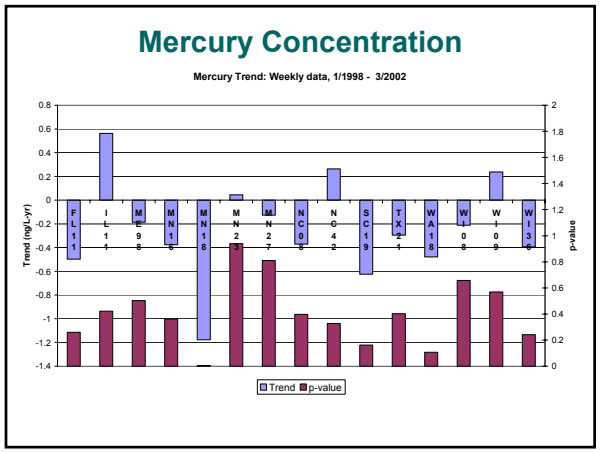

Nitrate Concentration
Nitrate (Data : Trend Comparison)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99
Kendall Trends Test
Seasonal Kendall Trends Test


Nitrate Concentration
Nitrate Differences (collocated sites, monthly averages)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99


Ammonium Concentration
Ammonium Trend (monthly averages, 1/94 - 9/02)
Trend (µeq/L-yr)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Kendall trends test
seasonal Kendall trends test
statistically significant
not statistically significant
AR02
MN16
OH49
UT99
MO03
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99


Ammonium Concentration
Ammonium (Data : Trend Comparison)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO33
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99
Kendall Trends Test
Seasonal Kendall Trends Test


Ammonium Concentration
Ammonium Differences (collocated sites, monthly averages)
Median Relative Difference
80%
70%
60%
50%
40%
30%
20%
10%
0%
AR02
MN16
OH49
UT99
MO33
SC06
MD13
NE99
MT00
IL63
NY65
VT99
OR09
FL14
VA28
WY95
MA08
NH02
CO08
CA99


Mercury Concentration
Mercury Trend: Weekly data, 1/1998 - 3/2002
Trend (ng/L-yr)
p-value
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
FL11
IL11
ME98
MN16
MN18
MN20
MN27
NC05
NG42
SC19
TX21
WA12
WI08
WI09
WI36
Trend
p-value