The National Trends Network Database:

Data Validation Coding & the Use of Site History at the Central Analytical Laboratory

Validation Codes

Used primarily to identify samples that have been compromised

Types:

1) Sampling Protocol Codes (SP codes)
- Reflect precipitation collector malfunctions

Validation Codes

Used primarily to identify samples that have been compromised

Types:

1) Sampling Protocol Codes (SP codes)
- Reflect precipitation collector malfunctions

2) Screening Level Codes (SL Codes)
- Reflect gross contamination based on operator/lab remarks & site history

Sampling Protocol Codes (SP Codes)

"U" Wet-side bucket exposed to > 6 hours of dry deposition over the duration of the sampling period (3.9% in 2004)

Sampling Protocol Codes (SP Codes)

"U" Wet-side bucket exposed to > 6 hours of dry deposition over the duration of the sampling period (3.9% in 2004)

"B" Wet-side bucket exposed to *all* deposition over the entire sampling period (0.9% in 2004)

Sampling Protocol Codes (SP Codes)

"U" Wet-side bucket exposed to > 6 hours of dry deposition over the duration of the sampling period (3.9% in 2004)

"B" Wet-side bucket exposed to *all* deposition over the entire sampling period (0.9% in 2004)

Blank Wet-side bucket exposed to < 6 hours of dry deposition over the duration of the sampling period (95.2% in 2004)

Sampling Protocol Codes (SP Codes)

"U" Wet-side bucket exposed to > 6 hours of dry deposition over the duration of the sampling period (3.9% in 2004)

"B" Wet-side bucket exposed to *all* deposition over the entire sampling period (0.9% in 2004)

Blank Wet-side bucket exposed to < 6 hours of dry deposition over the duration of the sampling period (95.2% in 2004)

"Q" Quality assurance sample

Screening Level Codes (SL Codes)

"F" Gross mishandling in the field (0.3% in 2004)

Screening Level Codes (SL Codes)

"F" Gross mishandling in the field (0.3% in 2004)

"L" Gross mishandling in the lab (0.0% in 2004)

Screening Level Codes (SL Codes)

"F" Gross mishandling in the field (0.3% in 2004)

"L" Gross mishandling in the lab (0.0% in 2004)

"C" *Contaminated sample that exhibits anomalous chemistry compared to Site History distributions* (8.1% in 2004)

Screening Level Codes (SL Codes)

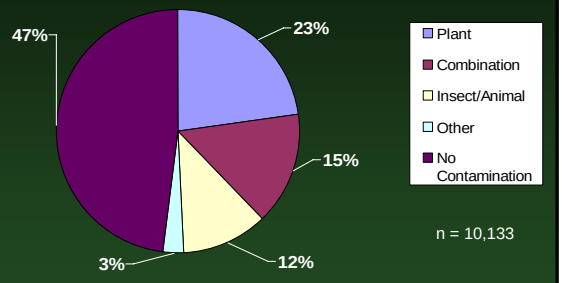
"F" Gross mishandling in the field (0.3% in 2004)

"L" Gross mishandling in the lab (0.0% in 2004)

"C" *Contaminated sample that exhibits anomalous chemistry compared to Site History distributions* (8.1% in 2004)

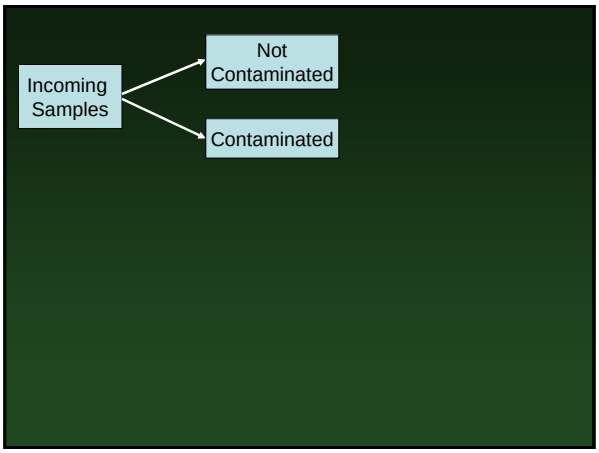
Blank (91.6% in 2004)

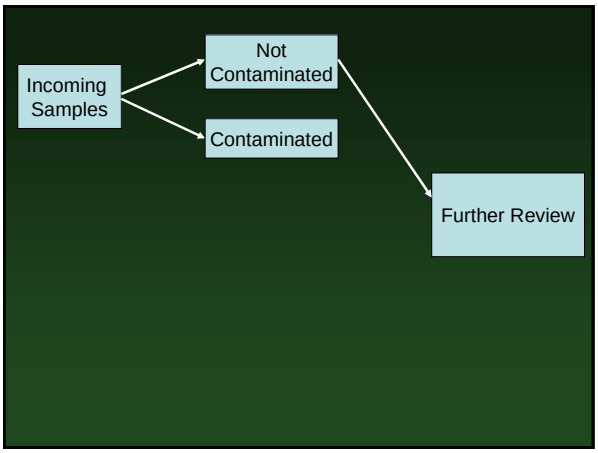
NTN Wet Sample Contamination, 2004

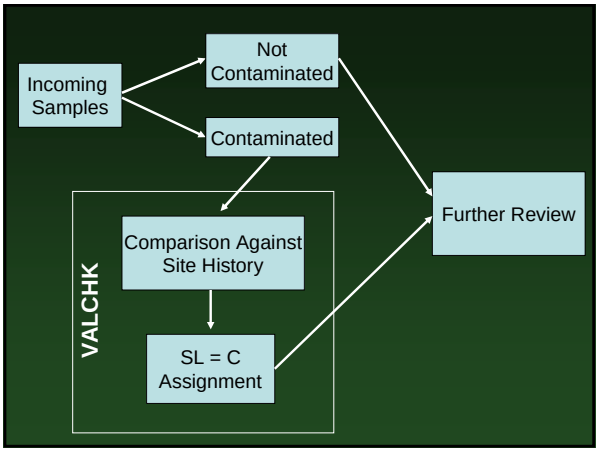


n = 10,133

15.6% of contaminated samples are SL = C







NTN Site History

- Includes, for each site from start-up to present, all Wet-type (>35ml) samples with Blank SP & SL codes

NTN Site History

- Includes, for each site from start-up to present, all Wet-type (>35ml) samples with Blank SP & SL codes
- Calculates descriptive statistics for these samples including percentile distributions for all analytes

NTN Site History

- Includes, for each site from start-up to present, all Wet-type (>35ml) samples with Blank SP & SL codes
- Calculates descriptive statistics for these samples including percentile distributions for all analytes
- Updated quarterly

VALCHK - Site History and SL Coding

- All analyte values from incoming *contaminated* samples are compared against a site's history to determine if outliers exist

VALCHK - Site History and SL Coding

- All analyte values from incoming *contaminated* samples are compared against a site's history to determine if outliers exist
- Each analyte value is automatically assigned a score based on its position within the sites historical distribution of values for that analyte

VALCHK - Site History and SL Coding

- All analyte values from incoming *contaminated* samples are compared against a site's history to determine if outliers exist
- Each analyte value is automatically assigned a score based on its position within the sites historical distribution of values for that analyte
- An SL code of "C" is assigned if the sum of scores is > or = 4.0

VALCHK Scoring

Concentration vs. Site History	pH & Conductance	Other Analytes
> Maximum	1	2
≥ 90th	0.5	1
≤ 10th	0.5	0
< Minimum	1	0

Hypothetical VALCHK Scoring Scenarios

	Sample A	B	C	D
pH	0.5			1
Cond.			1	
Ca	1			
Mg			1	
K		2		
Na				1
NH_4	2			1
NO_3				1
Cl				1
SO_4		2		
PO_4			2	
Sum	3.5	4.0	4.0	5.0
SL Code	blank	C	C	C

Concerns with the Use of Site History in SL Coding

- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?

Concerns with the Use of Site History in SL Coding

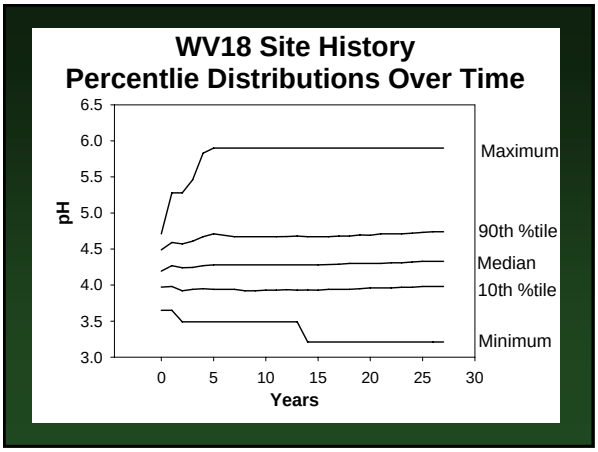
- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?
- Analyte distributions for new sites have small sample sizes with potentially narrow distributions.
  Do young, narrowly distributed site histories affect SL coding?

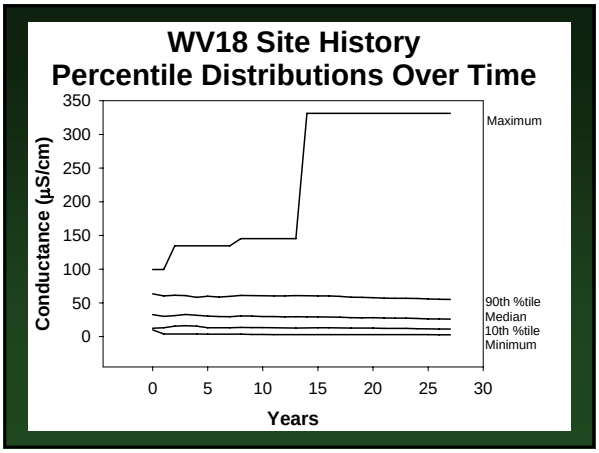
Concerns with the Use of Site History in SL Coding

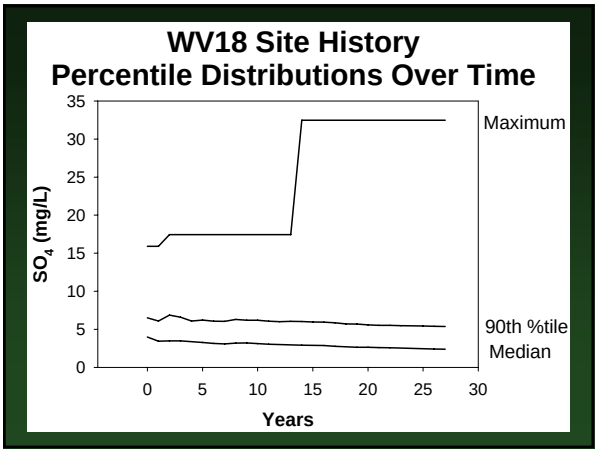
- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?
- Analyte distributions for new sites have small sample sizes with potentially narrow distributions.
  Do young, narrowly distributed site histories affect SL coding?
- Sites start-up in different seasons. In the spring samples may contain more contaminants.
  Does start-up season affect SL coding?

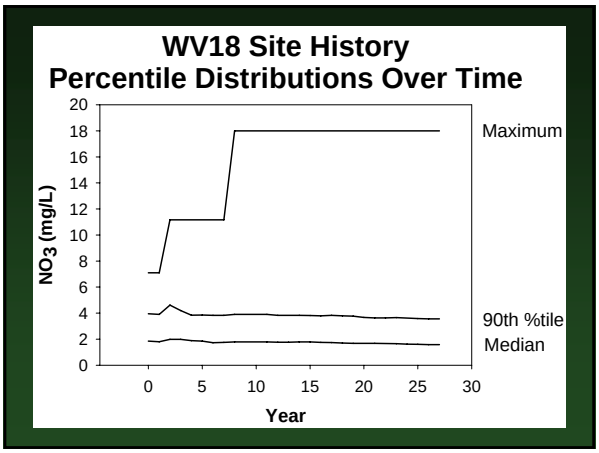
Concerns with the Use of Site History in SL Coding

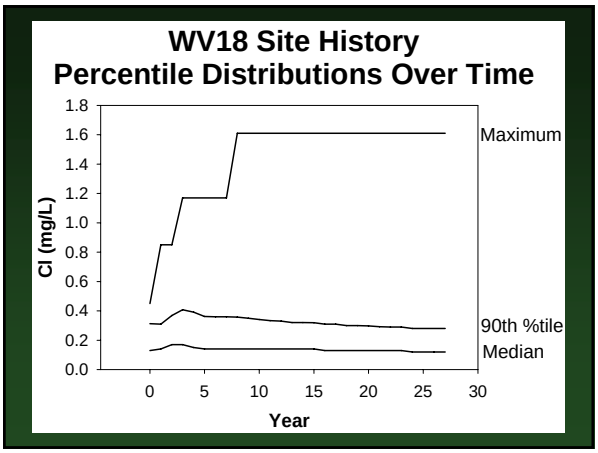
- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?
- Analyte distributions for new sites have small sample sizes with potentially narrow distributions.
  Do young, narrowly distributed site histories affect SL coding?
- Sites start-up in different seasons. In the spring samples may contain more contaminants.
  Does start-up season affect SL coding?

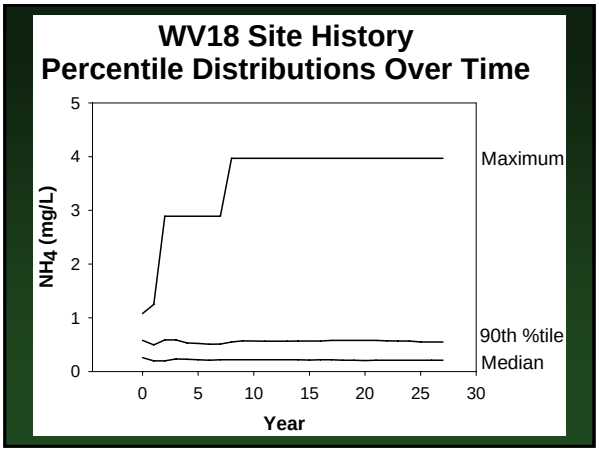


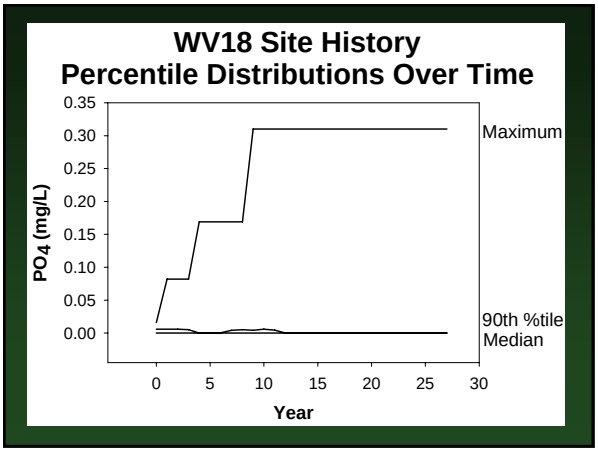


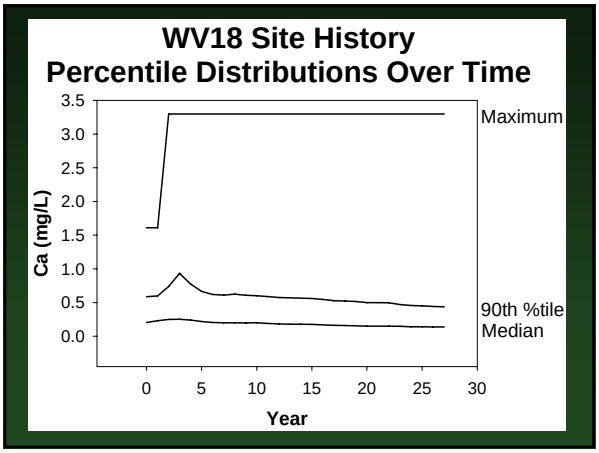


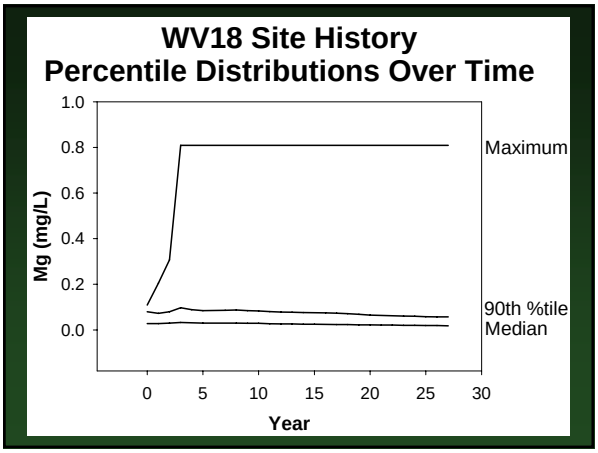

WV18 Site History
Percentile Distributions Over Time
NO3 (mg/L)
Year
Maximum
90th %tile
Median

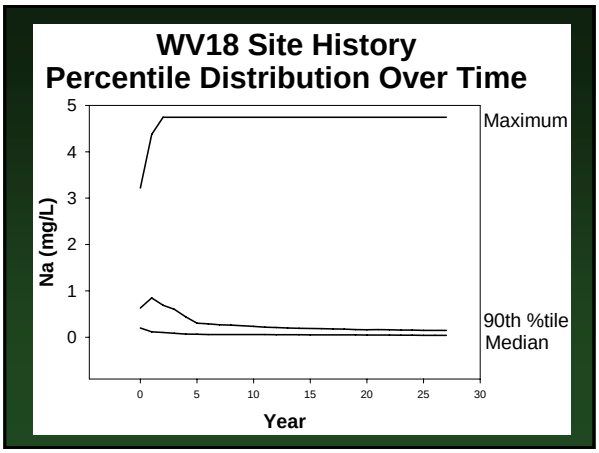

WV18 Site History
Percentile Distributions Over Time
Cl (mg/L)
Year
Maximum
90th %tile
Median

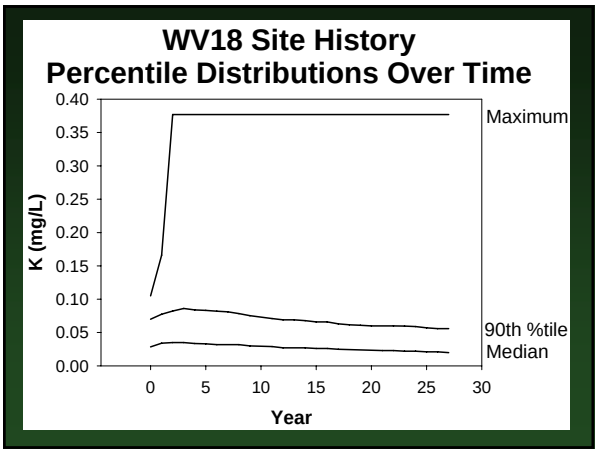

WV18 Site History
Percentile Distributions Over Time
NH4 (mg/L)
Year
Maximum
90th %tile
Median


WV18 Site History
Percentile Distributions Over Time
PO4 (mg/L)
Year
Maximum
90th %tile
Median


WV18 Site History
Percentile Distributions Over Time
Ca (mg/L)
Year
Maximum
90th %tile
Median


WV18 Site History
Percentile Distributions Over Time
Mg (mg/L)
Year
Maximum
90th %tile
Median





The Maximum Problem

The Maximum Problem

- Maxima are unidirectional

The Maximum Problem

- Maxima are unidirectional
- Older sites have maxima that are often decades old, yet still serve as bench marks in scoring outliers

The Maximum Problem

- Maxima are unidirectional
- Older sites have maxima that are often decades old, yet still serve as bench marks in scoring outliers
- Contaminated samples that are non-representative of site precipitation chemistry may not be flagged as such because analyte concentrations that are otherwise anomalous fail to achieve new maxima

The Maximum Solution?

The Maximum Solution?

- Floating Site Histories

The Maximum Solution?

- Floating Site Histories
- Site histories based on “recent” data

The Maximum Solution?

- Floating Site Histories
- Site histories based on “recent” data
- But, “dry” sites may be at a disadvantage since site history sample size may be far smaller than that of a “wet” site

The Maximum Solution?

- Floating Site Histories
- Site histories based on “recent” data
- But, “dry” sites may be at a disadvantage since site history sample size may be far smaller than that of a “wet” site
- Site histories based on last *x* number of samples may alleviate this problem

The Maximum Solution?

- Floating Site Histories
- Site histories based on “recent” data
- But, “dry” sites may be at a disadvantage since site history sample size may be far smaller than that of a “wet” site
- Site histories based on last *x* number of samples may alleviate this problem
- Or…

The Maximum Solution? (continued)

The 99th percentile

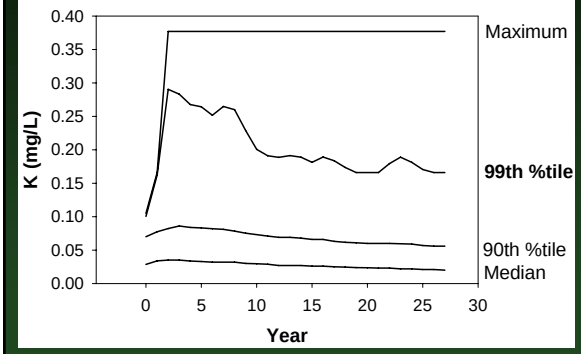
The Maximum Solution? (continued)

- The 99th percentile could replace the maximum as a new bench mark for scoring outliers

The Maximum Solution? (continued)

- The 99th percentile could replace the maximum as a new bench mark for scoring outliers
- The 99th percentile is bidirectional and will track average trends

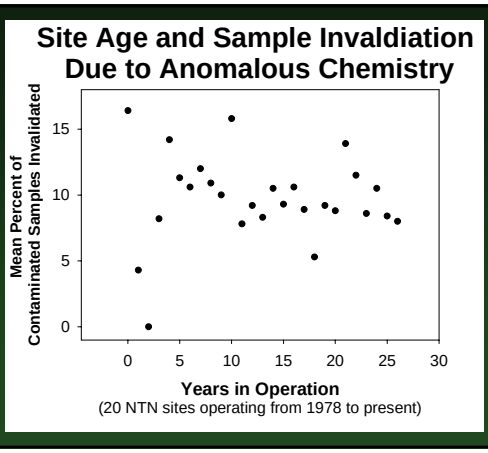
The 99th Percentile



Concerns with the Use of Site History in SL Coding

- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?
- Analyte distributions for new sites have small sample sizes with potentially narrow distributions.
  Do young, narrowly distributed site histories affect SL coding?
- Sites start-up in different seasons. In the spring samples may contain more contaminants.
  Does start-up season affect SL coding?

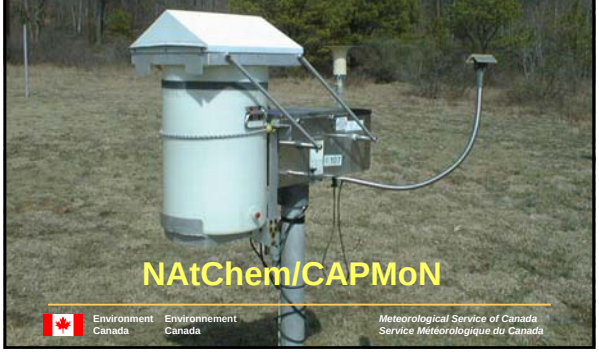
Site Age and Sample Invaldiation Due to Anomalous Chemistry



Concerns with the Use of Site History in SL Coding

- Many sites have histories > 25 years old.
  Do ever-expanding analyte distributions affect SL coding?
- Analyte distributions for new sites have small sample sizes with potentially narrow distributions.
  Do young, narrowly distributed site histories affect SL coding?
- Sites start-up in different seasons. In the spring samples may contain more contaminants.
  Does start-up season affect SL coding?

Evaluation of Outliers and Extreme Values

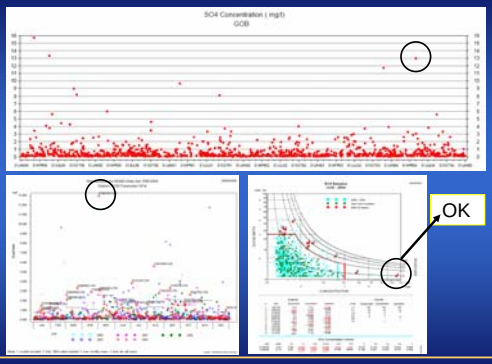


NAtChem/CAPMoN

Five steps for every species for every site

1. Inspect Time Series plots
2. Inspect Concentration/Precipitation/Deposition plots
3. Inspect Superimposed Time Series plots
4. For the outliers on the plots, evaluate other species concentrations, ion balance values and field and lab comments
5. Apply appropriate valid or invalid flag

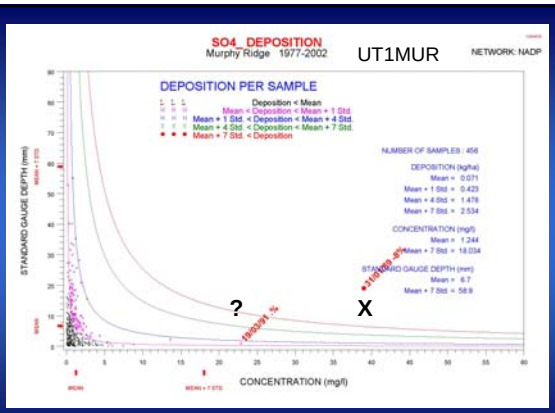
Another Example – Deemed Valid

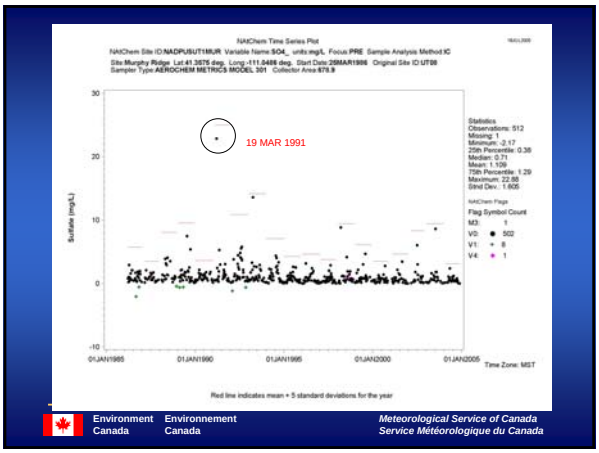


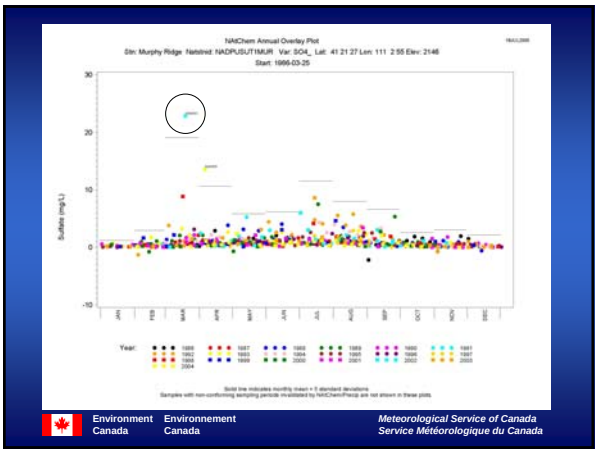
Examples of NADP Outliers

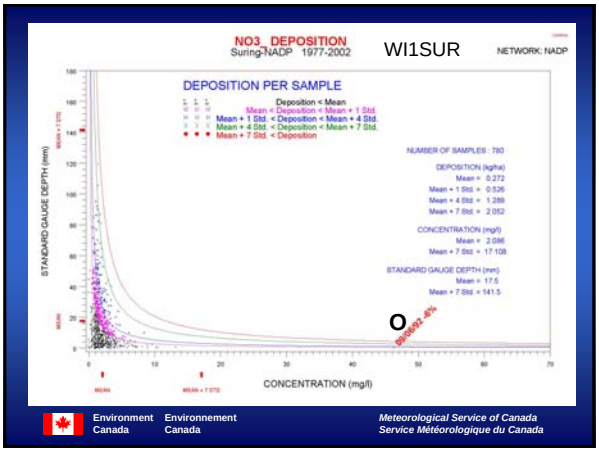
$SO_4^=$ OUTLIERS
UT08 (Murphy Ridge)

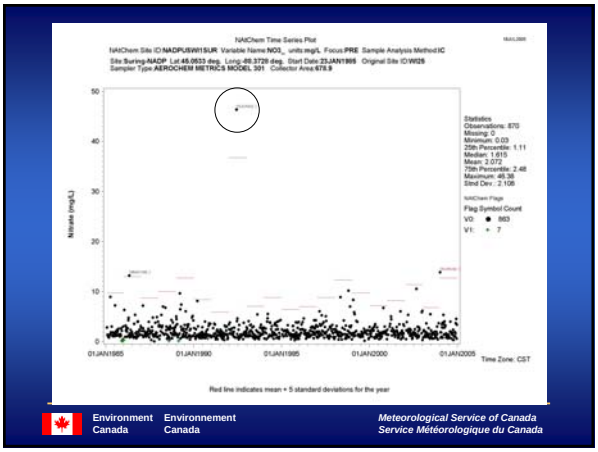
Elevation = 2146 m

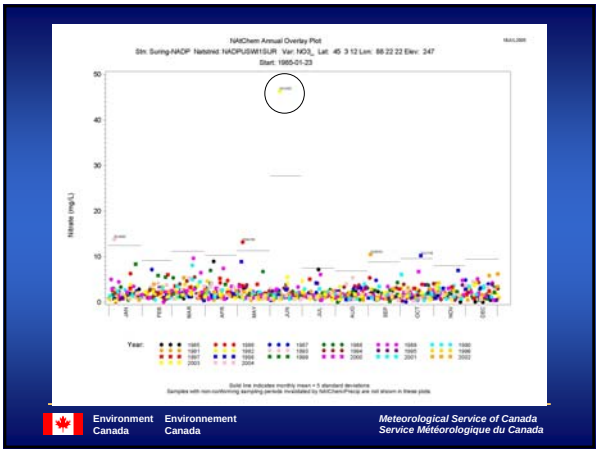


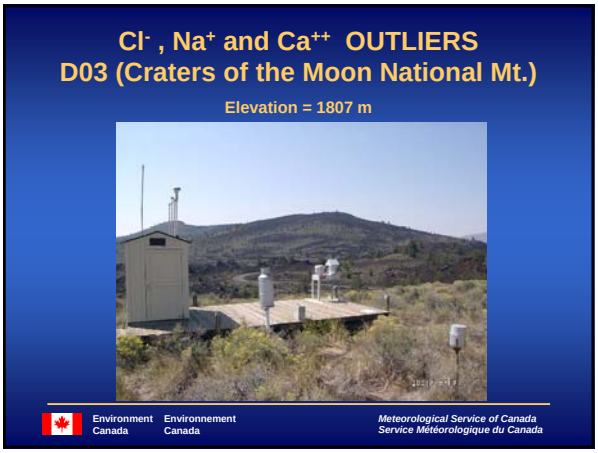

19 MAR 1991
Red line indicates mean + 5 standard deviations for the year
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

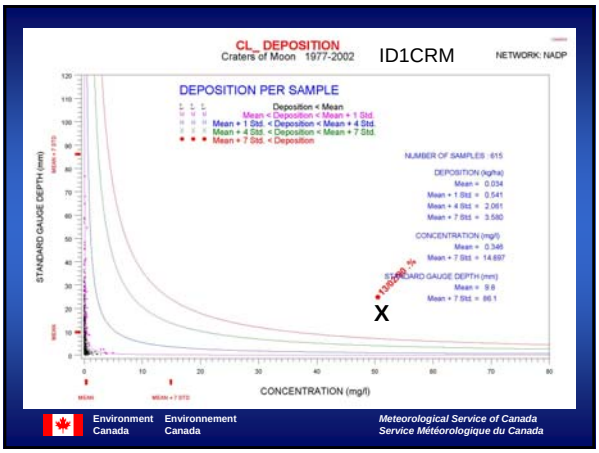

Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

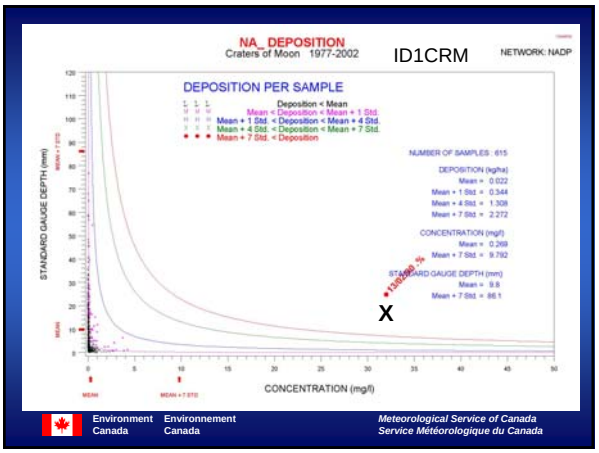

NO3_ DEPOSITION
Suring-NADP 1977-2002
WI1SUR
NETWORK: NADP
DEPOSITION PER SAMPLE
STANDARD GAUGE DEPTH (mm)
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

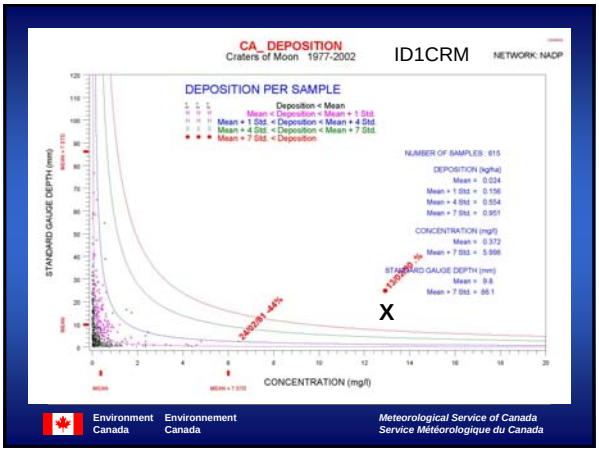

Red line indicates mean + 5 standard deviations for the year
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

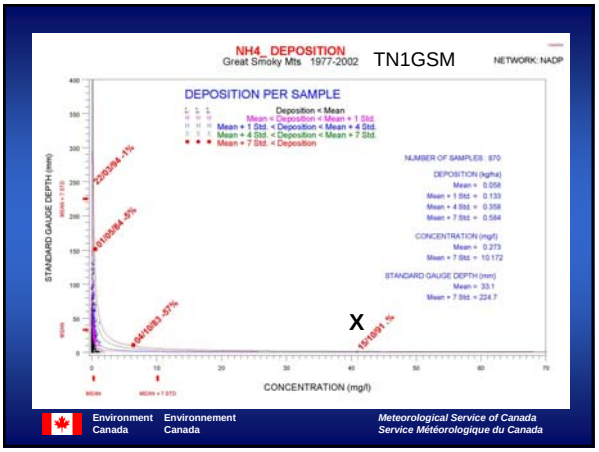

Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

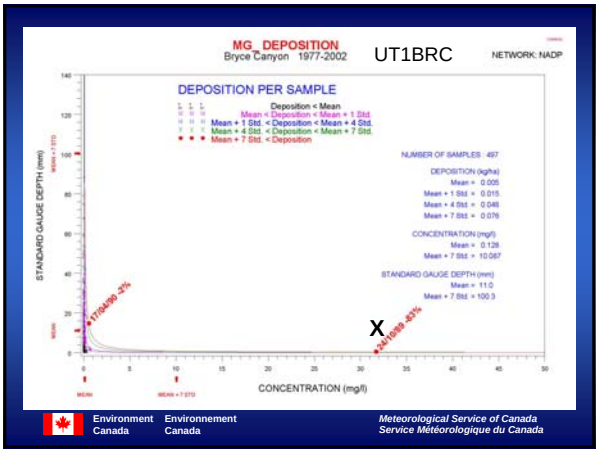

Cl- , Na+ and Ca++ OUTLIERS
D03 (Craters of the Moon National Mt.)
Elevation = 1807 m
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

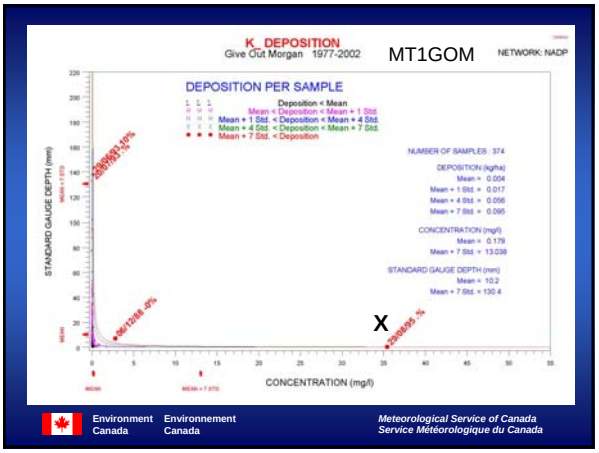

CL_ DEPOSITION
Craters of Moon 1977-2002
ID1CRM
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

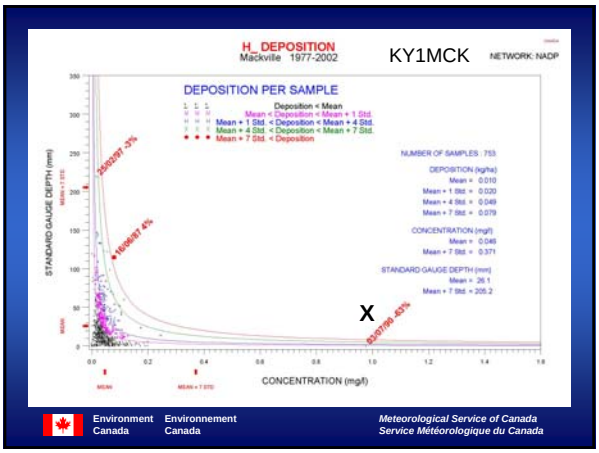

NA_ DEPOSITION
Craters of Moon 1977-2002
ID1CRM
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada


CA_ DEPOSITION
Craters of Moon 1977-2002
ID1CRM
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

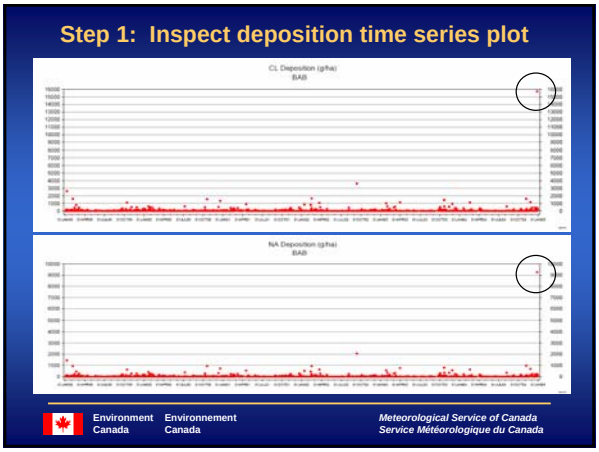

NH4_ DEPOSITION
Great Smoky Mts 1977-2002
TN1GSM
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

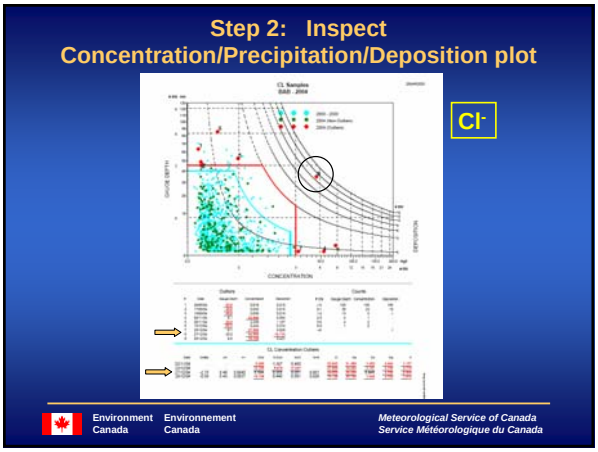

MG_ DEPOSITION
Bryce Canyon 1977-2002
UT1BRC
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

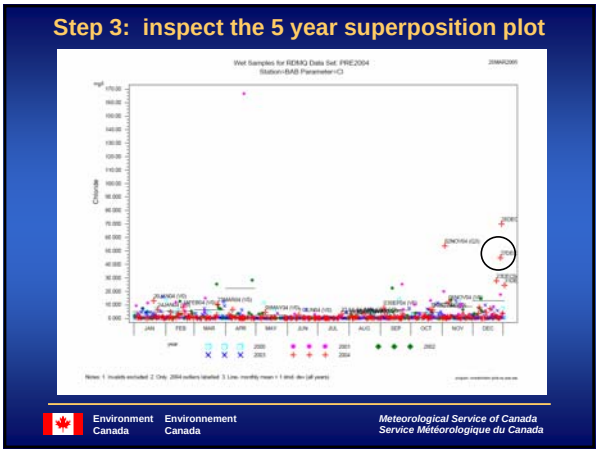

K_ DEPOSITION
Give Out Morgan 1977-2002
MT1GOM
NETWORK: NADP
DEPOSITION PER SAMPLE
CONCENTRATION (mg/l)
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

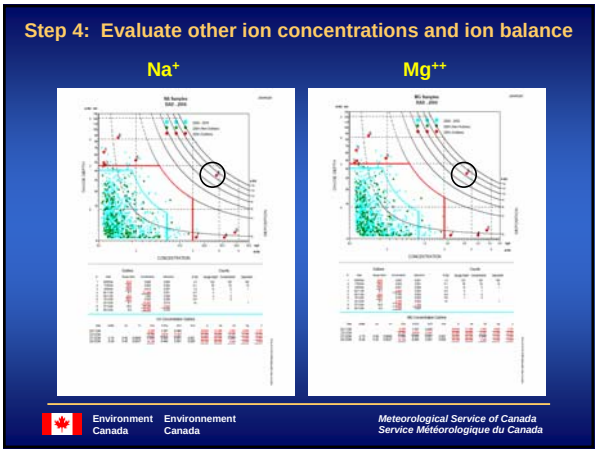

H_ DEPOSITION
Mackville 1977-2002
KY1MCK
NETWORK: NADP
DEPOSITION PER SAMPLE
STANDARD GAUGE DEPTH (mm)
CONCENTRATION (mg/l)
X
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada


END
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

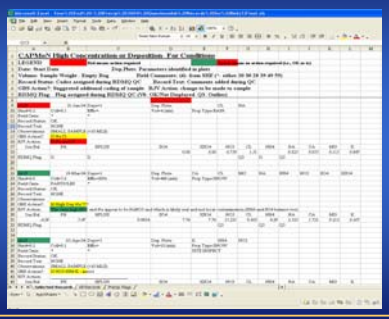

Step 1: Inspect deposition time series plot
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada


Step 2: Inspect Concentration/Precipitation/Deposition plot
Cl-
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada


Step 3: inspect the 5 year superposition plot
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

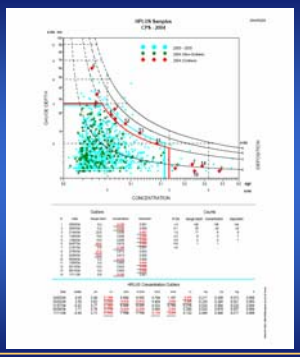

Step 4: Evaluate other ion concentrations and ion balance
Na+
Mg++
Environment Canada
Environnement Canada
Meteorological Service of Canada
Service Météorologique du Canada

Step 5: Assign a Valid or Invalid Flag (Valid in this case because the high $SO_4^{=}$, Cl^{-}, Na^{+} and Mg^{++} values were due to sea salt in a big storm)



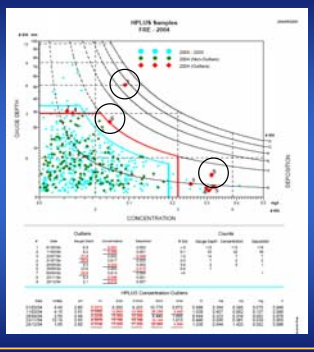
Environment Canada Environnement Canada — *Meteorological Service of Canada Service Météorologique du Canada*

A case with no outliers



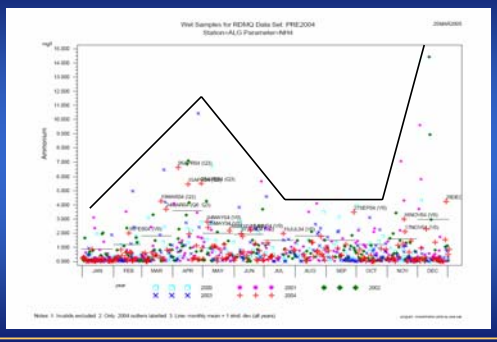
Environment Canada Environnement Canada — *Meteorological Service of Canada Service Météorologique du Canada*

An example of 3 points to be investigated



Environment Canada Environnement Canada — *Meteorological Service of Canada Service Météorologique du Canada*

An example of the importance of evaluating outliers with respect to seasonality and other years



Environment Canada Environnement Canada — *Meteorological Service of Canada Service Météorologique du Canada*

Summary of Method

- Outliers are identified through inspection of the three types of plots
- Once identified, an outlier value is assessed with respect to its position on the Concentration-Precipitation-Deposition plot, the seasonality of the ion, the other ions, the ion balance, and lab/field comments
- A datum is flagged as invalid only if there is clear evidence of contamination or sampling problems

Environment Canada Environnement Canada — *Meteorological Service of Canada Service Météorologique du Canada*