ASSESSMENT OF MEASUREMENT ERROR IN CSN DATA USING COLLOCATED SAMPLERS

KELSEY HADDAD, LI DU, JAY TURNER

Washington University in St. Louis

National Ambient Air Monitoring Conference

August 11-14, 2014

MEASUREMENT ERROR: WHY DO WE CARE?

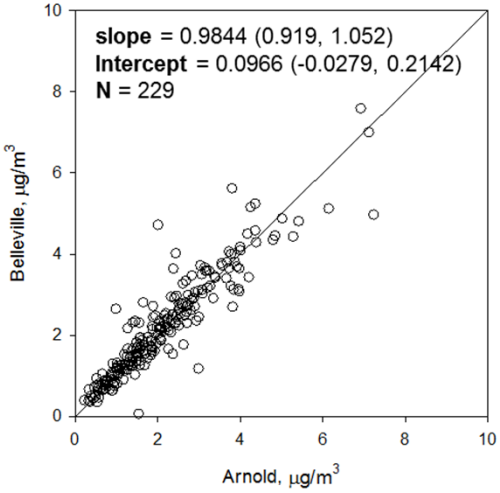
MEASUREMENT ERROR: WHY DO WE CARE?

- **Source Apportionment Modelling**

MEASUREMENT ERROR: WHY DO WE CARE?

- **Source Apportionment Modelling**

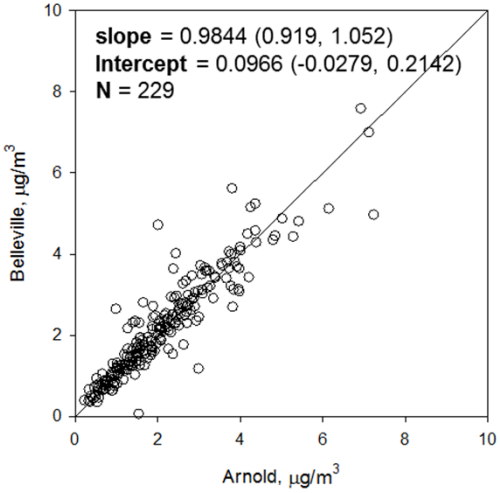
- **Interpretation of Spatiotemporal Variability**



MEASUREMENT ERROR: WHY DO WE CARE?

- **Source Apportionment Modelling**

- **Interpretation of Spatiotemporal Variability**



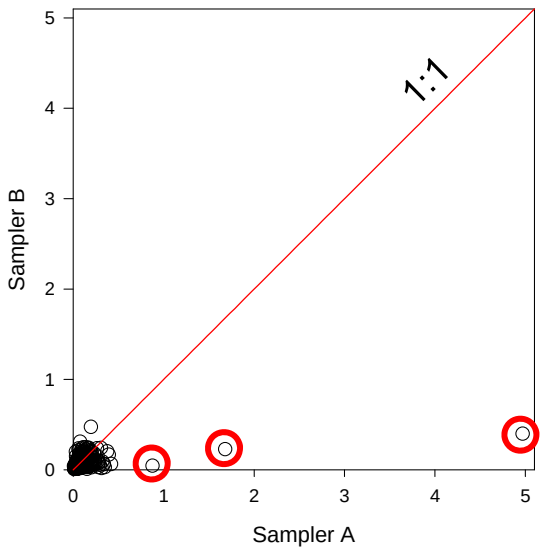
- **Network Operations**

COLLOCATED SAMPLING SITES IN THE CSN NETWORK

MANIFESTATIONS OF MEASUREMENT ERROR

MANIFESTATIONS OF MEASUREMENT ERROR

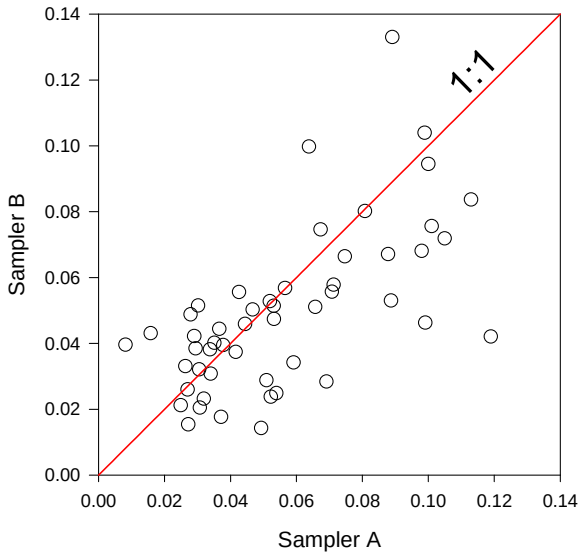
- **Outliers**



MANIFESTATIONS OF MEASUREMENT ERROR

- **Outliers**

- **Systematic Bias**

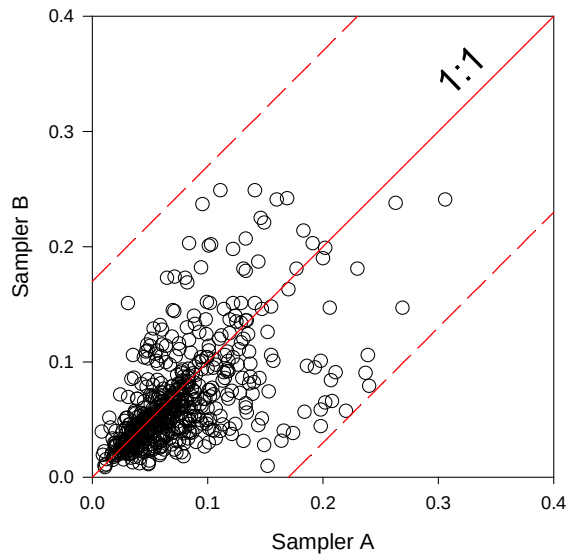


MANIFESTATIONS OF MEASUREMENT ERROR

- **Outliers**

- **Systematic Bias**

- **Random Noise**



MANIFESTATIONS OF MEASUREMENT ERROR

- **Outliers**

- **Systematic Bias**

- **Random Noise**

PRECISION AND UNCERTAINTY

PRECISION AND UNCERTAINTY

Top-Down Approach:

Sample Specific Bias (D_i) = $\frac{(C_{i1} - C_{i2})/\sqrt{2}}{\bar{C}_i}$

C_{i1} = sample *i*, primary
C_{i2} = sample *i*, collocate

Hyslop, N. P. and W. H. White (2009). *J. Air & Waste Manage. Assoc.* **59**(9): 1032-1039.
Hyslop, N. P. and W. H. White (2008). *Atmos. Environ.* **42**(11): 2691-2705.

PRECISION AND UNCERTAINTY

Top-Down Approach:

Sample Specific Bias (D_i) = $\frac{(C_{i1} - C_{i2})/\sqrt{2}}{\bar{C}_i}$

C_{i1} = sample *i*, primary
C_{i2} = sample *i*, collocate

Root Mean Square (RMS) Precision (%) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n} D_i^2} \times 100$

Hyslop, N. P. and W. H. White (2009). *J. Air & Waste Manage. Assoc.* **59**(9): 1032-1039.
Hyslop, N. P. and W. H. White (2008). *Atmos. Environ.* **42**(11): 2691-2705.

PRECISION AND UNCERTAINTY

Top-Down Approach:

Sample Specific Bias (D_i) = $\frac{(C_{i1} - C_{i2})/\sqrt{2}}{\bar{C}_i}$

C_{i1} = sample *i*, primary
C_{i2} = sample *i*, collocate

Root Mean Square (RMS) Precision (%) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n} D_i^2} \times 100$

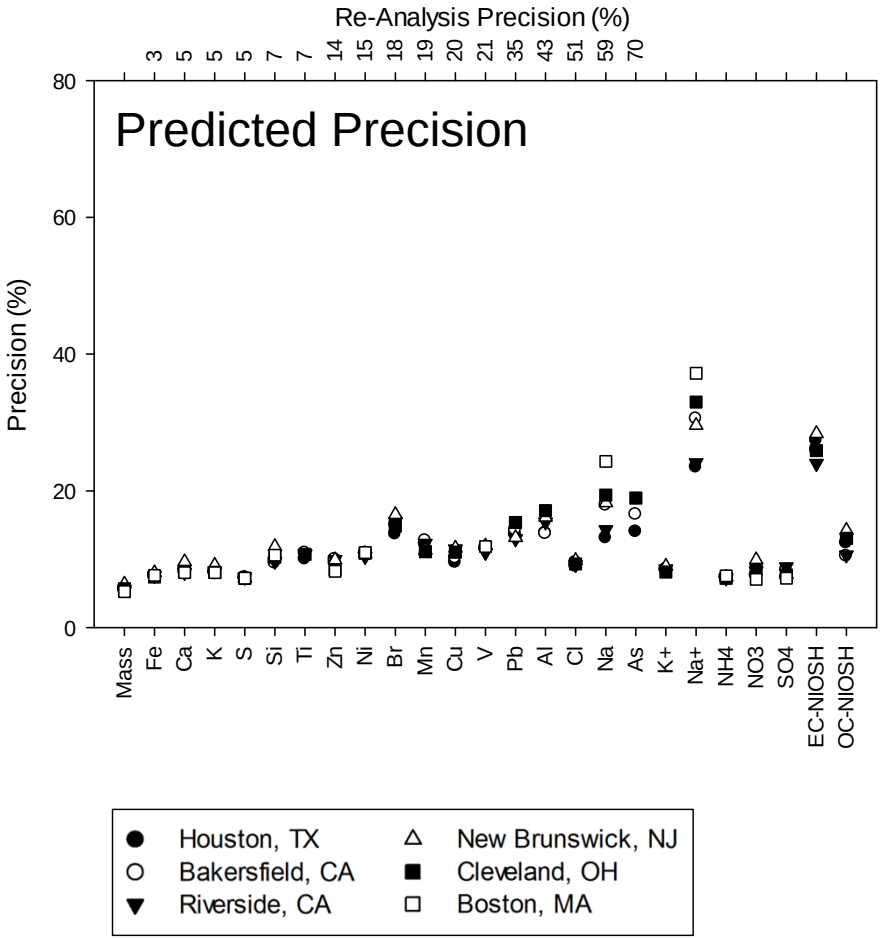
Bottom-up Approach:

Predicted Precision (%) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n} \frac{(U_{C_{i1}}^2 + U_{C_{i2}}^2)/2}{\bar{C}_i}} \times 100$

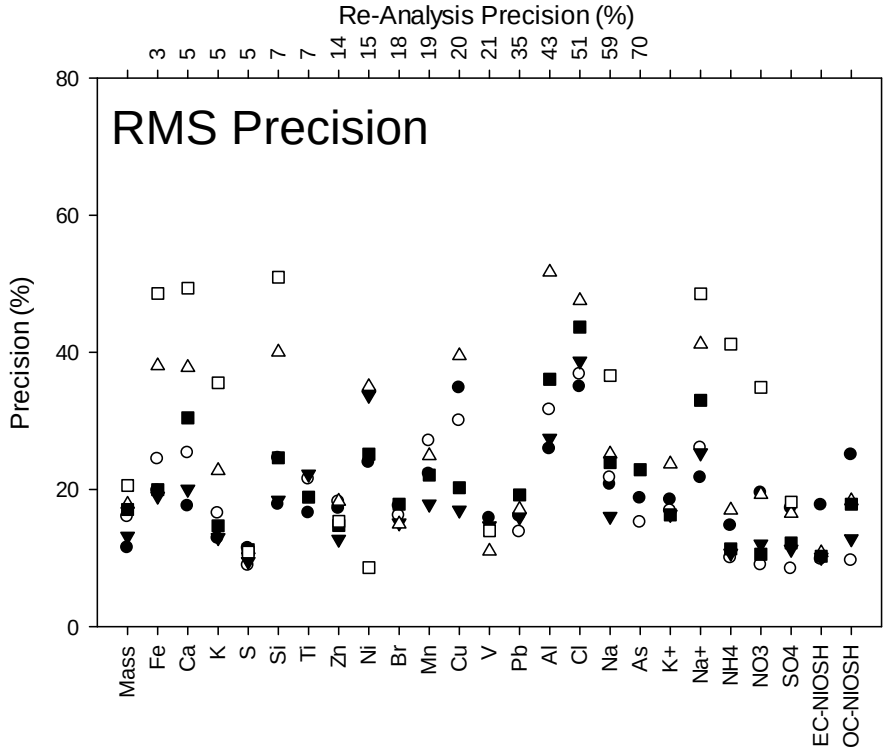
Hyslop, N. P. and W. H. White (2009). *J. Air & Waste Manage. Assoc.* **59**(9): 1032-1039.
Hyslop, N. P. and W. H. White (2008). *Atmos. Environ.* **42**(11): 2691-2705.

Precision Comparison between Collocated Samplers

Bottom-up:

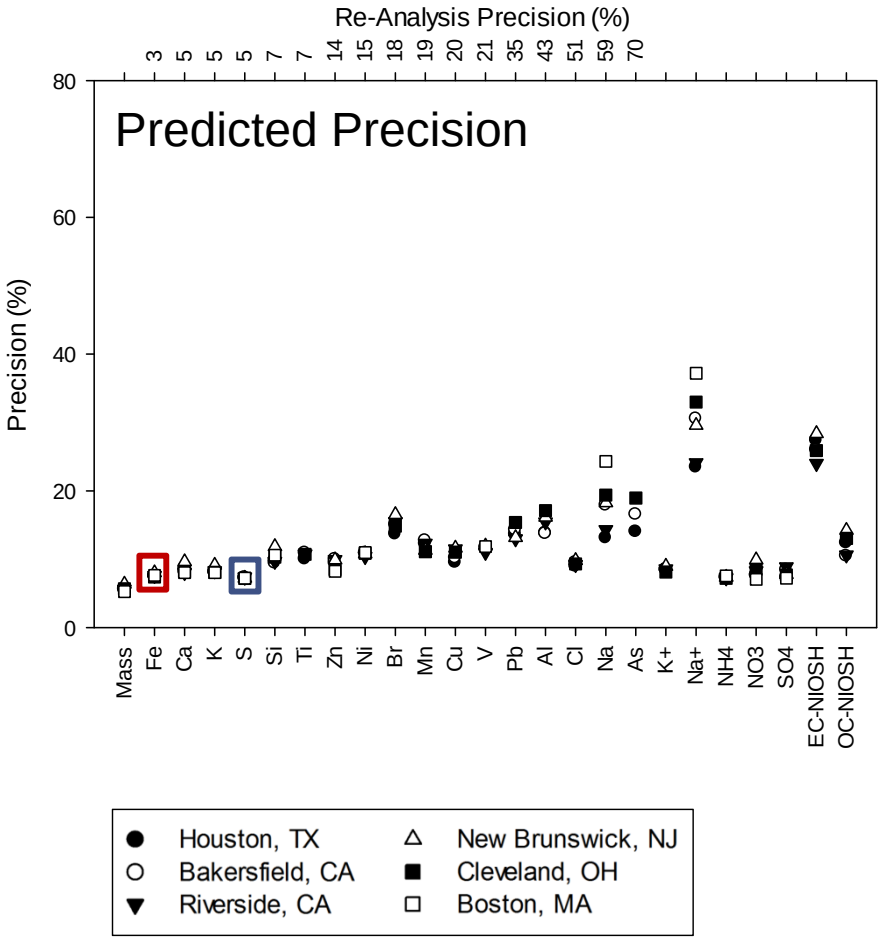


Top-Down:

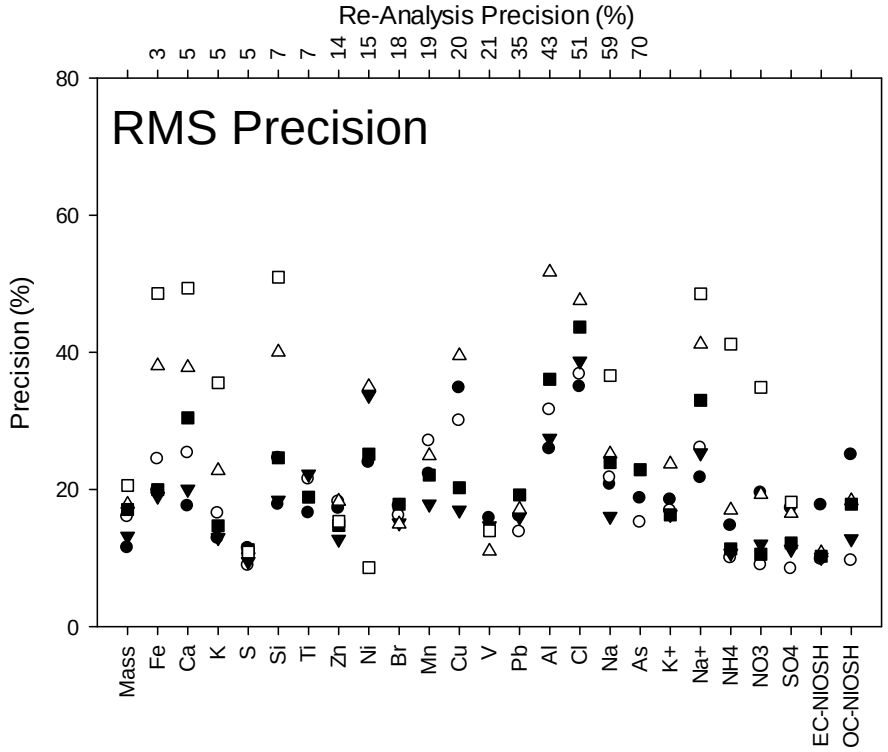


Precision Comparison between Collocated Samplers

Bottom-up:

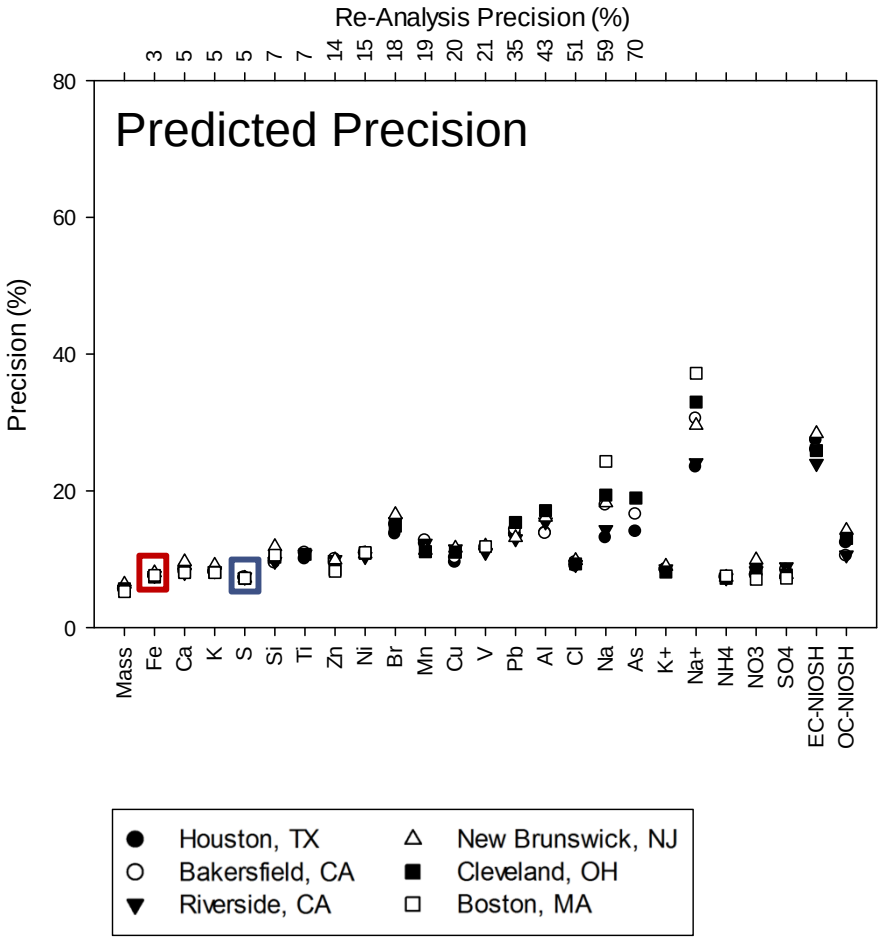


Top-Down:

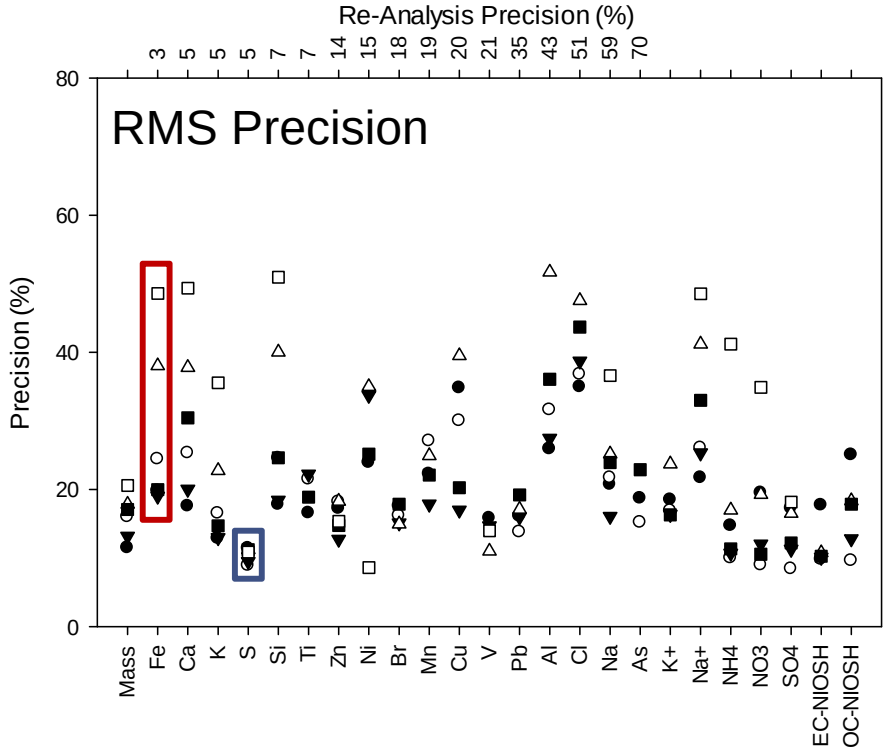


Precision Comparison between Collocated Samplers

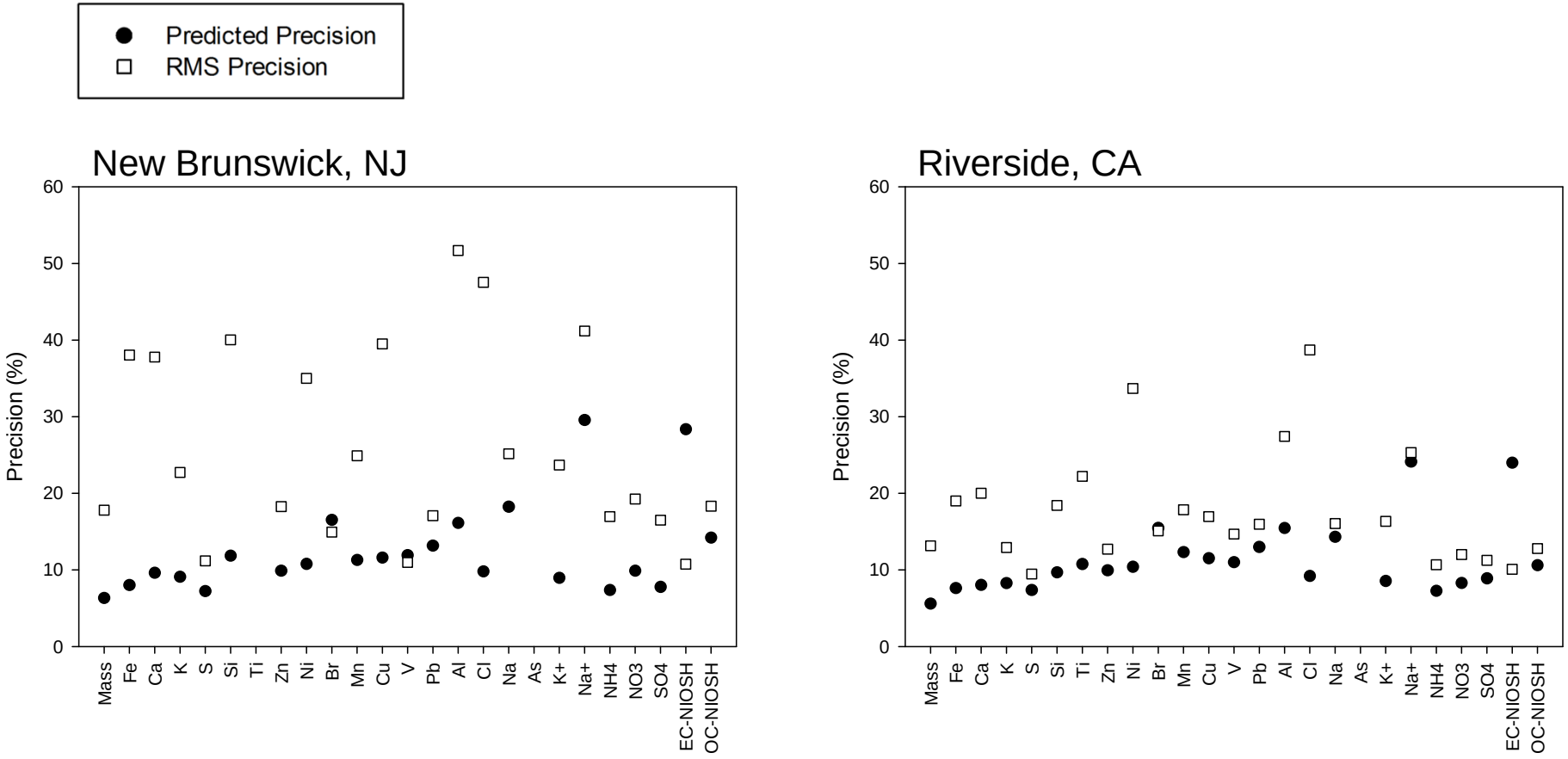
Bottom-up:



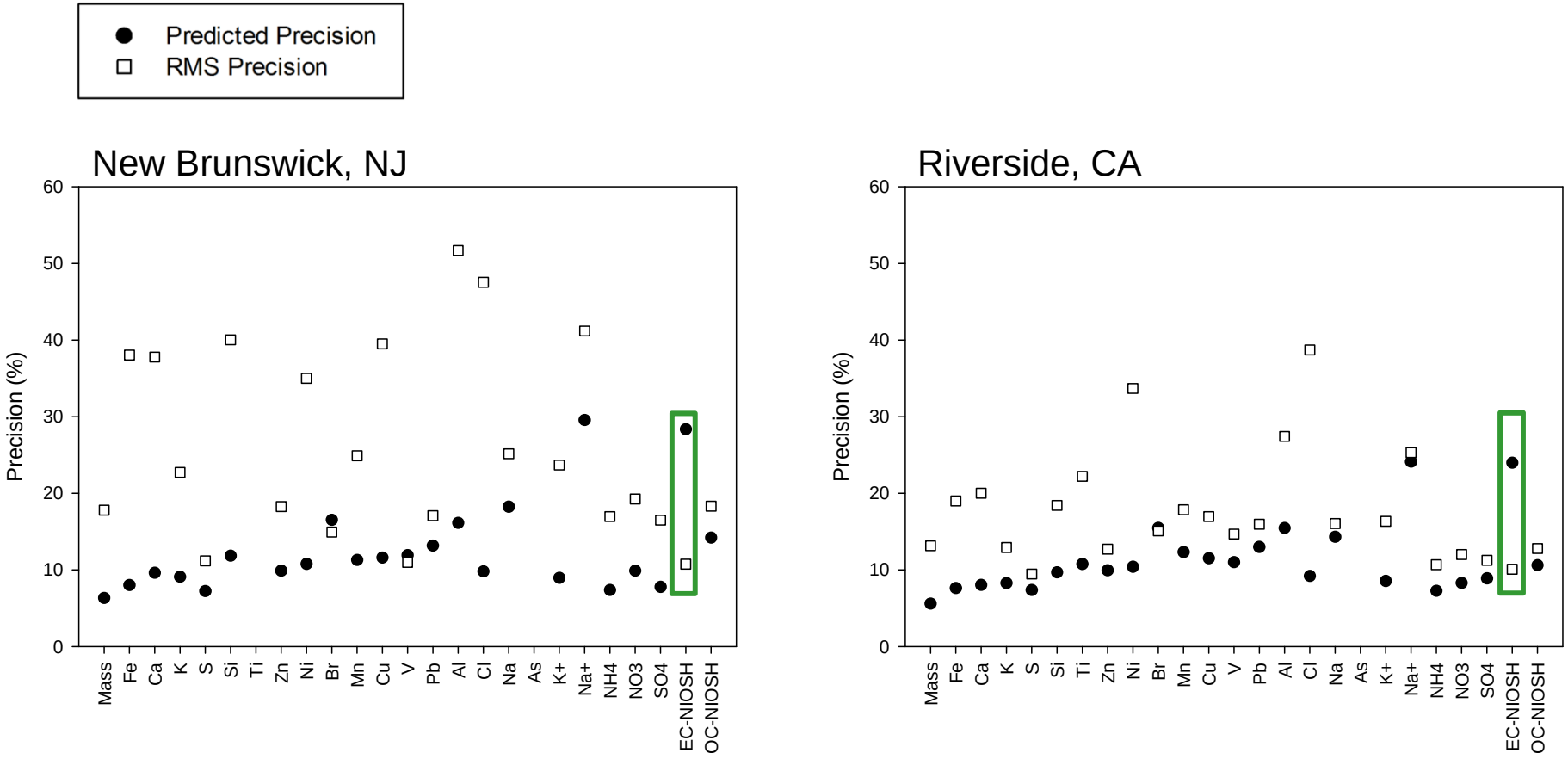
Top-Down:



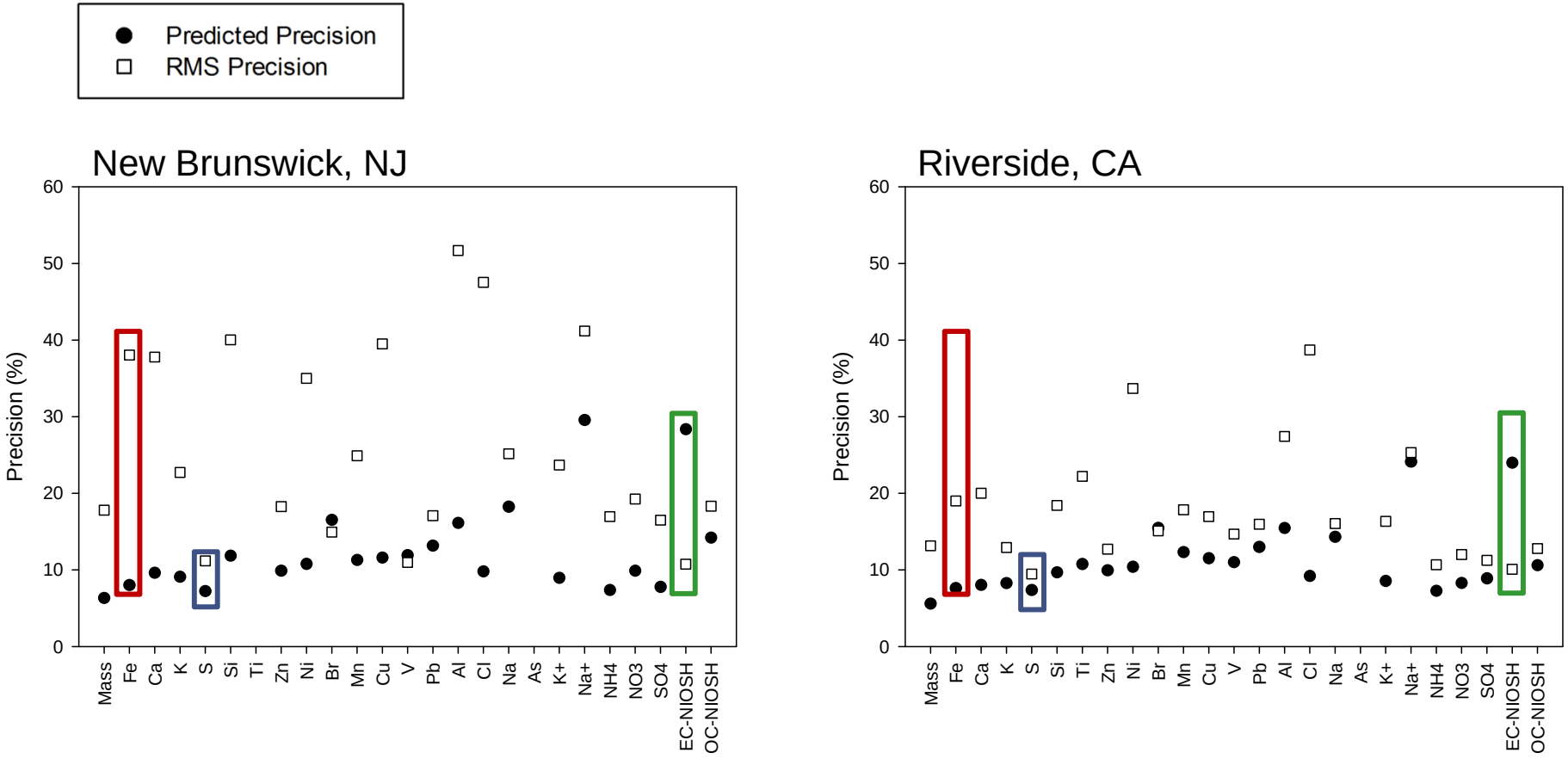
RMS Precision vs. Predicted Precision



RMS Precision vs. Predicted Precision



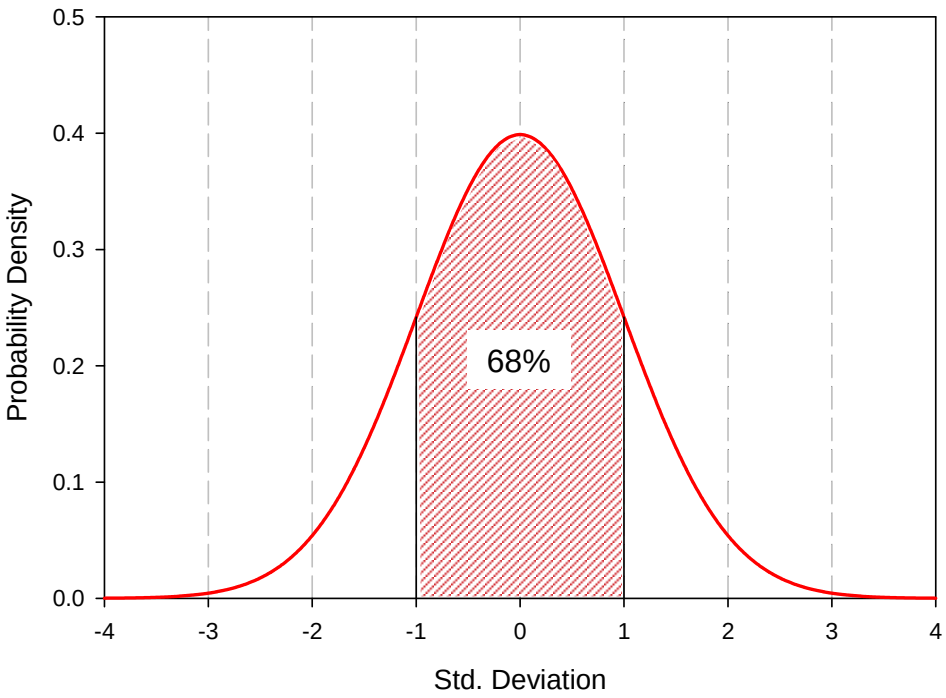
RMS Precision vs. Predicted Precision

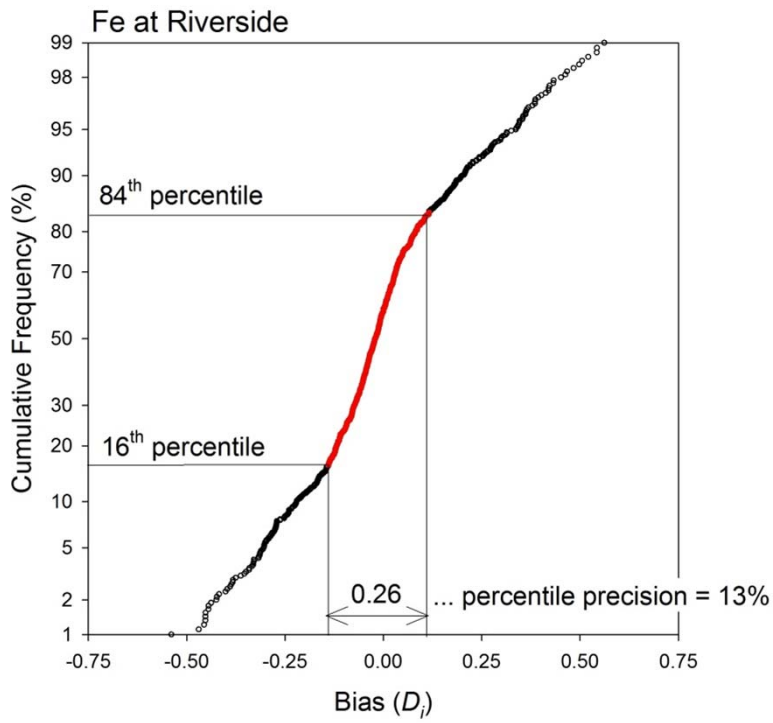


PERCENTILE PRECISION

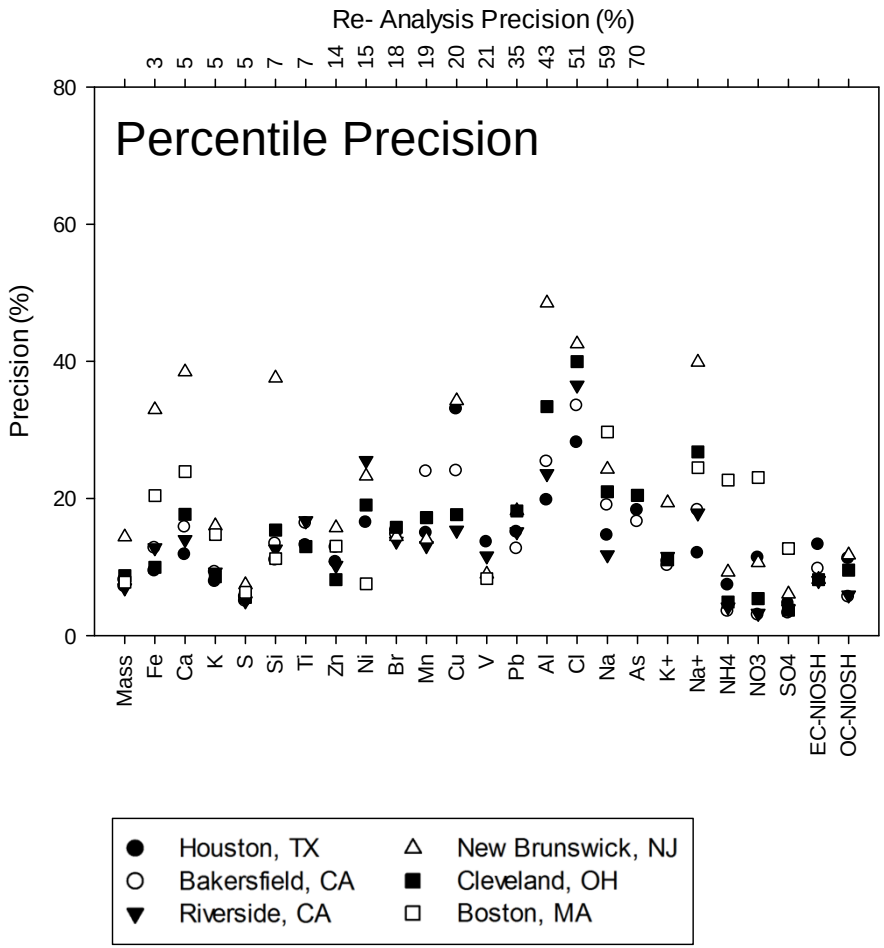
$$\textbf{Percentile Precision (\%)} = \frac{1}{2}\left[P_{84}(D_i) - P_{16}(D_i)\right] \times 100$$

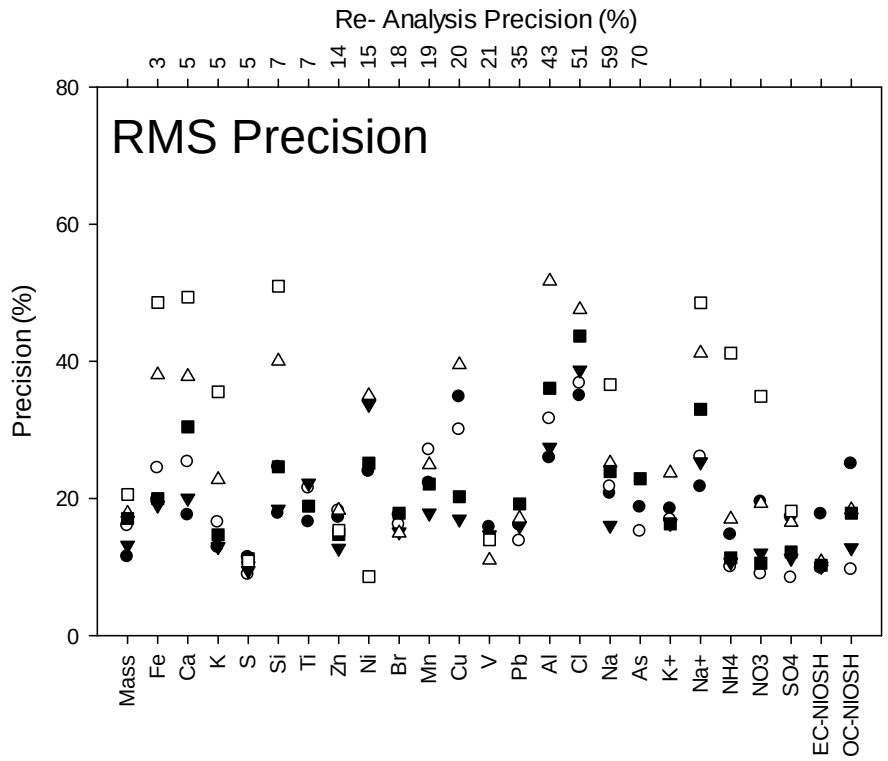
Hyslop, N. P. and W. H. White (2009). *J. Air & Waste Manage. Assoc.* **59**(9): 1032-1039.





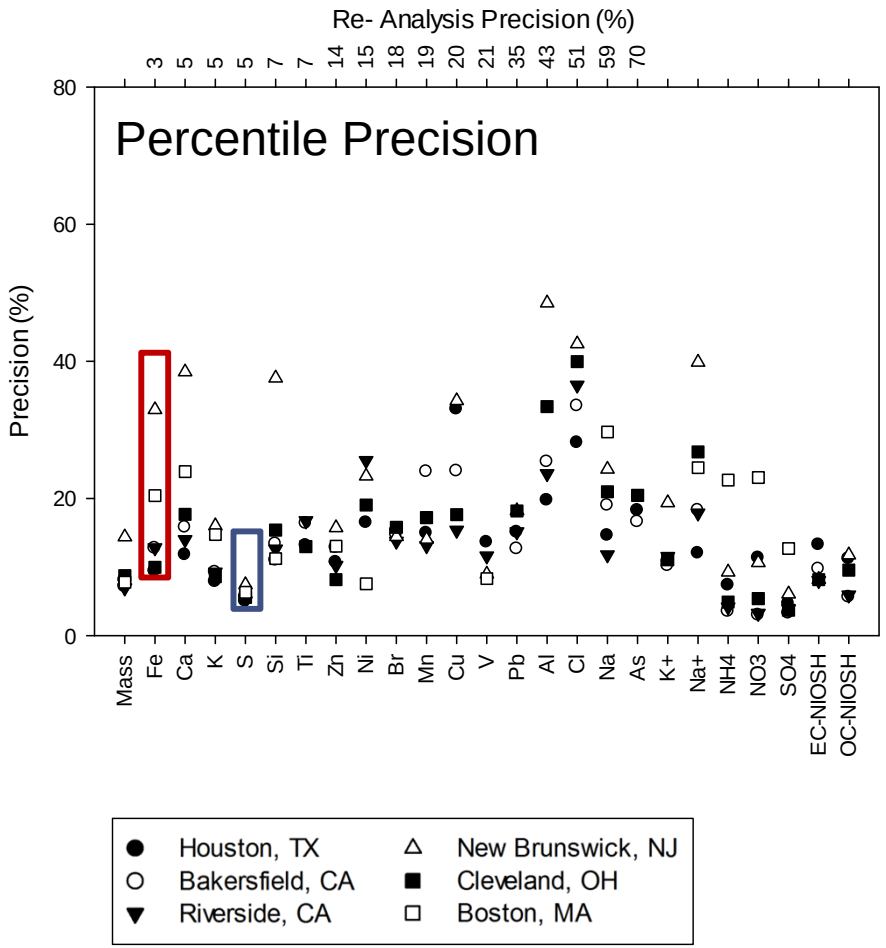
Precision Comparison between Collocated Samplers



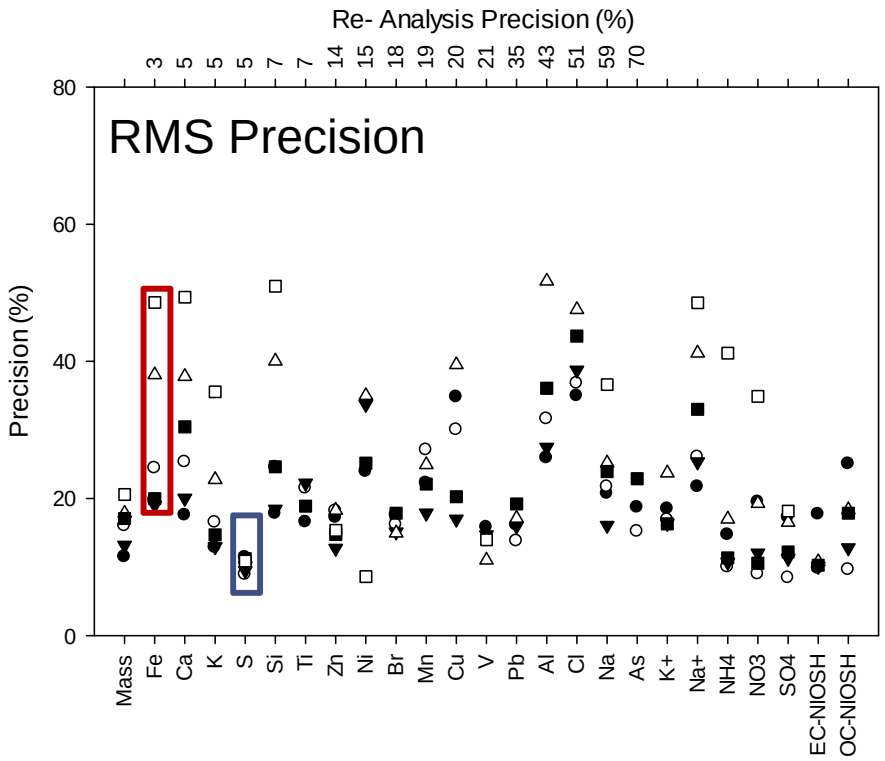


Precision Comparison between Collocated Samplers

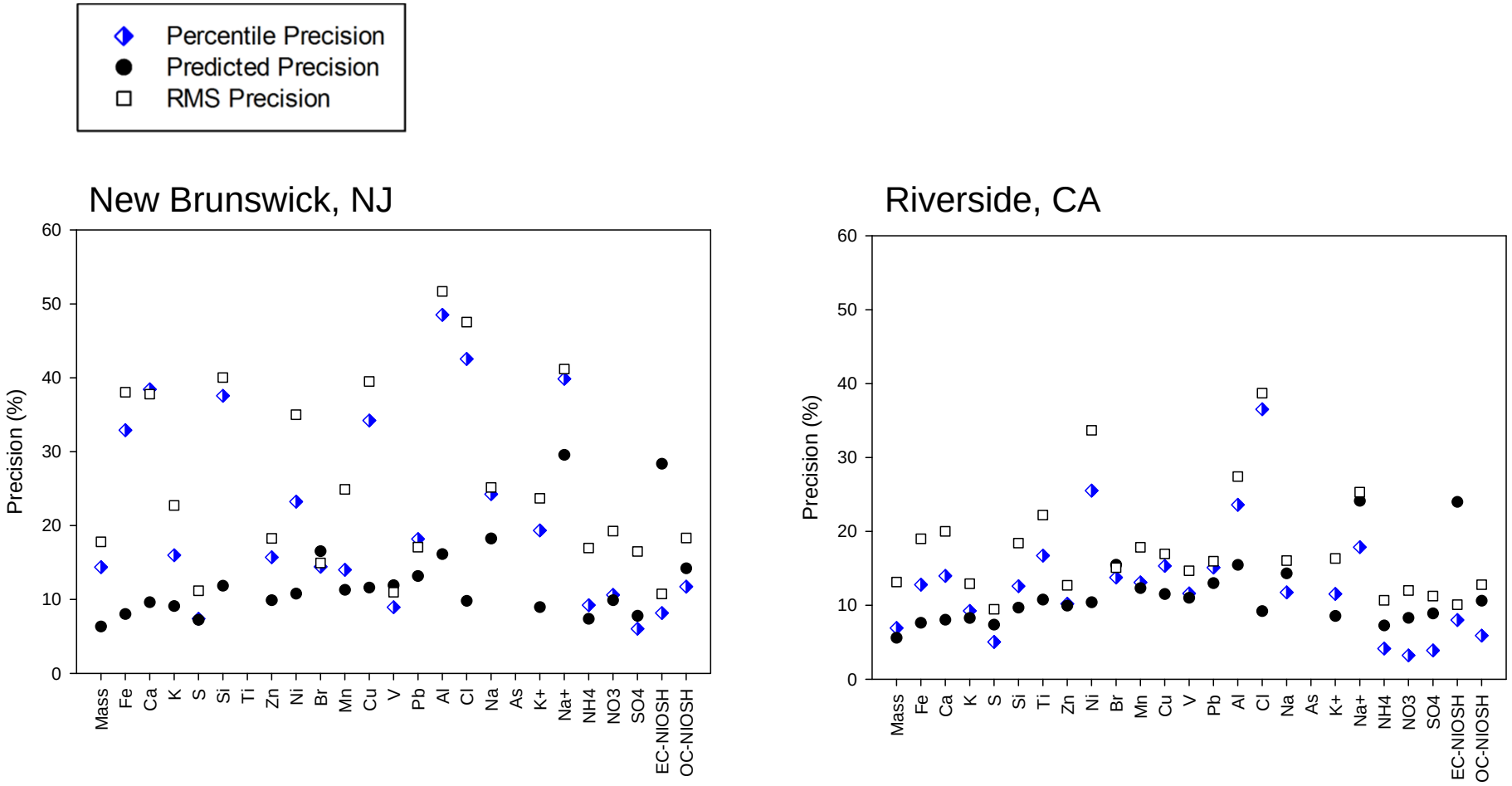
Bottom-up:



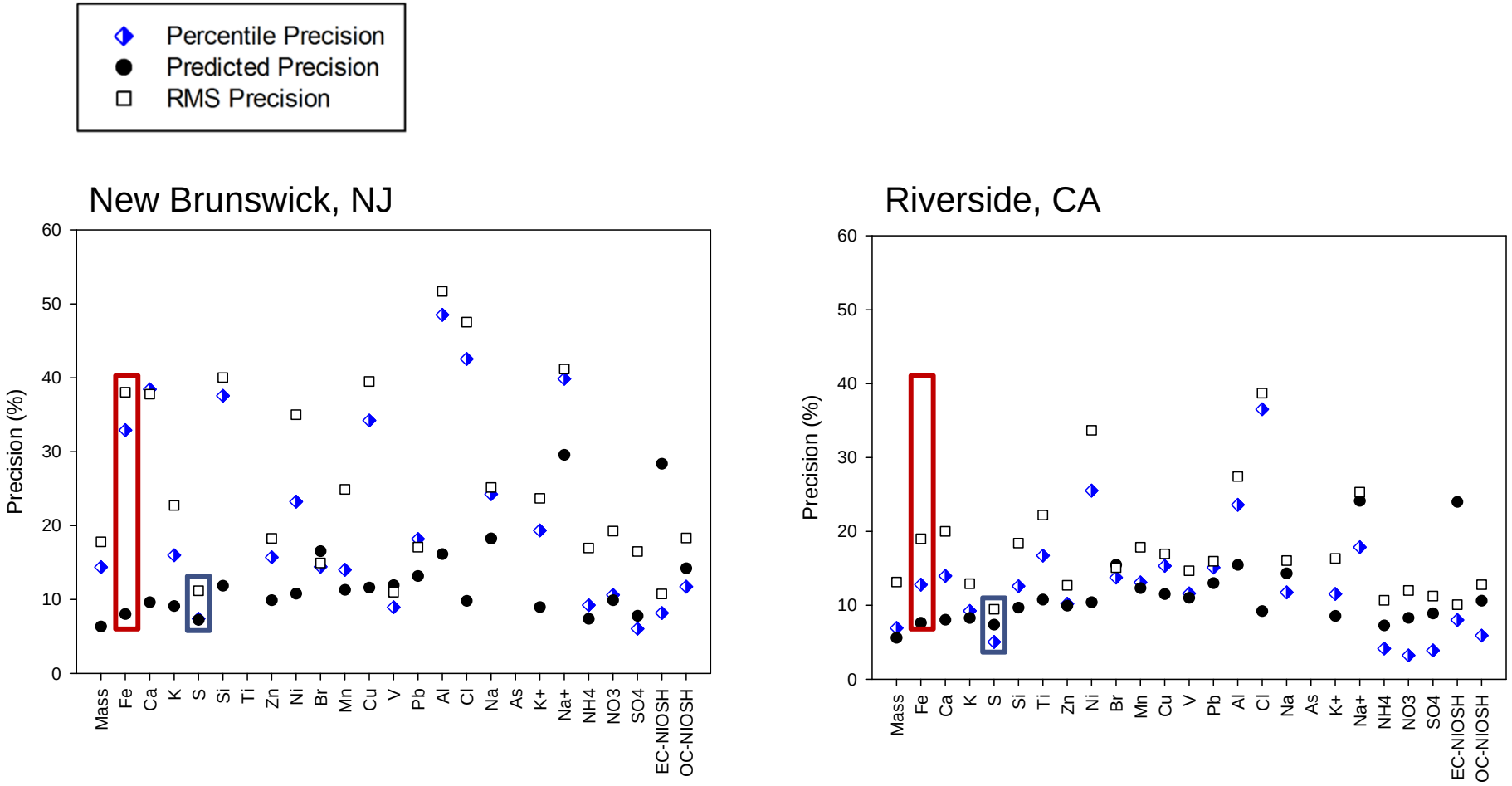
Top Down:



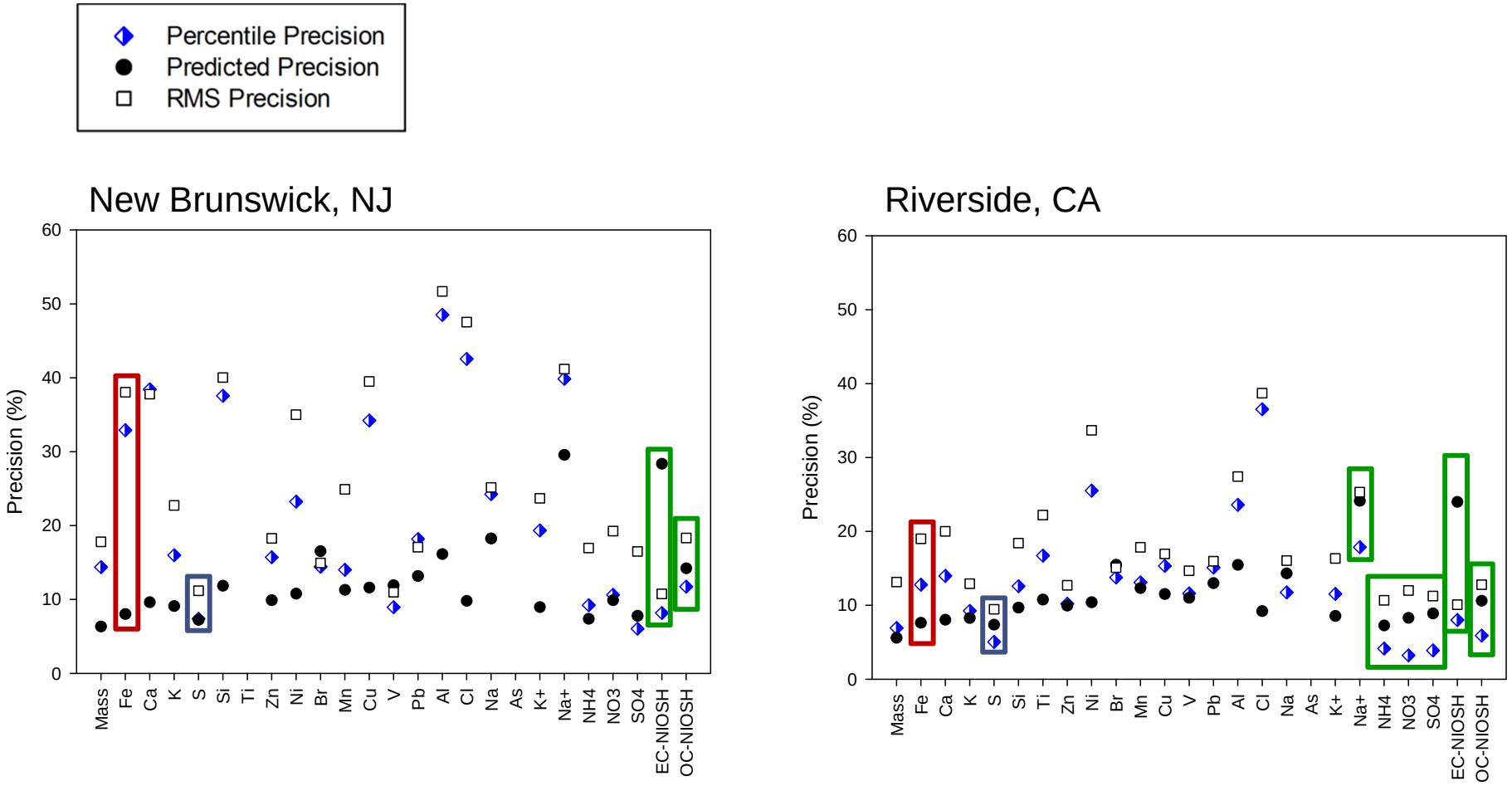
RMS, Percentile, and Predicted Precision



RMS, Percentile, and Predicted Precision



RMS, Percentile, and Predicted Precision



ASSESSMENT OF MEASUREMENT ERROR THUS FAR:

For most (but not all!) XRF species…

- **Predicted precision underestimates the RMS precision which uses all data**
- **Predicted precision also underestimates the percentile precision which uses the "best" 2/3 of the data (e.g., no influence from outliers)**
- **To what extent does <u>systematic bias</u> influence the collocated data precision estimates?**

QUANTIFYING BIAS

QUANTIFYING BIAS

- **Sample Specific Bias (D_i)=** $\frac{(C_{i1}-C_{i2})/\sqrt{2}}{\bar{C}_i}$

QUANTIFYING BIAS

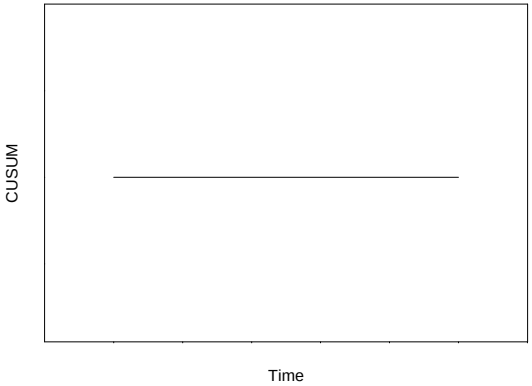
- **Sample Specific Bias (D_i)=** $\frac{(C_{i1}-C_{i2})/\sqrt{2}}{\bar{C}_i}$
- **Overall Bias (%) =** $\frac{1}{n}\sum_{i=1}^{n} D_i \times 100$

QUANTIFYING BIAS

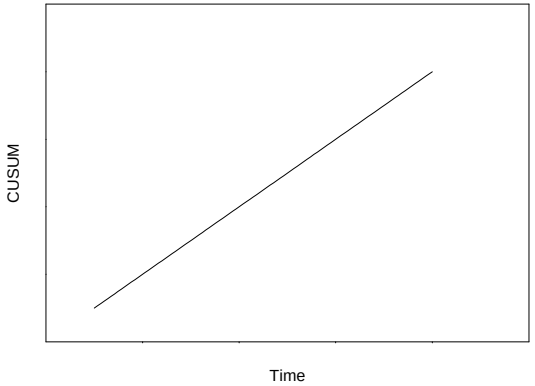
- **Sample Specific Bias (D_i)=** $\frac{(C_{i1}-C_{i2})/\sqrt{2}}{\bar{C}_i}$
- **Overall Bias (%) =** $\frac{1}{n}\sum_{i=1}^{n} D_i \times 100$
- **Cumulative Summation (CUSUM) Plots-** sum of the sample specific bias over first *k* samples; captures temporal variability of the bias

QUANTIFYING BIAS

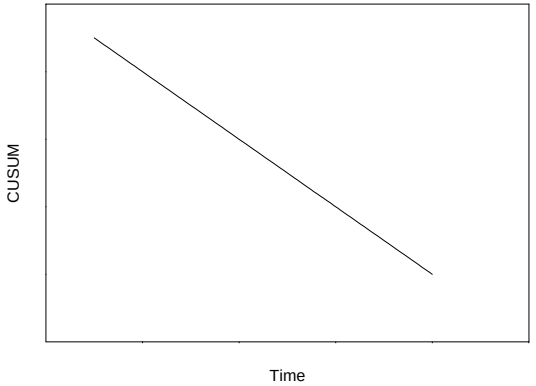
- **Sample Specific Bias (D_i)=** $\frac{(C_{i1}-C_{i2})/\sqrt{2}}{\bar{C}_i}$
- **Overall Bias (%) =** $\frac{1}{n}\sum_{i=1}^{n} D_i \times 100$
- **Cumulative Summation (CUSUM) Plots-** sum of the sample specific bias over first *k* samples; captures temporal variability of the bias



Horizontal: No difference between samplers

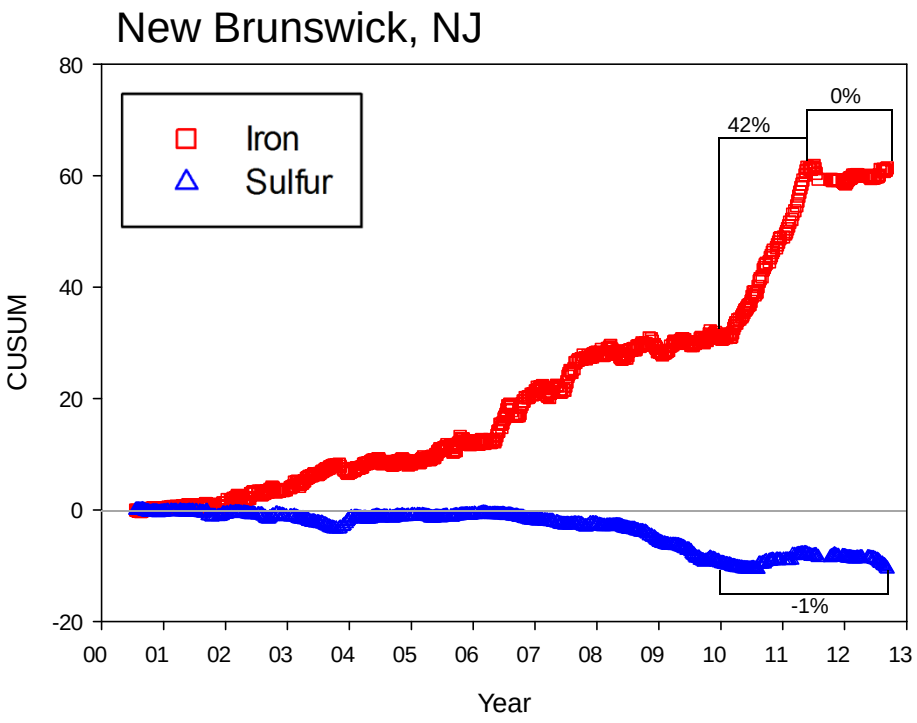


Positive Slope: Sampler A greater than Sampler B

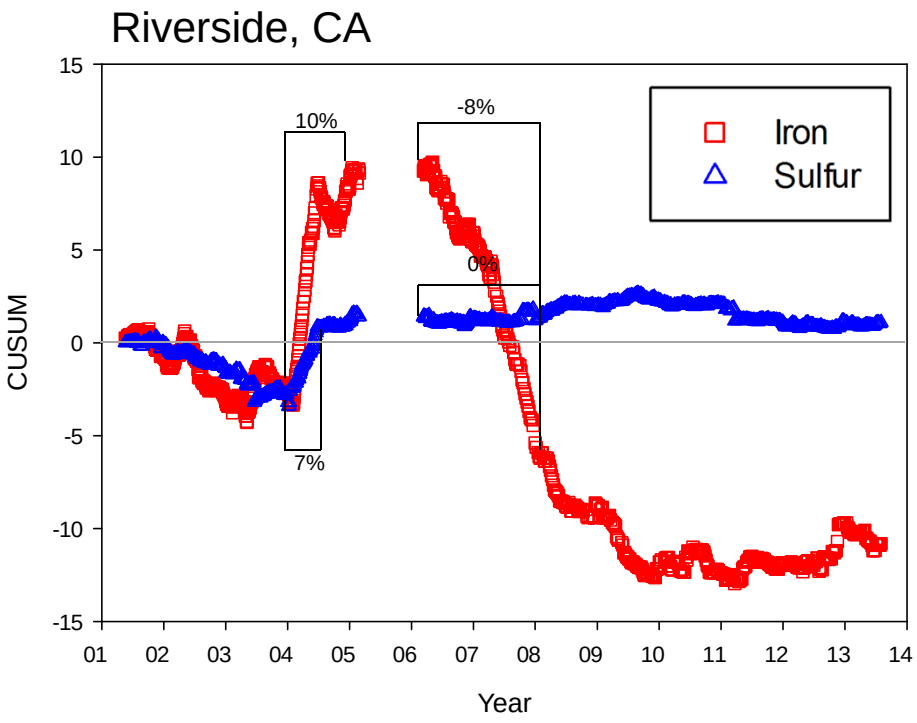


Negative Slope: Sampler B greater than Sampler A

TEMPORAL PATTERNS IN THE BIAS



Overall Bias:
Iron: 10%
Sulfur: -2%



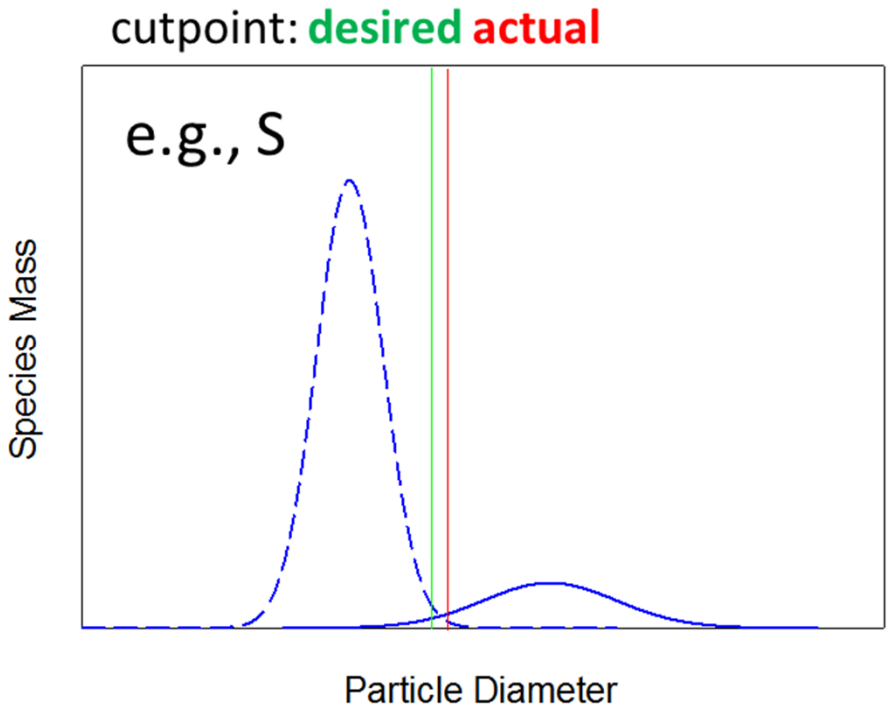
Overall Bias:
Iron: -1%
Sulfur: 0%

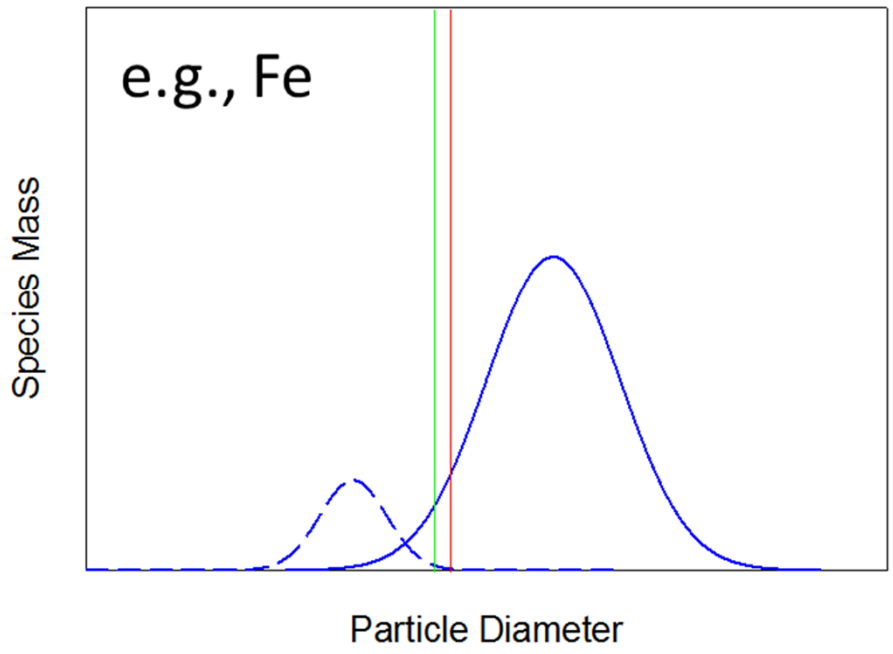
SOURCES OF PERSISTENT BIAS

- **Cleaning and maintenance**
- **Size discrimination**
- **Flow rate error**

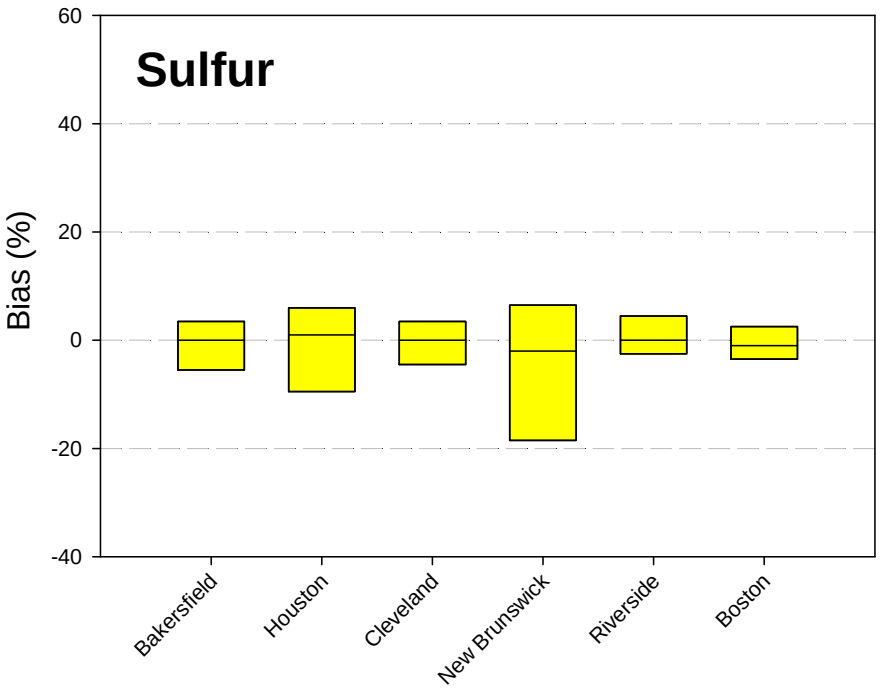
SOURCES OF PERSISTENT BIAS

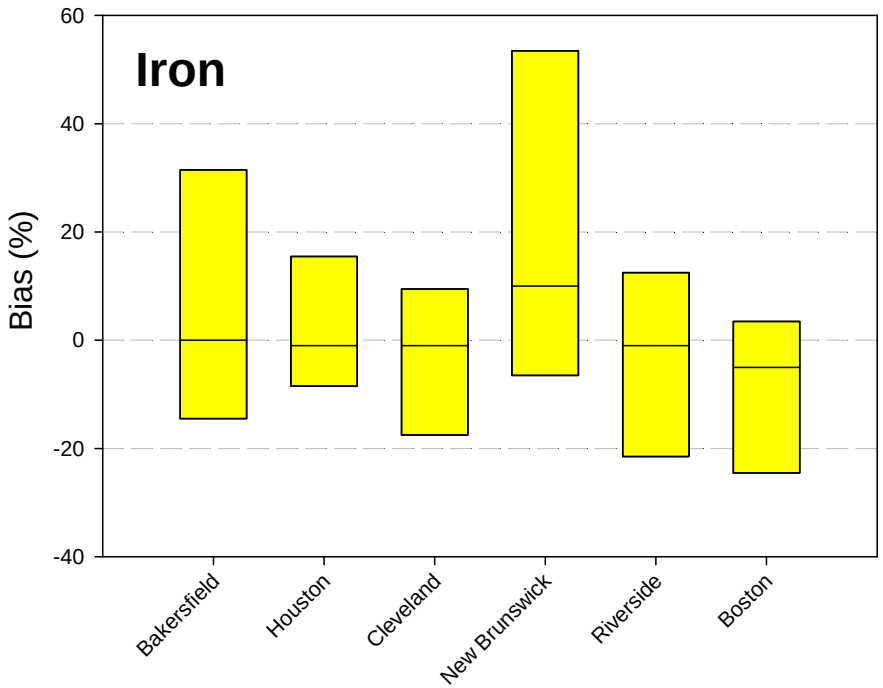
- **Cleaning and maintenance**
- **Size discrimination**
- **Flow rate error**

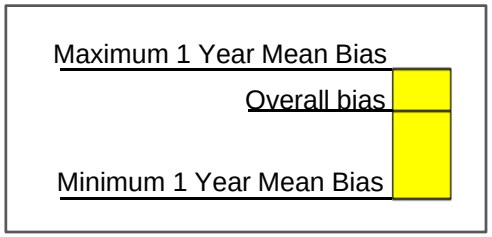




EXTENT OF PERSISTENT BIAS







*Censored to include: only values 3X the method detection limit; data points from days where iron, mass, and sulfur have recorded values at all three sites; n>80 within the 365 day window

CONCLUSIONS

- **For most species, the uncertainty estimates are missing one or more components of measurement error**
- **Periods of persistent bias, sometimes large, are observed throughout the network**
- **For long time series, the overall bias may not adequately characterize the influence of bias on precision estimates**
- **Future work: Identify the major drivers of bias to help optimize network operations**

ACKNOWLEDGEMENT

McDonnell International Scholars Academy: Boeing Corporation Fellow and U.S. Scholar Program