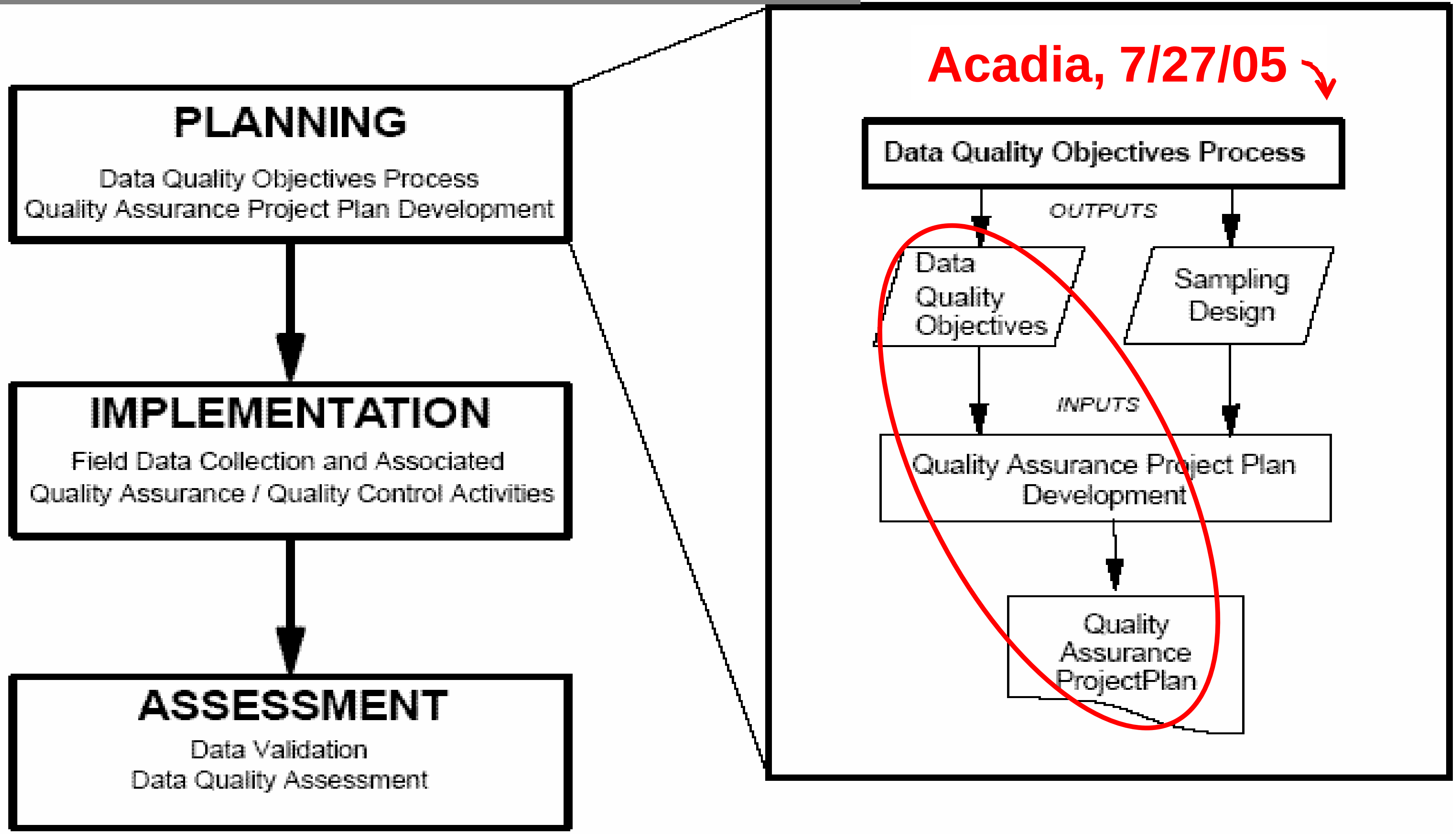
Overarching QA Issues

- **Guidance and documentation**
  - **Measurement (MQOs) and data quality objectives (DQOs)**
  - Metadata
  - Up-to-date and accurate standard operating procedures (SOPs)
- **Data validation**
  - **Objective data quality standards and data validation procedures**
  - Meeting data delivery schedule
  - Multi-tiered flagging system (NARSTO flagging scheme)
- Data management
  - Robust archival system for all key dataset stages
  - Tracking of changes to data and data processing procedures
  - Data reproducibility


IN ORDER TO AVOID SHODDY MISTAKES, EVERYTHING WE DO FROM NOW ON WILL BE PART OF A DOCUMENTED PROCESS.
scottadams@aol.com
www.dilbert.com
WHAT DOCUMENTED PROCESS DID YOU USE TO DECIDE WHAT DOCUMENTED PROCESS TO USE?
© 2005 Scott Adams, Inc./Dist. by UFS, Inc.
7-2-05
OR IS THIS ONE OF THOSE SHODDY MISTAKES I KEEP HEARING ABOUT?

United States Environmental Protection Agency | Office of Research and Development Washington, D.C. 20460 | EPA/600/R-96/055 September 1994

GUIDANCE FOR THE DATA QUALITY OBJECTIVES PROCESS

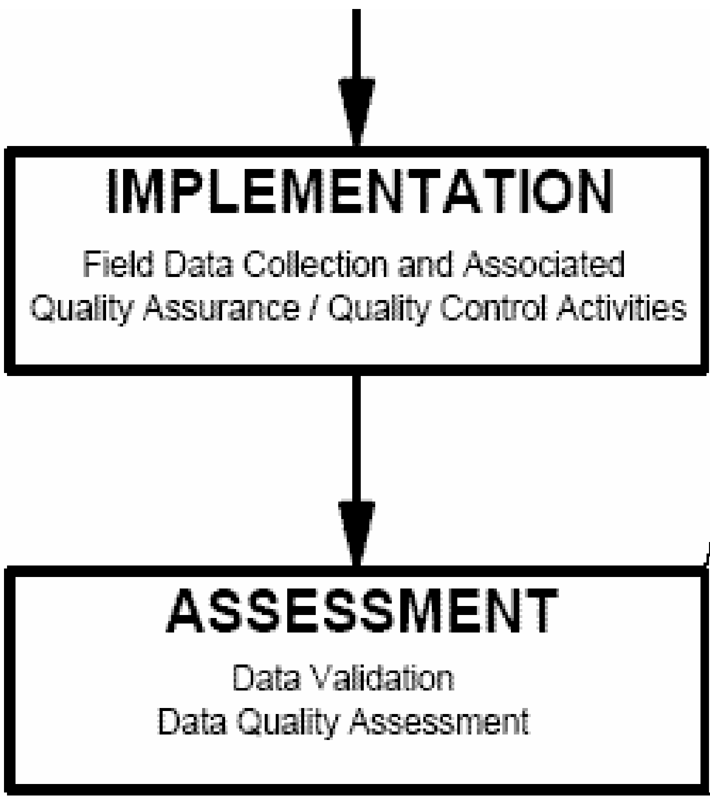


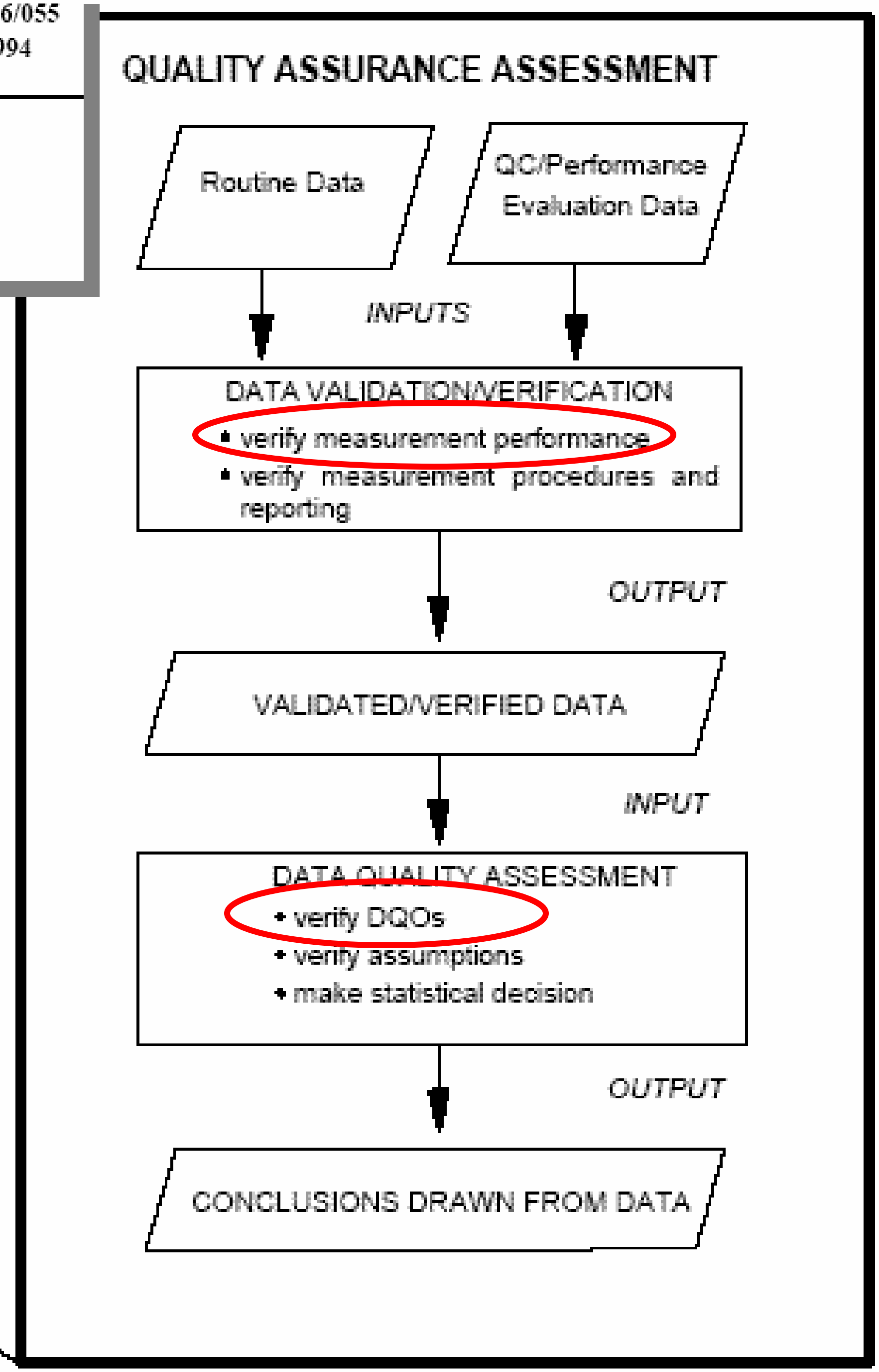
United States Environmental Protection Agency

Office of Research and Development Washington, D.C. 20460

EPA/600/R-96/055 September 1994

GUIDANCE FOR THE DATA QUALITY OBJECTIVES PROCESS





IMPROVE QAPP
Project Management
Section—Page: 4—30 of 52
Revision: 0.0
Date: March 2002

"The primary Data Quality Objective (DQO) for IMPROVE is to be able to measure a 5-percent change in bext in 5 years. The effect of individual components on bext depends on the site. … The DQO for IMPROVE will therefore require that a **5-percent change in five years** in each of the major components of sulfate, organic carbon, and soil must also be achieved.

…

The Measurement Quality Objectives (MQO's) should either reflect the DQO or be sufficiently stringent so that the DQO wi…

…

How reliably?

Without fail?(!)

Table 4. IMPROVE Measurement Quality O… Minimum Quantifiable Limit.

Method	Parameters	Precision*	Accuracy	MQL
Gravimetric	Mass	± 5 µg	± 5 µg	300 ng/m^3
XRF	Elements Fe to Pb	± 5%	± 5%	0.05 - 0.18 ng/m^3
PIXE	Elements S to Mn	± 5%	± 5%	1 - 4 ng/m^3
	Element Na	± 5%	± 10%	20 ng/m^3
PESA	Elemental H	± 5%	± 5%	4 ng/m^3
IC	NO_3, SO_4, NH_4	± 5%	± 5%	10 - 30 ng/m^3
	NO_2, Cl	± 5%	± 5%	60 - 100 ng/m^3
TOR	Organic Carbon	± 5%	± 5%	250 ng/m^3
	Elemental Carbon	± 10%	± 5%	100 ng/m^3

Quality Assurance Project Plan
Chemical Speciation
of PM2.5 Filter Samples

←STN

Prepared for:
U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

Date: January 15, 2004
Page 16 of 99

"A key conclusion of the DQO study was that the statistical power to detect concentration trends in the chemical speciation data is relatively insensitive to measurement error, up to about twice the level seen in the IMPROVE Washington, D.C., data.

This is because the 'uncontrollable' error components, which are primarily due to natural day-to-day variation in pollutant levels, dominate the random errors that limit the ability of the statistical analysis to detect a trend."

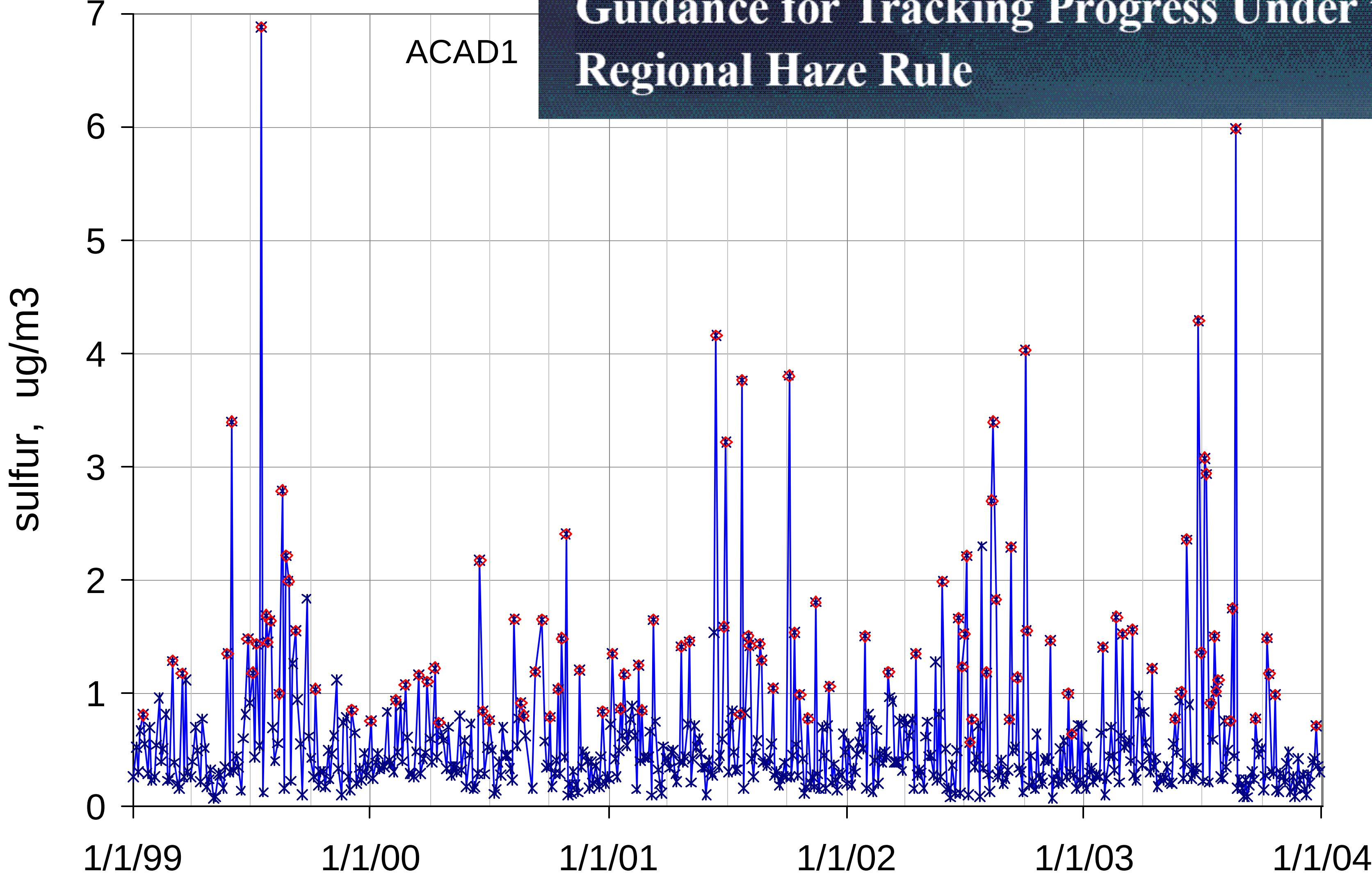
"Table A.7.3 shows the number of years worth of data necessary to detect a **5 percent annual trend** using the IMPROVE data set in conjunction with the regression model assuming one in three day sampling."

Date: January 15, 2004

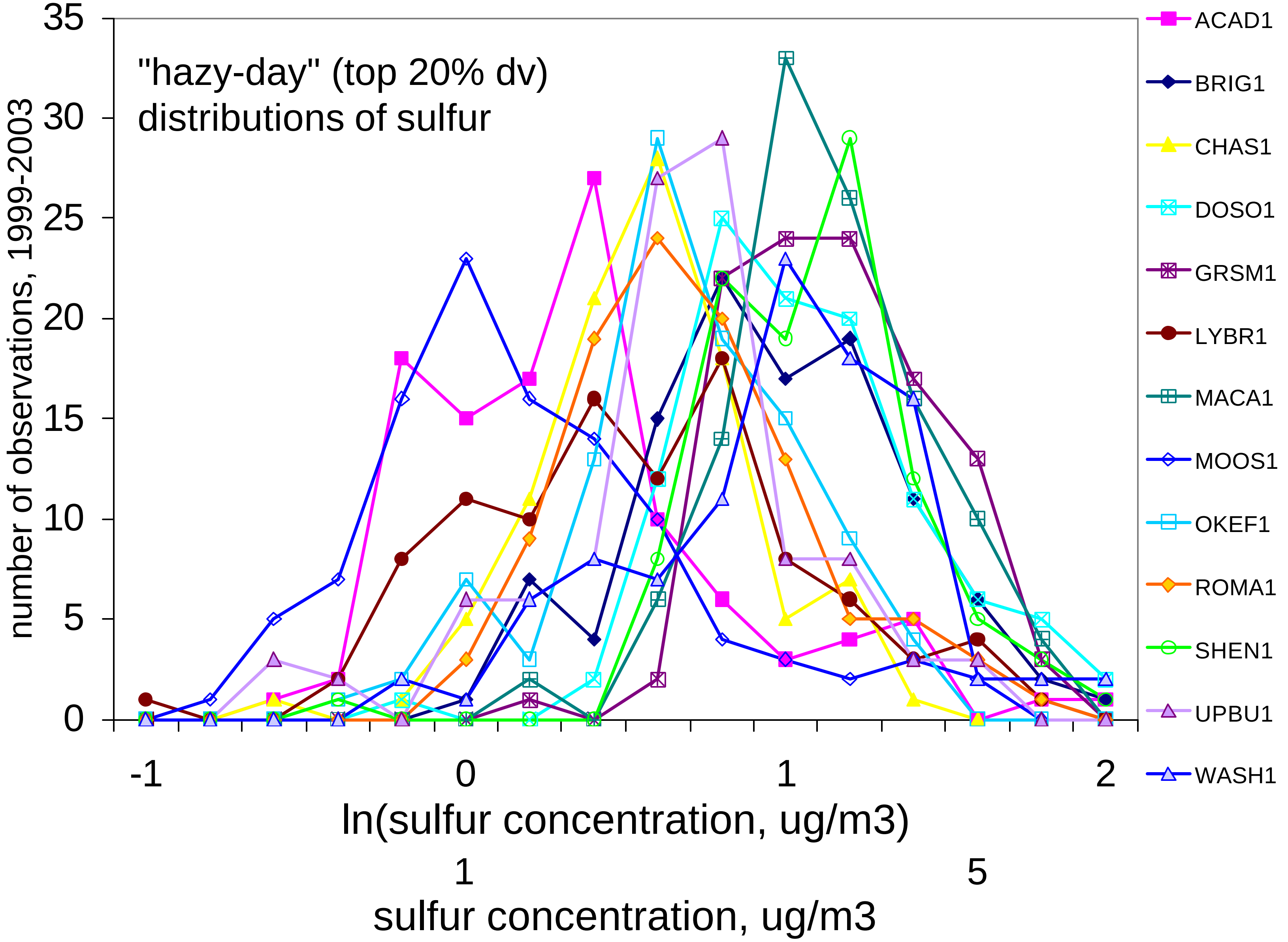
Page 17 of 99

Table A.7.3 Years Required to Achieve 0.8 Power With 1-in-3-day Sampling

Species	Measurement Error = IMPROVE Error × 1
Sulfate	4.1 years
Nitrate	6.3 years
Calcium	4.1 years
Total Carbon	3.4 years

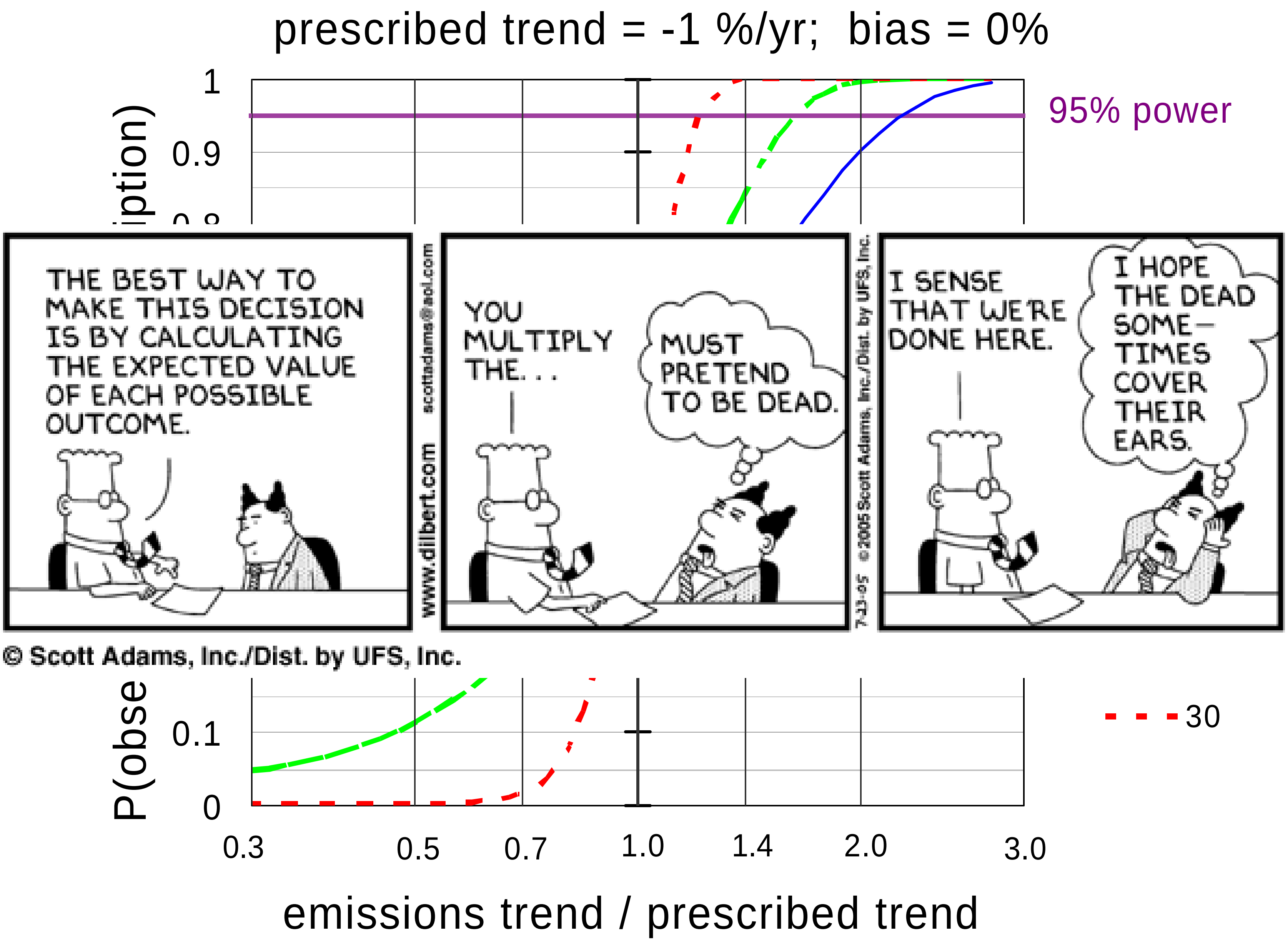


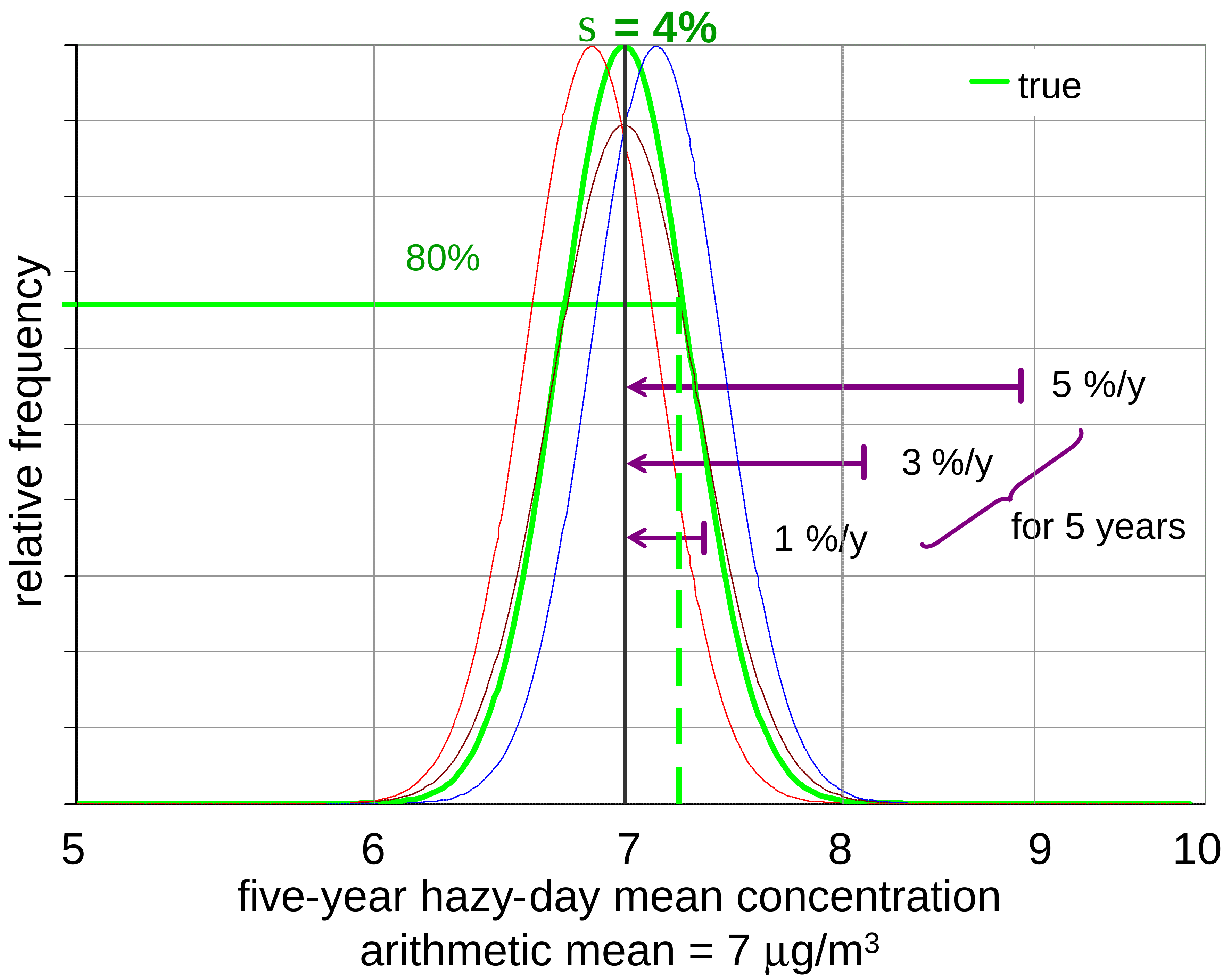
Sulfur concentrations at Acadia. The haziest 20% of days in each calendar year are highlighted in red, based on deviews downloaded from CIRA.


"hazy-day" (top 20% dv)
distributions of sulfur
number of observations, 1999-2003
0
5
10
15
20
25
30
35
-1
0
1
2
ln(sulfur concentration, ug/m3)
1
5
sulfur concentration, ug/m3
ACAD1
BRIG1
CHAS1
DOSO1
GRSM1
LYBR1
MACA1
MOOS1
OKEF1
ROMA1
SHEN1
UPBU1
WASH1

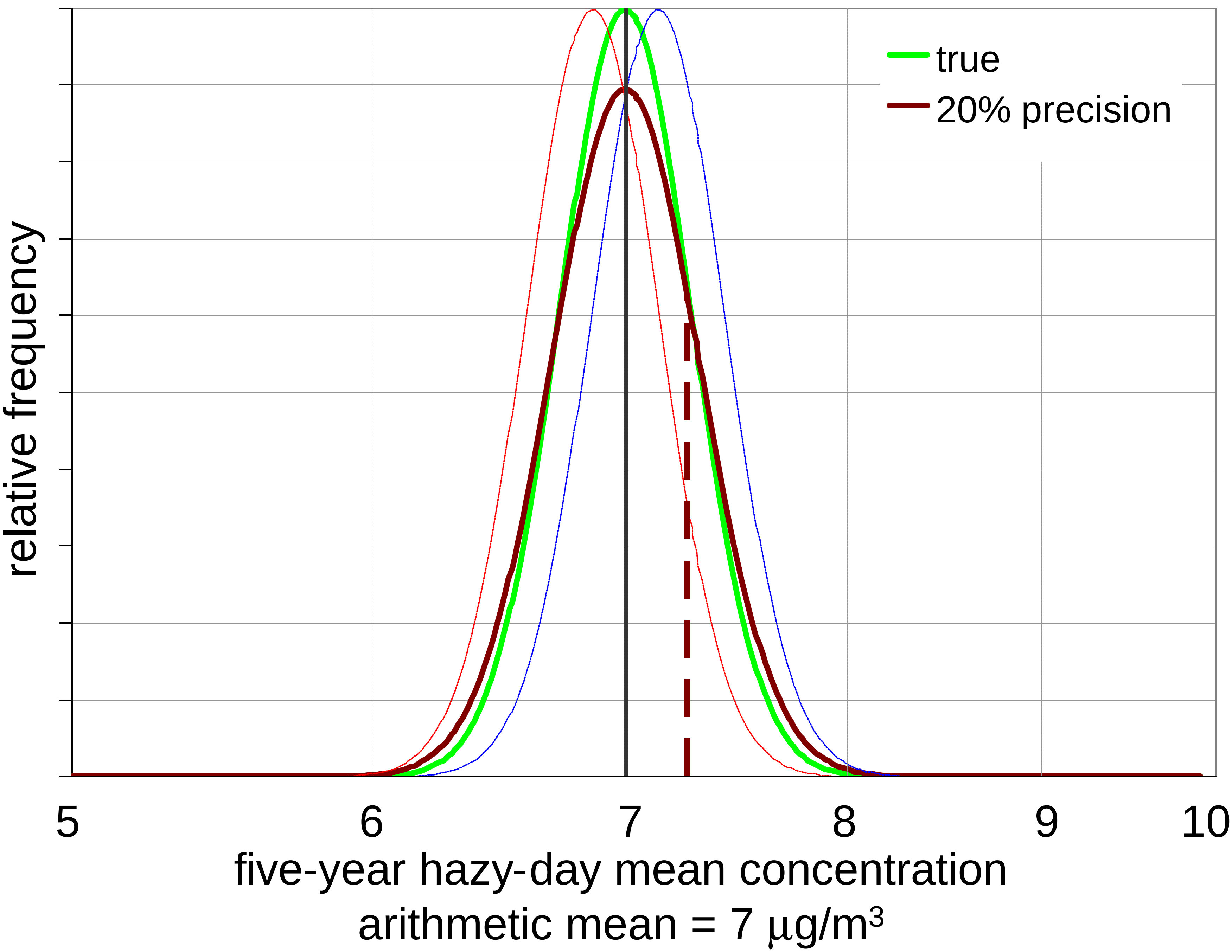
EUS site	**geom**	**arith**
ACAD1	4.8%	6.1%
BRIG1	4.0%	3.9%
CHAS1	3.4%	3.3%
DOSO1	3.6%	3.9%
GRSM1	2.9%	2.9%
LYBR1	5.3%	5.6%
MACA1	3.0%	3.0%
MOOS1	5.0%	6.2%
OKEF1	4.2%	3.5%
ROMA1	3.7%	4.2%
SHEN1	3.5%	3.5%
UPBU1	4.4%	4.2%
WASH1	4.6%	4.5%
median	**4.0%**	**3.9%**

Alternative estimates of the sampling uncertainty in 5-year sulfur means from hazy days in 1999-2003.

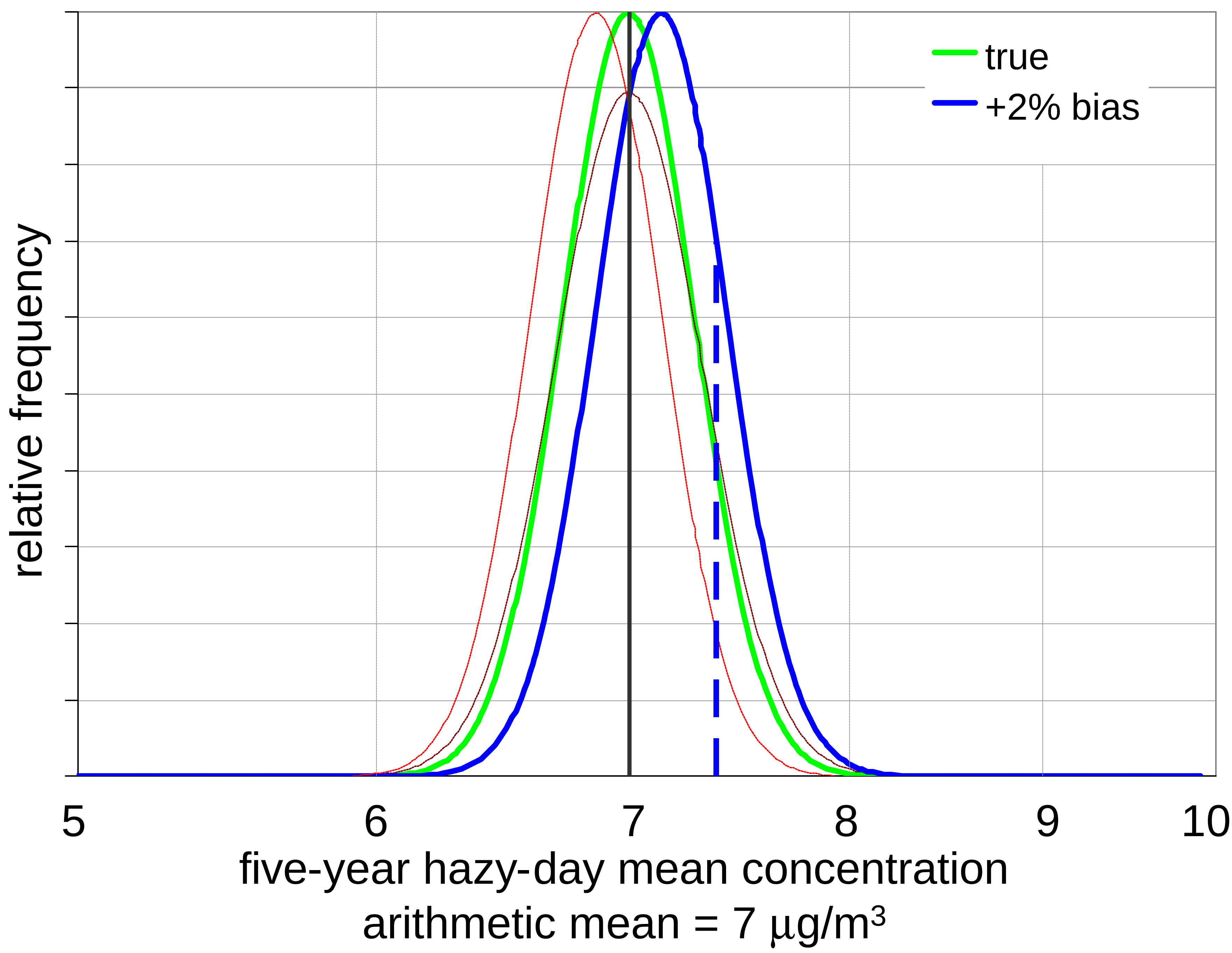

prescribed trend = -1 %/yr; bias = 0%
1
0.9
0.1
0
95% power
30
P(obse
iption)
0.3
0.5
0.7
1.0
1.4
2.0
3.0
emissions trend / prescribed trend
THE BEST WAY TO MAKE THIS DECISION IS BY CALCULATING THE EXPECTED VALUE OF EACH POSSIBLE OUTCOME.
YOU MULTIPLY THE. . .
MUST PRETEND TO BE DEAD.
I SENSE THAT WE'RE DONE HERE.
I HOPE THE DEAD SOME-TIMES COVER THEIR EARS.
scottadams@aol.com
www.dilbert.com
7-13-05 © 2005 Scott Adams, Inc./Dist. by UFS, Inc.

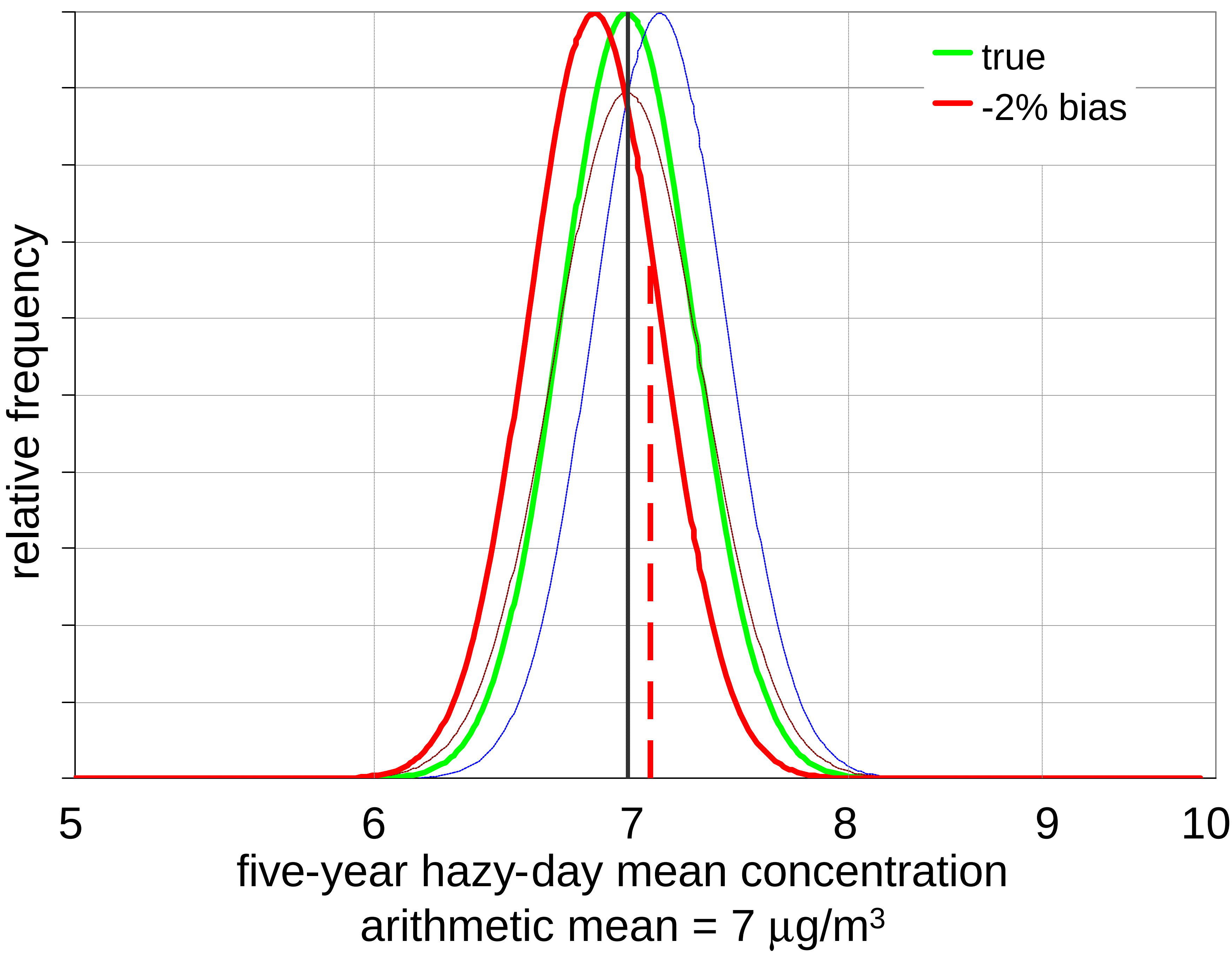


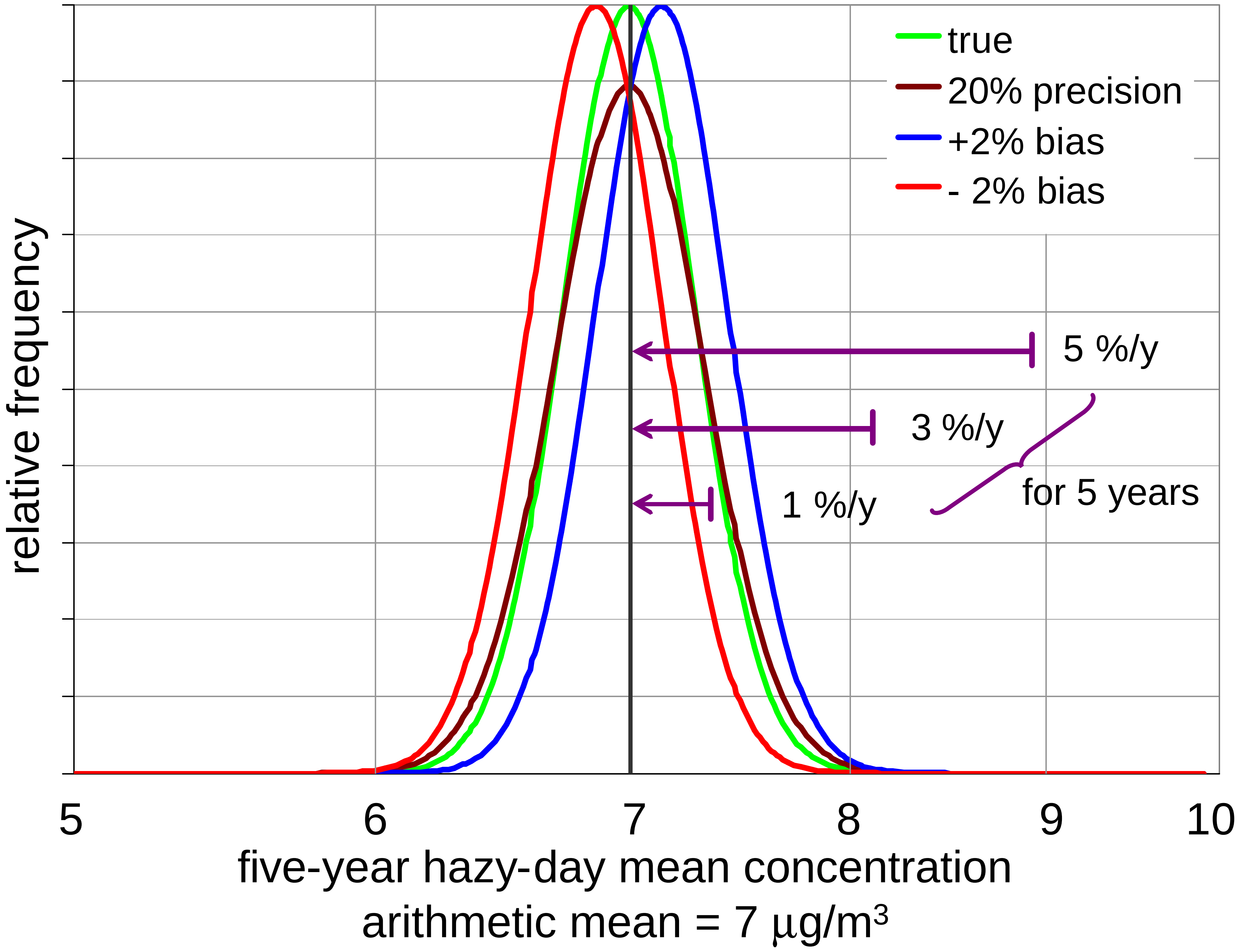
s = 4%
true
80%
5 %/y
3 %/y
1 %/y
for 5 years
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m³



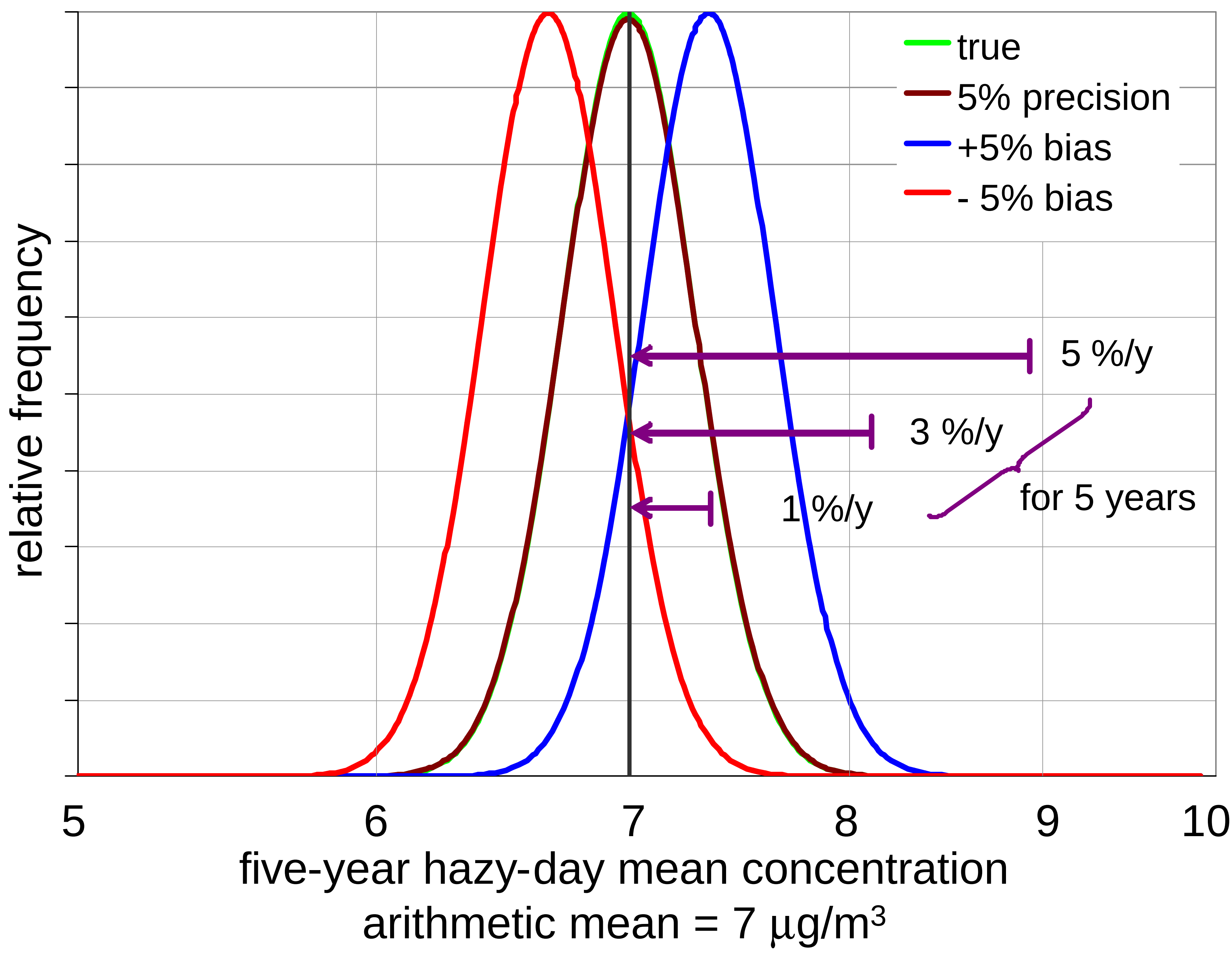
true
20% precision
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m³


true
+2% bias
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m³

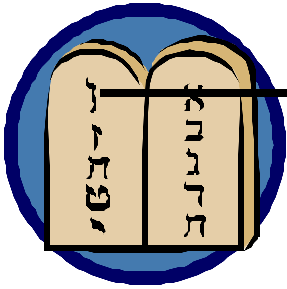

true
-2% bias
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m³



true
20% precision
+2% bias
- 2% bias
5 %/y
3 %/y
1 %/y
for 5 years
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m3



true
5% precision
+5% bias
- 5% bias
5 %/y
3 %/y
1 %/y
for 5 years
relative frequency
5
6
7
8
9
10
five-year hazy-day mean concentration
arithmetic mean = 7 µg/m³



Focus on bias*,
not imprecision**.

(If tracking progress is your objective.)

* “calibration drift”; “slow-changing”

** “random noise”; “uncorrelated”

Our QAPP specifies “precision” and “accuracy”.

Table 4. IMPROVE Measurement Quality Objectives of Precision, Accuracy, and Minimum Quantifiable Limit.

Method	Parameters	Precision*	Accuracy	MQL
Gravimetric	Mass	± 5 µg	± 5 µg	300 ng/m^3
XRF	Elements Fe to Pb	± 5%	± 5%	0.05 - 0.18 ng/m^3
PIXE	Elements S to Mn	± 5%	± 5%	1 - 4 ng/m^3
	Element Na	± 5%	± 10%	20 ng/m^3
PESA	Elemental H	± 5%	± 5%	4 ng/m^3
IC	NO_3, SO_4, NH_4	± 5%	± 5%	10 - 30 ng/m^3
	NO_2, Cl	± 5%	± 5%	60 - 100 ng/m^3
TOR	Organic Carbon	± 5%	± 5%	250 ng/m^3
	Elemental Carbon	± 10%	± 5%	100 ng/m^3

4.6.1 Precision Objectives

Precision is a measure of mutual agreement among individual measurements of the same property. The overall precision will be determined by collocated sampler

intercomparisons. Collocated sampler intercomparisons involving pairs of measurements can be used in three ways. The first way is to calculate the precision directly for parameters where the concentrations are generally sufficiently large so that counting statistics and uncertainties in blank subtraction are negligible. The second use of collocated comparisons is verify that the uncertainties in the database adequately characterize the measured differences. This will be used for trace elements and variables with lower concentrations. Finally, the third use of collocated measurements is to monitor the individual parameters used in the calculation of individual concentrations, such as volume uncertainty, analytical calibration uncertainty.

The standard procedure is to calculate the precision for a given variable directly. The uncertainty for a given concentration has components that are proportional to the concentration (flow rate and analytical calibration uncertainties), and components that are independent of concentration or depend on the square root of the concentration (artifact and statistical uncertainties). The precision can be defined either as the absolute precision or relative precision. The absolute precision equation requires that the uncertainty be predominantly independent of concentration, which is never valid for aerosol measurements. The relative precision equation requires that the uncertainty be predominantly proportional to uncertainty. Fortunately, there are many cases where this is valid. The equation for the paired relative precision is given by:

$$P_{rel}^2 = \frac{2}{n}\sum\frac{(x_i - y_i)^2}{(x_i + y_i)^2} , \qquad \text{(Equation 3)}$$

where x and y are the paired measurements.

The collocated measurements should be used to estimate relative precision only when the concentrations are sufficiently large so that the uncertainty in artifact subtraction or the statistical uncertainty are negligible. Examples of where relative precison is appropriate are: sulfate, nitrate, and sulfur at all except a few very clean sites, the major soil elements at many sites, carbon at many sites, zinc and lead at many sites. The range of validity can be extended to other sites and variables by selecting only those cases where the concentration is large. For example, sodium, chlorine, and chloride can be characterized if only samples with major marine influence are considered. The precision calculation is rarely appropriate are for most trace elements. For example, the selenium precision would be different for most Eastern sites than for most Western sites simply because the concentrations are much larger. A precision calculation that depends on the ambient concentrations is of little value in characterizing the system.

The second way to use collocated measurements is to verify the uncertainties calculated for each measurement based on both collection and analysis. With this method, the results are independent of the ambient concentrations. This method compares the differences in concentrations with the predicted uncertainties in the two concentrations. The appropriate statistic is the goodness-of-fit (chi-square) parameter, which compares the measured differences with the calculated uncertainties of the concentrations. For a comparison using two collocated samplers, the pairwise goodness-of-fit parameter, χ^2, is given by:

$$\chi^2 = \frac{1}{n}\sum\frac{(x_i - y_i)^2}{\sigma_x^2 + \sigma_y^2} \qquad \text{(Equation 4)}$$

where x and y are the paired concentrations and σ_x and σ_y are the corresponding calculated uncertainties. If the value of χ^2 is near unity, then the calculated uncertainties correctly predict the measured differences. Values significantly larger than unity for identical samplers indicate either that the calculated uncertainties are too small. Either the precision estimates in collection and analytical calibration are too small or there are other sources of uncertainty that should be included in the calculation. Values significantly smaller than unity indicate that the uncertainty estimates are too small. If value of χ^2 differs from unity, the parameters used to calculate individual uncertainties should be re-evaluated.

IMPROVE will have collocated samplers at approximately 4% of the sites. Rather than having a complete collocated sampler at six sites, there will be a single collocated module at twenty-four sites. These will be equally divided between the four module types.

In the database, it is necessary to calculate the uncertainty of each concentration. This will depend on the precision for various components of the concentration calculation, such as volume and analytical calibration. The uncertainty in a given concentration is a combination of several factors:

- The precision in the volume (primarily from the precision in the flow rate). This enters as a relative uncertainty (e.g. 3%) for every parameter.
- The calibration and normalization precision of the analytical instrument. This enters as a relative uncertainty (e.g. 3%) for all methods except gravimetric analysis.
- The statistical counting uncertainty for analytical methods that count events in a spectral peak. This enters as a constant uncertainty (e.g. 5 μg) divided by the volume for all elemental parameters.

- The standard deviation of the field blanks (ions) or secondary filters (carbon), when an artifact is subtracted from the measurement. This includes any constant uncertainty in the analysis plus the variability of the artifact. This enters as a constant uncertainty (e.g. 5 μg) divided by the volume.
- The standard deviation of the gravimetric control filters. This enters as a constant uncertainty (e.g. 5 μg) divided by the volume.

The relative precision in the flow rate can be estimated from flow audits and from an examination of possible sources of variability.

The relative precision in analytical calibration and normalization can be estimated by replicate analysis of the same filter. For analytical methods that have inherent uncertainty in counting statistics, such as elemental analysis, caution must be exercised to avoid commingling calibration precision and statistical uncertainty. The calibration precision can only be estimated from variables in which the statistical uncertainty can be neglected. For example, with x-ray analyses, the calibration precision can only be estimated from a few elements that are present in high areal concentrations.

The method for determining analytical precision for each method is listed in Table 5.

Table 5. Methods for determining Analytical Precision

Method	Precision Method
Gravimetric	standard deviation of control filters
HIPS	replicate analyses conducted in previous session
XRF	replicate analyses conducted in previous session
PIXE	replicate analyses conducted in previous session
PESA	replicate analyses conducted in previous session
IC	standard deviation of QA standards, replicate analysis of eluent
TOR	analysis of additional punch in filter

The collocated samplers can check the validity of the precisions various components. For example, if the χ^2 calculated in the above equation is consistently greater than one, then the flow rate estimate of 4% is too low. The measurements can then be used to adjust the parameters used in the uncertainty calculations.

In general, it is not possible to define the precision of the concentration of a given variable as a single number. For most variables, the precision obtained by collocated sampling at one site and season cannot be extended to a different site and season. IMPROVE has tried to overcome this limitation by calculating the uncertainty for every concentration and providing this uncertainty in the database. The parameters that are

involved are derived from collocated measurements, the actual possible mechanisms for error propagation, and from other tests. It is recommended that the data users use the calculated uncertainties rather than a fixed fractional or absolute uncertainty for the parameter.

Internal Redundancy. It should be noted that speciated sampling with independent modules and redundant measurements permits a modified form of precision comparison through intercomparisons for each sample. Although this involves some assumptions not required in direct collocated intercomparisons, the advantage is that can be performed on every sample rather than on some much smaller number, such as 4%. The comparison of equivalent parameters using independent modules is performed for every sample during the data validation procedure. The sulfur–sulfate comparison is especially helpful because the concentrations are generally large enough so that the uncertainties from counting statistics and artifact subtraction are negligible. In addition to being used to monitor the calibration/normalization precisions of both IC and PIXE, the comparison performs several additional functions.

- It provides a quality control check for Modules A and B for every sample.
- It monitors the flow rate calibration for Modules A and B.
- It provides a measure of accuracy for the measurement of sulfur and sulfate by independent analytical methods.

Two checks between Modules A and C are possible. Organics mass can be estimated from elemental hydrogen measured on the Teflon A filter, with the assumption that the hydrogen is all from either organic matter or ammonium sulfate. The estimate is too small if the sulfate is acidic or marine and too large if there is significant nitrate relative to sulfate. The first comparison is between organics mass by hydrogen with organic mass by carbon. The second is between the coefficient of absorption measured on the Teflon A filter with light-absorbing carbon measured on the quartz C filter. While less precise than the sulfur–sulfate comparison, these do provide a useful quality control check between Modules A and C for every sample.

Thus the internal comparisons allow a check of all three $PM_{2.5}$ modules for every sample. The emphasis here is on quality control, while collocated precision tests are strictly quality assurance. However, the inter-module comparisons do provide very robust data sets for quality assurance.

Precision MQO's and the overall DQO. There are two possible ways to characterize the uncertainty of the mean of a given set of measurements: (1) calculate the standard error of the concentrations, and (2) calculate the uncertainty by propagating the individual measurement uncertainties. For IMPROVE, the statistic of concern is the

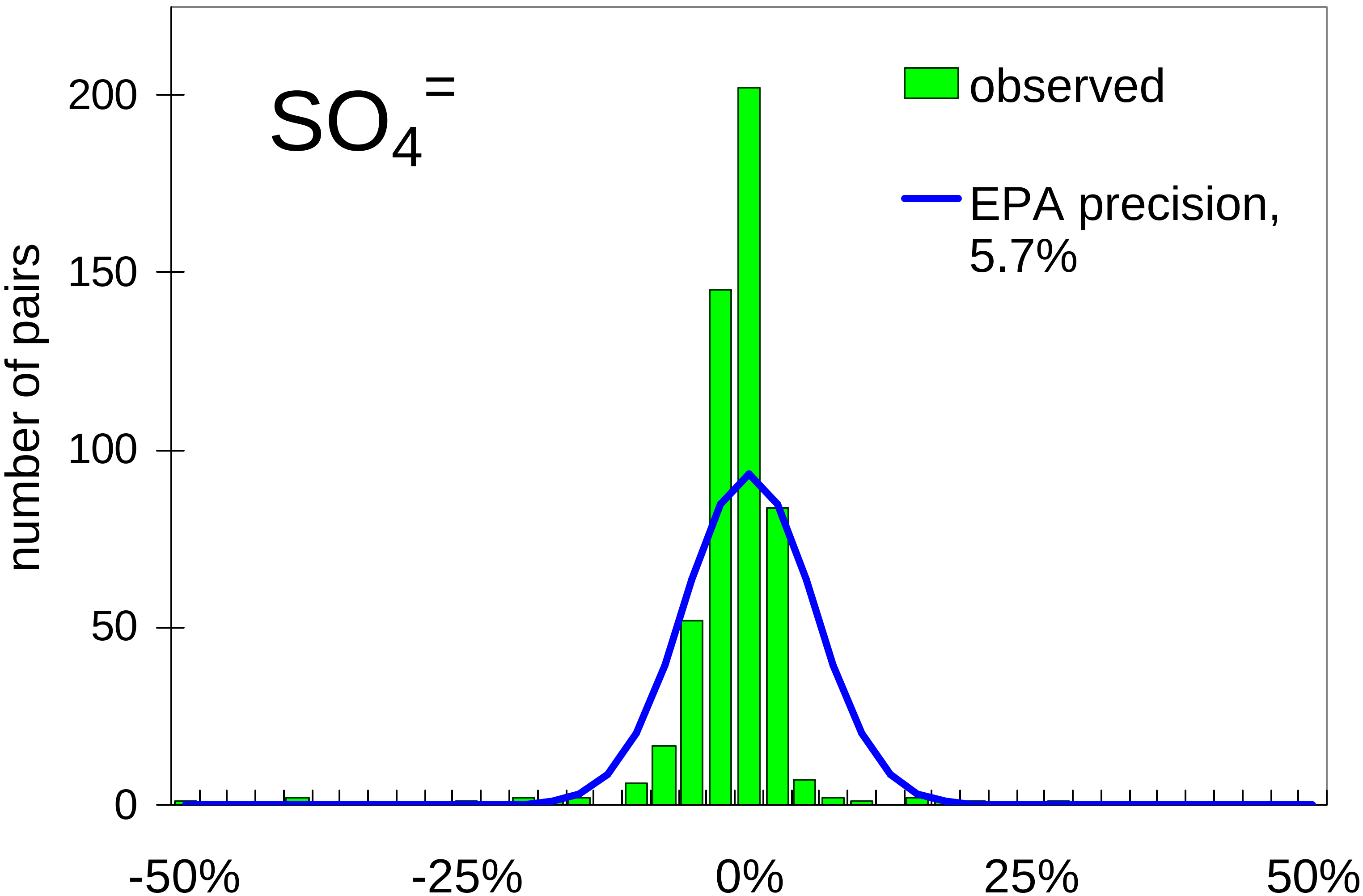
average value of the reconstructed extinction parameter for the 20% worst days or 20% cleanest days. For sampling every third day, each group corresponds to about 24 samples. In practice, the standard error of the values in the group is about an order or magnitude larger than the uncertainty that would be obtained by propagation of the 24 individual uncertainties. Consider the specific case of sulfur concentrations at one site for the calendar year 1998. The 24 lowest concentrations in ng/m³ were: 44, 44, 67, 67, 119, 122, 143, 149, 151, 163, 200, 205, 213, 217, 221, 221, 222, 224, 232, 234, 234, 242, 242. The corresponding uncertainties (between 5 and 7%) were: 3, 3, 5, 5, 7, 8, 9, 8, 9, 10, 11, 11, 12, 13, 12, 12, 12, 12, 12, 15, 13, 13, 13, 14. The average is 175 ng/m³. The standard error is 14 ng/m³ (8%). However, the propagated uncertainty is only 19 ng/m³ (1%). For this same data set, the 24 highest concentrations gave the same general result that the standard error (in this case, 14%) was much larger than the propagated precision (in this case, also 1%). The appropriate estimate of the uncertainty should be the standard error. Note that in both cases, the standard error is associated with the variation in the ambient concentrations and is not affected in any significant way by the reasonable measurement uncertainty of 5%. This example shows is that the precision MQO's of about 5% sufficient to not adversely affect the overall DQO of calculating trends based on groups of data.

4.6.2 Accuracy Objectives

Accuracy is the agreement between a sample reading and the true value of the sample. Accuracy is determined by comparing instruments to reference standards traceable to the National Institute of Standards and Technology (NIST). The reference standard for flow rate and each analytical method is listed in Table 6.

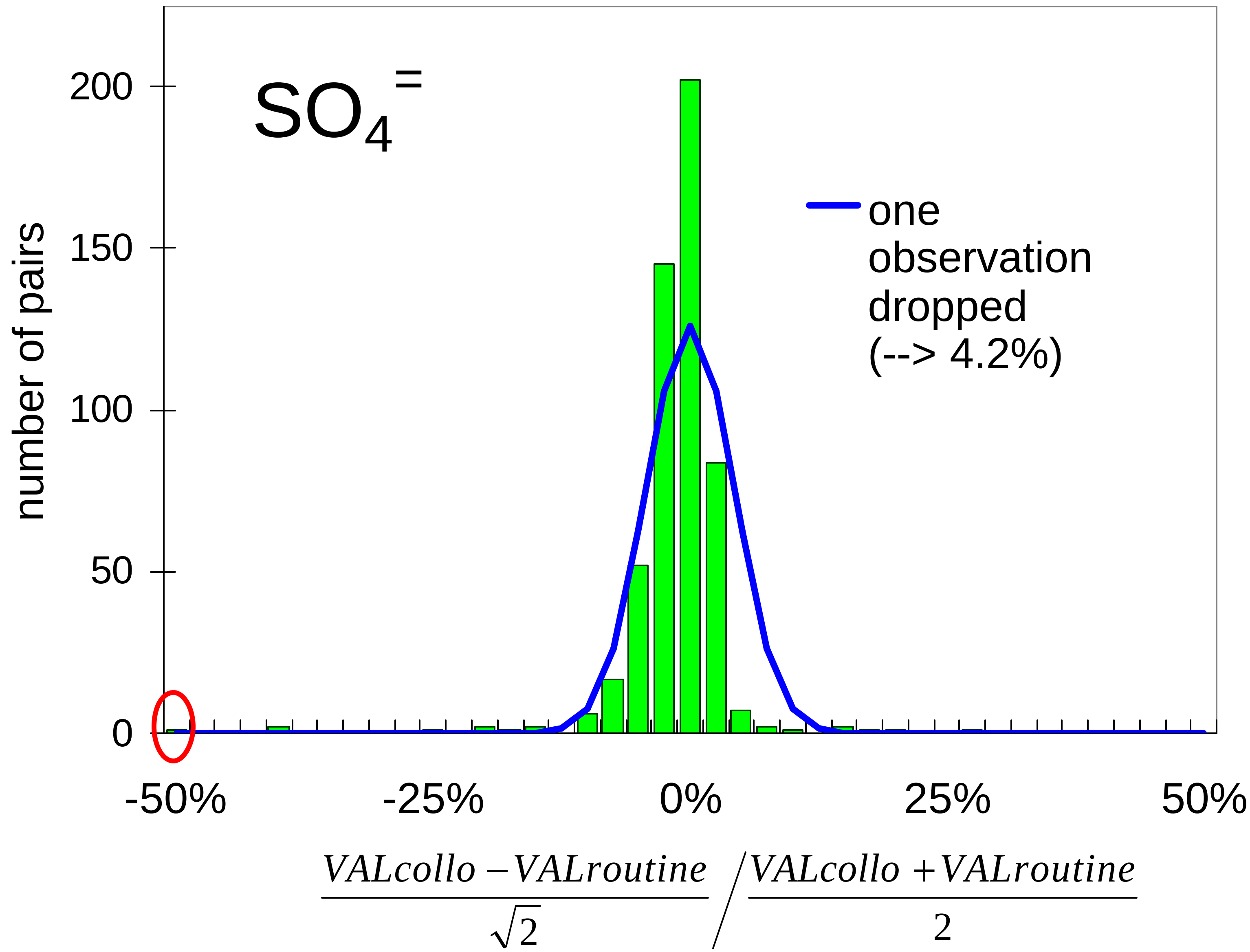
Table 6. Flow Rate and Analytical Accuracy

Method	Accuracy Reference
Flow Rate	NIST-traceable Dry-Cal Nexus DC-2 Flow-Calibrator
Gravimetric	NIST-traceable Class 1.1 (Class M) mass standards
HIPS	Reference Integrating Sphere system, which is itself calibrated using NIST-traceable absorption standards
XRF	17 commercial NIST traceable elemental standards
PIXE	27 commercial NIST traceable elemental standards
PESA	mylar standards of known area weighed on a balance calibrated by NIST-traceable standards
IC	NIST-traceable solutions of each ion
TOR	CH_4 gas, CO_2 gas, samples spiked with KHP and sucrose, none NIST-traceable.

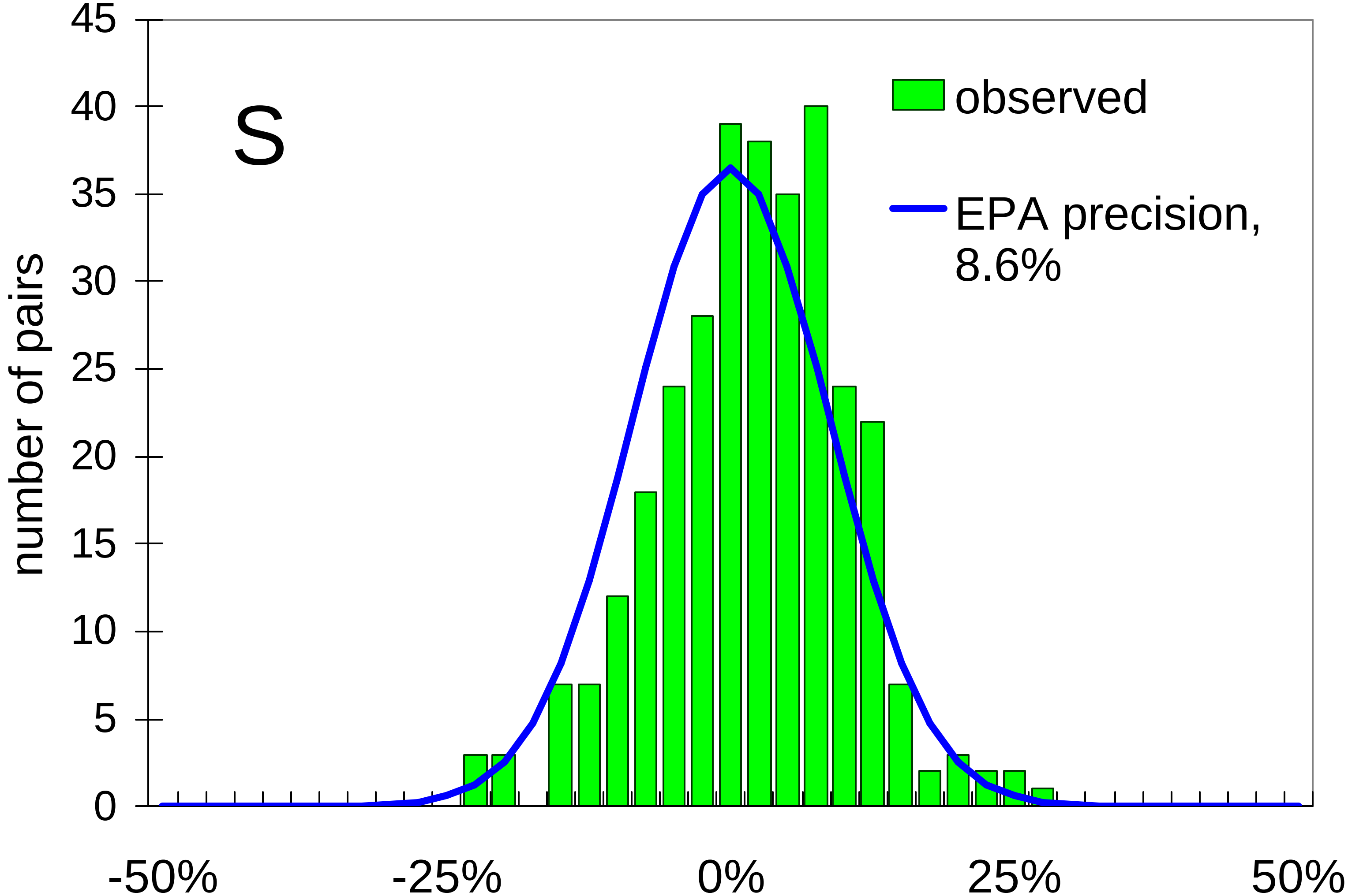


$$\frac{VALcollo - VALroutine}{\sqrt{2}} \bigg/ \frac{VALcollo + VALroutine}{2}$$

Collocated B modules, BIBE, FRRE, GAMO, LAVO, MACA, 6/03 – 8/04; 530 sample pairs.



$SO_4^=$
one observation dropped (--> 4.2%)
number of pairs
200
150
100
50
0
-50%
-25%
0%
25%
50%
$\frac{VALcollo - VALroutine}{\sqrt{2}} \Big/ \frac{VALcollo + VALroutine}{2}$



$$\frac{VALcollo - VALroutine}{\sqrt{2}} \Bigg/ \frac{VALcollo + VALroutine}{2}$$

Collocated A modules,MEVE, OLYM, PMRF, SAFO, TRCR, 10/03 – 8/04; 317 sample pairs.

$$s_{RHR} = s_{obs} / \sqrt{n}$$

$$\cong 5.7\% / \sqrt{100} < 0.6\%,\ SO_4^{=}$$

$$\cong 8.6\% / \sqrt{100} < 0.9\%,\ S$$

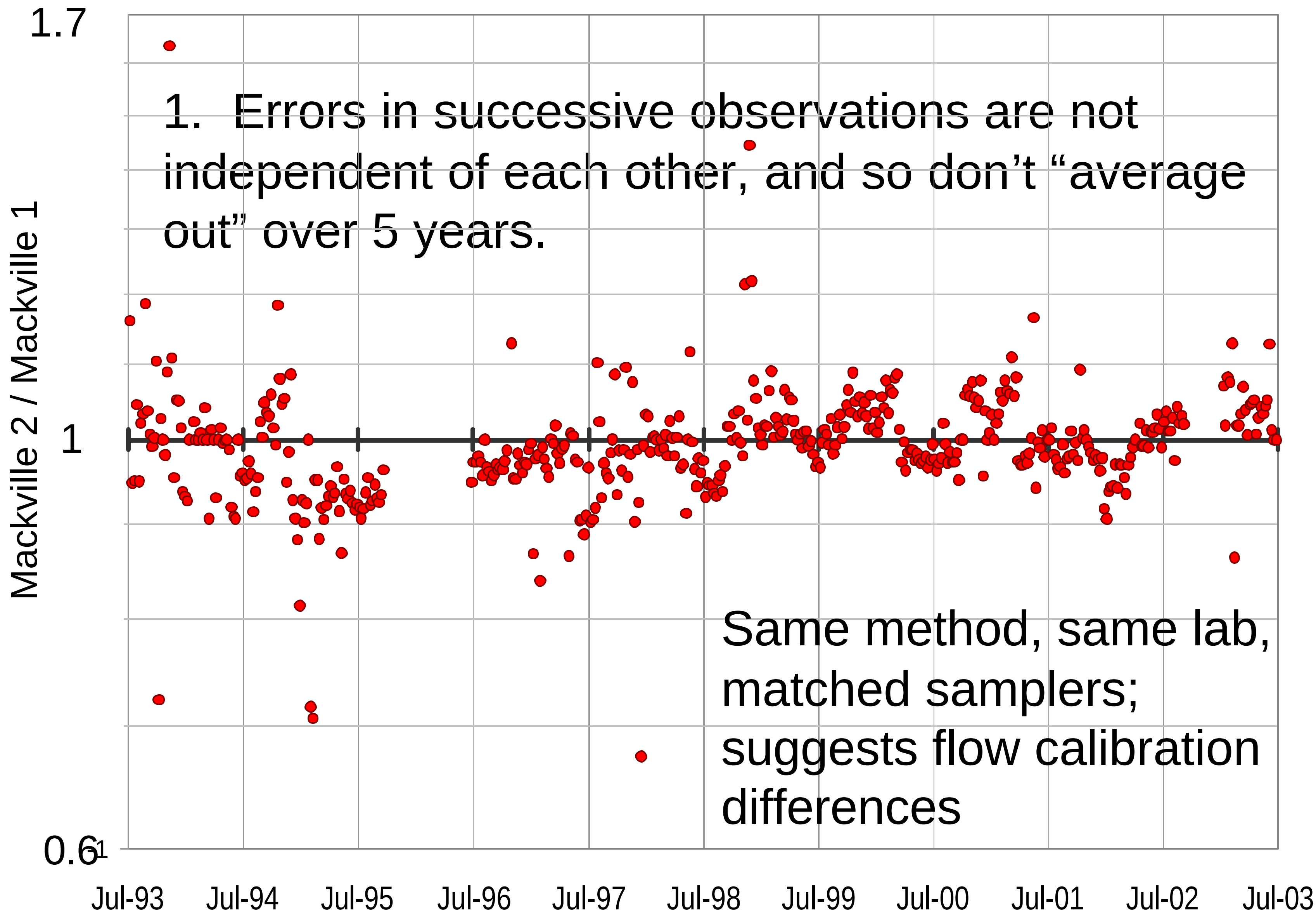
$$\sqrt{0.6\%^2 + 0.9\%^2} < 1\%, \quad SO_4^{=}\text{-}3S$$

Five-year means (µg/m³) from hazy days in 1999 – 2003.

EUS site	m(**SO_4**)	m(**3xS**)	D
ACAD1	4.90	4.84	-1.2%
BRIG1	8.85	8.42	-5.0%
CHAS1	5.89	5.56	-5.7%
DOSO1	9.31	9.15	-1.7%
GRSM1	9.96	9.91	-0.5%
LYBR1	5.97	5.95	-0.3%
MACA1	9.55	9.71	1.7%
MOOS1	4.16	4.22	1.4%
OKEF1	6.27	6.20	-1.1%
ROMA1	6.91	6.55	-5.4%
SHEN1	9.81	9.32	-5.1%
UPBU1	6.38	6.34	-0.6%
WASH1	9.01	8.62	-4.3%

$$\sqrt{\frac{1}{n}\sum \Delta^2} =$$ **3.3%**

Why >> 1%?



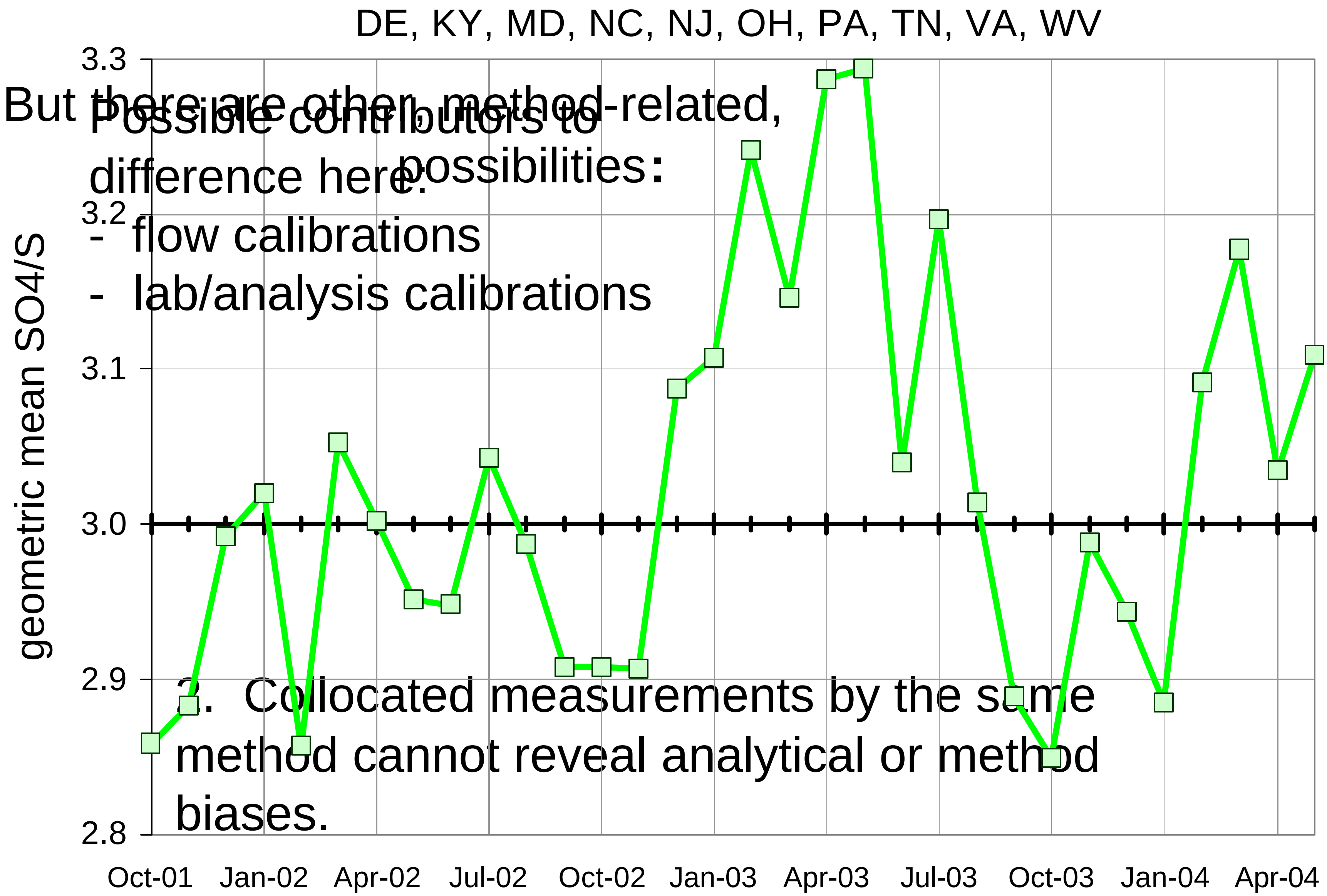
Collocated weekly CASTNet sulfate measurements at Mackville, KY.

But there are other, method-related, possibilities:

Possible contributors to difference here:

- flow calibrations
- lab/analysis calibrations

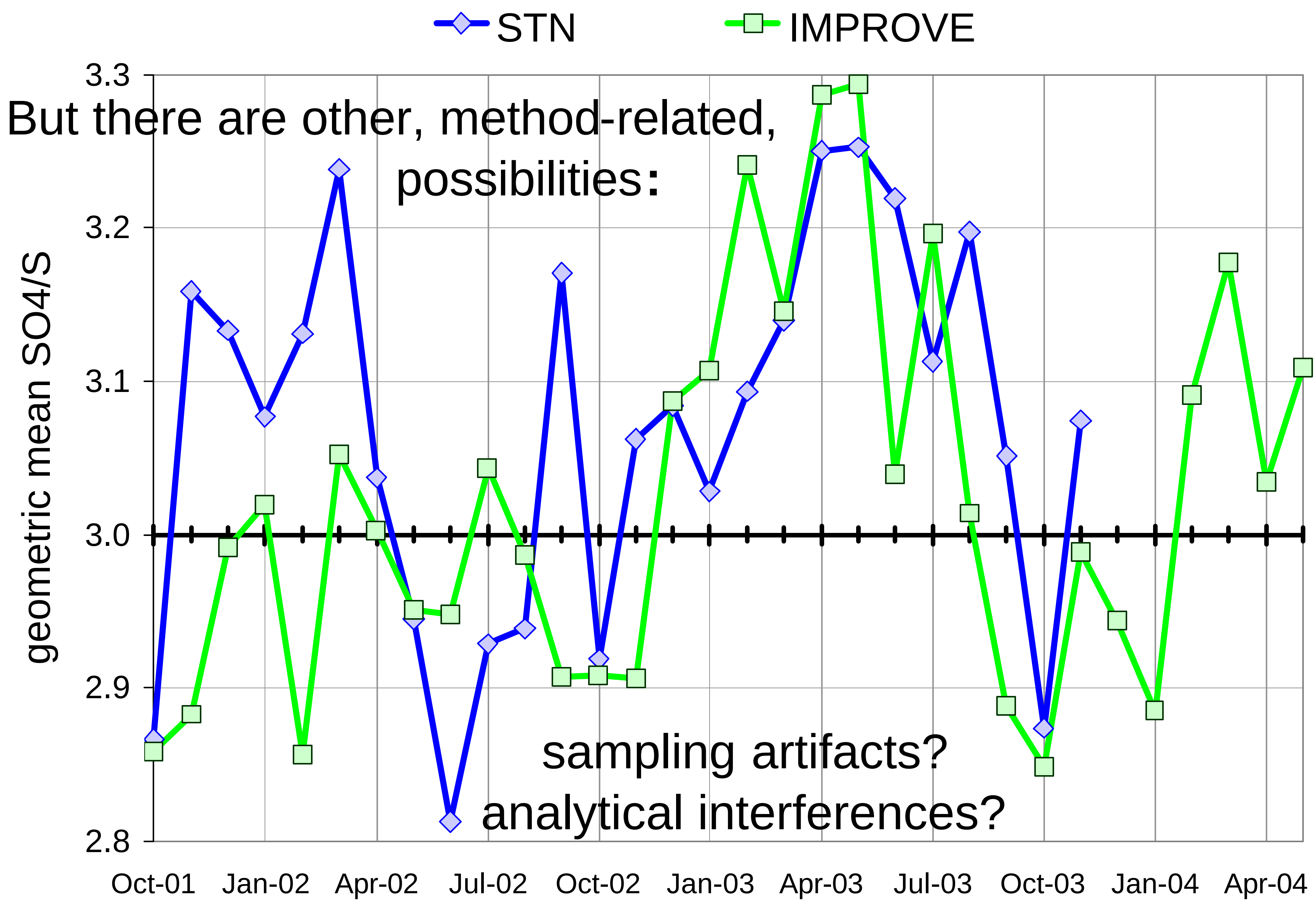
3. Collocated measurements by the same method cannot reveal analytical or method biases.



Monthly geometric mean $SO_4^{=}$/S over all IMPROVE sites in DE, KY, MD, NC, NJ, OH, PA, TN, VA, WV.

(Requires $SO_4^{=}$ & 3S >1 $\mu g/m^3$.)

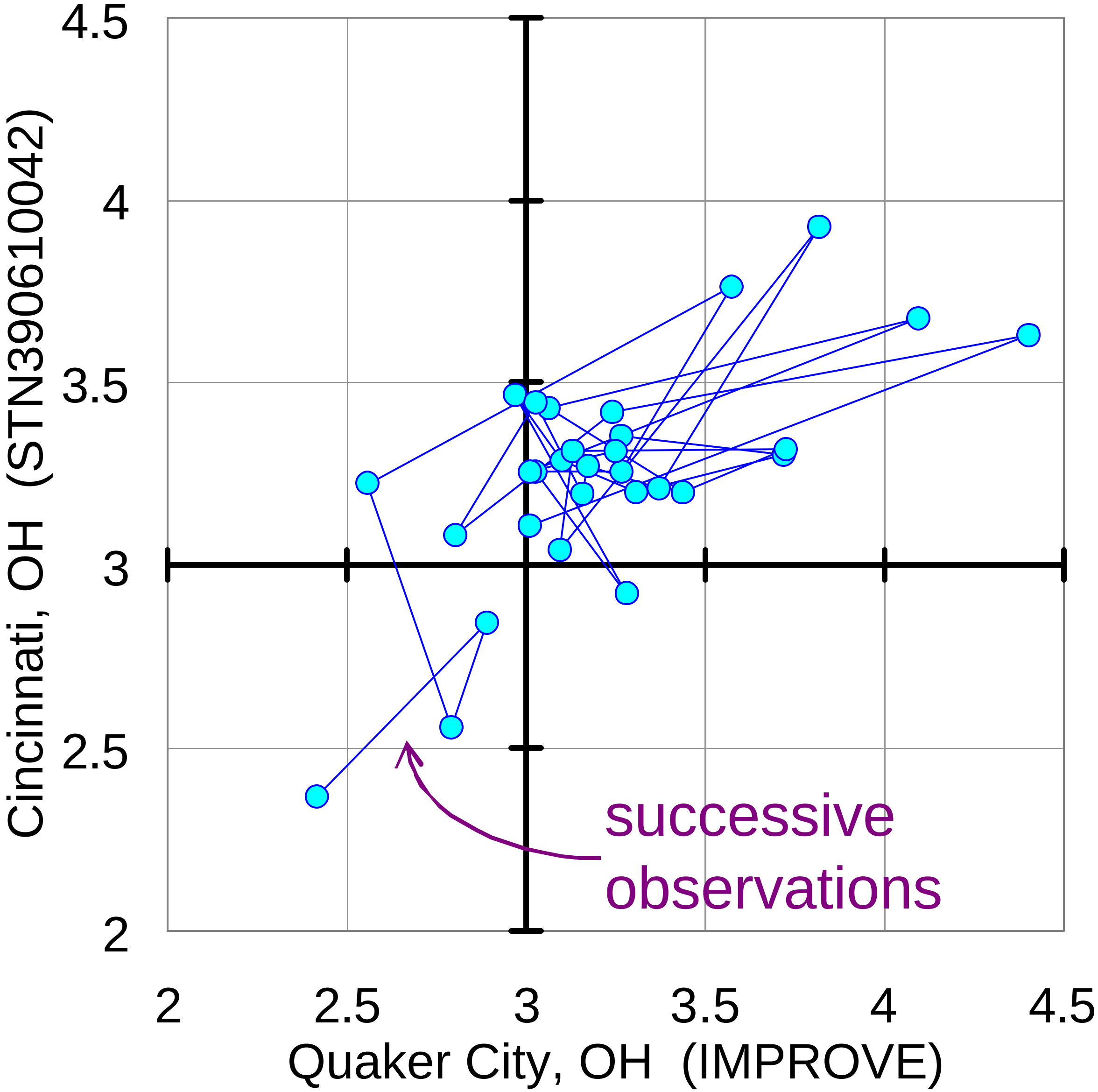
But there are other, method-related, possibilities:



sampling artifacts?
analytical interferences?

Monthly geometric mean $SO_4^{=}$/S over all IMPROVE and STN sites in DE, KY, MD, NC, NJ, OH, PA, TN, VA, WV.

(Requires $SO_4^{=}$ & 3S >3 µg/m^3 for STN, $SO_4^{=}$ & 3S >1 µg/m^3 for IMPROVE.)



$SO_4^{=}/S$

Aug 2002
- Nov 2003:

30 24h samples with
$SO_4^{=} > 3$ ug/m^3
and
S > 1 ug/m^3

at both locations

There are occasional (this pairing is not typical) hints that “it’s not all happening in the lab.”

Where does this leave us?

4.6.1 Precision Objectives

Precision is a measure of mutual agreement among individual measurements of the same property. The overall precision will be determined by collocated sampler

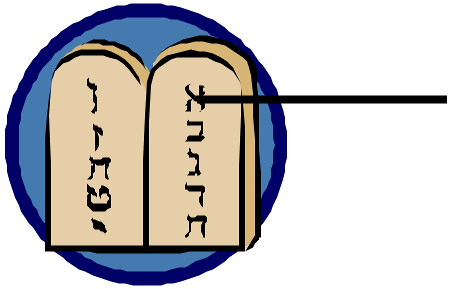
A pair of samplers is all we need to determine the overall precision of our measurement system, but we have no way to determine our overall bias or accuracy.

4.6.2 Accuracy Objectives

Accuracy is the agreement between a sample reading and the true value of the sample. Accuracy is determined by comparing instruments to reference standards traceable to the National Institute of Standards and Technology (NIST). The reference standard for flow rate and each analytical method is listed in Table 6.

The best we can hope to do is to stabilize our bias, to ensure that measurements at different locations at different times are comparable.

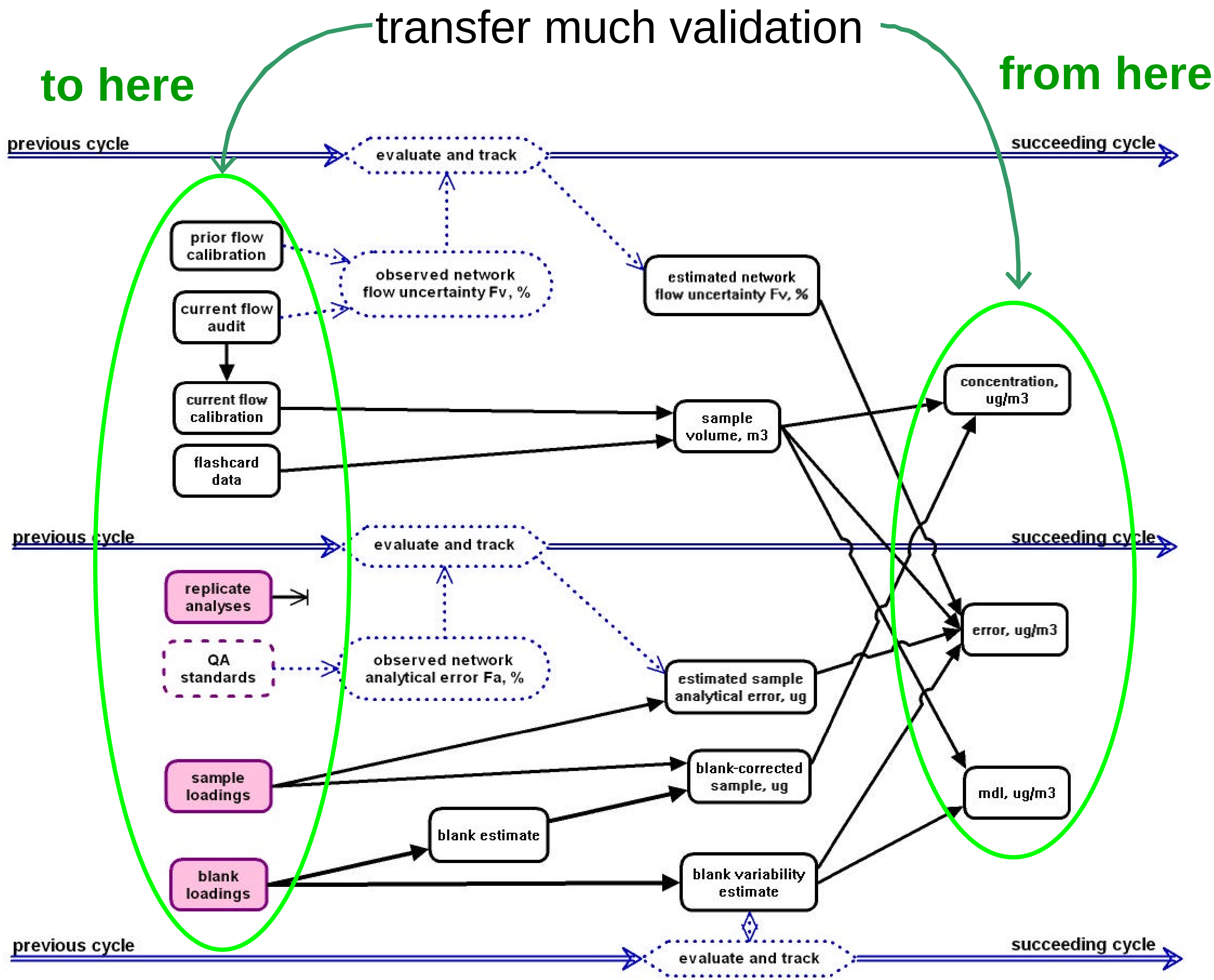
(As EPA does with the $PM_{2.5}$ FRM!)



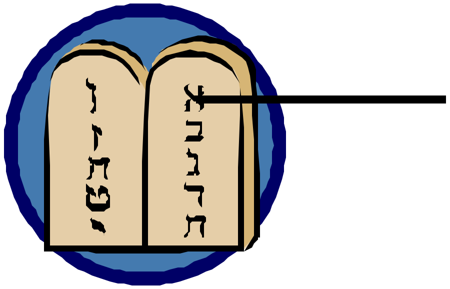
Focus on tolerances, not outliers.

Track and analyze routine ancillary data (*e.g.* flow and analytical calibrations, field and lab blank values) with no diminimus criteria for entry: no error or deviation is "negligible."

Use large-scale statistics to detect "unremarkable" patterns or changes over time in the ancillary data**;** unlike concentrations, they are not "contaminated" by the natural noise of the atmosphere.

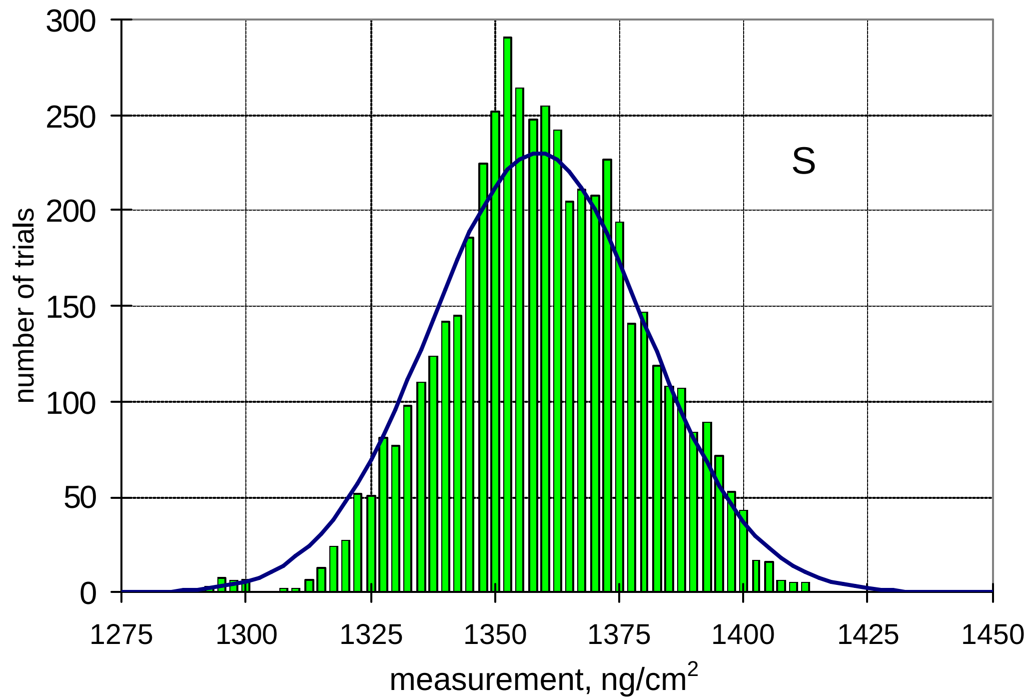


transfer much validation
to here
from here
previous cycle
evaluate and track
succeeding cycle
prior flow calibration
observed network flow uncertainty Fv, %
estimated network flow uncertainty Fv, %
current flow audit
current flow calibration
sample volume, m3
concentration, ug/m3
flashcard data
previous cycle
evaluate and track
succeeding cycle
replicate analyses
error, ug/m3
QA standards
observed network analytical error Fa, %
estimated sample analytical error, ug
sample loadings
blank-corrected sample, ug
mdl, ug/m3
blank estimate
blank loadings
blank variability estimate
previous cycle
evaluate and track
succeeding cycle



Focus on tolerances, not outliers. -- continued

Accept that the normal distribution is a continuum, with tails! Accept that some outliers will defy "explanation."



Utilize rich flag fields to identify and annotate observations taken outside established tolerances.