Title

Estimating measurement uncertainty in an ambient sulfate trend

Permalink

https://escholarship.org/uc/item/7wv3832x

Journal

Atmospheric Environment, 39(36)

ISSN

1352-2310

Authors

White, Warren H.
Ashbaugh, Lowell L
Hyslop, Nicole P
et al.

Publication Date

2005-11-01

Peer reviewed

Estimating measurement uncertainty in an ambient sulfate trend

W.H. White, L.L. Ashbaugh, N.P. Hyslop, and C.E. McDade   revised 8/1/2005

Crocker Nuclear Laboratory, University of California, One Shields Avenue, Davis, CA 95616; white@crocker.ucdavis.edu

ABSTRACT

This paper examines the effects of sampling and analytical error on time trends derived from routine monitoring. Our analysis is based on actual concentration differences observed among three long sulfate series recorded by collocated and independent measurements at Shenandoah National Park. Five-year sulfate trends at this location are shown to include a one-sigma uncertainty of about 1 %/year from measurement error alone. This is significantly more than would be estimated under naïve statistical assumptions from the demonstrated precision of the measurements. The excess uncertainty arises from subtle trends in the errors themselves.

KEYWORDS

IMPROVE, CASTNet, Shenandoah, sulfate, trend, error, uncertainty

INTRODUCTION

Tracking improvements in visibility under the Regional Haze Rule (USEPA, 2003) is a key objective of the IMPROVE network (Interagency Monitoring of Protected Visual Environments; IMPROVE, 2005a). The Rule seeks restoration of natural visibility conditions through steady improvements over the next six decades, and requires that implementation activities be verified to yield progress in the actual atmosphere. Documenting change on this time scale entails particle measurements that will support accurate comparisons between different eras, even as monitoring methods evolve (White,

1997; Weatherhead et al., 1998). Such absolute measurement stability is a requirement not encountered in the shorter, more intensive field campaigns typically mounted to support source apportionment and model validation studies.

It is easy to argue on theoretical grounds that measurement errors should have little impact on trend estimates. Random errors are as likely to cancel as to reinforce each other in successive observations, and the estimated imprecision of an average accordingly declines with the square root of the number of observations going into it. Fixed biases are even less of a concern: zero offsets cancel when observations from different periods are subtracted, and scaling factors cancel when ratios are taken.

The problem with theoretical analyses is that actual measurement series generally include non-ideal errors. Figure 1 shows a record of observed differences between collocated CASTNet (Clean Air Science and Trends Network; USEPA, 2005) monitors at Mackville, KY. (Such an extended series is not yet available for IMPROVE, which began routine collocated sampling only in 2003.) The Mackville measurements show errors that are neither independent in successive observations nor constant across all observations; instead, monitor 1 tends to read higher than monitor 2 for several successive months and then read lower for the next several months, with no regular periodicity. Such a pattern could arise from flow recalibrations and other adjustments made to individual samplers in the course of routine field operations (*cf.* Sickles and Shadwick, 2002).

This paper offers a case study of the trend uncertainty produced by such actual patterns of measurement error. Data are taken from Shenandoah National Park in northern Virginia, where IMPROVE and CASTNet have conducted collocated particle monitoring since 1988. Analyses focus on the sulfate fraction, for which the networks together yield three independent series of measurements. These redundant determinations allow the uncertainties of measurement to be isolated from those of meteorological variability and incomplete sampling.

MEASUREMENT ERROR

Concentration trends are naturally modeled in exponential terms, as linear trends in the logarithm of concentration (USEPA, 1998, 2003). The logarithmic transformation facilitates the statistics of trend estimation, converting weather's multiplicative effects on concentrations into additive "noise" that is more easily removed by standard techniques. In this framework it becomes necessary to characterize measurement error in terms of its effects on the logarithm of concentration.

When C is the measured value of a true concentration C_{true}, we define the measurement error to be

$$e \equiv \ln(C) - \ln(C_{true})$$

$$= \ln(C/C_{true}). \tag{1}$$

We are ignorant of C_{true} in general, and thus of e, but we can make two measurements of the same air parcel to determine their combined error:

$$\ln(C_2/C_1) = (\ln(C_{true}) + e_2) - (\ln(C_{true}) + e_1)$$

$$= e_2 - e_1. \tag{2}$$

The series $\ln(C_2/C_1)$ from the collocated CASTNet monitors at Mackville has already been presented, in Figure 1's implicitly logarithmic plot.

Figure 2a re-plots the data from Figure 1 as a function of concentration levels, represented horizontally by the geometric mean of the two measurements. The observed scatter shows little dependence on concentration, indicating that multiplicative errors (*e.g.* from flow measurement and aliquot preparation) dominate additive errors (*e.g.* from blank determination) in these data. Our log ratios $e = \ln(C/C_{true})$ can thus be plausibly modeled as random samples from a common distribution. The breadth of this distribution, typically as characterized by the standard deviation, represents the precision of the measurement.

Collocated measurements, such as those recorded at Mackville, are widely regarded as the "gold standard" for estimating precision. Under the plausible assumptions that $\text{var}(e_1) = \text{var}(e_2)$ (the measurement systems are identical) and that $\text{cov}(e_1, e_2) \geq 0$

3

(individual components of error are either decoupled or shared between the systems), we can bound precision in terms of observable differences between the measurements (Barlow, 1989):

$$rms(e) \equiv \sqrt{\frac{1}{n}\sum_1^n e^2} = \sqrt{\mathrm{var}(e) + \bar{e}^2}$$

$$\geq \sqrt{\mathrm{var}(e)} = \sqrt{\frac{\mathrm{var}(e_1) + \mathrm{var}(e_2)}{2}}$$

$$\geq \sqrt{\frac{\mathrm{var}(e_1) - 2\,\mathrm{cov}(e_1,e_2) + \mathrm{var}(e_2)}{2}} = \sqrt{\frac{\mathrm{var}(e_2 - e_1)}{2}}$$

$$= \frac{1}{\sqrt{2}} \sqrt{\mathrm{var}(\ln(C_2/C_1))}. \quad\quad\quad [3]$$

It should be emphasized that [3] provides only a lower bound for the error: equality holds only if the errors e_1 and e_2 are unbiased ($\bar{e}_1 = \bar{e}_2 = 0$) and completely uncorrelated with each other ($\mathrm{cov}(e_1,e_2) = 0$).

The point to be made in the remainder of this section is that collocated measurements do not, in general, reveal all of the contributions to measurement error that can affect trend estimates. The standard deviation of the logarithmic ratio, $SDLR = \sqrt{\mathrm{var}(\ln(C_2/C_1))}$, provides a useful metric for this discussion but differs from more standard metrics in which collocated precisions are usually reported. Our analysis of trend uncertainty in subsequent sections is unaffected by this difference, as it rests on the underlying data series rather than any precision metric. The following digression is thus provided only for context. Collocated precision is most commonly characterized in terms of the relative difference $RD = 2\frac{C_2 - C_1}{C_2 + C_1}$, which approaches the logarithmic ratio quadratically ($RD - \ln(C_2/C_1) = O(RD^2)$) at small values (Whittaker and Watson, 1969). The U.S. Environmental Protection Agency (USEPA, 1997) defines precision as $\frac{1}{\sqrt{2}}rms(RD)$, which is 6.6% for the Mackville data. Sickles and Shadwick (2002) analyzed the Mackville data through September 2000 and reported a median absolute value of 3.3% for RD after correcting the individual values for the median

4

overall bias. Illustrated in Figure 2b, these statistics differ in predictable directions from

$\frac{1}{\sqrt{2}} SDLR = 0.049$ because they give more (USEPA) or less (Sickles and Shadwick) weight

to the largest disagreements.

Data from a temporary overlap between two IMPROVE sampling systems illustrate how shared error sources can hide in collocated measurements. IMPROVE installed new sampling systems throughout the network during the summer of 2000, and the old and new versions of the system were run concurrently for several months at Shenandoah National Park and two other locations to establish their equivalence. Figure 3 shows 24h sulfate concentrations derived from the March-August collocation at Shenandoah. Each system includes three independent fine-particle sampling modules, one for gravimetric and elemental analyses, another for ionic analysis, and the third for carbon analysis. On the assumption that all fine-particle sulfur is in the +6 oxidation state (Charlson *et al*., 1978), sulfate mass concentrations $[SO_4^=]$ can be taken directly from the ion measurement or calculated as $3[S]$ from the elemental measurement and ratio of molar weights.

The overlapping measurements at Shenandoah left us with four independent measurements of what should be the same sulfate concentration: $[SO_4^=]_1$, $[SO_4^=]_2$, $3[S]_1$, and $3[S]_2$, where the subscripts denote measurements by sampler versions 1 and 2. Suppose, contrary to the point we seek to establish, that the error in each measurement is completely uncorrelated with all the others. Assuming the two versions of the sampler to be functionally equivalent, as they are designed to be, [3] then supplies the error variance corresponding to

each of the chemical analyses: $\mathrm{var}(e_{SO4}) = \frac{1}{2}\mathrm{var}(\ln\left([SO_4]_2 \middle/ [SO_4]_1\right)) = 0.044^2$ and

$\mathrm{var}(e_{3S}) = \frac{1}{2}\mathrm{var}(\ln\left(3[S]_2 \middle/ 3[S]_1\right)) = 0.064^2$. From these uncertainties, [2] and [3] predict how

well the two chemical analyses will agree in same version of the sampler:

$\frac{1}{2}\mathrm{var}(\ln\left([SO_4] \middle/ 3[S]\right)) = \frac{1}{2}\mathrm{var}(e_{SO4} - e_{3S}) = \frac{\mathrm{var}(e_{SO4}) + \mathrm{var}(e_{3S})}{2} = \frac{0.044^2 + 0.064^2}{2} = 0.055^2$. The

inter-analysis agreement actually observed is substantially poorer, however, with

$\frac{1}{2}\mathrm{var}(\ln\left([SO_4]_1 \middle/ 3[S]_1\right)) = 0.085^2$ and $\frac{1}{2}\mathrm{var}(\ln\left([SO_4]_2 \middle/ 3[S]_2\right)) = 0.102^2$. There are thus

differences that show up in the comparisons between sulfur and sulfate but do not show up as differences between version 1 and version 2 samplers. These must arise from errors that have the same effect on measurements by both samplers, such as a varying bias in one analytical method relative to the other.

The recognition that collocated measurements need not capture all contributors to measurement error is critical for trend analysis. Trends describe differences between observations made at different times, after all, while collocated measurements must coincide if atmospheric variations are not to confound their interpretation. The problem is that different errors vary on different time scales, ranging from sampling and analytical fluctuations that change from one sample to the next, through recurrent operations and calibration schedules in the field and laboratory, to the slow accumulation of technological and procedural improvements over many years. The demonstrated precision of a measurement at each point in time can thus give a misleading picture of how reliably comparisons can be made between different eras.

EXPERIMENTAL DATA

The IMPROVE network was designed to characterize regional haze, and collects 24h $PM_{2.5}$ samples on three different filter media for a variety of physical and chemical analyses. The data examined here are for sulfur (S) concentrations by x-ray fluorescence (XRF) and sulfate ($SO_4^=$) concentrations by ion chromatography (IC). Samples are collected on Teflon filters for XRF and on Nylon filters following carbonate denuders for IC, employing parallel sampling trains with separate inlets, cyclones, vacuum pumps, and critical orifices for flow control. All documentation and data are downloadable from IMPROVE (2005a).

Figure 4 compares the $[SO_4^=]$ and 3[S] series at Shenandoah. IMPROVE guidance generally recommends use of S data from the Teflon filter to track haze trends, viewing $SO_4^=$ from the Nylon filter as a quality-assurance byproduct of a sampling train designed to capture the nitrate ion (U.C. Davis, 1995). However it recommends use of $SO_4^=$ data from the Nylon filter in the early years at some eastern locations, when masks used with the Teflon filter appear to have caused occasional sampling artifacts at high loadings (IMPROVE, 2005b).

The mask at Shenandoah was removed on 19 April 1995, and our comparisons with CASTNet sulfate will accordingly employ IMPROVE [$SO_4^=$] before this date and IMPROVE 3[S] afterward. The hybrid series will be referred to as the <u>recommended</u> sulfate value for IMPROVE.

The IMPROVE sampling schedule changed in 2000 from twice weekly (Wednesdays and Saturdays) to every third day, to synchronize with regulatory compliance monitoring. A number of other changes have occurred through the years in details of sampling and analysis. The shift to every-third-day sampling was implemented with a new version of the IMPROVE sampler that records flow and temperatures continuously, permitting more accurate calculations of sample volume. Conventional copper-anode XRF replaced proton-induced x-ray excitation (PIXE) for analysis for S and other light elements in December 2001. Different labs and eluents have been used in different eras for the IC analysis, as have differing filter sizes and manufacturers. All these substitutions produced no evident discontinuities in the data record when introduced, but the possibility that subtle effects might accumulate through the years merits continuing scrutiny.

Observations are occasionally lost to operational problems, and some problems are more likely under certain atmospheric conditions. Figure 5 shows an example from 1998-99, when minor changes in the Nylon filter supply triggered sharp increases in clogging at eastern sites (IMPROVE, 2005c). Annual averages calculated from all valid [$SO_4^=$] data were biased low in 1998-99 simply because sample recovery rates were higher on days with lower concentrations, independent of any measurement error. We can eliminate the need to consider such sampling issues by limiting all comparisons to paired individual measurements, avoiding the aggregation of unmatched observations.

Although the IMPROVE $SO_4^=$ and S data come from independent analyses on different filters, both measurements pass through a common system of quality checks (which include comparing [$SO_4^=$] with 3[S]) and reporting conventions. CASTNet monitors dry deposition at Shenandoah in collocation with IMPROVE, and the CASTNet measurements provide a fully independent record at this system level. CASTNet uses filter packs to collect week-long acid gas and particle samples, and determines particle $SO_4^=$ from IC analysis of

the Teflon front filter. The CASTNet data are downloadable along with documentation from USEPA (2005).

The samples collected by CASTNet and IMPROVE differ in a number of ways: the CASTNet filter collects coarse as well as fine particles and samples from 10 rather than 3 m above ground level, at a flow rate of 1.5 rather than 23 lpm. These differences have limited implications for measured concentrations, because sulfate particles are predominantly fine, well-mixed spatially, and non-volatile. The more problematic incompatibility is that IMPROVE samples only 48 or 72h of the 168h CASTNet week. One approach to the mismatched CASTNet and IMPROVE sampling intervals has been to aggregate both measurements up to longer periods before comparison (Ames and Malm, 2001). To limit our comparisons to paired individual observations as suggested above, the present analysis instead exploits hourly visibility data to disaggregate the CASTNet measurements.

IMPROVE monitored total extinction at Shenandoah from June 1991 through 2001 with a long-path transmissometer, and has monitored particle scattering since September 1996 with an ambient nephelometer. As sulfate is the dominant contributor to light scattering and extinction there (Ferman *et al*., 1981; Weiss *et al*., 1982), these optical records offer a rough indication of daily variations in sulfate levels. We aggregate the data to 24h averages for comparability with the filter measurements, using only observations flagged as free from possible weather interference as described by Molenar (1997), Blandford (2004), and Mercer (2004). Because overnight relative humidities at Shenandoah commonly exceed levels flagged as possible weather interferences, we require a minimum of only six clear observations for a valid 24h average.

Figure 6 shows that much of the difference between the 24h IMPROVE and 168h CASTNet sulfate measurements is accounted for by observed changes in daily light extinction. Sulfate concentrations from IMPROVE are most likely to exceed those from CASTNet on just those IMPROVE sampling days with higher extinction coefficients than the average for the CASTNet sampling week. The observed proportionality suggests scaling CASTNet sulfate concentrations by the measured 24h/168h extinction ratios to estimate "optically adjusted CASTNet" 24h concentrations:

$$SO_4^{24h} = \frac{B_{ep}^{24h}}{B_{ep}^{week}} SO_4^{week} \ .$$

The nephelometer data are used for the adjustment when they are available, as the transmissometer measurement includes scattering by gases as well as particles and is more difficult to calibrate in the field (Molenar, 2002). However any biases in the optical data, even those that might exhibit trends of their own (e.g. from the exclusion of high-humidity observations in a climate undergoing long-term change), will tend to cancel out since only a ratio of measurements taken within the same week enters the calculation.

Values smaller than 0.1 are plotted at 0.1 in Figure 6 and all subsequent figures, but these values are correctly entered in regression analyses and other statistics. With the exceptions of one outlying observation (4/22-29/97) dropped from the CASTNet data at Mackville and two successive IMPROVE Teflon filter observations (8/9/02 and 8/12/02) interchanged at Shenandoah in response to persuasive circumstantial evidence of a documentation error, all data are presented as downloaded from references 2 and 4 on 11/18/04.

TREND MODEL

The sensitivity of trend estimates to measurement error can be expected to depend on estimation methodology. Our analytical framework is adapted from a design study that was undertaken for the PM$_{2.5}$ Speciation Trends Network (STN; USEPA, 1998). This approach is more closely aligned with standard statistical models than are the calculations prescribed by the Regional Haze Rule, whose investigation is deferred to subsequent work.

Trends in concentration are typically sought, and interpreted, as evidence of trends in emissions. To this end, ambient concentrations are conveniently viewed as products

$$C_{meas} = \frac{C_{meas}}{C_{true}} \frac{C_{true}}{Q} Q \qquad\qquad [4]$$

of the emission rate Q, an atmospheric transport and transformation factor $X_{atmos} = C_{true}/Q$, and any error $\Phi_{error} = C_{meas}/C_{true}$ present in the measurement, where C_{meas} and C_{true} denote

the measured and true values. Standard statistical techniques are better suited to the logarithmic transform of the relationship,

$$\ln(C_{meas}) = \ln(\Phi_{error}) + \ln(X_{atmos}) + \ln(Q).$$ [5]

Emissions are modeled as the deterministic signal that trend analysis seeks to detect:

$$\ln(Q) = \ln(Q_0) + \beta_{trend}t.$$ [6]

Measurement errors are usually treated as random noise superimposed on a fixed calibration, which may itself be biased. As we are interested in the potential for this calibration to drift over time, we instead model measurement errors as

$$\ln(\Phi_{error}(t)) = \alpha_{calib}t + \delta_{noise},$$ [7]

where α_{calib} is constant and δ_{noise} is a random variable. Actual drift is more likely to occur in a series of discrete jumps, but our approximate representation should be adequate at multi-year time scales.

Atmospheric processes are assumed stationary over the long term, with regular cycles that follow the seasons and irregular fluctuations that reflect daily weather patterns. They are modeled as

$$\ln(X_{atmos}(t)) = \ln(F(season(t))) + \ln(X_0) + \delta_{weather},$$ [8]

where $season(t)$ is the fractional-year part of the time index and $\delta_{weather}$ is a random variable. $F(season(t))$ is the trend-adjusted 15-year geometric mean for each calendar month, plotted as a repeating cycle in Figure 4. These means were obtained from an ordinary least squares (OLS) fit of the regression model

$$\ln(C_{CASTNet}) = f_{jan} + \cdots + f_{dec} + bt$$ [9]

to the 15-year CASTNet record at Shenandoah. Because IMPROVE data were not used in this fit, we can continue to model them as random variables.

Equations [5-8] imply that the time series of measured concentrations can be modeled as a linear regression of log concentrations on linear time:

$$\ln(C_*(t)) = a + bt + \varepsilon,$$ [10]

where

$C_*(t) = C_{meas}(t)/F(season(t))$ is the deseasonalized concentration,

$a = \ln(X_0 Q_0)$ is the expected log-concentration at the initial emissions rate,

$b = \alpha_{calib} + \beta_{trend}$ is the annual trend, and

$\varepsilon = \delta_{noise} + \delta_{weather}$ is daily fluctuation not predicted by season or trend.

Figure 7 shows the deseasonalized series $C_{meas}/F(season)$ obtained for Shenandoah when C_{meas} is taken from measured $[SO_4^=]$ or 3[S] as recommended by IMPROVE. Linear regression on model [7] yields a 15-year trend of $b^{OLS} = -2.0$ %/year, indicated by the dashed line. With a standard error of 0.4 %/year, this is statistically indistinguishable from the slope of -1.9 %/year that results from regression of raw CASTNet data on model [9]. As a point of comparison, total SO$_2$ emissions in the eastern U.S. are estimated to have declined between 1988 and 1999 by -1.8 %/year (Malm et al., 2002).

STATISTICAL ANALYSIS

The preceding section established that observed trends $b^{OLS} \cong \alpha_{calib} + \beta_{trend}$ in sulfate concentrations can conflate real trends in emissions with artifacts arising from measurement drift. This section considers how to estimate such artifacts.

Figure 8 shows that measurement error is only a minor contributor to the total variance of daily concentrations. The 15-year geometric standard deviation of 24h $[SO_4]/3[S]$ ratios is 1.13 (*i.e.* the arithmetic standard deviation of $\ln([SO_4]/3[S])$ is $\ln(1.13)$); for comparison, the deseasonalized sulfate concentrations in Figure 6 have a 15-year geometric standard deviation of 2.0. The observed variance in $\ln([SO_4]/3[S])$ corresponds to a 9% (*i.e.* $\sqrt{(\ln(1.13)^2/2)}$) uncertainty in each of the two measurements if apportioned equally between them. This represents only 2% of the total variance in deseasonalized concentration.

The effect of measurement error on long-term trends is harder to extract from observations. We must consider the possibility that both sulfur and sulfate measurements drift over time:

$$\ln([SO_4]/C_{true}) = a_{SO4}t + \delta_{SO4},$$ [7a]

and $$\ln(3[S]/C_{true}) = a_{3S}t + \delta_{3S}.$$ [7b]

Regression on the model

$$\ln([SO_4]/3[S]) = (a_{SO4} - a_{3S})t + (\delta_{SO4} - \delta_{3S})$$ [11]

then yields an estimate for the relative drift $a_{SO4} - a_{3S}$.

The Shenandoah IMPROVE data show a 15-year trend of -0.4 %/year in $[SO_4]$ relative to 3[S], indicated by the dashed line in Figure 8. This slope is reassuringly small compared to the concentration trend of -2.0 %/year in Figure 7. It is of little value in bounding calibration drift, however, as it could result from the canceling of larger drifts in the individual measurements.

We can get a fuller picture of possible calibration drifts by examining how $a_{SO4} - a_{3S}$ varies when we look at different periods. Five years is sometimes viewed as the minimum interval in which one might hope to detect a meaningful change in regional concentrations (USEPA, 1998, 2003). Assuming as we do that sulfur exists in fine particles only as sulfate, the true value of the trend in $\ln([SO_4]/3[S])$ is of course zero in all periods. The second column of Table 1 lists the estimates obtained by regression on model [11] for each rolling 5-year period in the Shenandoah measurement record. The standard errors of the regression trend, shown in the third column, are adjusted for the effects of observed autocorrelation (Johnson, 1984). The comparison between the two IMPROVE measurement methods is tight enough to detect statistically significant trends in their disagreements. Although results from overlapping periods are clearly interdependent, they nevertheless shed light on overall variability.

If a_{SO4} and a_{3S} can be assumed uncorrelated with each other, then $\text{var}(a_{SO4} - a_{3S}) \cong \text{var}(a_{SO4}) + \text{var}(a_{3S})$ provides an indicator of calibration variations in 5-year concentration trends. The bottom row of Table 1 summarizes the uncertainties

estimated by apportioning these variations equally between both measurements. The

controlling uncertainty, indicated in bold, is the real variability of calibration trends across

the years. Some of these temporary excursions are attributable to problems that arose in one

measurement or the other and were then identified in the routine cross-checking made

possible by availability of the independent IC and XRF determinations (Eldred, 2001). Since

these known problems have been corrected in subsequent measurements, it can be argued

that they are of limited relevance to current operations. Under this interpretation, we could

expect to hold 5-year trend uncertainty below 1.3 %/year by always using the better of the

two measurements.

The better of the two measurements according to IMPROVE guidance is compared

with optically adjusted CASTNet data in Figure 9 and columns four and five of Table 1. As

suggested by the preceding argument, this comparison does yield smaller 5-year trends in

observed disagreements. (The overall trend is smaller as well, at -0.3 %/year.) Moreover,

the organizational independence of IMPROVE and CASTNet makes a very strong case for

statistical independence between the two measurements. The comparison suffers, however,

from a much weaker relationship between individual observations, evident in the scatter seen

in Figure 9. The individual 5-year trend estimates in column 5 accordingly have greater

standard errors, with a rms value of 1.8 %/year. This corresponds to estimation errors of

$1.8/\sqrt{2} = 1.3$ %/year in individual trends, making statistical noise rather than measurement

error the controlling uncertainty.

It should be emphasized that the scatter in our inter-network comparison results from

the conversion of CASTNet measurements to 24h "equivalents," not from any greater

imprecision in the measurements themselves. The optical adjustment does significantly

improve the bound obtained from the inter-network comparison, as shown by the much larger

numbers in columns 6 and 7 of Table 1 for direct comparison of weekly averages ("IMP/CN,

wk").

The most straightforward of the comparisons was the one shown in Figures 1 and 2,

involving collocated sampling at a location exposed to roughly the same regional haze as

Shenandoah. The identical instruments and replicate analyses at Mackville yield a 0.9

%/year uncertainty in 5-year trends, as shown in the last two columns of Table 1. This is only marginally below the 1.3-1.4 %/year values obtained from inter-method and inter-network comparisons at Shenandoah, and it seems reasonable to conclude that the measurement uncertainty achieved by current monitoring is somewhere around 1 %/year for 5-year sulfate trends. More precisely, we are unable to demonstrate that present uncertainty in 5-year sulfate trends is substantially below 1 %/year.

SUMMARY

Monitoring records for Shenandoah National Park suggest that 5-year sulfate trends estimated from existing historical data include a one-sigma uncertainty of about 1 %/year from measurement error alone. This estimate corresponds to a cumulative uncertainty of 5% concentration change over 5 years. It does not, however, imply an uncertainty of 10% over 10 years or 1% over a single year. Because calibrations against absolute standards can keep overall accuracy within set limits, the annualized effect of measurement error can be expected generally to decrease as the record lengthens.

Our earlier discussion of measurement error noted that the repeatability of contemporary measurements can be a misleading guide to trend precision, and we conclude with a numerical calculation to illustrate this point. Consider a Y-year series of unbiased measurements from every third day, with random errors and $E\%$ precision. We showed the artifact trend contributed by measurement error to be the slope b fitted by the regression

$$\ln\left(C_{meas}/C_{true}\right) = a + bt + \varepsilon .$$

Since $b = 0$ in the true relationship, the rms value of the regression estimates is the standard error s_b of the regression slope. This is given by (Glantz, 2002)

$$s_b = \frac{E}{s_t}\sqrt{\frac{1}{n-2}} \cong \frac{E}{Y/\sqrt{12}}\sqrt{\frac{1}{120Y}} = \frac{E}{Y\sqrt{10Y}} ,$$

where $s_t^2 \cong \frac{1}{Y}\int_{-Y/2}^{Y/2} t^2 dt = Y^2\!/12$ is the time variance and n the number of measurements. For $Y = 5$, this theoretical relationship yields $s_b \cong E/35$.

Naïve statistical calculations would thus suggest that measurements of 35% precision are good enough to determine a 5-year trend to within 1 %/year, and that the precisions demonstrated by current monitoring networks should hold uncertainties within much tighter bounds. It may be recalled that the observed 15-year log variance of IMPROVE 24h $[SO_4]/3[S]$ ratios implied a measurement precision of 9%, which would yield a naïve estimate for trend uncertainty of ¼ %/year. That larger trend discrepancies are observed in actual data series indicates the importance of slowly changing errors that are neither random nor fixed.

ACKNOWLEDGMENTS

REFERENCES

R.B. Ames and W.C. Malm, 2001. Comparison of sulfate and nitrate particle mass concentrations measured by IMPROVE and the CDN. *Atmospheric Environment 35*, 905-916.

R.J. Barlow, 1989. Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences, pp 8-18. John Wiley & Sons, Chichester.

J.C. Blandford, 2004. Nephelometer Data Reduction and Validation (IMPROVE Protocol). http://vista.cira.colostate.edu/improve/Publications/SOPs/ARS_SOP/2005/Neph/ Nephelometerdatareduction.pdf.

R.J. Charlson, D.S. Covert, T.V. Larson, and A.P. Waggoner, 1978. Chemical properties of tropospheric sulfur aerosols. *Atmospheric Environment 12*, 39-53.

U.C. Davis, 1995. IMPROVE Data Guide: A guide to Interpret Data.
http://vista.cira.colostate.edu/improve/Publications/OtherDocs/IMPROVEDataGuide/
IMPROVEDataGuide.htm.

R. Eldred, 2001. Sulfur-Sulfate History for IMPROVE. http://vista.cira.colostate.edu/
improve/Publications/GrayLit/006_Sulfur-Sulfate_History/ 006_sulfur-sulfate_history.htm.

M.A. Ferman, G.T. Wolff, and N.A. Kelly, 1981. The nature and sources of haze in the
Shenandoah Valley/Blue Ridge Mountains area. *J. Air Pollution Control Association 31*,
1074-1081.

S.A. Glantz, 2002. Primer of Biostatistics, Fifth Edition, p 270. McGraw-Hill, New York.

IMPROVE, 2005a. Interagency Monitoring of Protected Visual Environments.
http://vista.cira.colostate.edu/improve.

IMPROVE, 2005b. Underestimation of Sulfur Concentrations During High Loadings and
Humidity Conditions in the Eastern US.
http://vista.cira.colostate.edu/improve/Data/QA_QC/Issues/S_underestimation.htm.

IMPROVE, 2005c. Significant Loss of Ion Concentrations in 1998 Due to Clogging of
Module B Nylon Filters.
http://vista.cira.colostate.edu/improve/Data/QA_QC/Issues/IonLoss.htm.

J. Johnston, 1984. Econometric Methods, pp 310-312. McGraw-Hill Book Co., New York.

W.C. Malm, B.A. Schichtel, R.B. Ames, and K.A. Gebhart, 2002. *J. Geophysical Research
107(D22)*, 4627, doi:10.1029/2002JD002107.

G.S. Mercer, 2004. Transmissometer Data Reduction and Validation (IMPROVE Protocol).
http://vista.cira.colostate.edu/improve/Publications/SOPs/ARS_SOP/2005/Trans/
Transmissometerdatareduction.pdf.

J.V. Molenar, 1997. Analysis of the Real World Performance of the Optec NGN-2 Ambient Nephelometer. In: Visual Air Quality: Aerosols and Global Radiation Balance, 243-265, Air and Waste Management Association, Pittsburgh, Pa.

J. Molenar, 2002. Warnings on the use of transmissometer data to analyze long-term visibility trends.   http://vista.cira.colostate.edu/improve/Publications/GrayLit/ TransDataUseWarning/TransDataUseWarning.htm.

J.E. Sickles II and D.S. Shadwick, 2002. Precision of atmospheric dry deposition data from the Clean Air Status and Trends Network, *Atmospheric Environment 36*, 5671-5686.

USEPA, 1997. Ambient Air Quality Surveillance for Particulate Matter, 40 CFR Part 58, pp. 71-72. http://www.epa.gov/ttn/oarpg/t1/fr_notices/pm-mon.pdf

USEPA, 1998. Data Quality Objectives for the Trends Component of the $PM_{2.5}$ Speciation Network. http://www.epa.gov/ttnamti1/files/ambient/pm25/spec/dqo3.pdf.

USEPA, 2003. Guidance for Tracking Progress under the Regional Haze Rule, EPA-454/B-03-004. http://www.epa.gov/ttn/oarpg/t1/memoranda/rh_tpurhr_gd.pdf.

USEPA, 2005. Clean Air Science and Trends Network. http://www.epa.gov/castnet.

E.C. Weatherhead, G.C. Reinsel, G.C. Tiao, X.-L. Meng, D. Choi, W.-K. Cheang, T. Keller, J. DeLuisi, D.J. Wuebbles, J.B. Kerr, A.J. Miller, S.J. Oltmans, and J.E. Frederick, 1998. Factors affecting the detection of trends: statistical considerations and applications to environmental data. *Journal of Geophysical Research 103,* 17149-17161.

R.E. Weiss, T.V. Larson, and A.P. Waggoner, 1982. *In situ* rapid-response measurement of $H_2SO_4/(NH_4)_2SO_4$ aerosols in rural Virginia. *Environmental Science & Technology 16*, 525-532.

W.H. White, 1997. Deteriorating air or improving measurements? -- on interpreting concatenate time series. *Journal of Geophysical Research 102*, 6813-6821.

E.T. Whittaker and G.N. Watson, 1969. A Course in Modern Analysis, 4th ed, pp 95-96. Cambridge University Press, Cambridge.

	Shenandoah, VA						Mackville, KY	
	SO$_4$/3S		IMP/OACN, 24h		IMP/CN, wk		CN2/CN1	
Period	decrease/yr	+/-	decrease/yr	+/-	decrease/yr	+/-	decrease/yr	+/-
6/88 – 5/93	0.3%	0.6%			-6.6%	2.3%		
6/89 – 5/94	-0.5%	0.7%			-5.0%	2.1%		
6/90 – 5/95	0.1%	0.7%			-1.7%	1.9%		
6/91 – 5/96	2.3%	0.8%			1.0%	1.8%		
6/92 – 5/97	2.8%	0.7%	-2.8%	2.3%	-0.8%	1.6%		
6/93 – 5/98	2.6%	0.7%	-0.9%	2.1%	1.7%	1.9%	0.5%	1.0%
6/94 – 5/99	2.0%	0.6%	0.4%	1.9%	3.9%	1.9%	-1.7%	0.8%
6/95 – 5/00	0.8%	0.5%	-1.0%	1.7%	2.3%	2.1%	-2.1%	0.8%
6/96 – 5/01	-0.6%	0.6%	0.7%	1.5%	2.2%	2.1%	-1.5%	0.7%
6/97 – 5/02	-1.5%	0.7%	1.5%	1.4%	-1.0%	2.1%	-0.3%	0.6%
6/98 – 5/03	-3.5%	0.8%	0.3%	1.5%	-3.1%	1.8%	0.1%	0.6%
$rms/\sqrt{2}$	**1.3%**	0.5%	0.9%	**1.3%**	**2.3%**	1.4%	**0.9%**	0.5%

Table 1: Five-year trends in measurement differences.
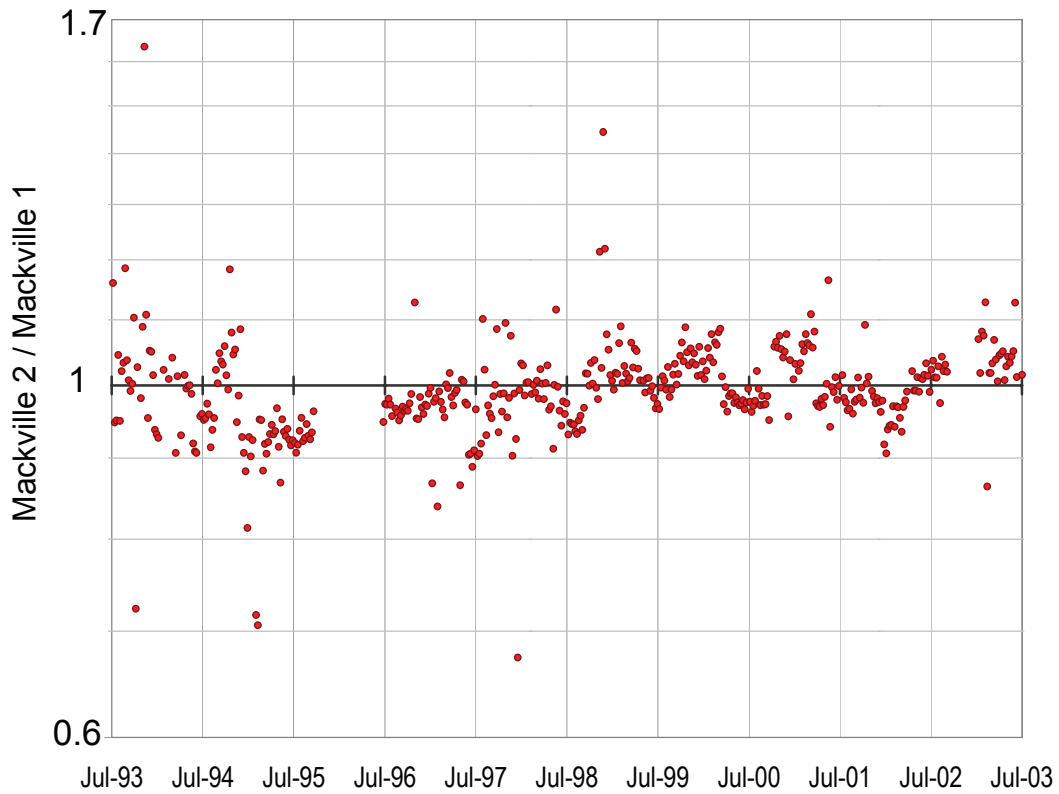
Figure 1: Collocated weekly CASTNet sulfate measurements at Mackville, Kentucky. Each point represents the ratio of ambient sulfate concentrations determined for two 168h samples collected by identical filter sampling trains. Operations were suspended during the 1995-96 federal budget impasse.
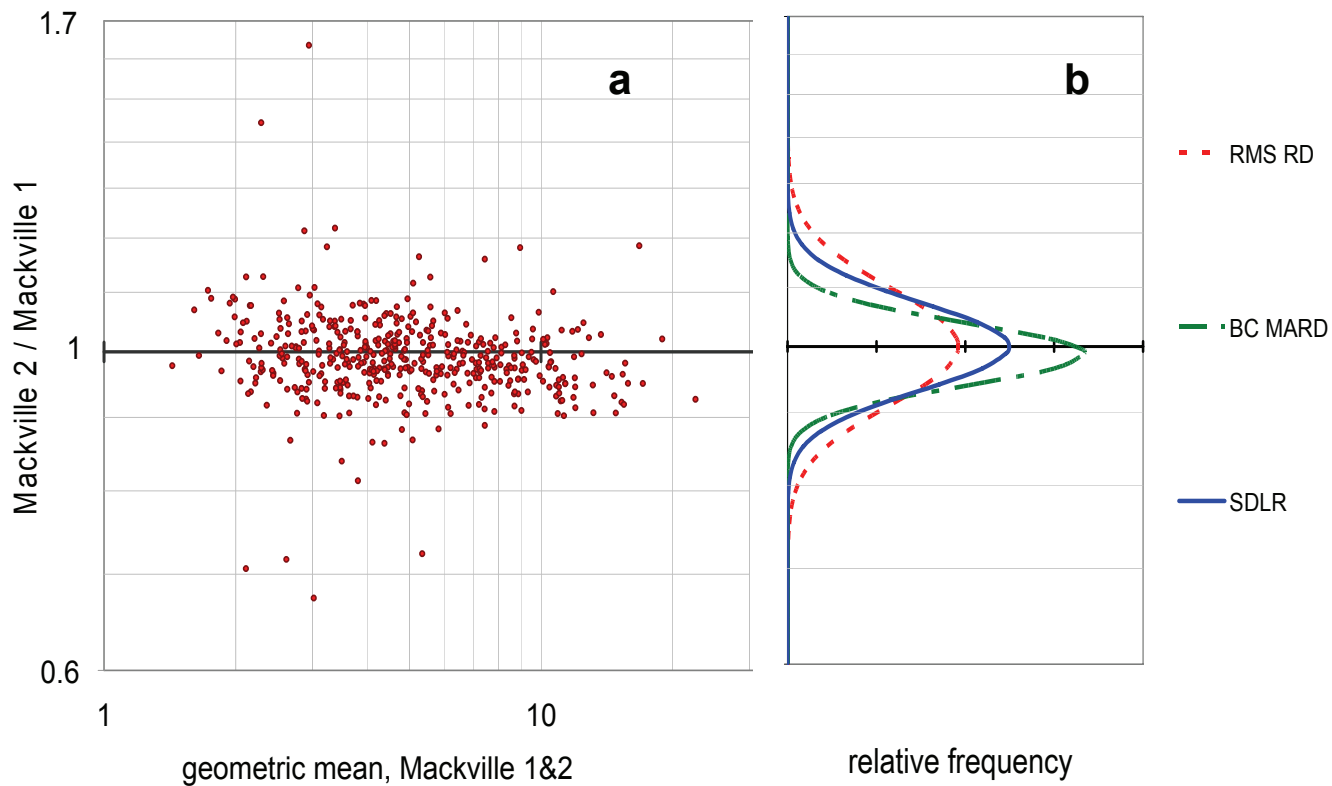
Figure 2: (a) Collocated weekly CASTNet sulfate measurements as in Figure 1.
(b) Model distributions corresponding to three different precision metrics discussed in text:
root mean square relative difference (RMS RD), blank-corrected median absolute relative
difference (BC MARD), and standard deviation of the logarithmic ratio (SDLR).
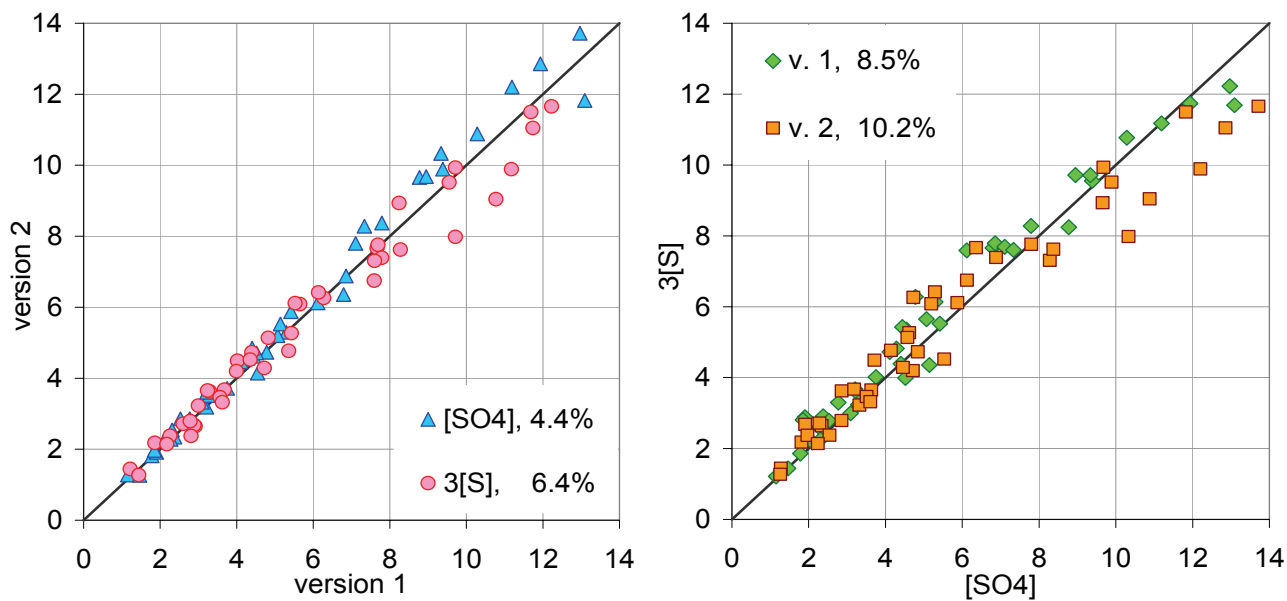
Figure 3: IMPROVE 24h sulfate concentrations (μg/m^3) from four independent sampling modules at Shenandoah National Park, March-August 2000. Legends give the precision metric $SDLR/\sqrt{2}$ described in the text for the indicated pair-wise comparisons.

Figure 4: Sulfate concentrations at Shenandoah National Park (Virginia). Individual points indicate 24h IMPROVE measurements of $SO_4^=$ and 3S; the repeating curve shows a seasonal cycle independently derived from CASTNet observations.
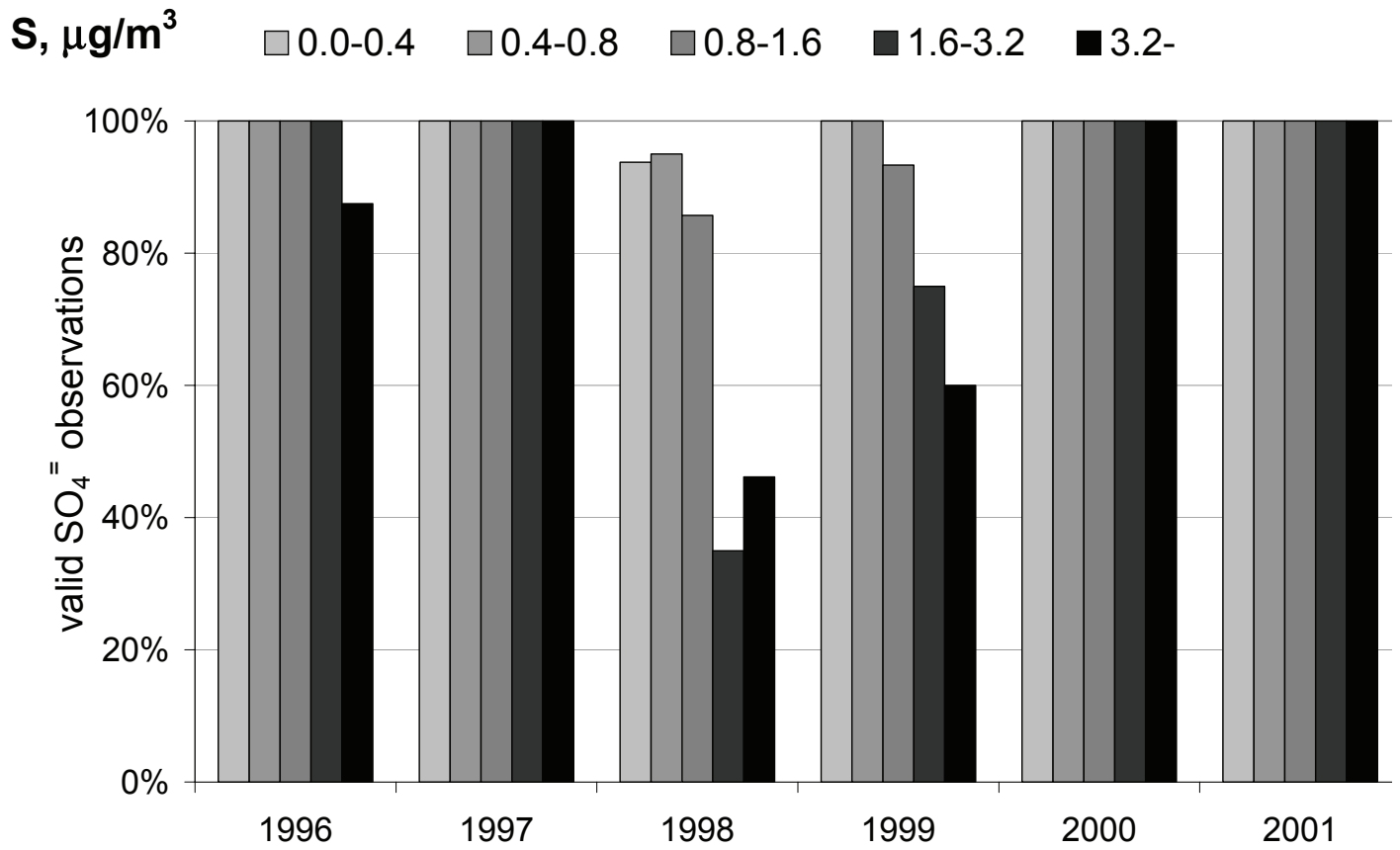
Figure 5: Data recovery rates at Shenandoah for $SO_4^=$ measured by IC, as a function of concurrent S concentrations measured by XRF.
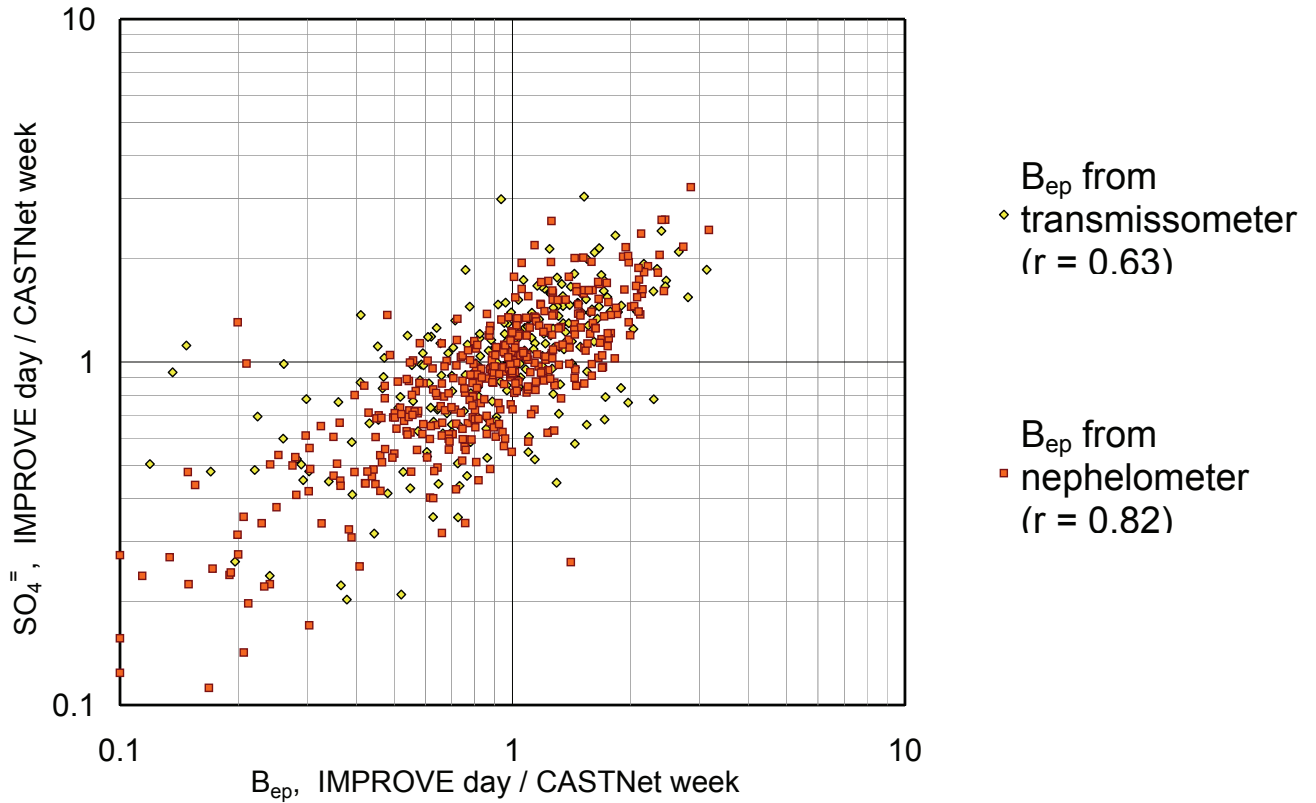
Figure 6: Intra-week correlation at Shenandoah between sulfate and extinction by particles, for weeks with CASTNet sulfate concentrations above 2 ug/m^3. The vertical axis shows the ratio of a 24h sulfate concentration, as measured by the recommended IMPROVE method, to that week's sulfate concentration measured by CASTNet. The horizontal axis shows the corresponding ratio of particle light extinction, estimated from transmissometer measurements of total extinction or nephelometer measurements of particle scattering.
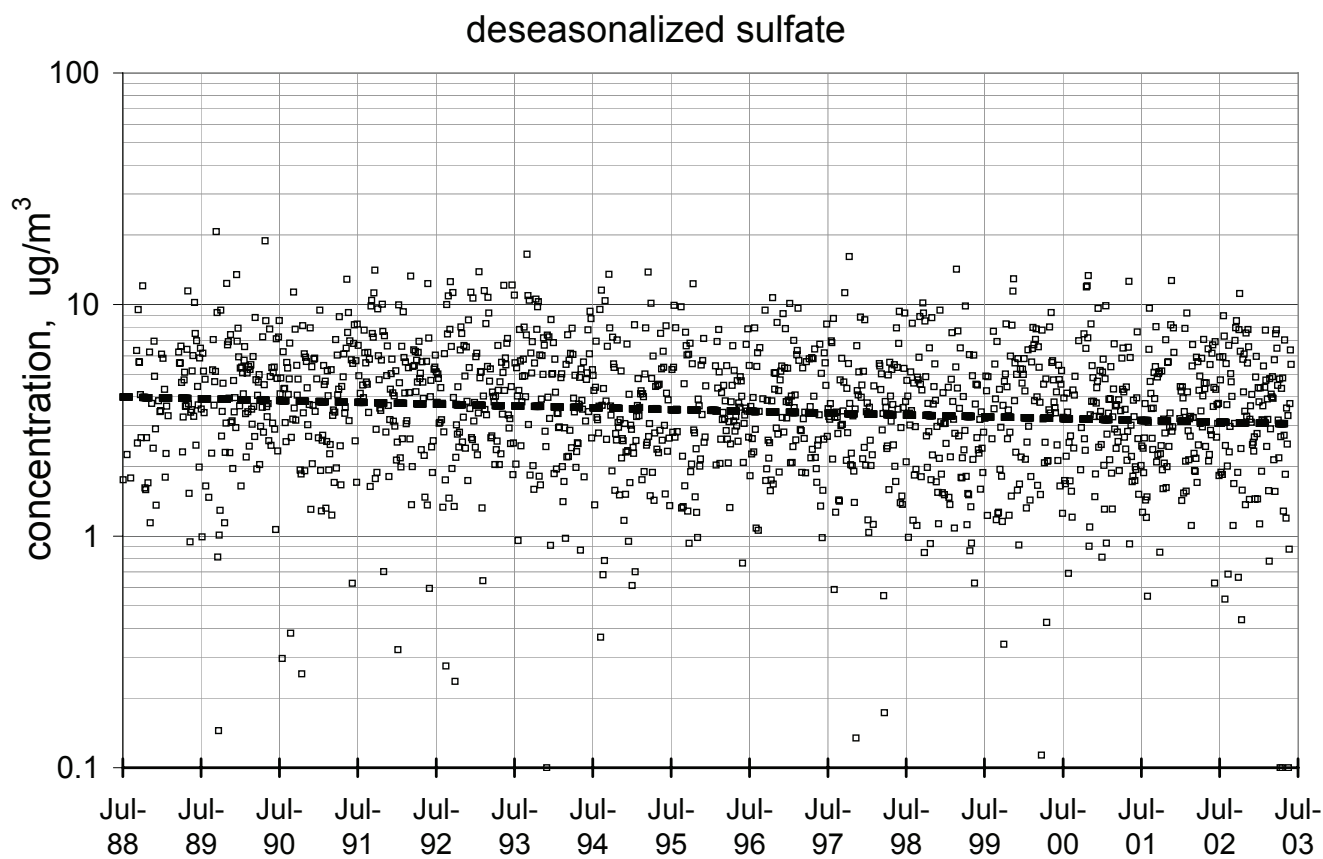
Figure 7: Deseasonalized IMPROVE sulfate concentrations at Shenandoah National Park. Individual points indicate 24h measurements of $[SO_4^=]$ before 19 April 1995 and of 3[S] afterward, as recommended by IMPROVE guidance.
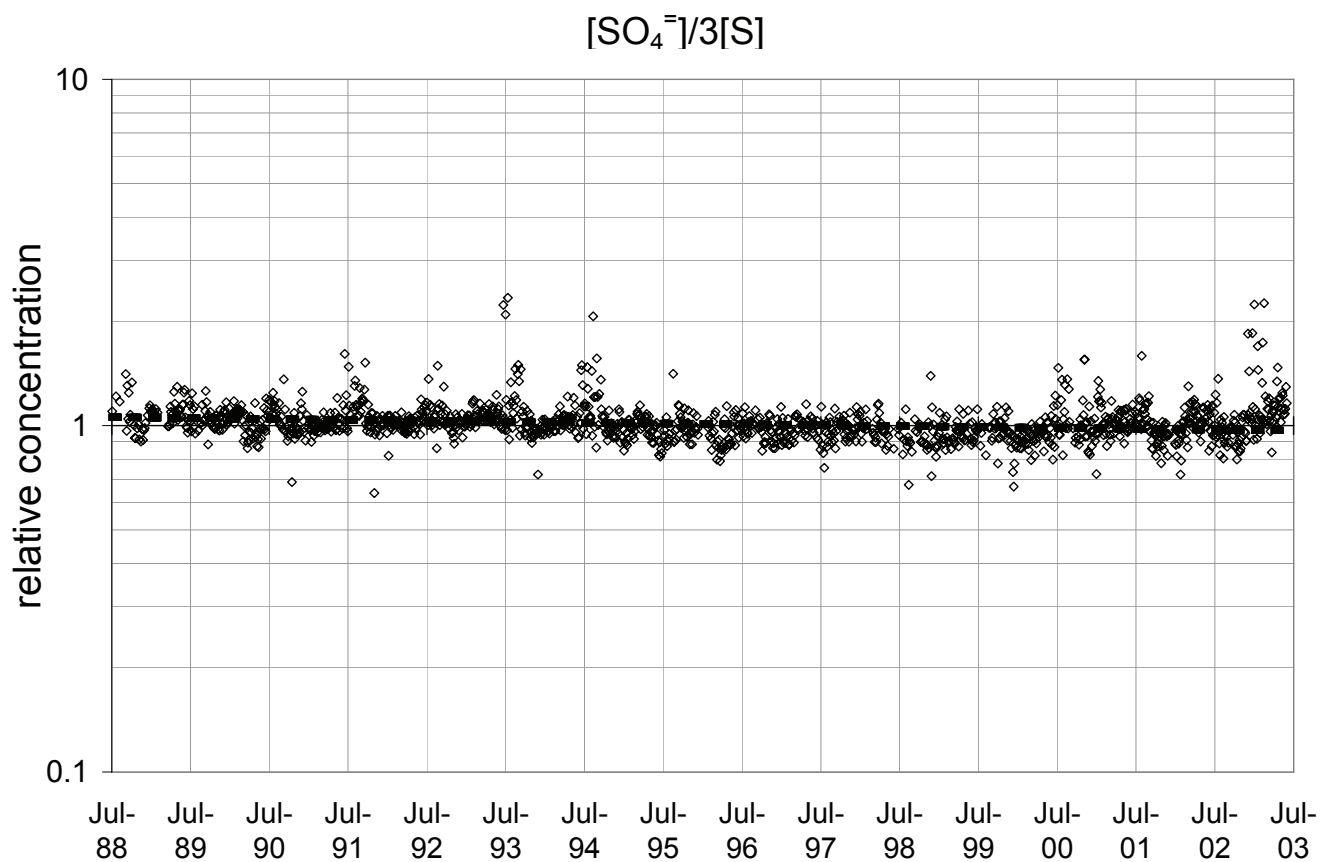
Figure 8: IMPROVE sulfate concentration ratios at Shenandoah National Park. Each point indicates a ratio $[SO_4^=]/3[S]$ of independent 24h measurements by two different methods.
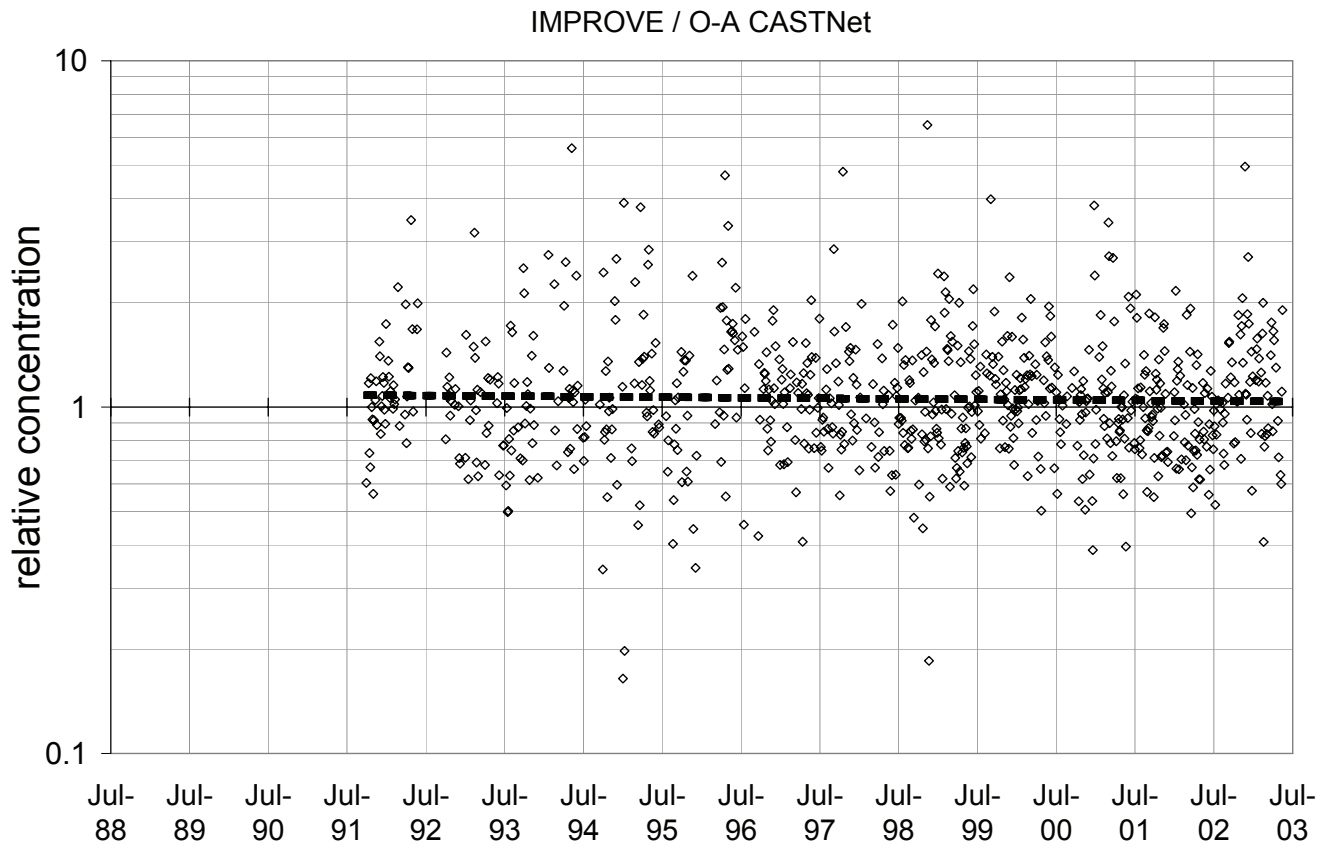
Figure 9: Ratios of recommended IMPROVE and optically-adjusted CASTNet values for 24h sulfate concentrations at Shenandoah National Park.