Environmental Science & Technology

ARTICLE

pubs.acs.org/est

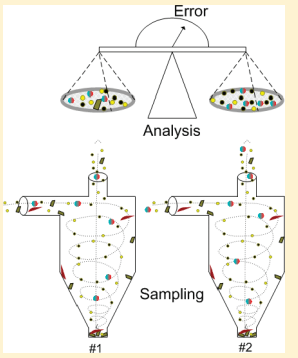
Identifying Sources of Uncertainty from the Inter-Species Covariance of Measurement Errors

Nicole P. Hyslop* and Warren H. White

Crocker Nuclear Laboratory, University of California, Davis, One Shields Ave, Davis, California 95616, United States

Supporting Information

ABSTRACT: A standard metric of measurement precision in environmental monitoring is the variance of differences between duplicate (collocated) samples. With duplicate measurements of multiple species, we can extend this variance analysis to include the interspecies covariance of differences between duplicate samples; these covariances can provide clues about the sources of error. We illustrate the potential of such an analysis with atmospheric aerosol measurements from two national air quality monitoring networks: Interagency Monitoring of Protected Visual Environments (IMPROVE) and Speciation Trends Network (STN). These aerosol "speciation" networks provide the multivariate data sets needed to characterize error covariance by operating duplicate samplers at several of their monitoring locations and analyzing both the collected aerosol samples for multiple species. We observe covariance among the measurement differences for multiple species in both networks. The covariance among measurement differences for soil-derived elements suggests an error associated with the particle size discrimination step in sampling, which is not currently included in either network's uncertainty estimates. The multivariate statistical analyses of aerosol speciation data performed by standard source apportionment models assume that measurement errors in different species are independent of each other; the present analysis invalidates this assumption for several species measured by IMPROVE and STN.



INTRODUCTION

Assessing measurement error is a nontrivial component of any serious measurement program. Identifying and assessing all possible sources of error makes heavy demands on human expertise and judgment; retrospective analyses of measurement uncertainty estimates for now well-established physical constants have shown a consistent bias toward underestimating actual errors.[1,2] Errors arise from all aspects of a measurement process including preparation, collection, analysis, and data reduction.[7] Errors in the individual steps of the measurement process compound into errors in the calculated quantity. Identifying the sources of error becomes more difficult as the number of steps in the process increases.

Many measurement programs use duplicate (collocated) measurements to quantify a portion of the total measurement uncertainty.[3−6] Duplicate measurements offer closure to the uncertainties predicted by propagating the uncertainties from individual measurement steps, much as a total mass measurement offers closure to a mass budget based on measurements of individual chemical species. Analyses of collocated sampling data from the Interagency Monitoring of Protected Visual Environments (IMPROVE) and Speciation Trends Network (STN) show that the observed measurement uncertainties are higher than predicted by the networks for several species.[4,5]

Errors in sample collection and data reduction can have a common effect on the values for all analytes reported from a given sample. Similarly, systemic analytical errors in techniques such as chromatography and spectrometry can affect multiple analytes in common. Other analytical errors may affect a subset of mutually interfering species or only an individual species, depending on the error mechanism. There can thus be multiple levels of common errors nested within a data set; each species measured in a sample is likely to have some unique and some shared components of error. Correlations among errors in different species may provide valuable information about the sources of error because an error that is correlated among different species presumably arises from a common source.

This paper extends the ideas underlying collocated-precision calculations to the multivariate context, developing a framework for the interpretation of correlations among errors in different species. The analysis is applied to atmospheric aerosol speciation measurements to make the ideas concrete but is easily adapted to a variety of other contexts. The analysis reveals significant interspecies correlations among the errors in routine IMPROVE and STN measurements. A simplified model is developed to distinguish the individual sources of error.

DATA

Collocated measurements from IMPROVE and STN are used in this analysis. Both networks measure concentrations of $PM_{2.5}$ (particulate matter with diameters less than 2.5 μm) by collecting 24-h filter samples every three days. The two networks use similar equipment and techniques. IMPROVE and STN collect $PM_{2.5}$

Received: August 13, 2010
Accepted: March 24, 2011
Revised: January 28, 2011
Published: April 11, 2011

on Teflon filters that are weighed for mass and analyzed by X-ray fluorescence (XRF) for most elements with atomic numbers between Na and Zr; IMPROVE also measures H by Proton Elastic Scattering (PESA). Precision estimates based on the collocated data have been published for both networks.[4−6]

IMPROVE is a cooperative measurement effort designed to characterize current visibility conditions in scenic areas (primarily National Parks and Forests). IMPROVE operates approximately 170 sites across the United States (US), seven of which have duplicate samplers. This analysis uses data collected from 2005 through June 2009 at the seven collocated sampling sites: Mesa Verde (MEVE) National Park (NP) in CO, Olympic NP (OLYM) in WA, Phoenix (PHOE) in AZ, Proctor Maple Research Facility (PMRF) in VT, Sac and Fox Nation (SAFO) in KS, Trapper Creek (TRCR) in AK, and Saint Marks (SAMA) in FL. The IMPROVE data were downloaded from http://vista.cira.colostate.edu/views/ in April 2010.

STN is designed to support the National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ and provides data on the chemical composition of $PM_{2.5}$.[14] STN operates 54 sites across the US, six of which have duplicate samplers. This analysis uses data collected from 2005 through January 2010 at the six collocated sampling sites: Bakersfield, CA; Riverside, CA; New Brunswick, NJ; Cleveland, OH; Boston, MA; and Houston, TX.[4] The STN data were downloaded from http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm/ in June 2010.

Many of the elements measured by XRF are present only at very low concentrations. To avoid the distractions introduced by missing data, this analysis is limited to nine species (eight elements and mass) that are commonly detected in IMPROVE and five species (four elements and mass) that are commonly detected in STN. Fewer elements are used for STN because its higher limits of detection translate into fewer detected concentrations.[15]

Two screening criteria were applied to the data. The first criterion is intended to minimize the uninformative noise associated with measurements below the detection limits; only data greater than the detection limit, defined as the critical limit, were used in the analysis.[15] For STN, the reported method detection limits were used, and for IMPROVE, the 95th percentile field blank concentrations were used to screen the data. Se and V are most affected by this criterion and strict adherence would eliminate an unacceptably high number of data records; therefore, instead of completely eliminating all records with data below the detection limit, concentrations were considered equal when both observations were below the detection limit (i.e., the collocated concentration ratios were set equal to one, $c_2/c_1 = 1$, where subscripts 1 and 2 distinguish the collocated samples). If the differences are biased, this treatment may affect the error estimates, but if the differences are unbiased this treatment will have no effect on the error estimates. The second data screening criterion focuses the analysis on the representative bulk of the observations; we do not want outliers to drive the statistics. For each species the collocated concentration ratios were screened to retain only those between the first and 99th percentiles, thereby eliminating the worst outliers. The cumulative effect of this criterion is to eliminate 5−10% of the measurement days, because the outlying ratios occur on different days for different species, and the covariance is only calculated for days when all selected species pass the screening criteria.

METHODS

Errors and Observed Differences. Uncertainties and errors are commonly given in relative terms, with the expectation that their magnitudes are generally proportional to the concentration. The relative differences observed between collocated measurements can be expressed in various ways. The U.S. Environmental Protection Agency regulatory guidance specifies the ratio of the difference to the mean as the basis for reporting collocated precision statistics.[16] A more tractable metric for the analyses undertaken here will be the natural log of the collocated concentration ratio, $\ln(c_2/c_1) \equiv \Delta$. The natural log transformation helps to normalize the distribution of ratios, which is skewed to the right because it is limited by a value of zero on the left. For simplicity, Δ will be referred to here as the measurement *difference*. Δ reflects the combined effect of errors in both of the collocated measurements. If the true concentration, C, is known, and e_i and ε_i are defined as the absolute and relative errors in c_i, respectively, the ratio would provide an estimate of the true relative error,

$$\ln\left(\frac{c_i}{C}\right) = \ln\left(\frac{C+e_i}{C}\right) = \ln(1+\varepsilon_i) \approx \varepsilon_i \qquad (1)$$

The small-error approximation $\ln(1 + \varepsilon) \cong \varepsilon$ is justified by previous analyses that showed relative errors (ε_i) ranging from 0.05 to 0.17, depending on the species.[4−6] Using approximation 1 to interpret the ratio of the two measurements yields the difference between the two measurements' errors

$$\Delta \equiv \ln\left(\frac{c_2}{c_1}\right) = \ln\left(\frac{c_2}{C}\frac{C}{c_1}\right) = \ln\left(\frac{c_2}{C}\right) - \ln\left(\frac{c_1}{C}\right)$$
$$\approx \varepsilon_2 - \varepsilon_1 \qquad (2)$$

Collocated differences do not capture the total error in an individual measurement because any errors that affect both samplers similarly escape detection. As shown in (2), the true concentrations cancel out in our formulation. This is an inherent limitation of using collocated precision determinations to determine error. For example, if the flow meter used to calibrate both samplers is biased, both samplers will have a flow rate bias that is not captured because the errors cancel out in the calculation of differences. Similarly, if both samples are analyzed on the same mis-calibrated laboratory instrument, the resulting analytical errors will cancel out. Collocated measurements only reveal differences between the two measurements and therefore understate total errors. This paper focuses on determining if there are common errors among the various species measured by each sampler; the errors common among different species will be revealed with our analysis as long as the errors are not also common to both samplers. For the remainder of this paper, "common errors" refer to errors that are common among elements and not to errors that are common between the collocated samplers, which cannot be detected by this analysis.

To begin the analysis, we visually inspect the relationships between the measurement differences, Δx, at all the collocated sites for a range of species x. Figure 1 shows a scatter plot matrix of Δx for five species measured in (a) IMPROVE and (b) STN. Figure 1 shows some weak relationships among the differences for most species in the IMPROVE network and some stronger relationships among Fe, Ca, and Si in both networks.

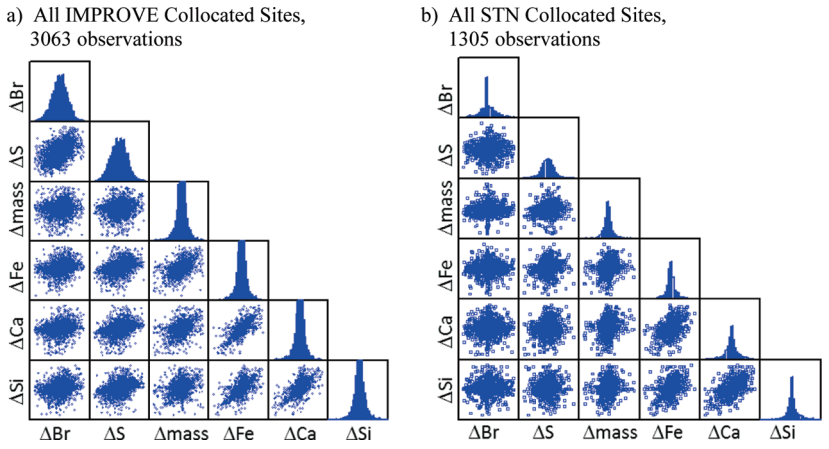


Figure 1. Scatterplot matrix of observed differences, $\Delta x \equiv \ln(c_{x,2}/c_{x,1})$ where x = Br, S, mass, Fe, Ca, and Si, from the (a) IMPROVE and (b) STN collocated measurements. Along the diagonal are histograms of the observed differences for each species. Off the diagonal are scatterplots of the observed differences in one species versus the observed differences for another species. For example a point in the lower left graph represents the observed difference in duplicate Si measurements versus the observed difference in the duplicate Br measurements from a single day.

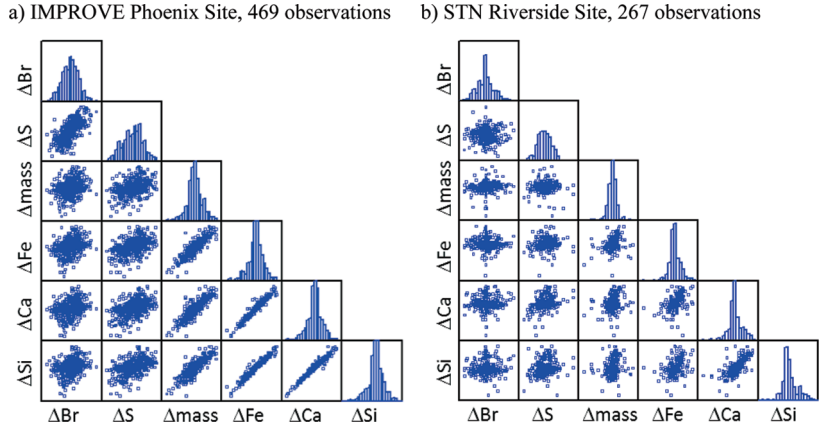


Figure 2. Scatterplot matrix of observed differences, $\Delta x \equiv \ln(c_{x,2}/c_{x,1})$, in the collocated concentrations at the (a) IMPROVE Phoenix site and (b) STN Riverside site.

It is important to bear in mind that Figure 1 plots concentration differences (our indication of measurement error) and not the concentrations themselves. We expect associations among different species' concentrations; stagnation promotes the accumulation of all species' concentrations, for example, just as ventilation and dilution lower all species' concentrations. However, measurement errors in different species' concentrations are usually modeled as random, unrelated to actual atmospheric conditions and independent of each other; this model is inconsistent with the interspecies associations seen between measurement differences in Figure 1.

The associations between different species' measurement differences are more pronounced at some individual sites; Figure 2 shows the scatter plot matrix for the (a) IMPROVE Phoenix and (b) STN Riverside sites. At these two sites, Fe, Ca, and Si differences show strong correlations and additionally correlate with mass concentration differences. The Phoenix site is very dusty and has high concentrations of soil-related elements, such as Fe, Ca, and Si, which constitute a large fraction of the mass. Once again, correlations among Fe, Ca, and Si concentrations are to be expected because these elements are all associated with soil dust, but such concentration associations should not lead to measurement error associations.

Next, we quantify the relationships to further understand their significance. The relationships in Figures 1 and 2 can be expressed with two common statistical quantities: covariances or correlation coefficients. Covariance is an absolute measure of the shared variation between two species, $\mathrm{cov}(\Delta x,\Delta y) = [\Sigma_{i=1}^{n}(\Delta y_i - \overline{\Delta y})(\Delta x_i - \overline{\Delta x})]/(n-1)$. The correlation coefficient, r, divides the covariance by the standard deviations of the two species to create a normalized measure of the shared variation between two species, $r = [\mathrm{cov}(\Delta x,\Delta y)]/[(s_{\Delta x}{}^2 \cdot s_{\Delta y}{}^2)^{1/2}]$, where $s_{\Delta y}{}^2 = [\Sigma_{i=1}^{n}(\Delta y_i - \overline{\Delta y})^2]/(n-1)$. Covariances are more useful for analysis, because the covariance contributions from independent sources of error are additive, but correlation coefficients are more familiar so they are presented first.

Table 1a presents the correlation matrix for the IMPROVE collocated sites, corresponding to Figure 1. The correlations show a couple interesting features that are straightforward to interpret. The most obvious of these, already evident in Figure 1, is that the dust elements (Fe, Ca, and Si) have higher correlation coefficients than any other species in Table 1a. The weakest correlations are between ΔBr and Δmass (0.096) and ΔS and Δmass (0.144); all species are expected to share an error related to slight differences in sample volume between the two samplers,

Table 1. Correlation, Covariance, and Relative Error Matrices of Measurement Differences for All IMPROVE Collocated Sites

(a) correlation matrix

	ΔBr	ΔS	Δmass	ΔFe	ΔCa	ΔSi
ΔBr	1.000					
ΔS	0.523	1.000				
Δmass	0.096	0.144	1.000			
ΔFe	0.261	0.378	0.470	1.000		
ΔCa	0.236	0.347	0.421	0.758	1.000	
ΔSi	0.243	0.347	0.328	0.663	0.679	1.000

(b) covariance matrix

	ΔBr	ΔS	Δmass	ΔFe	ΔCa	ΔSi
ΔBr	0.008					
ΔS	0.002	0.003				
Δmass	0.001	0.001	0.006			
ΔFe	0.003	0.003	0.005	0.020		
ΔCa	0.003	0.003	0.005	0.017	0.026	
ΔSi	0.004	0.003	0.004	0.017	0.019	0.031

(c) shared relative error matrix

	E_{Br}	E_S	E_{mass}	E_{Fe}	E_{Ca}	E_{Si}
E_{Br}	6%					
E_S	3%	4%				
E_{mass}	2%	2%	5%			
E_{Fe}	4%	4%	5%	10%		
E_{Ca}	4%	4%	5%	9%	11%	
E_{Si}	4%	4%	5%	9%	10%	12%

and these weak correlations are likely indicative of the volume error.

Alternatively, covariances can be used to quantify the relationships in Figures 1 and 2. Recall that $\Delta x \cong \varepsilon_{x,2} - \varepsilon_{x,1}$ includes two measurement errors, one error associated with the measurement from each sampler and collocated measurements of the same species overlook any errors shared by both measurements. It is reasonable to assume that the unshared errors in the collocated measurements are statistically independent of each other, so that $\mathrm{cov}(\varepsilon_{x,1},\varepsilon_{x,2}) = \mathrm{cov}(\varepsilon_{y,1},\varepsilon_{y,2}) = 0$, and are drawn from a common distribution of potential errors. With these assumptions, the covariance between any two species' measurement differences simplifies to

$$\mathrm{cov}(\Delta x, \Delta y) = \mathrm{cov}(\varepsilon_{x,2} - \varepsilon_{x,1}, \varepsilon_{y,2} - \varepsilon_{y,1}) \cong \mathrm{cov}(\varepsilon_{x,1}, \varepsilon_{y,1}) + \mathrm{cov}(\varepsilon_{x,2}, \varepsilon_{y,2}) \cong 2\mathrm{cov}(\varepsilon_x, \varepsilon_y) \quad (3)$$

Table 1b shows the covariance matrix for all the IMPROVE sites; since $\mathrm{cov}(\Delta x,\Delta x) \cong 2\,\mathrm{cov}(\varepsilon_x,\varepsilon_x) = 2\mathrm{var}(\varepsilon_x)$, the diagonal entries provide estimates of the total error variance for each species.

The physical, as opposed to statistical, significance of the relationships between the observed differences is hard to judge directly from the covariance matrix, whose dimensions are squared differences. Although these are the dimensions in which individual errors are additive, it is in terms of the relative errors themselves that most of us maintain our expectations. The square root of the variance (standard deviation) is a widely used measure of error. The next section will illustrate that the covariances can be interpreted as the variances of shared errors. Table 1c shows the relative error estimates obtained by taking the square roots of one-half the observed covariances, $E_{obs} = \{[(\mathrm{cov}(\Delta x,\Delta y))/(2)]^{1/2}\} \times 100\%$. The same patterns are observed in this relative error matrix as in the correlation and covariance matrices but the numbers are more familiar.

MODEL

Both IMPROVE and STN have existing models for the error distributions but the models are not well suited to the present analysis. A simple model of the error in individual measurements involving only three sources of error will be developed here to help interpret the correlations among the measurement differences. In keeping with our characterization of error in relative terms, the effect of each error is multiplicative. In general, some errors are proportional to the magnitude of the concentration while other errors are absolute or independent of the measured concentration.[5] For example, calibrations typically include slope (span) and offset (zero) terms; error in the slope results in a multiplicative error whereas error in the intercept results in an absolute error. At high concentrations, the absolute errors become relatively insignificant and the multiplicative errors dominate the overall error. As mentioned above, we excluded concentrations below the detection limit, where absolute errors dominate, from this analysis.

The measurement of $PM_{2.5}$ species concentrations can be conceptualized as a series of three operations, each subject to error. First a volume of air is pulled through a cyclone inlet designed to remove particles with aerodynamic diameters above 2.5 μm from the air stream; errors in this operation result in more or less particulate matter being collected on the filter than intended. Next, the mass of species collected on the filter is determined, and there is error associated with this analysis. Lastly, the species concentration in the sampled air is calculated by dividing the mass by the sample volume, which is measured with some error. The following notation will facilitate the dissection of the total concentration measurement error into its component parts: x, species; V, true volume of sampled air; v, volume measured for sampled air; L_x, true mass of particulate x in the sampled air; $L_{x,2.5} = F_x L_x$, true mass of $PM_{2.5}$ x in the sampled air; $\Lambda_x = f_x L_x$, true mass of x collected on the filter; λ_x, mass of x measured on the filter.

In the above terms, the measured and true concentrations for $PM_{2.5}$ x in the sampled air can be written as $c_x = \lambda_x/v$ (measured) and $C_{x,2.5} = L_{x,2.5}/V$ (true). The relative error in the measured concentration is then $c_x/C_{x,2.5} = (\lambda_x/L_{x,2.5})(V/v)$. Note that $L_{x,2.5} = F_x L_x = (F_x(\Lambda_x/f_x))$, so the factor $\lambda_x/L_{x,2.5} = (\lambda_x/\Lambda_x)(f_x/F_x)$ is the product of the relative errors (λ_x/Λ_x) in the analytical determination and (f_x/F_x) in the particle size classification. The total concentration error is thus the product of three multiplicative errors: $c_x/C_{x,2.5} = (\lambda_x/\Lambda_x)(f_x/F_x)(V/v)$. Logarithms convert this decomposition into a sum, which lends itself more directly to classical statistical manipulations, and the small-error approximation $\ln(\theta) \cong \theta - 1$ once again allows these log ratios to be interpreted directly in terms of relative error

$$\varepsilon_x \cong \ln\left(\frac{c_x}{C_{x,2.5}}\right) = \ln\left(\frac{\lambda_x}{\Lambda_x}\right) + \ln\left(\frac{f_x}{F_x}\right) + \ln\left(\frac{V}{v}\right) \cong \varepsilon_{an,x} + \varepsilon_{size,x} + \varepsilon_{vol}. \quad (4)$$

Volume errors (ε_{vol}) result primarily from errors in the sample flow rate measurements. Sample flow rate measurement errors equally affect every species measured on a sample because volume is used to calculate all species concentrations. Departures of flow rate from nominal do affect size discrimination, but uncertainties in measuring flow do not; non-nominal flows can be accurately measured and nominal flows can be inaccurately measured. The volume error, ε_{vol}, will accordingly be modeled as a random variable that is statistically independent of all other uncertainties.

Table 2. Covariance Matrix for Two Generic Haze Species *a,b* and Two Generic Dust Species *c,d* with Each Covariance Broken down into Its Model Components

		HAZE SPECIES		DUST SPECIES	
		Δ_a	Δ_b	Δ_c	Δ_d
HAZE	Δ_a	$2(\mathrm{var}(\varepsilon_{an,a})+\mathrm{var}(\varepsilon_{vol}))$			
	Δ_b	$2\,\mathrm{var}(\varepsilon_{vol})$	$2(\mathrm{var}(\varepsilon_{an,b})+\mathrm{var}(\varepsilon_{vol}))$		
DUST	Δ_c	$2\,\mathrm{var}(\varepsilon_{vol})$	$2\,\mathrm{var}(\varepsilon_{vol})$	$2(\mathrm{var}(\varepsilon_{an,c})+\mathrm{var}(\varepsilon_{vol})+\mathrm{var}(\varepsilon_{size}))$	
	Δ_d	$2\,\mathrm{var}(\varepsilon_{vol})$	$2\,\mathrm{var}(\varepsilon_{vol})$	$2(\mathrm{var}(\varepsilon_{vol})+\mathrm{var}(\varepsilon_{size}))$	$2(\mathrm{var}(\varepsilon_{an,d})+\mathrm{var}(\varepsilon_{vol})+\mathrm{var}(\varepsilon_{size}))$

For each species measured, there is error associated with the analytical technique. The errors in elements determined by XRF are assumed to be statistically independent although there is evidence, outside the scope of this paper, that spectral interferences cause some dependencies. In addition, we assume the analytical errors are causally unrelated to particle size discrimination errors although it is known that X-ray self-attenuation is a function of particle size. The relative analytical error, $\varepsilon_{an,x} = \ln(\lambda_x/\Lambda_x)$, will be modeled as an independent random variable.

The IMPROVE and STN samplers employ cyclone separators designed to trap coarse particles and pass particles with diameters less than 2.5 μm through to the sample filter. Size discrimination errors ε_{size} can be caused by flow rates that depart from nominal, particles breaking apart in the cyclone, or particles bouncing off the sides of the cyclone. The IMPROVE samplers use a critical orifice to set the flow rate; this passive control allows flow rates to decrease as the filter loads with particulate matter, and the decreased flow rates allow slightly larger particle diameters to survive the cyclone . The STN samplers have active flow control which minimizes deviations from the nominal flow rate. The magnitude of ε_{size} is affected by the ambient particle-size distribution; this error will be more obvious at sites heavily influenced by coarse particles. Like the other errors, ε_{size} will be modeled as an independent random variable.

The size discrimination error does not affect all particles. Consider an idealized mix of "haze" particles supplying all of one group of elements and "dust" particles supplying all of another. All haze particles are assumed small enough that the sampling of haze elements is insensitive to variations in the cyclone cut-point. The dust particles are allowed to vary in composition and size from sample to sample, but all dust particles in a given sample are assumed to have the same elemental composition. Under these assumptions, the error resulting from cut-point variations is identically zero for haze elements in all samples and takes a common value $\varepsilon_{size} \equiv \ln(f/F)$ for all dust elements in any individual sample. Total $PM_{2.5}$ mass includes both haze and dust particles, and the impact of the size discrimination error on mass depends on the mix.

The model for the total error is the sum of three independent random variables, $\varepsilon_x = \varepsilon_{an,x} + \varepsilon_{vol} + \varepsilon_{size}$. As mentioned previously, if the individual sources of error affecting a species are independent, the individual error variances are additive. Similarly the error covariance of a pair of species is the sum of their shared-error variances as illustrated in Table 2 for haze elements *a,b* and dust elements *c,d*. Fitting the model described above to this matrix provides estimates of the total error variances for each species and, by comparison of off-diagonal terms, estimates of $\mathrm{var}(\varepsilon_{vol})$, $\mathrm{var}(\varepsilon_{size})$, and $\mathrm{var}(\varepsilon_{an,x})$.

RESULTS

The interpretation of a covariance matrix in terms of error components is illustrated with site-specific data from collocated IMPROVE and STN samplers. The elements in this analysis can be grouped by particle size under three categories: "haze", "dust" and "both" haze and dust. Br, H, S, V, and Se tend to be associated with particles in the <1 μm size range and are categorized as "haze".[17−19] In the absence of industrial sources, the majority of Fe, Ca, Si, and Ti mass arises from soil dust in the >1 μm size range, and these elements are categorized as "dust".[17,18] Total $PM_{2.5}$ mass can contain both "haze" and "dust" elements, and so is categorized as "both".

Table 3 shows the total and shared error estimates for each IMPROVE collocated site. Uncertainty ranges are omitted from Table 3 for visual clarity but are provided in the Supporting Information. Results for the TRCR site are not included because it has a relatively small number of records meeting the detection limit screening criteria (138 records). Generally, the median uncertainty in the error estimates is ±0.4%; a small number of the error estimates (less than 10%) have uncertainties larger than ±1% and most of these high uncertainties are associated with errors in V attributed to measurements near the limit of detection. Dashes in cells indicate instances where the square root is undefined because the covariance is negative. These negative values predominantly occur with the trace elements V and Se.

The diagonal terms in Table 3, IMPROVE total error estimates, show interesting patterns by site. First, Br, S, and H are well-measured haze elements at all the sites with median total errors of 6%, 3%, and 7%, respectively. V and Se error estimates show more variation by site. V total error is lowest at OLYM, because OLYM experiences relatively high V concentrations, likely from oil burning in the ocean shipping industry,[20−22] and the (relative) analytical precision is better at these higher concentrations.[5,6] The differences in Se total error are likely related to the range of concentrations measured at the individual sites. The behavior of the total error estimates for mass and the dust elements splits the sites into two groups; PHOE and MEVE total errors are twice as high as the other sites for these four species. PHOE and MEVE are situated in dry, dusty environments so the high errors are associated with high concentrations; this suggests that the errors are not analytical, because (relative) analytical errors tend to decrease with increasing concentrations.

Table 3. Estimates of Total (Diagonal) and Shared (Off-Diagonal) Errors for Each IMPROVE Collocated Monitoring Site, E_{obs} = $[(\mathrm{cov}\,(\Delta x,\Delta y))/2]^{1/2} \times 100\%$[a]

		Haze					Both	Dust		
Site	Species	E_{Br}	E_S	E_H	E_V	E_{Se}	E_{Mass}	E_{Fe}	E_{Ca}	E_{Ti}
Mesa Verde National Park, Colorado (MEVE), 356 data records	E_{Br}	6%								
	E_S	3%	4%							
	E_H	3%	2%	7%						
	E_V	-	2%	-	26%					
	E_{Se}	5%	3%	-	-	13%				
	E_{mass}	3%	2%	2%	3%	4%	7%			
	E_{Fe}	5%	5%	5%	5%	6%	8%	13%		
	E_{Ca}	5%	5%	6%	4%	6%	8%	14%	16%	
	E_{Ti}	5%	4%	5%	6%	5%	8%	13%	14%	15%
Phoenix, Arizona (PHOE), 449 data records	E_{Br}	4%								
	E_S	4%	4%							
	E_H	3%	4%	8%						
	E_V	4%	4%	6%	15%					
	E_{Se}	3%	2%	-	-	14%				
	E_{mass}	3%	3%	5%	8%	-	7%			
	E_{Fe}	4%	5%	6%	11%	-	9%	13%		
	E_{Ca}	4%	5%	7%	12%	-	9%	14%	15%	
	E_{Ti}	4%	5%	7%	12%	-	9%	13%	14%	14%
Olympic National Park, Washington (OLYM), 348 data records	E_{Br}	5%								
	E_S	3%	3%							
	E_H	2%	2%	8%						
	E_V	3%	3%	2%	9%					
	E_{Se}	5%	3%	3%	-	16%				
	E_{mass}	1%	1%	2%	1%	1%	4%			
	E_{Fe}	3%	3%	2%	4%	4%	3%	8%		
	E_{Ca}	4%	3%	3%	3%	4%	2%	6%	8%	
	E_{Ti}	4%	4%	3%	3%	7%	2%	8%	8%	16%
Proctor Maple Research Facility, Vermont (PMRF), 389 data records	E_{Br}	6%								
	E_S	3%	3%							
	E_H	3%	2%	7%						
	E_V	-	3%	-	19%					
	E_{Se}	5%	3%	3%	1%	12%				
	E_{mass}	-	1%	1%	2%	0%	4%			
	E_{Fe}	4%	3%	2%	4%	2%	2%	7%		
	E_{Ca}	3%	3%	3%	1%	3%	2%	5%	7%	
	E_{Ti}	4%	3%	3%	-	2%	3%	6%	6%	14%
Sac and Fox Nation, Kansas, (SAFO), 339 data records	E_{Br}	6%								
	E_S	3%	4%							
	E_H	3%	3%	7%						
	E_V	2%	3%	4%	27%					
	E_{Se}	4%	3%	3%	3%	7%				
	E_{mass}	1%	1%	2%	1%	1%	3%			
	E_{Fe}	3%	3%	3%	5%	4%	3%	7%		
	E_{Ca}	3%	3%	4%	6%	4%	3%	7%	8%	
	E_{Ti}	4%	4%	4%	2%	3%	2%	6%	7%	9%
St. Marks, Florida (SAMA), 417 data records	E_{Br}	6%								
	E_S	3%	3%							
	E_H	3%	3%	7%						
	E_V	2%	3%	3%	14%					
	E_{Se}	4%	3%	3%	3%	8%				
	E_{mass}	2%	1%	2%	-	1%	3%			
	E_{Fe}	3%	3%	3%	3%	3%	2%	7%		
	E_{Ca}	3%	3%	2%	3%	3%	1%	5%	7%	
	E_{Ti}	4%	4%	4%	3%	3%	2%	6%	5%	11%

[a] Uncertainty ranges are provided in the Supporting Information.

The off-diagonal terms in Table 3, IMPROVE shared error estimates, vary across species indicating that there are multiple sources of error affecting the measurements, as described by the model. The off-diagonal terms in the lower right corner, "dust" components vs each other, are higher than in the rest of the matrix at each site; our model suggests these high shared errors are associated with a size discrimination error. In fact, the shared errors account for essentially all the total error for each dust element. As with the total error estimates, the magnitude of the shared relative errors among the dust elements and mass varies by site and the highest shared errors are associated with PHOE and MEVE. The high shared errors among the soil elements at these two sites partially extends to the haze elements, suggesting that these haze elements are sometimes found in particles with diameters >1 μm.[19,23−27] The smallest shared errors are among the haze elements and mass. The covariance model in Table 2 suggests attributing a minimum, or baseline, shared error to the volume measurement, whose uncertainty has been independently assessed at 3%. The shared errors are less uniform than this would suggest, however; at four sites (OLYM, PMRF, SAFO, and SAMA), the errors shared between mass and haze elements are smaller than the errors shared among different haze elements. This suggests the possibility of an error common to the element measurements that does not affect the mass measurements, perhaps a nonuniformity in the sample deposit. Nonuniform deposits would not affect the mass measurement because the entire filter is analyzed, but could cause a shared error in the XRF and PESA measurements because only a portion of the filter is exposed to the exciting beam of X-rays or protons.

Table 4 shows the total and shared error estimates for the three STN collocated sites that have over 200 data records meeting the screening criteria. Results for all six STN sites along with uncertainty ranges are included in the Supporting Information. The STN total error estimates are generally higher than the IMPROVE estimates for the same species, particularly for Br. The S and mass total errors are similar at all the sites, whereas the dust elements show more variation by site. Similar to IMPROVE, the dust elements have higher shared error estimates than the other species. Unlike IMPROVE, the shared errors between mass and the haze elements do not appear to be any different from the shared errors among haze elements. The STN XRF instruments compensate for possible deposit inhomogeneity by spinning the filters during the analysis.

DISCUSSION

The preceding analysis has revealed a major source of measurement error that is not currently addressed by the uncertainty models for either IMPROVE or STN. Measurements in both networks exhibit errors shared across the soil elements that are predominantly associated with particle diameters >1 μm. The interspecies correlation of these errors suggests a common error

Table 4. Estimates of Total (Diagonal) and Shared (Off-Diagonal) Errors for Each STN Collocated Monitoring Site, $E_{obs} = [(\mathrm{cov}\,(\Delta x,\Delta y))/2]^{1/2} \times 100\%$[a]

		Haze		Both	Dust		
Site & Sampler		E_{Br}	E_S	E_{Mass}	E_{Fe}	E_{Ca}	E_{Si}
Bakersfield, California 258 data records	E_{Br}	22%					
	E_S	3%	5%				
	E_{mass}	-	3%	12%			
	E_{Fe}	2%	5%	6%	19%		
	E_{Ca}	5%	6%	7%	12%	20%	
	E_{Si}	-	4%	5%	12%	14%	18%
Cleveland, Ohio 226 data records	E_{Br}	21%					
	E_S	3%	7%				
	E_{mass}	3%	4%	12%			
	E_{Fe}	4%	5%	7%	16%		
	E_{Ca}	5%	4%	9%	15%	26%	
	E_{Si}	6%	1%	8%	11%	19%	24%
Riverside, California 267 data records	E_{Br}	21%					
	E_S	3%	5%				
	E_{mass}	5%	0%	10%			
	E_{Fe}	-	4%	7%	18%		
	E_{Ca}	6%	6%	7%	12%	19%	
	E_{Si}	5%	4%	6%	11%	15%	18%

[a] Uncertainty ranges are provided in the Supporting Information.

in the misclassification of particle size. At sites with high soil dust concentrations, the apparently incomplete exclusion of large particles affects measurements of mass; at particularly dusty sites it also contributes to error in species typically associated with smaller particles, such as V, Se, H, S, and Br. The current IMPROVE and STN uncertainty estimates do not include a term representing size discrimination error; the strong correlations among the errors in the dust-related elements suggest that a size discrimination error term should be added to the uncertainty models.

Concentration associations are exploited in multivariate source apportionment approaches, where the ambient sample is represented mathematically as a varying mix of chemically distinct effluents from a collection of source types with stable chemical compositions. This interpretation assumes that measurement errors, being random and hence independent of each other, can only obscure such associations and can never artificially inflate them, but this analysis shows that measurement errors can artificially inflate the associations. Standard models for such analyses allow users to input estimated measurement uncertainties in individual species concentrations, but neglect the possibility of correlations between errors in different species.[8,9] Tests on real and simulated data show a complex dependence of the results on both real and estimated uncertainties.[10−12] Covarying measurement errors are particularly relevant for factor-analytical approaches, which infer the chemical profiles of sources from the observed covariance structure of species concentrations.[9,13]

Considerable attention has been given in the literature on source apportionment and other applications to the effects on multivariate statistical analyses of measurement errors in speciated aerosol data. Positive matrix factorization was from the beginning developed to accommodate heteroscedastic errors, by allowing for the specification of sample-specific measurement uncertainties.[9] A number of recent studies have examined the complications introduced by serial correlation of measurement errors.[13] The likely effects of the correlations examined here, between measurement errors in different species, have by contrast largely escaped such attention. The only published consideration we have found is in Christensen and Gunst,[10] who included interspecies correlations in some of the simulated data they generated to test versions of the chemical mass balance approach, none of which accounted for these correlations. Their simulated correlations were $r = 0$ or 0.3 across all species, well below some of the pairwise values exhibited in Table 1. A detailed analysis of the implications of covarying errors for source apportionment is beyond the scope of this paper, but our results show this to be a clear need.

■ ASSOCIATED CONTENT

Supporting Information. Uncertainties for the total and shared error estimates presented in Tables 3 and 4. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: nmhyslop@ucdavis.edu.

■ ACKNOWLEDGMENT

This work was supported by the United States National Park Service Contract C2350-04-0050. The authors thank Tony Wexler and two anonymous reviewers for their input.

■ REFERENCES

(1) Henrion, M.; Fischhoff, B. Assessing uncertainty in physical constants. *Am. J. Phys.* **1986**, *54* (9), 791–798.

(2) Chen, G.; Gott, J. R.; Ratra, B. Non-Gaussian error distribution of Hubble constant measurements. *Publ. Astron. Soc. Pac.* **2003**, *115* (813), 1269–1279.

(3) Dutton, S. J.; Schauer, J. J.; Vedal, S.; Hannigan, M. P. $PM_{2.5}$ characterization for time series studies: Pointwise uncertainty estimation and bulk speciation methods applied in Denver. *Atmos. Environ.* **2009**, *43* (5), 1136–1146.

(4) Flanagan, J. B.; Jayanty, R. K. M.; Rickman, E. E., Jr.; Peterson, M. R. $PM_{2.5}$ speciation trends network: Evaluation of whole-system uncertainties using data from sites with collocated samplers. *Air Waste Manage. Assoc.* **2006**, *56*, 492–499.

(5) Hyslop, N. P.; White, W. H. An evaluation of interagency monitoring of protected visual environments (IMPROVE) collocated precision and uncertainty estimates. *Atmos. Environ.* **2008**, *42* (11), 2691–2705.

(6) Hyslop, N. P.; White, W. H. Estimating precision using duplicate measurements. *Air Waste Manage. Assoc.* **2009**, *59*, 1032–1039.

(7) International Organization for Standardization (ISO). Guide to the Expression of Uncertainty in Measurement. 1995, Switzerland.

(8) Watson, J. G.; Cooper, J. A.; Huntzicker, J. J. The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos. Environ.* **1984**, *18* (7), 1347–1355.

(9) Paatero, P.; Tapper, U. Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126.

(10) Christensen, W. F.; Gunst, R. F. Measurement error models in chemical mass balance analysis of air quality data. *Atmos. Environ.* **2004**, *38*, 733–744.

(11) Christensen, W. F.; Schauer, J. J. Impact of species uncertainty perturbation on the solution stability of positive matrix factorization of

atmospheric particulate matter data. *Environ. Sci. Technol.* **2008**, *42*, 6015–6021.

(12) White, W. H.; Macias, E. S. On measurement error and the empirical relationship of atmospheric extinction to aerosol composition in the non-urban west. In *Visibility Protection*, An APCA International Specialty Conference,Bhardwaja, P. S., Ed.; Air & Waste Management Association: Pittsburg, PA, 1987.

(13) Pollice, A. Recent statistical issues in multivariate receptor models. *Environmetrics* **2009**10.1002/env.

(14) $PM_{2.5}$ and Chemical Speciation General Information. http://www.epa.gov/ttn/amtic/specgen.html.

(15) Hyslop, N. P.; White, W. H. An empirical approach to estimating detection limits using collocated data. *Environ. Sci. Technol.* **2008**, *42* (14), 5235–5240.

(16) Code of Federal Regulations (CFR). Revised Requirements for Designation of Reference and Equivalent Methods for PM2.5 and Ambient Air Quality Surveillance for Particulate Matter; Final Rule, 1997. Code of Federal Regulations Part IV Environmental Protection Agency 40 CFR Parts 53 and 58. 1997, http://www.epa.gov/ttn/oarpg/t1/fr_notices/pm-mon.pdf, pp 71−72.

(17) Seinfeld, J. H.; Pandis, S. N. *Atmospheric Chemistry and Physics*, 2nd ed.; John Wiley & Sons: New York, 2006.

(18) Finlayson-Pitts, B. J. Pitts, J. N. *Chemistry of the Upper and Lower Atmosphere*; Academic Press: New York, 2000.

(19) Pacyna, J. M.; Source inventories for atmospheric trace metals. In *Atmospheric Particles*, IUPAC Series on Analytical and Physical Chemistry of Environmental Systems, Vol. 5; Harrison, R. M., van Grieken, R. E., Eds.; Wiley: Chichester, U.K., 1998.

(20) Ault, A. P.; Gaston, C. J.; Wang, Y.; Dominguez, G.; Thiemens, M. H.; Prather, K. A. Characterization of the Single Particle Mixing State of Individual Ship Plume Events Measured at the Port of Los Angeles. *Environ. Sci. Technol.* **2010**, *44* (6), 1954–1961.

(21) Agrawal, H.; Malloy, Q. G. J.; Welch, W. A.; Miller, J. W.; Cocker, D. R., III In-use gaseous and particulate matter emissions from a modern ocean going container vessel. *Atmos. Environ.* **2008**, *42* (21), 5504–551010.1016/j.atmosenv.2008.02.053.

(22) Isakson, J.; Persson, T. A.; Selin Lindgren, E. Identification and assessment of ship emissions and their effects in the harbour of Goteborg, Sweden. *Atmos. Environ.* **2001**, *35* (21), 3659–366610.1016/S1352-2310(00)00528-8.

(23) Bukowiecki, N; Hill, M; Gehrig, R.; Zwicky, C. N.; Lienemann, P.; Hegedüs, F.; Falkenberg, G.; Weingartner, E.; Baltensperger, U. Trace metals in ambient air: Hourly size-segregated mass concentrations determined by synchrotron-XRF. *Environ. Sci. Technol.* **2005**, *39* (15), 5754–5762.

(24) Dundar, M. F. Vanadium concentrations in settled outdoor dust particles. *Environ. Monit. Assess.* **2006**, *123*, 345–350.

(25) Salma, I.; Maenhaut, W.; Zaray, G. Comparative study of elemental mass size distributions in urban atmospheric aerosol. *J. Aerosol Sci.* **2002**, *33*, 339–356.

(26) Utsunomiya, S.; Jensen, K. A.; Keeler, G. J.; Ewing, R. C. Direct identification of trace metals in fine and ultrafine particles in the detroit urban atmosphere. *Environ. Sci. Technol.* **2004**, *38* (8), 2289–2297.

(27) Narita, Y.; Tanaka, S.; Santoa, S. J. A study on the concentration, distribution, and behavior of metals in atmospheric particulate matter over the North Pacific Ocean by using inductively coupled plasma mass spectrometry equipped with laser ablation. *J. Geophys. Res.* **1999**, *204* (D21), 26859–26866.