APPENDIX E: AN ASSESSMENT OF MEASUREMENT ERRORS IN THE IMPROVE AND STN NETWORKS FROM IN-NETWORK AND CROSS-NETWORK COLLOCATED DATA AND THE ESTIMATED COMPARABILITY OF DATA COLLECTED FROM THE TWO NETWORKS

INTRODUCTION

The IMPROVE network has ~160 sites located primarily in remote rural areas where speciated fine aerosol mass concentrations are monitored. The $PM_{2.5}$ chemical Speciation Trends Network (STN) also monitors speciated fine aerosol mass concentrations at ~300 sites located primarily in urban and suburban areas. The collocation of IMPROVE and STN sites in select urban and rural locations allows for analysis of the intercomparability between the two monitoring networks.

The STN was established to provide nationally consistent data for the urban and suburban areas of the United States for the assessment of trends and to provide a basic, long-term record of the constituents of $PM_{2.5}$ in urban areas in support of the $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS) [EPA speciation guidance, 1999]. The STN has a core set of 54 National Ambient Monitoring Stations, 6 of which have collocated samplers, and an additional ~250 State and Local Air Monitoring Stations. The STN was established in a manner such that it has similar measurements to those collected by IMPROVE—both networks collect 24-hour samples on appropriate filter media on a 1-in-3-day sampling schedule for quantifying $PM_{2.5}$ mass and its chemical constituents. Measurements from collocated IMPROVE and STN sites are expected to be the same within combined measurement uncertainty for most parameters.

Integrating data from the two networks provides more complete information on the spatial and temporal distributions of $PM_{2.5}$ aerosol mass and its major constituents throughout the United States. However, to do this meaningfully requires an understanding of the comparability of the data collected from the two networks. The in-network collocated data from IMPROVE and STN, as well as the cross-network collocated data, were explored to develop estimates on the comparability of multiyear mean concentrations between sites from the two networks. The IMPROVE in-network collocated data were examined to 1) determine if the observed relative measurement errors were consistent with idealized random errors and 2) establish the typical uncertainty in mean concentrations calculated from the collocated IMPROVE measurements. The discussion of the in-network IMPROVE collocated data was included to provide context for analysis of the cross-network collocated data. The cross-network IMPROVE and STN collocated data were explored with the same objectives as the in-network IMPROVE collocated data, with appropriate statistical modifications to account for the expected differences in measurement errors between the two networks.

Finally, the characteristic uncertainties observed in the in-network IMPROVE and in-network STN collocated data were used to calculate the expected characteristic uncertainties in the cross-network collocated data. The objective of this step was to evaluate the observed cross-network measurement uncertainty in context of what we would expect if the only information we had was the uncertainties observed in the individual networks.

E.1 STN MEASUREMENTS

The STN measurements are briefly described to aid in the interpretation of analyses of the collocated IMPROVE and STN samplers; the reader is referred to the STN documentation for further information on their methods: http://www.epa.gov/ttn/amtic/speciepg.html. The IMPROVE and STN networks use different samplers and different standard operating procedures for sample collection and analysis and maintain independent quality assurance programs. However, the two networks use similar analytical techniques to measure a similar set of parameters. Both networks utilize gravimetric analysis for quantification of $PM_{2.5}$ mass, IC for NO_3^- and $SO_4^=$, and XRF for elements including S, Al, Fe, Ca, Si, and Ti.

The two networks use different operationally defined thermal/optical procedures for determining the organic carbon (OC) and elemental carbon (EC) fractions of total carbon (TC). As previously noted in Chapter 1, IMPROVE utilizes the TOR method, whereas STN utilizes the thermal optical transmittance (TOT) NIOSH 5040 method. The two methods utilize different temperature profiles and charring-correction methods—by light reflectance in the case of IMPROVE and transmittance in the case of STN. In a controlled method comparison study, the TOR correction method was found to produce higher EC values than the TOT correction method [Chow et al., 2004]. Temperature protocols can further impact the OC/EC split point, with the higher-temperature/shorter-residence-time temperature profile used in the study enhancing the discrepancy in EC values when comparing the TOR and TOT corrections [Chow et al., 2004]. In summary, these two techniques are not expected to result in comparable OC/EC splits because of the above methodological distinctions but should have comparable TC results [Chow et al., 2004]. Another important distinction between the IMPROVE and STN carbon measurements is that IMPROVE data are routinely blank corrected, whereas STN data are not. As of 2006, STN will begin transitioning all sites to IMPROVE samplers and IMPROVE TOR analysis for carbon measurements. The transition will only affect carbon measurements and is expected to take a couple of years.

STN employs a variety of samplers, including the MetOne spiral ambient speciation sampler (SASS), the Anderson reference ambient air sampler (RASS), the URG mass aerosol speciation sampler (MASS), and the R&P 2300 [EPA speciation guidance, 1999; Rao et al., 2003]. The IMPROVE-STN collocated samplers, as well as STN's in-network collocated samplers, only include the SASS, RASS, and MASS samplers. The MASS sampler is the dominant sampler across the network. The SASS and RASS systems are similar to the IMPROVE sampler in that they utilize a Teflon filter for $PM_{2.5}$ mass and elements, a quartz filter for carbon measurements, and a nylon filter for ions. The SASS sampler is also like the IMPROVE sampler in that each channel of measurement has an independent size-selective inlet. In contrast, the RASS sampler has a single inlet and two size-selective cyclones shared by the four sampling channels.

Unlike the IMPROVE RASS and SASS samplers, the MASS samplers are designed to be operated in pairs, with the MASS 400 operating with a sodium-carbonate coated denuder followed by a two-stage Teflon and nylon filter pack and the MASS 450 operating with a singlestage quartz filter pack. The intended advantage of the MASS system is that it allows for the separate quantification of aerosol nitrate from the Teflon filter and volatilized nitrate from the backup nylon filter. The disadvantage is the loss of ammonium nitrate from the Teflon filter prior to the IC analysis while under vacuum during the preceding XRF analysis. Each of the sampler systems operates at a significantly different flow rate and filter face velocity that can impact the collection of semivolatile species.

E.2 STN OC BLANK CORRECTION PROCEDURE UTILIZED IN THIS REPORT

STN routinely collects both field and trip blanks in support of the carbon analyses. The distinction between field and trip blanks is the temporary unpacking and mounting of field blanks. Field blanks are not retained in the sampler during sampler operation as in IMPROVE's procedures. The field blanks are collected at a frequency of 10% and the trip blanks at a frequency of 3% of routine samples [Flanagan and Peterson, 2002]. While blank filters are routinely collected by STN, the reported OC, EC, and TC values are not blank corrected. Research Triangle Institute (RTI), the primary STN contractor, conducted an analysis of the field and trip blanks collected from 9 February 2000 through 31 March 2002; in that study they found that the field and trip blanks were comparable and that each sampler type required a separate blank correction factor [Research Triangle Institute, 2002]. The network-wide estimates for the blank correction factors for TC for each of the major STN sampler types from this study were utilized for the analyses in this report but not in the manner recommended by RTI in their initial report [2002]. More recent studies by RTI have indicated that it is more appropriate to only blank correct the OC fraction using the reported TC blank correction factor [Rao, T., Personal Communication, 2006]. We are thus following this recommended procedure for blank correcting the STN data. The following analyses of the cross-network collocated data are performed utilizing both the uncorrected and the blank-corrected OC data. The blank-corrected data will be referred to as adjusted OC (OC.adj). Chapters 2 and 3, which include analysis of STN data, utilize adjusted OC data.

E.3 METHODOLOGY USED FOR ESTIMATING THE COMPARABILITY OF MULTIYEAR MEAN CONCENTRATIONS

E.3.1 Introduction

The goal of these analyses is to answer the question: From a usability standpoint, how comparable are the IMPROVE and STN data? Other studies conducted by Rice and Camalier, currently still in progress [Rice, J. , personal communication, 2006], are focused on analyzing the cross-network collocated data set with the aim of better understanding the differences in methods and equipment that cause divergences in the resultant data set. Those analyses, when available, will complement those reported here and aid in efforts aimed at increasing comparability between the two networks, with changes in either procedure or equipment, without jeopardizing comparability within the networks with the past data record.

The specific arena this report aims to address is the comparability of collocated multiyear mean concentrations for each of the measurements utilized in the IMPROVE reconstructed fine mass (RCFM) model, calculated utilizing all reported concentration values. The purpose of this analysis is to provide an adequate framework for the spatial and temporal trends examined in

Chapters 3 and 4 using data from both the IMPROVE and STN networks. The in-network and cross-network collocated data are explored individually and together to gain a better understanding of the observed versus expected comparability of individual sample pairs as well as paired mean concentrations. Paired mean values were composed of between 4 months and 2+ years of data, depending on the length of the available data record; the IMPROVE-STN collocated samplers all had record lengths of at least 1 year and generally 2+ years. As will be elaborated on in the following sections, the analyses are designed to address the following questions:

1. How comparable are mean concentrations from collocated samplers for each of the measurements included in the IMPROVE RCFM model, utilizing all reported values for just the IMPROVE network?
2. How comparable are multiyear mean concentrations from both IMPROVE and STN collocated sites?
3. Are the combined relative errors observed in the cross-network collocated data consistent with what we would expect, given the relative precisions observed in each network?

E.3.2 Statistical Framework

Since the analyses in Chapters 2 and 3 make use of all reported values including those below minimum detection limits (mdl), all statistics calculated in this analysis also include below-mdl values. Sample pairs where the parameter was not detected by the XRF system for one or both of the samples, and therefore reported as having 0 mass concentrations, were excluded from the relative difference, d_i (equation 1, Table E.1), calculations but included in the mean concentration estimates as 0.

The observed root mean square (rms) relative difference, rms(d) (equation 2, Table E.1), of the collocated measurements provides an estimate of the typical magnitude of the combined measurement error, if one can assume that the samples have been drawn independently from the same distribution. The reasoning is as follows: Let Z_i be the true atmospheric value, so that X_i-Z_i and Y_i-Z_i are the errors in the two measurements. Then algebra gives

$$d_i = \frac{X_i - Z_i}{Z_i} - \frac{Y_i - Z_i}{Z_i} + O\left(\left(\frac{X_i - Z_i}{Z_i}\right)^2, \left(\frac{Y_i - Z_i}{Z_i}\right)^2\right) \cong \frac{X_i - Z_i}{Z_i} - \frac{Y_i - Z_i}{Z_i}$$

if the relative errors are small. More algebra gives

$$\frac{1}{n}\sum_i d_i^2 \cong \frac{1}{n}\sum_i \left(\frac{X_i - Z_i}{Z_i}\right)^2 + \frac{1}{n}\sum_i \left(\frac{Y_i - Z_i}{Z_i}\right)^2 - \frac{2}{n}\sum_i \left(\frac{X_i - Z_i}{Z_i}\right)\left(\frac{Y_i - Z_i}{Z_i}\right) \quad [*]$$

If the measurements are unbiased, so the expected values of *X* and *Y* are both *Z*, then the expected values of the three terms on the right-hand side are the variance of the relative error in

X, the variance of the relative error in *Y*, and the covariance of the relative errors in *X* and *Y*. Because the errors in the paired samples are assumed to be independent, the third term is 0. In the case of equivalent measurements, such as in-network collocated data, the measurements are assumed to have identical error variances, and the first and second terms are equal. The relative precision of measurement *X* is defined here as the square root of the variance in its relative error (equation 3, Table E.1). Thus, in the case of equivalent measurements (in-network collocated data), the relative precision of a parameter is found by taking the square root of half the combined variance observed in the collocated data:

$$\frac{1}{n}\sum_i d_i^{\,2} \cong 2(precision)^2 \text{, or } precision = \sqrt{\frac{1}{n}\sum_i d_i^{\,2} \Big/ 2} \; *$$

* This is the same as equation 3 above

For nonequivalent measurements, such as cross-network data, the assumption of identical error variances is unlikely to be valid, and therefore additional partitioning of the combined relative error, rms(d), into relative precision is neither easy nor straightforward.

It is assumed in these analyses that the relative precision for a parameter provides a useful measure of the total uncertainty, e.g., sampling and analytical errors, in the measurements. However, White et al. [2005] have demonstrated that in-network relative precisions calculated from collocated data do not capture nonconstant biases that affect both members of the sample pair such as analytical calibration drifts. Thus the relative precision of the parameter represents a lower estimate on the actual total measurement uncertainty when the measurements are intended to be used for trend analysis [White et al., 2005] and by extension any analysis that involves comparison of data from different time periods. Cross-network collocated data can provide an estimate of the impact of the nonconstant biases on the actual total measurement uncertainty; however, there is no a priori method for partitioning that additional uncertainty between the networks.

A further assumption is that the measurement errors should behave similarly to random noise, and thus the error in the mean concentration should be equal to $\sigma / \sqrt{n}$. If this assumption is reasonable, then the expected uncertainty for an average of n samples is given by E_rp($\overline{X}$) (equation 6, Table E.1). This assumption is explored by comparing the observed relative difference, O_rd($\overline{X}$) (equation 5, Table E.1), in the paired means (equation 4, Table E.1) and the expected relative precision of the mean E_rp($\overline{X}$). In unbiased measurements it should be highly unusual for O_rd($\overline{X}$)/2 to be more than 3*E_rp($\overline{X}$).

A final assumption is that the observed combined relative error, rms(d) (equation 2, Table E.1), in the cross network data should in the absence of significant hidden uncertainties (e.g., calibration biases) be comparable to the combined measurement error expected, given the relative precisions observed in the in-network collocated data for each network, E_rms(d) (equation 7, Table E.1). This assumption can be tested by comparing rms(d) to E_rms(d).

Table E.1. Statistics utilized in the analysis of in-network and cross-network collocated data for assessing measurement uncertainties and comparability of data collected by the IMPROVE and STN networks.

Equation Number	Description	Equation
1	The relative difference in a sample pair *where Yi and Xi represent paired samples	$d_i(X,Y) = \frac{(Yi - Xi)*2}{(Yi + Xi)}$ *
2	The root mean square relative difference for a population of collocated sample pairs	$rms(d) = \sqrt{(\frac{1}{n})\sum_{i=1}^{i=n}\left(d_i(X,Y)\right)^2}$
3	The relative precision for a population of sample pairs where the measurements are equivalent	$rp = \frac{1}{\sqrt{2}} rms(d)$
4	The paired mean concentrations *including only Xi and Yi where both Xi and Yi exist	$\overline{X} = mean(Xi)$ $\overline{Y} = mean(Yi)$ *
5	The observed relative difference in paired means *where $\overline{X}$ and $\overline{Y}$ represent the paired means	$O\_rd(\overline{X}) = \frac{(\overline{X} - \overline{Y})*2}{(\overline{X} + \overline{Y})}$ *
6	The expected relative precision in paired means	$E\_rp(\overline{X}) = \frac{rp}{\sqrt{n}}$
7	The expected rms difference in nonequivalent measurements *where X and Y represent the independent relative precisions calculated from in-network collocated data for IMPROVE and STN	$E\_rms(d_i) = \sqrt{rp_X^{\ 2} + rp_Y^{\ 2}}$

E.3.3 Data Sets Utilized in Analysis

The following data sets were used in these analyses:

1. In-network IMPROVE collocated data collected for ~0.3–1.5 years at a variety of IMPROVE sites utilizing extra sampler modules in most cases and a full collocated sampler at Phoenix from the installation date of the particular collocated module through 2004 (Chapter 1, Table 1.3)

2. In-network STN reported relative precisions based on data from the six STN sites with collocated samplers for the RCFM parameters for the MASS, RASS, and SASS sampler types collected from 2002–2004 [Rice, personal communication, 2006]

3. Cross-network IMPROVE-STN collocated data collected for ~2.5 years at six sites arranged in urban-rural site pairings in the Northwest, Southwest, and along the East Coast (Figure E.1) from October 2001 through December 2003

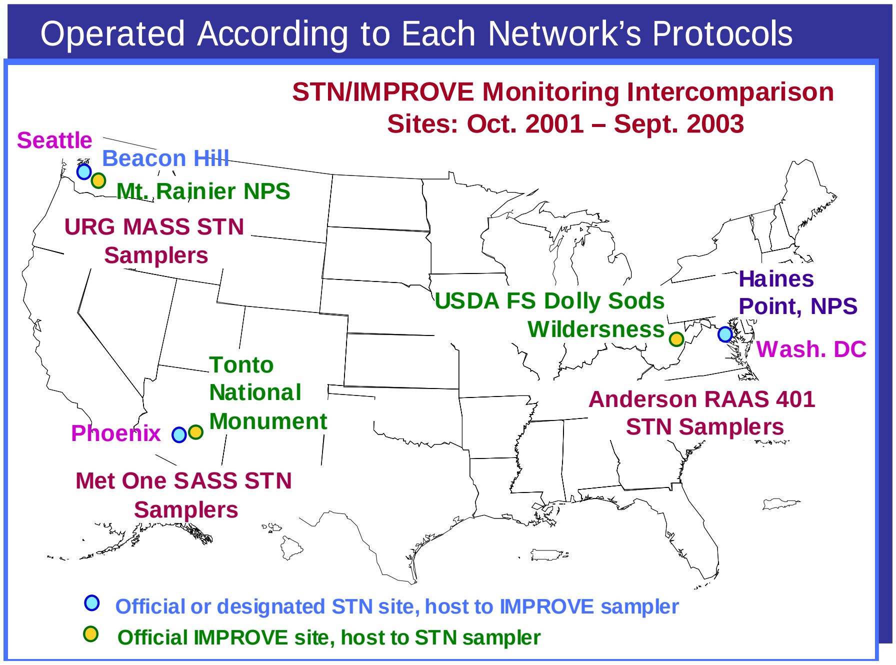


Figure E.1. Site map of the locations of the cross-network collocated IMPROVE and STN samplers. The host network is indicated by the color of the site marker on the map. The STN sampler type used for the regionally paired collocated samplers is also given on the map. The Washington, D.C., and Seattle sites are the same as the IMPROVE protocol sites WASH1 and PUSO1.

E.3.4 Outlined Approach

For the in-network IMPROVE collocated data, the steps taken to understand and estimate the typical comparability of mean concentrations within IMPROVE for each of the major parameters were

- Step 1. Explore the collocated data population to assess the following:
    - Do the collocated concentration values appear to have a 1:1 relationship?
- Step 2. Calculate the relative difference, d_i, for each collocated sample pair that reports a detected value and explore the population of relative measurement errors to assess if the relative measurements were consistent with idealized random errors:
    - Are there important relationships between the d_i and concentration?
    - Are there important relationships between the d_i and time?
    - Do the d_i as a population follow a symmetrical distribution?
    - Do the d_i as a population follow a normal distribution?
- Step 3a. Calculate the relative precision, rp, and also calculate the paired mean concentrations, $\overline{X}$ and $\overline{Y}$, with the data aggregated by parameter and also by parameter and site. Then calculate the observed relative difference, O_rd($\overline{X}$), in the paired means

and the expected relative precision of the means E_rp($\overline{X}$) and compare to assess the following:

- Are the observed differences in the mean values qualitatively comparable to the expected differences?
- Given the observed rp and O_rd($\overline{X}$), what is a conservative estimate of the comparability of multiyear mean concentrations from collocated monitors for each parameter?

For the cross-network IMPROVE and STN collocated data, Steps 1 and 2 were repeated. Step 3 was modified for the case of nonequivalent measurements.

- Step 3b. Calculate the rms relative difference, rms(d), and also calculate the paired mean concentrations, $\overline{X}$ and $\overline{Y}$, with the data aggregated by parameter and also by parameter and site. Then calculate the observed relative difference, O_rd($\overline{X}$), in the paired means to assess the following:
    - Given the observed rms(d) and O_rd($\overline{X}$),what is a conservative estimate of the comparability of multiyear mean concentrations from collocated monitors for each parameter?

Finally, to evaluate the performance observed in the cross-network data, the rms relative differences observed in the cross-network collocated data were compared for consistency with what would be expected given the relative precisions observed in each network individually. The concluding step in this analysis was accordingly

- Step 4. Calculate the expected rms difference, E_rms(d) from the in-network IMPROVE and STN relative precisions and qualitatively compare the E_rms(d) values to the rms(d) values observed in the cross-network data to assess the following:
    - Is the sum of the error variances observed in the in-network collocated data equal to the combined error variance, $rms(d)^2$, observed in the cross-network collocated data? That is, does rms(d)= $\sqrt{rp_X^{\ 2} + rp_Y^{\ 2}}$?

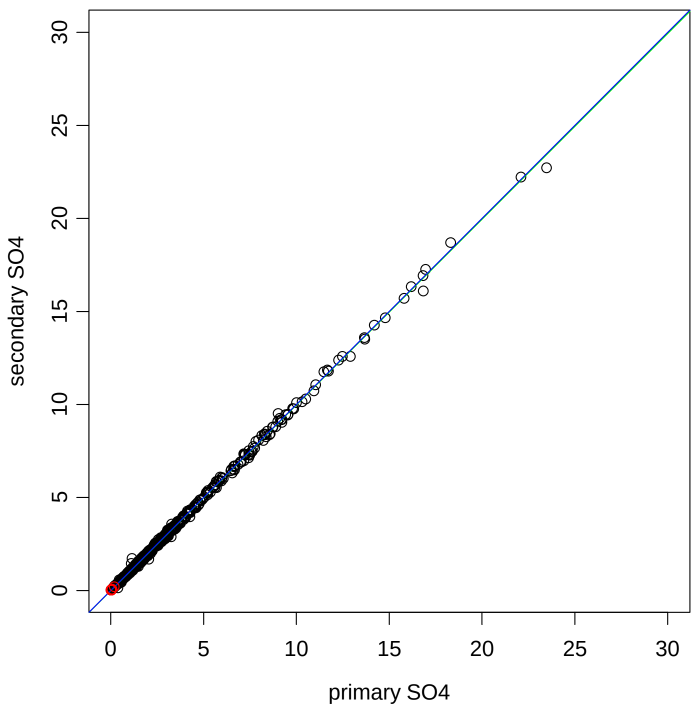
E.4 IMPROVE IN-NETWORK COLLOCATED DATA

Step 1

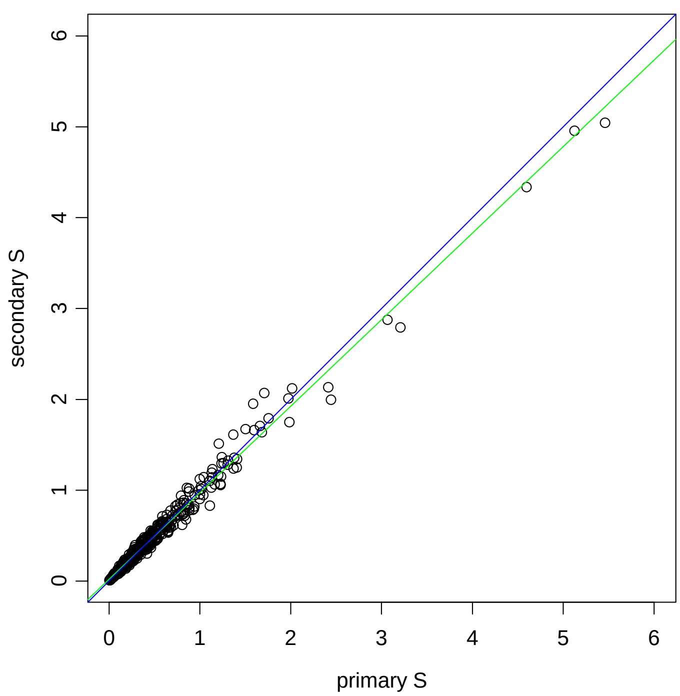
Discussion of the in-network IMPROVE collocated data is included here to provide context for the following discussion of the cross-network collocated data. The data were explored in a qualitative manner with the objective of providing a reference for evaluating the same relationships in the IMPROVE-STN collocated data. The in-network collocated data was collected at 24 sites utilizing a fifth sampling module that duplicates one of the four standard IMPROVE sampling modules (Chapter 1, Table 1.3). Start dates for the collocated sampler modules ranged from April 2003 to October 2004, and all valid data collected between the sampler start date and December 2004 were used in this analysis. Collocated measurements from a well-controlled study should exhibit a 1:1 relationship and mean values equivalent within measurement uncertainty. Scatter plots of the collocated data along with the ordinary least square (OLS) best fit line (green), the 1:1 line (blue), and the R^2 values for each of the major parameters were used to explore if the collocated data qualitatively displayed a 1:1 relationship (Table E.2 and Figure E.2, panels a–j). The correlation analysis and OLS regression line do not suggest that a 1:1 relationship is an unreasonable model for any of the major parameters utilized in the RCFM model. All of the major parameters have high R^2 values ranging from 0.75 for Al to 1.0 for SO_4 and NO_3, even with below-mdl and nondetected samples included in the correlation coefficient and regression calculations. Values for the OLS slope and intercept ranged from 0.82 to 1.01 and -0.01 to 0.03, respectively (Table E.2). Aluminum had the poorest relationship, and this is in large part due to the higher relative number of nondetects, as compared to the other soil elements, that relate to the difficulty in accurately identifying Al in the XRF spectrum.

Table E.2. OLS slope and intercept coefficients, the R^2 value, and the paired mean values for the collocated concentration values are given for each parameter. The summary statistics include all available collocated data, with each parameter having six contributing sites (Chapter 1, Table 1.3).

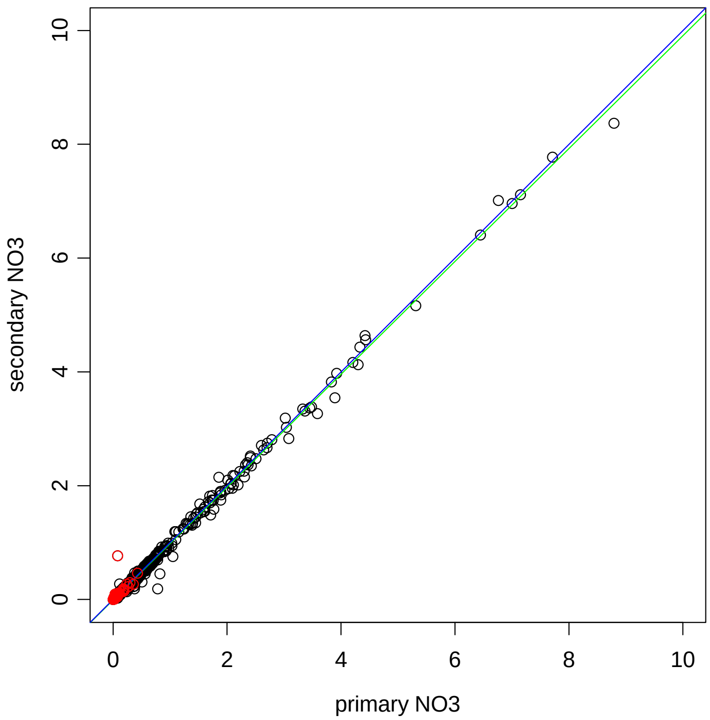
	Sampler 1 Mean μg/m³	Sampler 2 Mean μg/m³	Slope	Intercept	R^2
SO_4	0.0416	0.0409	1.00	-0.01	1.00
NO_3	0.0494	0.0519	0.99	-0.01	1.00
S	0.2375	0.2255	0.95	0.02	0.99
OC	0.0432	0.0439	1.01	0.00	0.97
Fe	0.4711	0.4612	0.96	0.00	0.91
Ti	1.1257	1.1440	0.95	0.00	0.90
EC	0.4150	0.4159	0.82	0.03	0.88
Si	0.1309	0.1322	0.91	0.01	0.87
Ca	2.1255	2.1128	0.99	0.00	0.86
Al	0.0034	0.0034	0.85	0.01	0.75

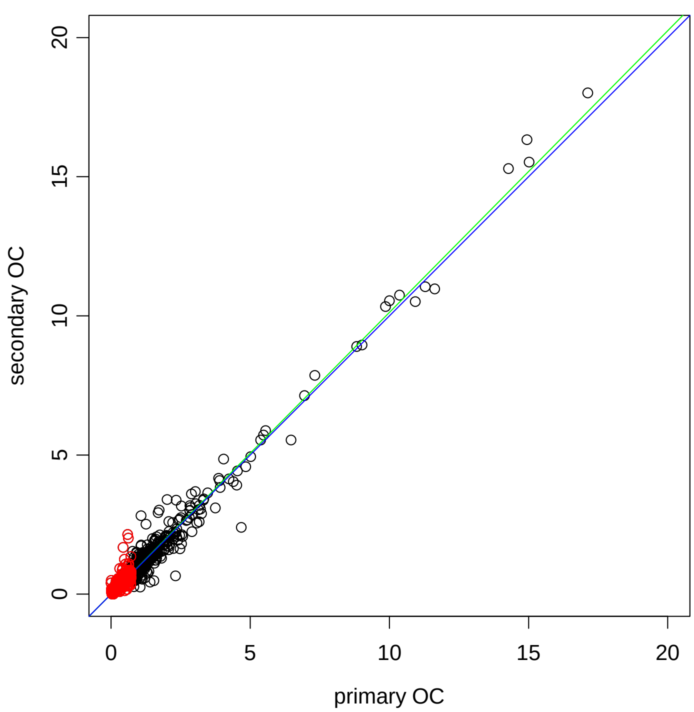


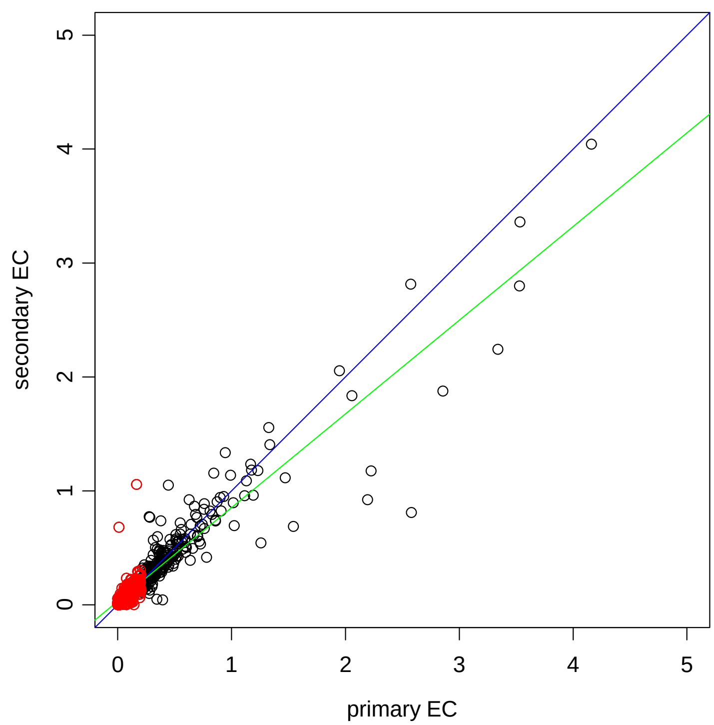
secondary SO4
primary SO4
0
5
10
15
20
25
30



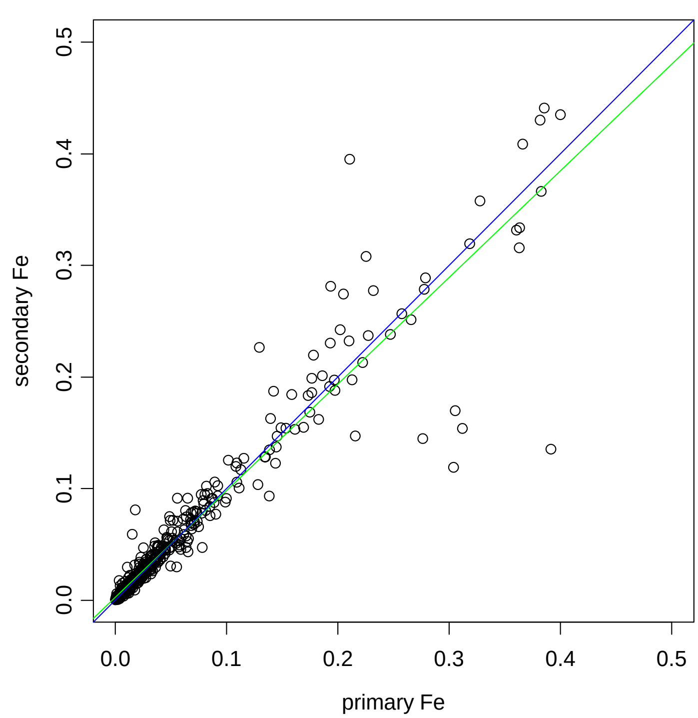
secondary S
primary S
0
1
2
3
4
5
6


secondary NO3
primary NO3
0
2
4
6
8
10

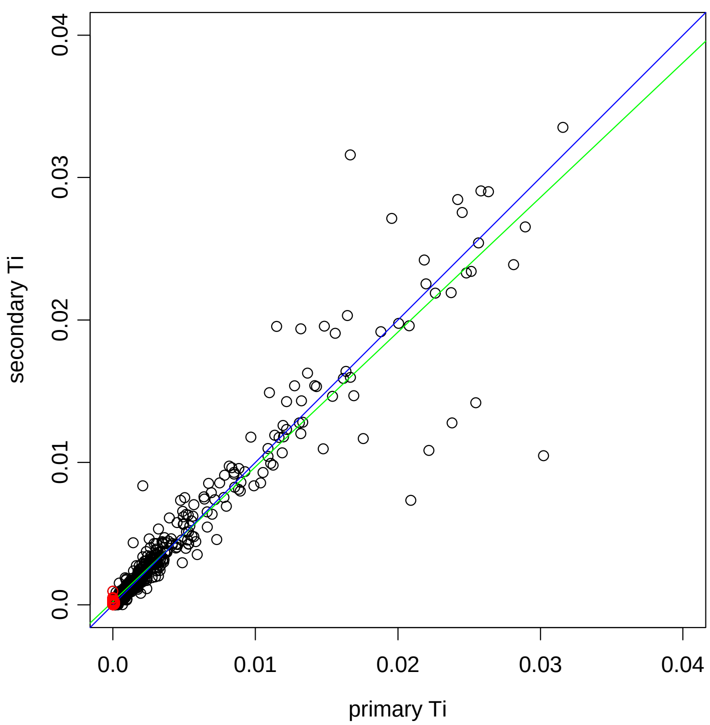

secondary OC
primary OC
0
5
10
15
20



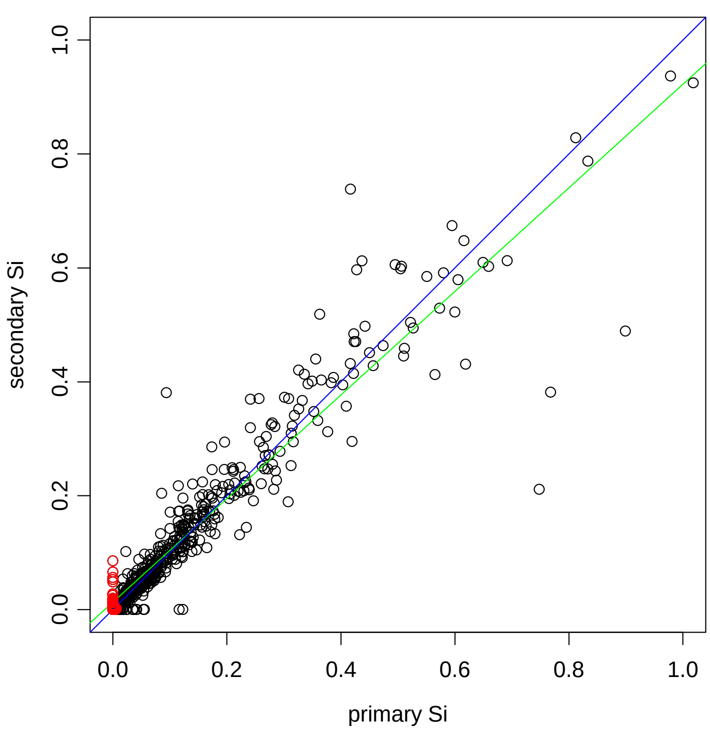
0
1
2
3
4
5
secondary EC
primary EC



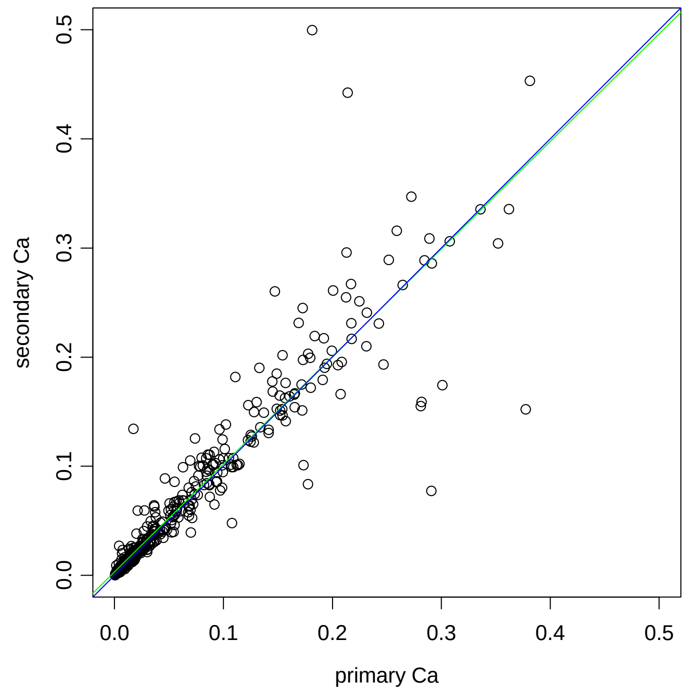
0.0
0.1
0.2
0.3
0.4
0.5
secondary Fe
primary Fe



secondary Ti
primary Ti
0.0
0.01
0.02
0.03
0.04



secondary Si
primary Si
0.0
0.2
0.4
0.6
0.8
1.0



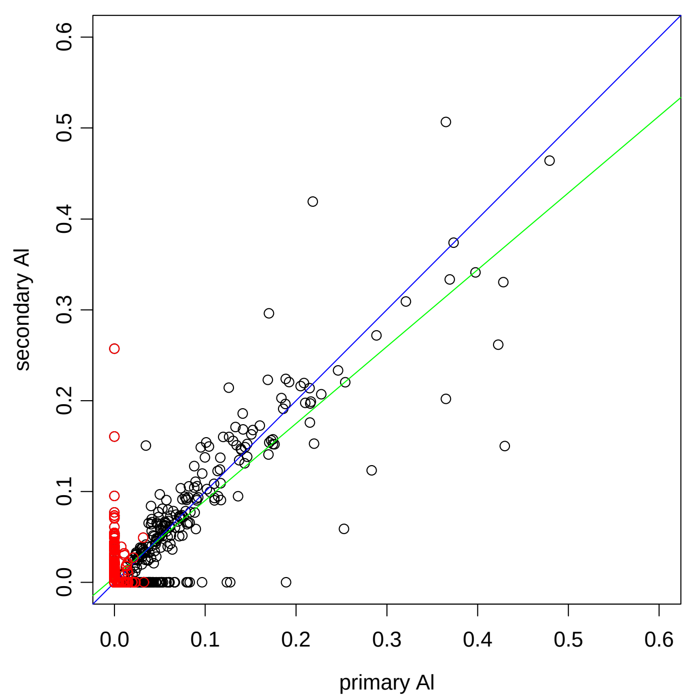


Figure E.2. a–j. Scatter plots of the collocated data from the IMPROVE network for the sites listed in Chapter 1, Table 1.3. The primary sampler module is plotted along the X axis and the secondary sampler module is plotted along the Y axis. Both the X and Y axes are in units of μg/m^3. Concentration values less than 3*mdl are displayed in red. The 1:1 line is shown in blue and the OLS regression line is shown in green.

Step 2

Idealized measurements are typically modeled as having random normal errors, and it is usually further assumed that the measurement errors are homoscedastic, meaning that the measurement errors are drawn from the same population over the full time period of the study and for the full range of reported values. Nonnormal measurement errors require the use of nonparametric statistics or transformations to normalize the errors when analyzing the data. In the case of heteroscedastic errors (nonconstant), statistical tests robust to varying errors must be used in all analyses, or the data must be transformed to sufficiently equalize the measurements across all dimensions of interest (e.g., time or concentration). Homoscedacity of the relative measurement errors is not expected for IMPROVE data since many of the samples are near (concentrations less than 10*mdl) the mdl, and relative measurement uncertainties increase rapidly as concentrations approach the mdl [Taylor, 1987]. The populations of relative differences, d_i, were explored to address if there were any important relationships between the d_i and concentration or time. The probability plots of the d_i were also qualitatively explored to assess if the measurement errors appear to be drawn from a single normal distribution. The results are included to illustrate why the data analyst must be cautious in the application of parametric statistical techniques or those sensitive to heteroscedastic errors to IMPROVE data rather than to illustrate any shortcoming in the IMPROVE program.

Examination of the scatter plots of the d_i versus the average concentration of the sample pairs (Figure E.3, panels a–j) shows that the median d_i value was near 0 for all major parameters, indicating no major biases between the collocated modules. Given that the data were produced using equivalent equipment and procedures, biases are not expected. With perhaps the exception of SO_4, the errors appear to be symmetrically distributed around the 0 d_i line. Most of the examined parameters appear to have higher variability in the relative differences at lower concentrations. This effect is most apparent in the blank-corrected parameters OC, EC, NO_3, and SO_4.

To confirm the relationship between d_i variability and concentration, the variability in the d_i for low and high concentration groups was compared for significant differences using the F-test for the equality of two standard deviations. The F-test is a parametric test that is not robust to nonnormality of the sample populations in the sense that the actual significance level of the test can be quite far from the intended significance level of the test [Hollander and Wolfe, 1999]. For example, a specified significance level of 0.05 might actually be as large as 0.166 or as small as 0.0056, and large sample size does not necessarily protect from this effect for some nonnormal populations [Hollander and Wolfe, 1999].

The d_i population was split into two subgroups, with the low concentration group being defined as having a concentration value from the primary sampler that was less than the median concentration value for that sampler; the high concentration group met the reverse criteria. The standard deviations for the two subgroups were compared via the F-test, and pairs that were significantly different at a specified alpha level of 0.05 are identified in bold in Table E.3. All parameters besides Ca showed significantly higher variability in the low concentration group. Furthermore, T tests allowing for different variability were performed to compare the mean values of the low and high concentration groups. The mean values for S, SO_4, EC, and OC were significantly different at a 0.05 alpha level.

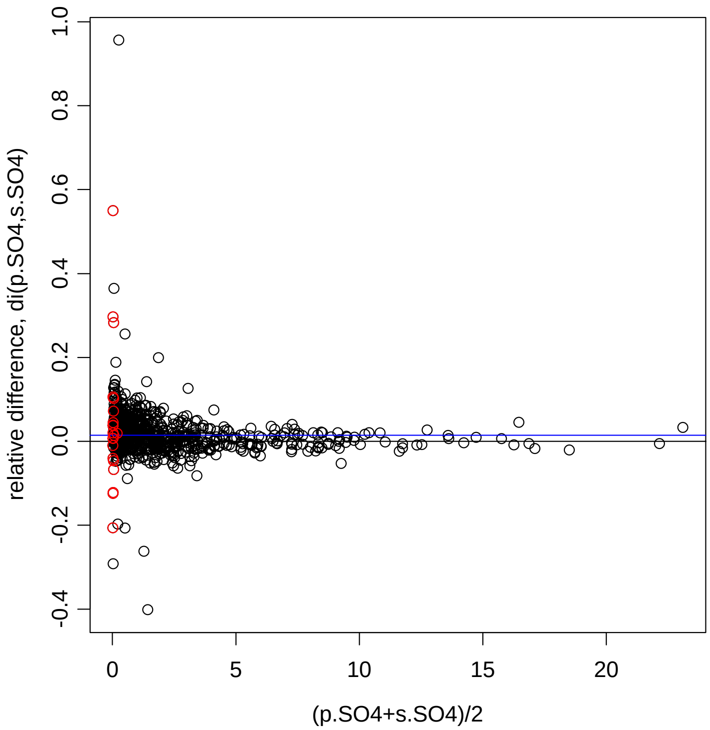
Changes over time in the precision of the measurements are suggested for some of the parameters in Figure E.4, panels a–j. The F-test for the equality of two standard deviations and the T-test with unequal variances were used to compare if the variability and central tendency of the d_i were significantly different in the first half of 2004 as compared to the second half (Table E.4). For S and SO_4, both the variability and central tendency of the d_i were significantly different between the two time periods. Additionally, the variability was significantly different between the two time periods for EC, Ca, Fe, Si, and Ti. Nitrate was the only parameter to show a significant difference in mean values without an accompanying significant difference in variability between the two time periods.

Table E.3. The mean, standard deviation, and number of valid values of the relative differences, d_i, in the low and high concentration groups are listed for the in-network collocated data. Pairs in bold are significantly different at a 0.05 signficance level, based upon T-tests allowing for different variability and F-tests for the equality of two standard deviations.

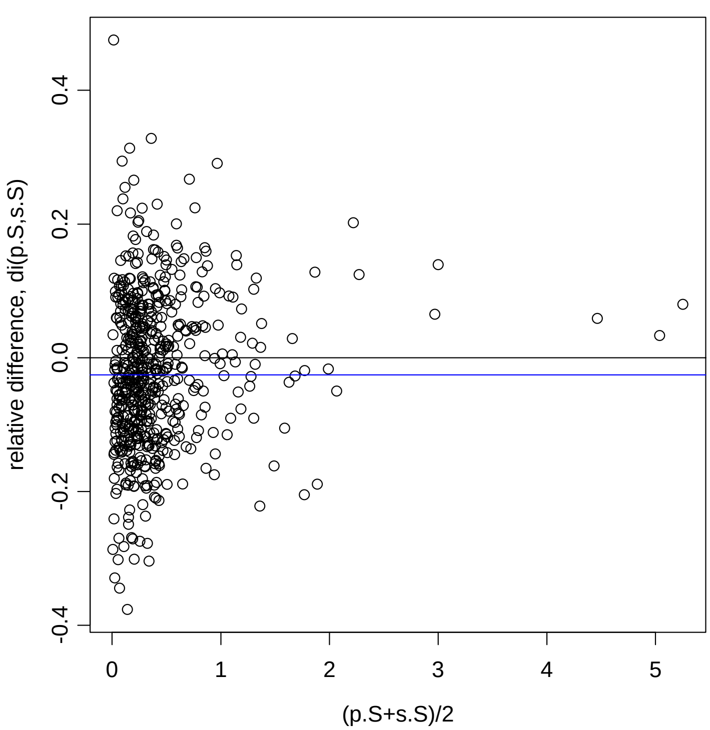
	S		**SO_4**		**NO_3**		**EC**		**OC**	
	Low	High	Low	High	Low	High	Low	High	Low	High
Mean	**-0.036**	**-0.005**	**0.031**	**0.007**	0.035	0.048	**-0.022**	**0.079**	**-0.074**	**0.025**
Standard Deviation	**0.119**	**0.107**	**0.075**	**0.040**	**0.300**	**0.118**	**0.499**	**0.300**	**0.412**	**0.236**
N	290	287	410	412	409	414	373	390	390	393
	Al		**Ca**		**Fe**		**Si**		**Ti**	
	Low	High	Low	High	Low	High	Low	High	Low	High
Mean	-0.070	-0.011	-0.075	-0.041	-0.046	-0.018	-0.008	-0.015	-0.022	-0.016
Standard Deviation	**0.404**	**0.273**	0.239	0.253	**0.280**	**0.219**	**0.309**	**0.236**	**0.327**	**0.248**
N	74	138	288	286	291	286	216	284	234	286

Table E.4. The mean, standard deviation, and number of valid values of the relative differences, d_i, in the first and second half of 2004 are listed for the in-network collocated data. Pairs in bold are significantly different at a 0.05 significance level based upon T-tests allowing for different variability and F-tests for the equality of two standard deviations.

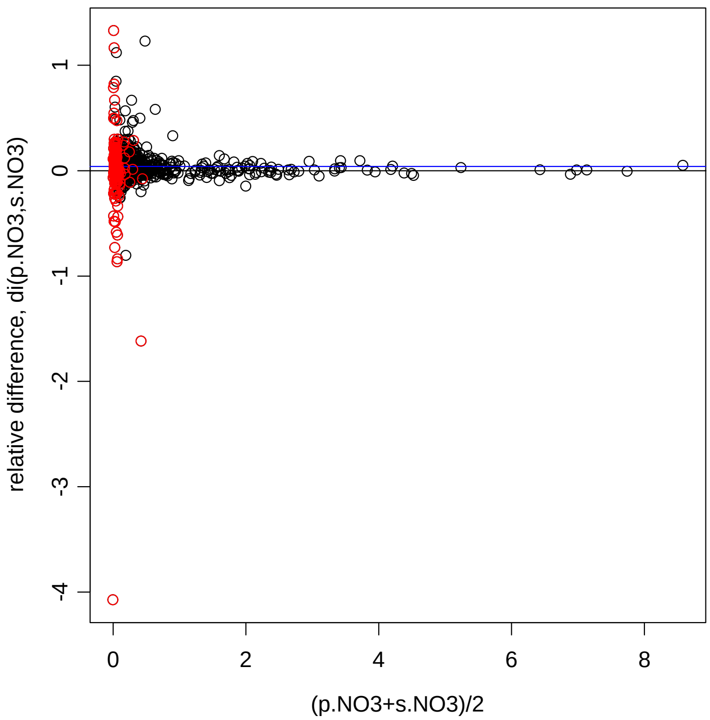
	S		**SO_4**		**NO_3**		**EC**		**OC**	
	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4
Mean	**-0.043**	**-0.003**	**0.012**	**0.023**	**0.034**	**0.068**	0.028	0.047	-0.020	-0.009
Standard Deviation	**0.119**	**0.107**	**0.039**	**0.034**	0.131	0.124	**0.372**	**0.433**	0.337	0.343
N	199	337	252	377	252	377	261	375	272	382
	Al		**Ca**		**Fe**		**Si**		**Ti**	
	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4	2004 Q1-Q2	2004 Q3-Q4
Mean	-0.043	-0.029	-0.057	-0.067	-0.039	-0.028	-0.019	-0.008	-0.010	-0.029
Standard Deviation	0.315	0.330	**0.220**	**0.264**	**0.209**	**0.280**	**0.241**	**0.289**	**0.250**	**0.297**
N	83	128	199	336	199	337	187	287	196	294

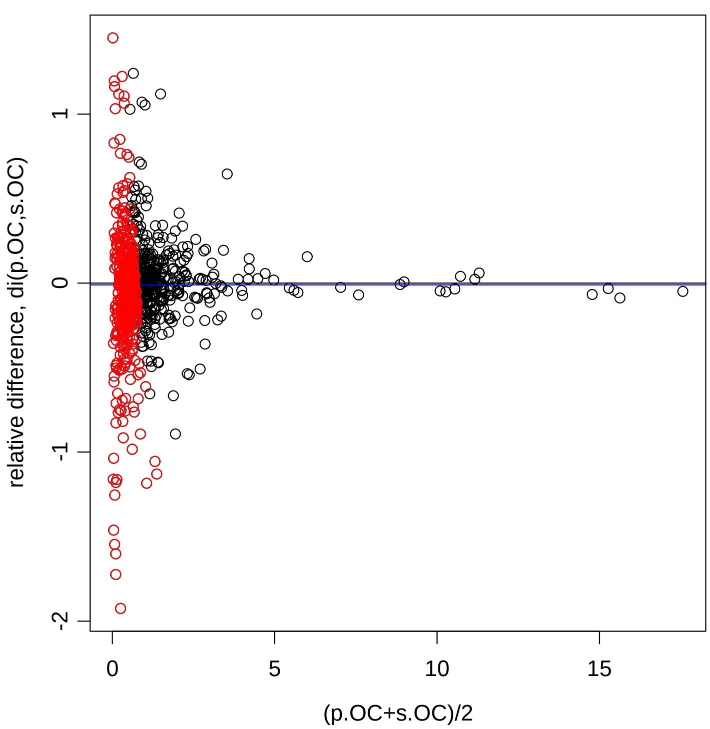


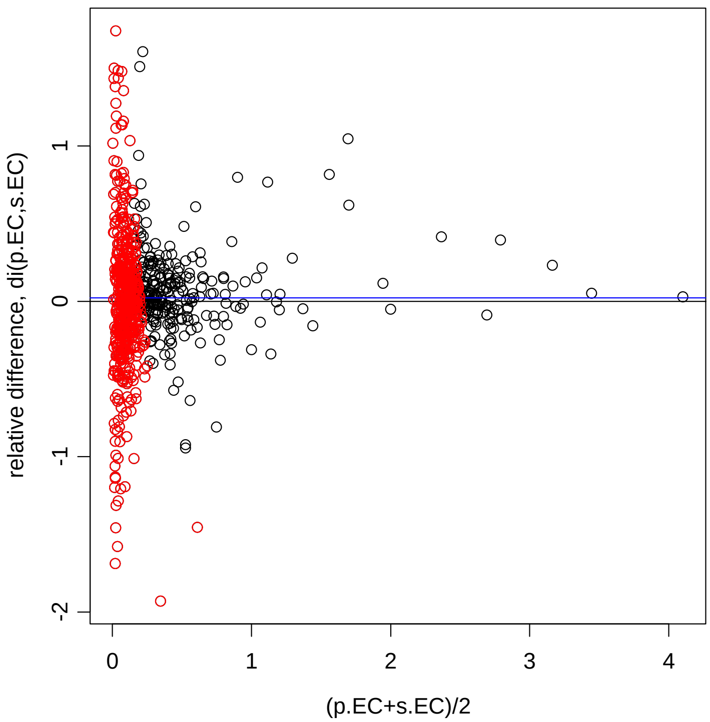
relative difference, di(p.SO4,s.SO4)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0
5
10
15
20
(p.SO4+s.SO4)/2



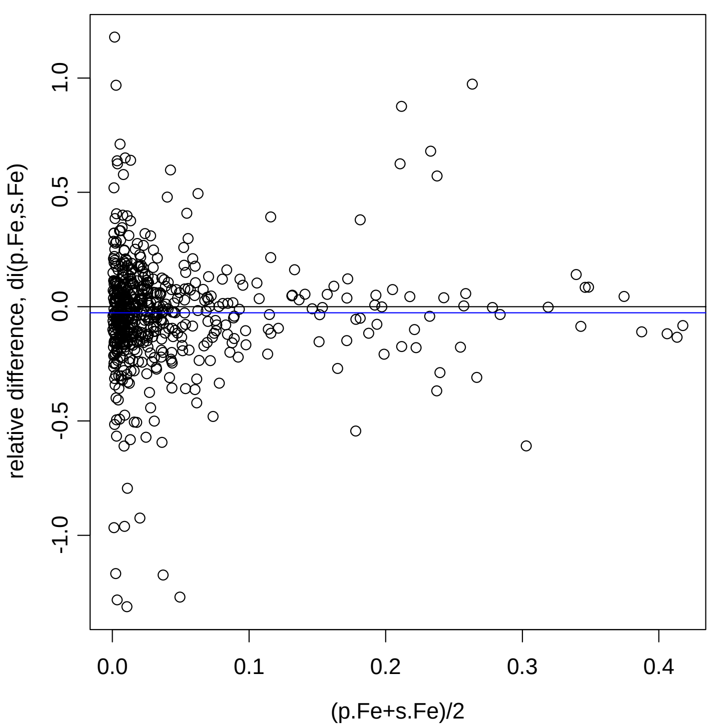
relative difference, di(p.S,s.S)
0.4
0.2
0.0
-0.2
-0.4
0
1
2
3
4
5
(p.S+s.S)/2


relative difference, di(p.NO3,s.NO3)
1
0
-1
-2
-3
-4
0
2
4
6
8
(p.NO3+s.NO3)/2

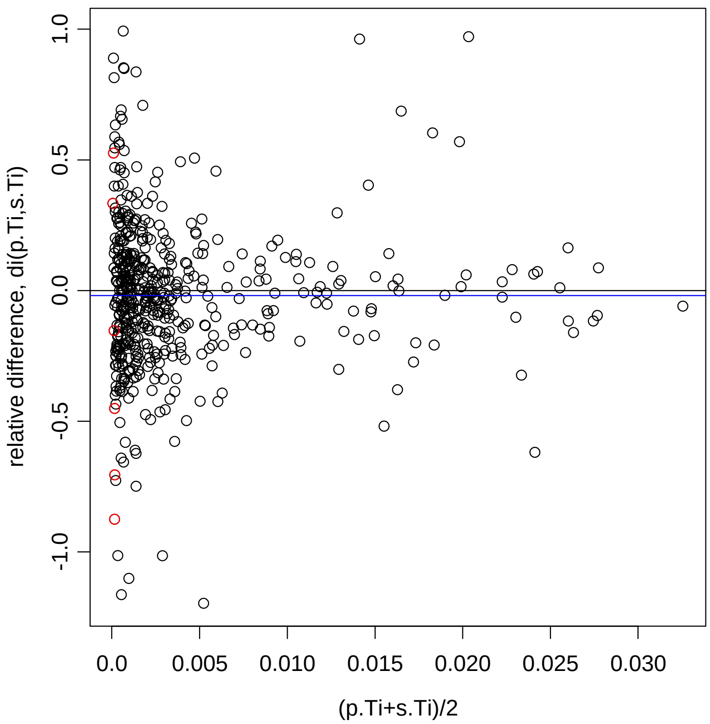

relative difference, di(p.OC,s.OC)
1
0
-1
-2
0
5
10
15
(p.OC+s.OC)/2



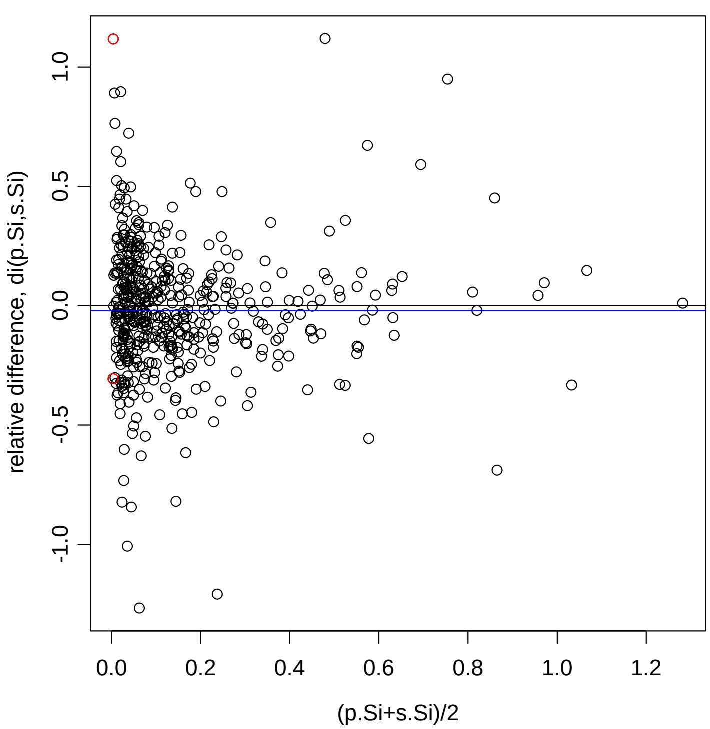
relative difference, di(p.EC,s.EC)
(p.EC+s.EC)/2



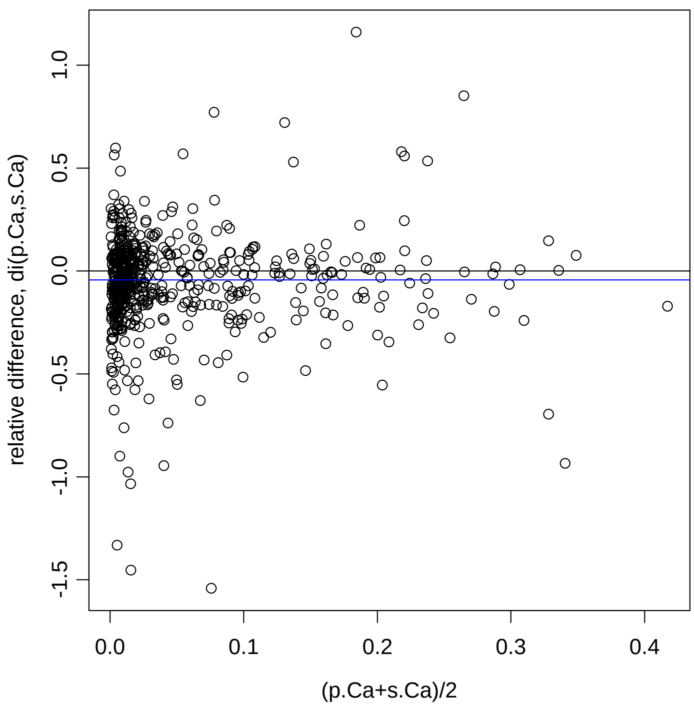
relative difference, di(p.Fe,s.Fe)
(p.Fe+s.Fe)/2



relative difference, di(p.Ti,s.Ti)
(p.Ti+s.Ti)/2
1.0
0.5
0.0
-0.5
-1.0
0.0
0.005
0.010
0.015
0.020
0.025
0.030



relative difference, di(p.Si,s.Si)
(p.Si+s.Si)/2
1.0
0.5
0.0
-0.5
-1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2



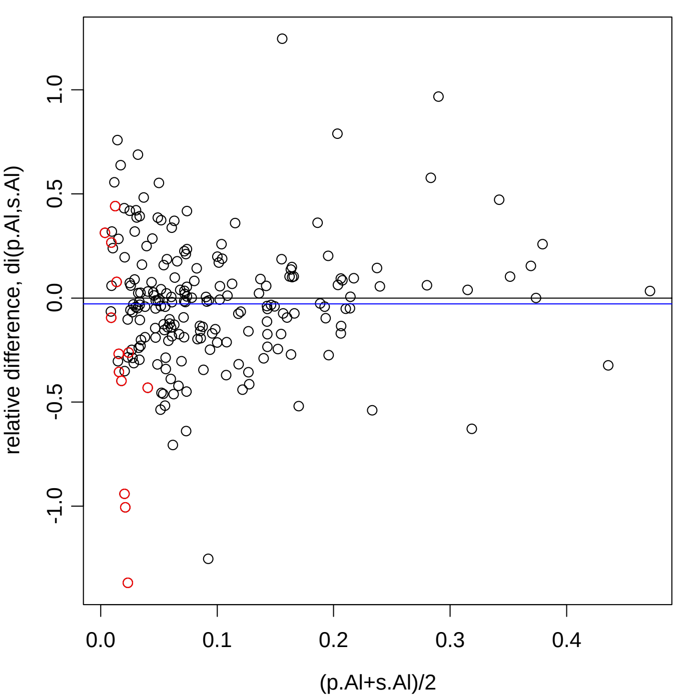
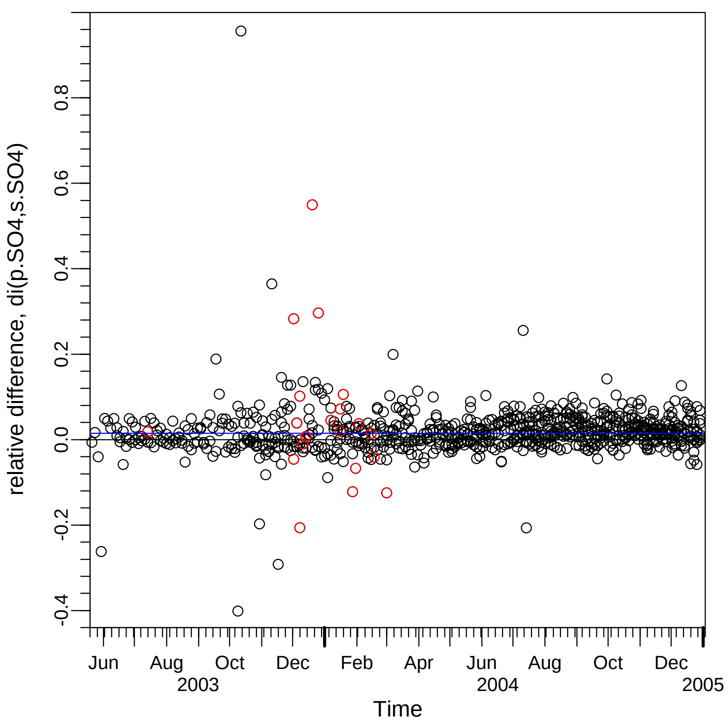
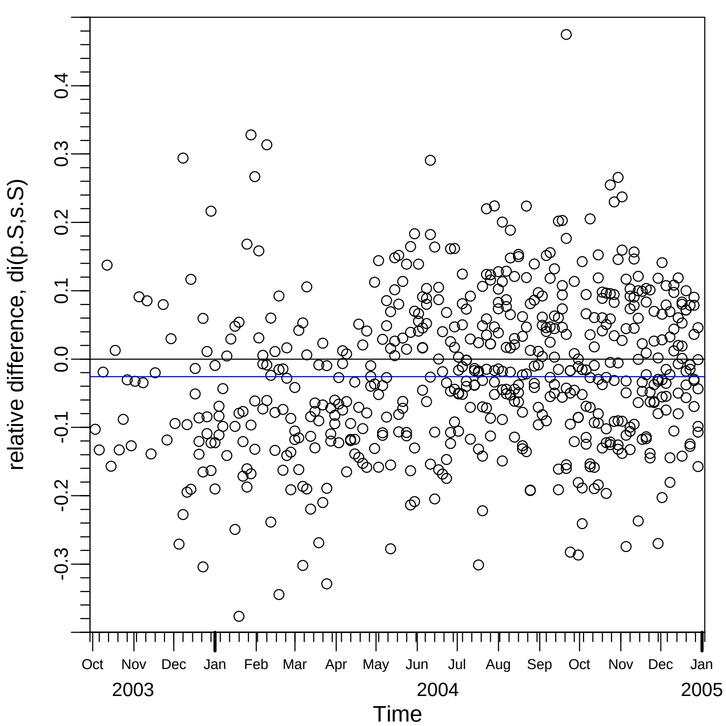
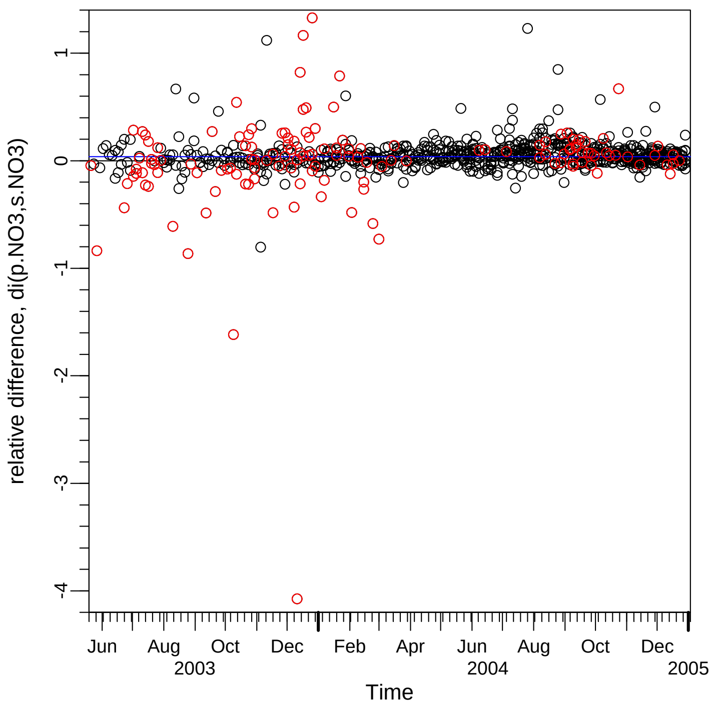
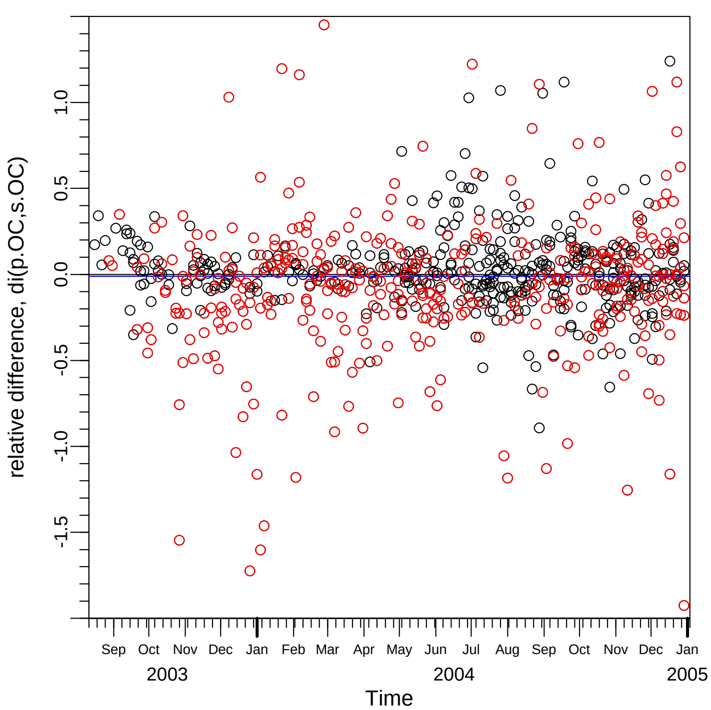


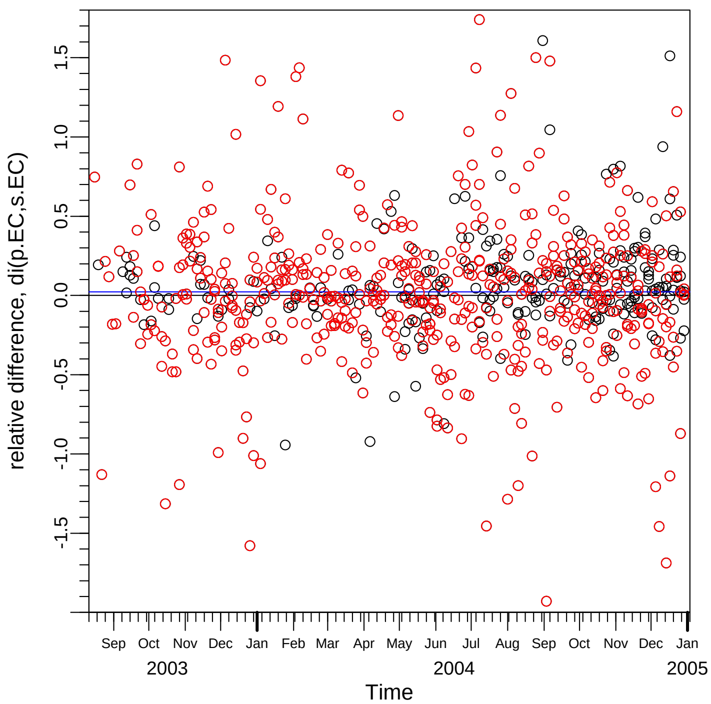
Figure E.3. a–j. Scatter plots of the relative differences of the paired samples, d_i, versus the average concentrations of the sample pairs for the collocated data from the IMPROVE network for the sites listed in Chapter 1, Table 1.3. The X axis has units of μg/m^3 and the Y axis is unitless. Concentration values less than 3*mdl are displayed in red. The blue line is the median d_i value.

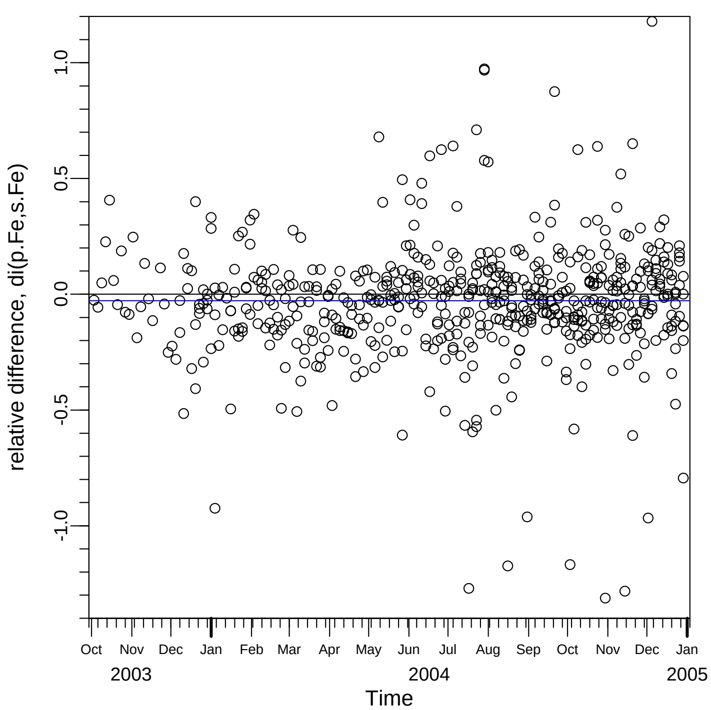

relative difference, di(p.SO4,s.SO4)
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
Jun
Aug
Oct
Dec
Feb
Apr
Jun
Aug
Oct
Dec
2003
2004
2005
Time

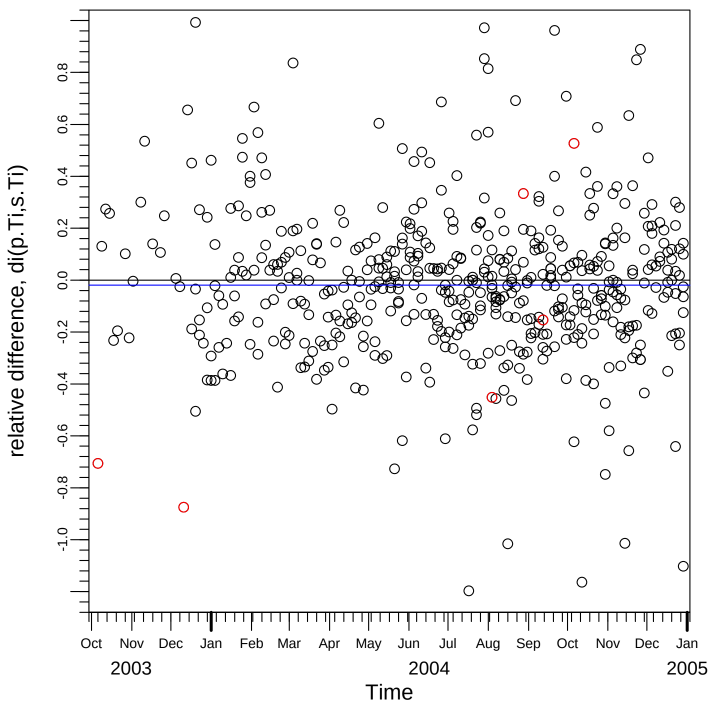

relative difference, di(p.S,s.S)
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2003
2004
2005
Time


relative difference, di(p.NO3,s.NO3)
1
0
-1
-2
-3
-4
Jun
Aug
Oct
Dec
Feb
Apr
Jun
Aug
Oct
Dec
2003
2004
2005
Time

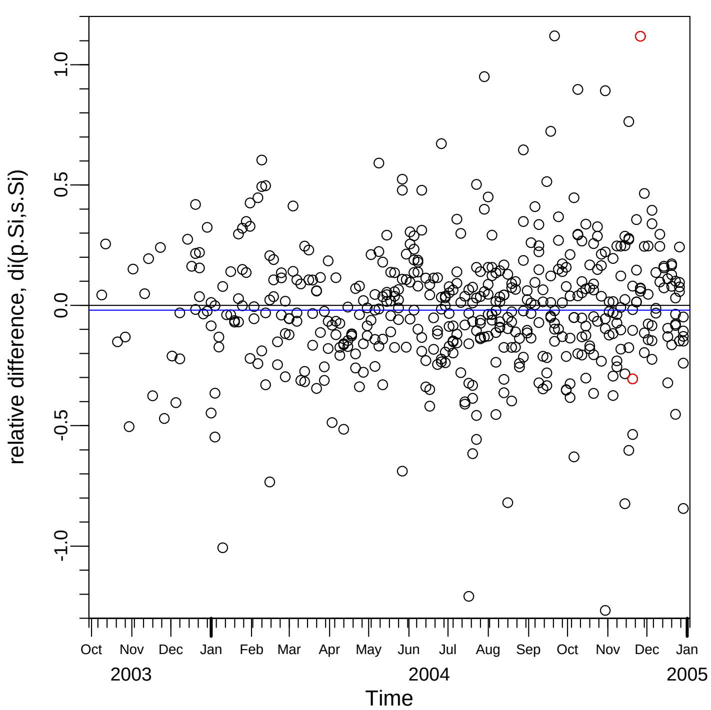

relative difference, di(p.OC,s.OC)
1.0
0.5
0.0
-0.5
-1.0
-1.5
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2003
2004
2005
Time


relative difference, di(p.EC,s.EC)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2003
2004
2005
Time

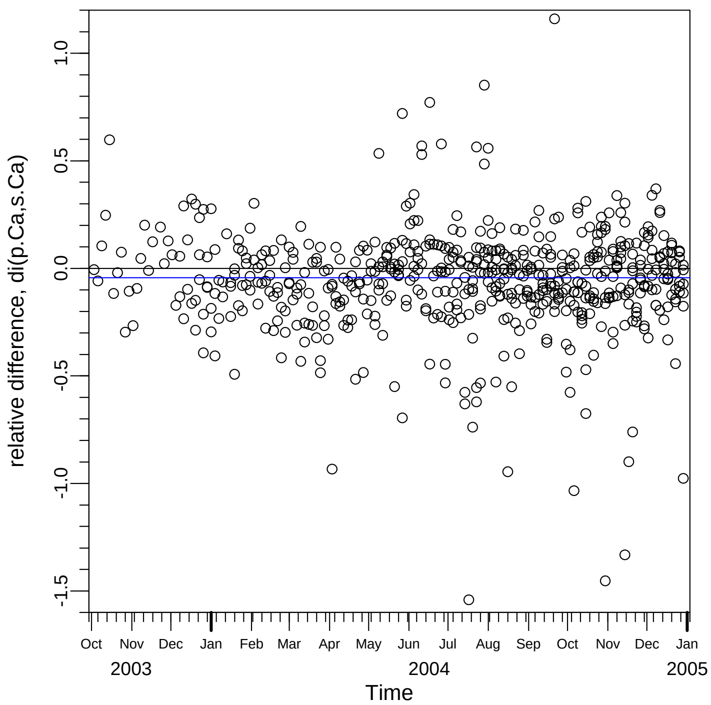

relative difference, di(p.Fe,s.Fe)
1.0
0.5
0.0
-0.5
-1.0
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2003
2004
2005
Time



relative difference, di(p.Ti,s.Ti)
Time
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2003
2004
2005
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0



relative difference, di(p.Si,s.Si)
Time
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan
2003
2004
2005
1.0
0.5
0.0
-0.5
-1.0



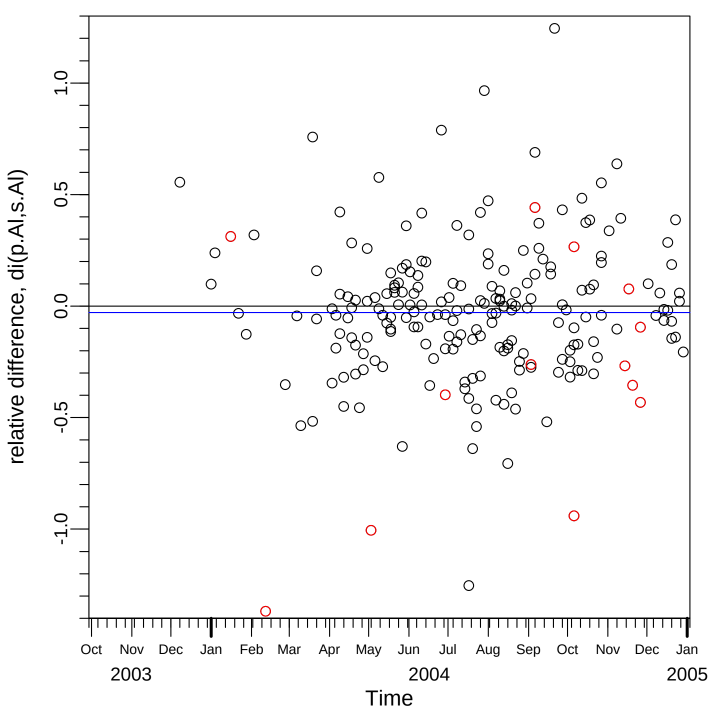
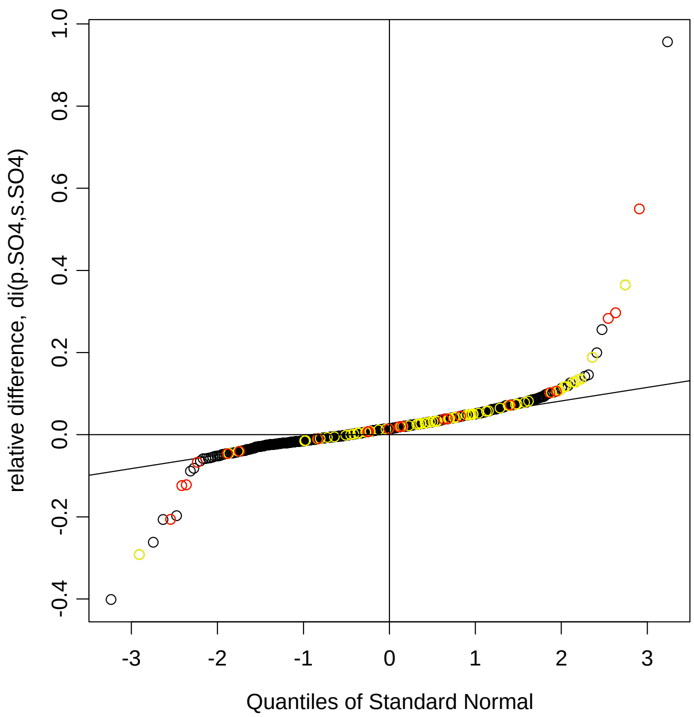
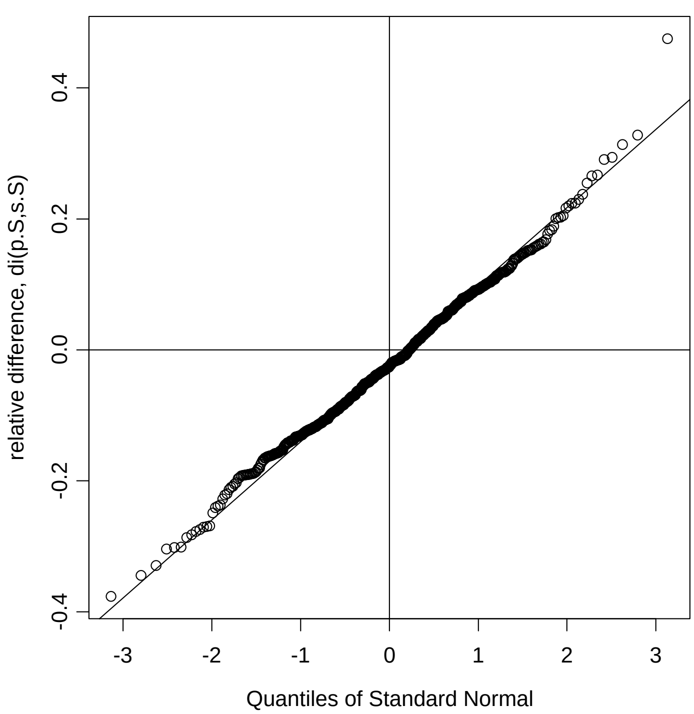


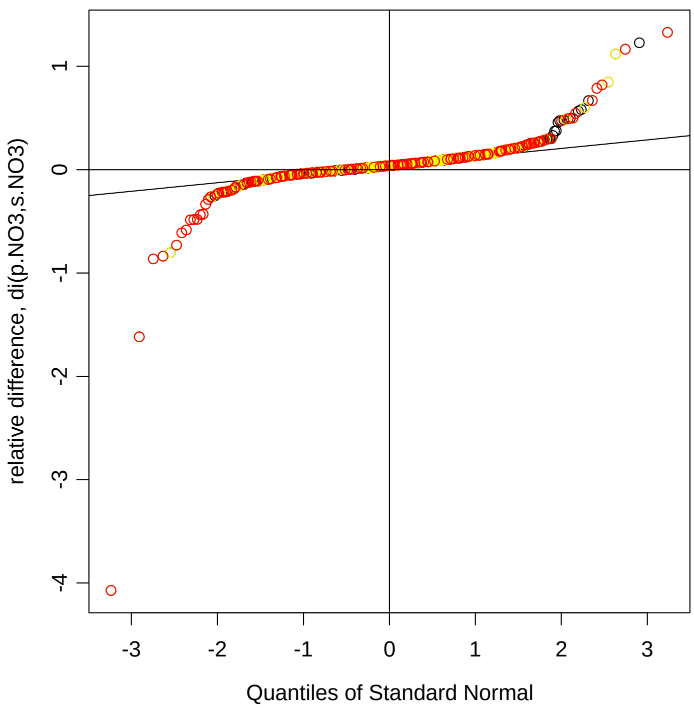
Figure E.4. a–j. Time series of the relative differences of the paired samples, d_i, for the collocated data from the IMPROVE network for the sites listed in Chapter 1, Table 1.3. The X axis has units of time and the Y axis is unitless. Concentration values less than 3*mdl are displayed in red. The blue line is the median d_i value.

Normal probability plots can be used to qualitatively test if the set of observations were drawn from a single normal population. If the d_i were drawn from a normal population, then the quantiles of the d_i population should fall along a straight line when plotted against the quantiles of the normal distribution as in the normal probability plots in Figure E.5, panels a–j [Rice, 1995]. The central tendency of the d_i population is centered approximately on 0 for all of the major parameters, where we would expect it to be for idealized measurements, again indicating no major biases between the collocated samplers. With the exception of S, the d_i populations for all major parameters clearly deviate from normality, with the tails of the distribution not lying along the straight-line fit. These deviations show that the extreme errors observed in the d_i population are larger than would be expected if all of the measurement errors were drawn from a single Gaussian distribution. This does not necessarily indicate that the error distribution at any fixed concentration is not Gaussian. For example, the high concentrations could all have errors drawn from a normal distribution with mean value $\mu = 0$ and standard deviation σ_1, $N(0, \sigma_1)$, and the lower concentrations have their errors drawn from a normal distribution with the same mean value and a larger standard deviation, $N(0, \sigma_2)$, where $\sigma_1 < \sigma_2$ and the result of grouping all these samples together would be a nonnormal error distribution.

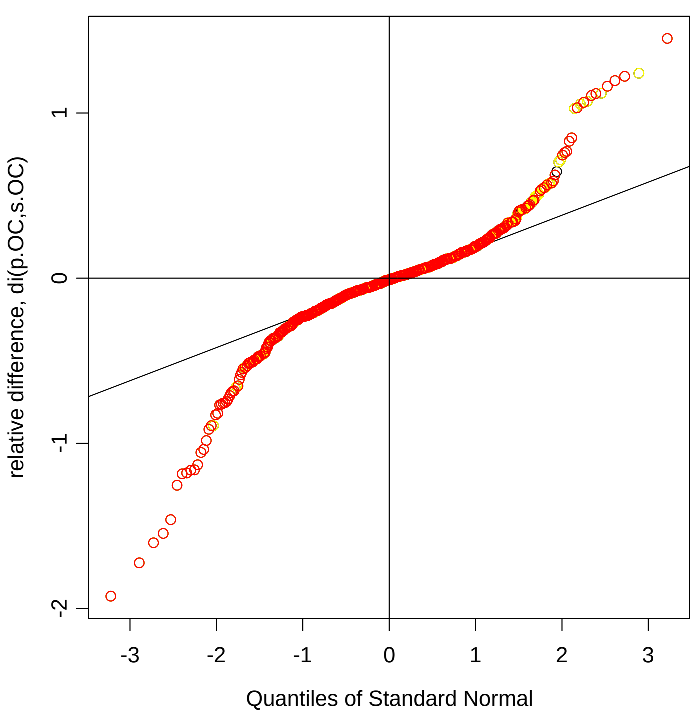
IMPROVE measurements clearly show deviations from the naïve ideal of independent and identically distributed normal measurement errors, and therefore the data analyst needs to be cautious in the use and interpretation of many of the common statistical techniques (e.g., OLS and T-tests) and should consider using techniques that are robust to these types of errors. Non-ideal measurement errors do not indicate a shortcoming that is in anyway unique to IMPROVE data; rather they are a product of trace chemical analysis.


relative difference, di(p.SO4,s.SO4)
Quantiles of Standard Normal
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3

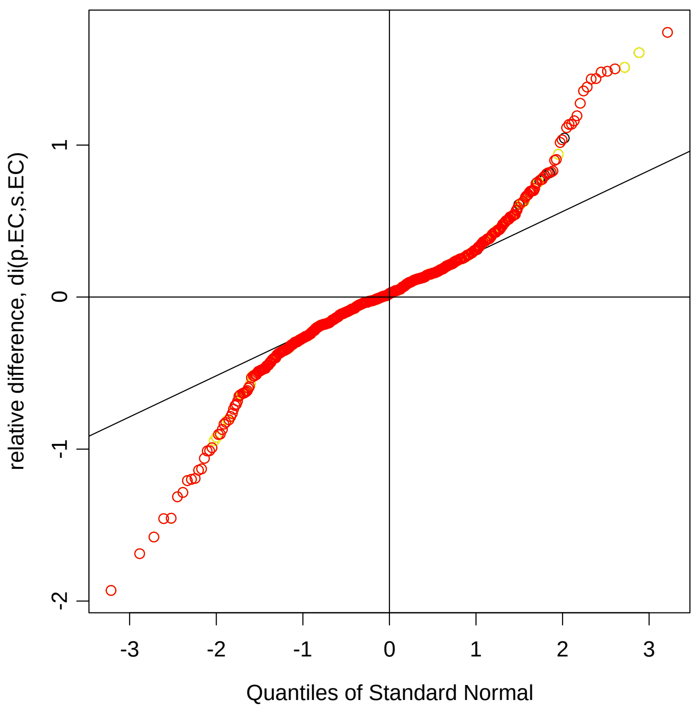

relative difference, di(p.S,s.S)
Quantiles of Standard Normal
-0.4
-0.2
0.0
0.2
0.4
-3
-2
-1
0
1
2
3

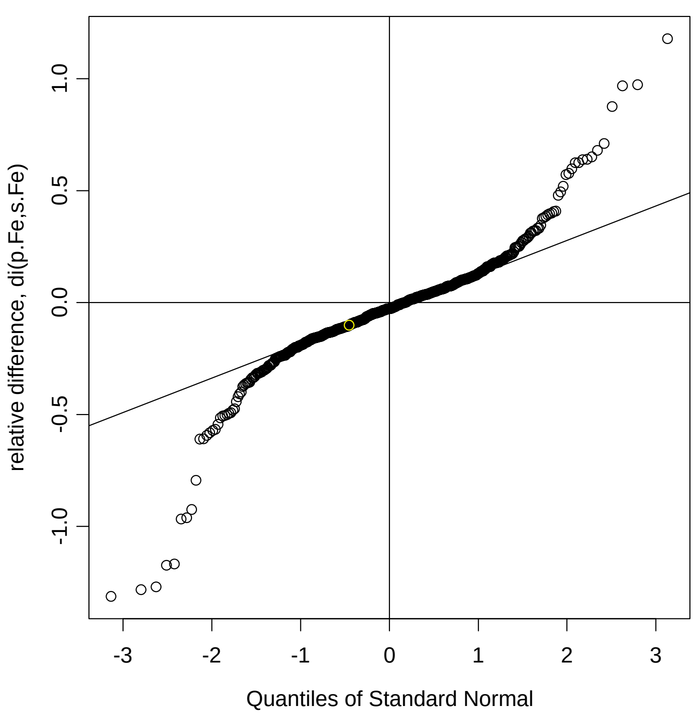


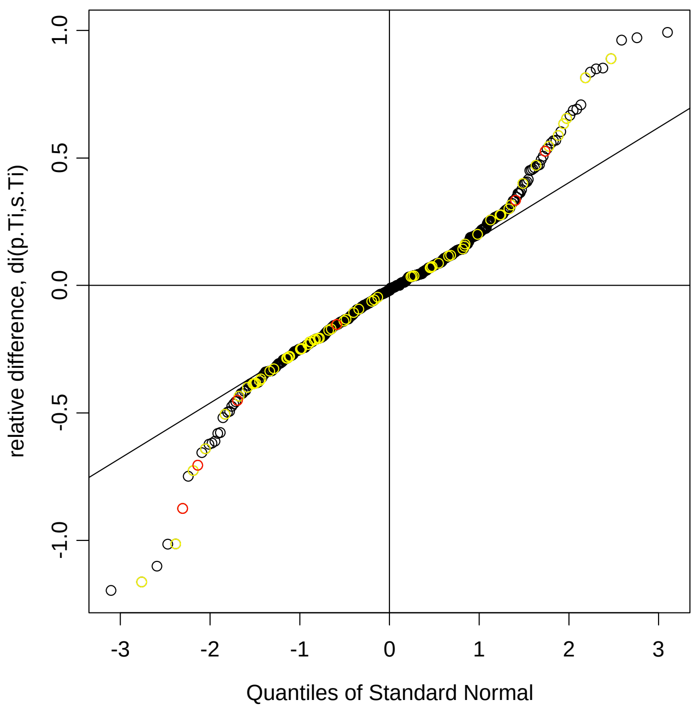
relative difference, di(p.NO3,s.NO3)
Quantiles of Standard Normal
1
0
-1
-2
-3
-4
-3
-2
-1
0
1
2
3

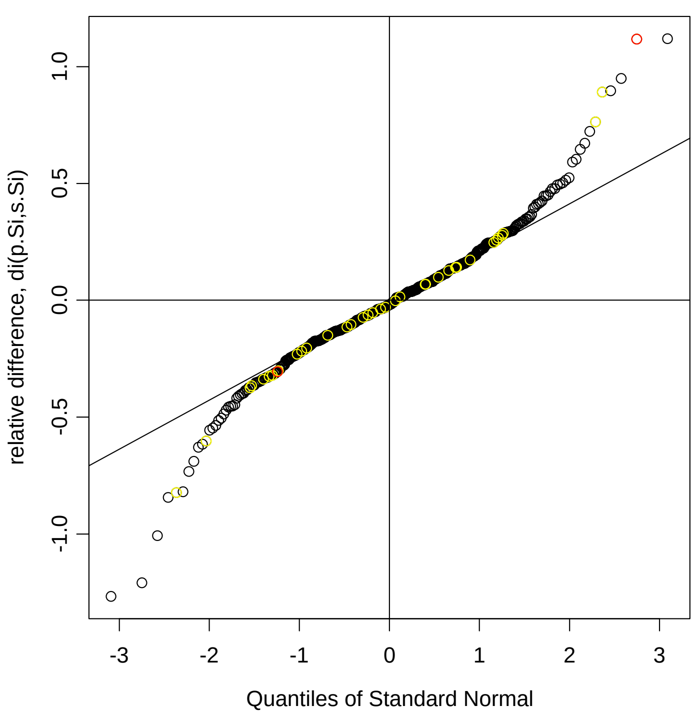


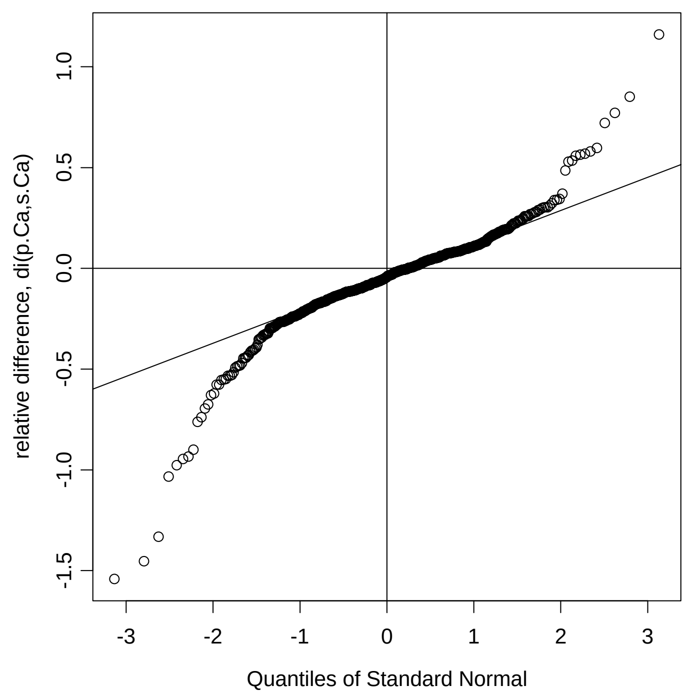
relative difference, di(p.OC,s.OC)
Quantiles of Standard Normal
1
0
-1
-2
-3
-2
-1
0
1
2
3


relative difference, di(p.EC,s.EC)
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3
-2
-1
0
1


relative difference, di(p.Fe,s.Fe)
Quantiles of Standard Normal
-3
-2
-1
0
1
2
3
-1.0
-0.5
0.0
0.5
1.0


relative difference, di(p.Ti,s.Ti)
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal


relative difference, di(p.Si,s.Si)
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
3
Quantiles of Standard Normal



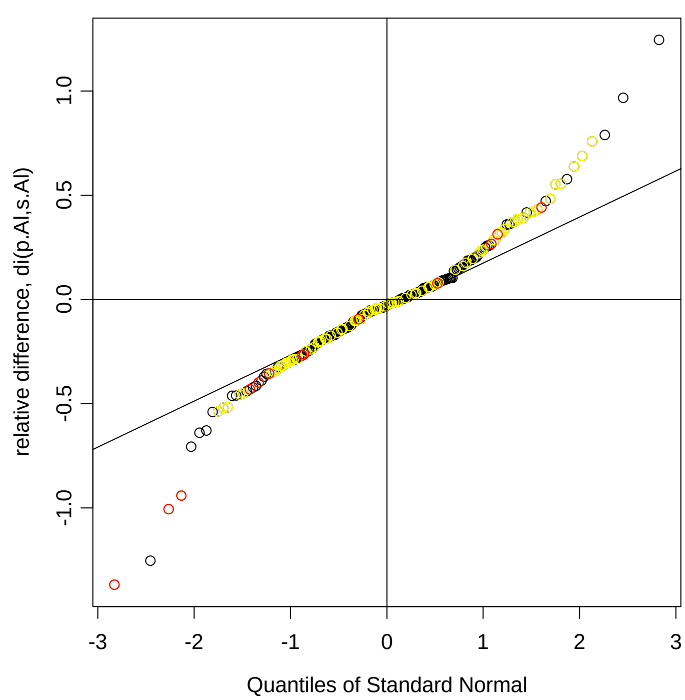


Figure E.5. a–j. Normal probability plots of the relative differences of the paired samples, d_i, for the collocated data from the IMPROVE network for the sites listed in Chapter 1, Table 1.3. The quantiles of the d_i are plotted against the quantiles of the normal distribution. A robust straight-line fit is plotted to aid the viewer in interpreting the linearity of the relationship. Values corresponding to concentrations less than 3*mdl and less than 10*mdl are shown in red and yellow, respectively.

Step 3a

The relative precision (rp) was calculated for each parameter from the in-network collocated data, with the data aggregated by parameter and also by parameter and site. The rp for the parameter-only aggregate (identified as network in Table E.5) is probably a more accurate representation of the characteristic relative precision of the parameter than the parameter-site aggregates because of the larger sample size. So the parameter-only aggregate values are used in all subsequent analyses, whereas the parameter-site aggregates are included to give a feeling for the range in relative precisions observed in specific subpopulations. The only parameters to have a characteristic relative precision for individual samples of less than 10% were S and SO_4. Nitrate, Ca, Fe, Si, and Ti had characteristic relative precisions in the 15–20% range, and Al, OC, and EC were in the 20–30% range.

Table E.5. Relative precisions observed in the in-network IMPROVE collocated data aggregated by parameter (network) and by parameter and site (identified by site).

Relative Precision								
	Network	**MEVE1**	**OLYM1**	**PMRF1**	**SAFO1**	**TRCR1**	**PHOE1**	
Al	23.0%	24.0%	24.1%	24.6%	21.1%	18.6%	22.9%	
Ca	17.9%	32.4%	11.8%	12.5%	15.8%	14.7%	19.1%	
Fe	17.9%	27.5%	17.3%	11.6%	13.6%	23.0%	18.5%	
Si	19.1%	25.5%	18.1%	18.1%	13.7%	21.5%	19.8%	
Ti	20.3%	26.0%	18.7%	20.4%	16.6%	24.1%	19.0%	
S	8.2%	7.2%	8.9%	7.5%	8.1%	9.7%	8.2%	
	Network	**BIBE1**	**BLMO1**	**FRRE1**	**GAMO1**	**LAVO1**	**PHOE1**	**MACA1**
NO_3	16.3%	14.9%	2.7%	4.6%	15.8%	25.7%	14.3%	8.5%
SO_4	4.5%	6.6%	3.2%	1.6%	5.5%	4.9%	3.4%	1.9%
	Network	**EVER1**	**HEGL1**	**HOOV1**	**MELA1**	**SAWE1**	**PHOE1**	**SENE1**
OC	24.0%	15.6%	11.1%	36.4%	18.5%	30.4%	12.6%	21.8%
EC	29.2%	28.7%	14.9%	38.8%	26.4%	19.0%	18.3%	31.4%

For low concentrations, the relative errors can be large, while the absolute difference between the sample pairs is small. The relative precision values reported in Table E.5 do not give any insight into the average absolute errors for a parameter, nor do they account for the concentration dependence of the relative precision estimates. Average absolute precision was calculated for each parameter using the formula

$$\text{Absolute precision}=\frac{Average(|\,Xi-Yi)\,|)}{2}$$

The relative precision was calculated for two subpopulations for each parameter—the low concentration subpopulation includes all sample pairs where the concentration on the primary sampler is below the 20^{th} percentile for that sampler, and the high concentration subpopulation includes all samples above the 20^{th} percentile. Table E.6 includes the average concentration for each sampler, the average absolute precision, the characteristic relative precision for the parameter using all concentration values, and the characteristic relative precision for the low and high concentration groups. The absolute error for S is actually higher than those for the soil elements, but because of the very low soil element concentrations, the relative errors are quite high. The relative errors for NO_3 had strong concentration dependence, with the relative error for the high concentration population approximately one-third of the low population group. The absolute error for NO_3 was actually lower than the absolute error for SO_4, which, at least in part because of its high concentrations, had the lowest relative precision. OC had both higher absolute and relative errors as compared to S, SO_4, and NO_3 for all populations examined. EC also had comparably high absolute errors, as compared to all noncarbon parameters, and high relative errors as compared to the nonsoil parameters. These values are included here to highlight these issues; the problem of integrating these issues into the uncertainty estimates for IMPROVE measurements will be left for future investigations.

Table E.6. Comparison of alternate calculations of measurement precision.

	Average Concentration Primary Sampler $\mu g/m^3$	**Average Concentration Secondary Sampler $\mu g/m^3$**	**Average Absolute Precision $\mu g/m^3$**	**Relative Precision for Whole Population**	**Relative Precision for the Lowest 20%**	**Relative Precision for the Highest 20%**
Al	0.0416	0.0409	0.0106	23.0%	37.6%	22.2%
Ca	0.0494	0.0519	0.0047	17.9%	21.4%	19.1%
Fe	0.0432	0.0439	0.0036	17.9%	24.5%	16.2%
Si	0.1309	0.1322	0.0141	19.1%	27.4%	18.0%
Ti	0.0034	0.0034	0.0003	20.3%	29.7%	17.5%
S	0.4150	0.4159	0.0181	8.2%	9.9%	7.5%
NO_3	0.4711	0.4612	0.0127	16.3%	29.6%	9.0%
SO_4	2.1255	2.1128	0.0224	4.5%	6.2%	1.6%
OC	1.1257	1.1440	0.0827	24.0%	38.7%	15.4%
EC	0.2375	0.2255	0.0269	29.2%	46.5%	21.6%

The expected uncertainty in the average concentrations, $E\_rp(\overline{X})$, calculated from the relative precisions and the observed relative differences, $O\_rd(\overline{X})$, in the corresponding paired averages are given in Table E.7. In unbiased equivalent measurements, $O\_rd(\overline{X})/2$ should rarely be greater than $3* E\_rp(\overline{X})$. This was the case for S, Fe, Si, Ti, and EC, which had no examples of absolute value $(O\_rd(\overline{X})/2)>3* E\_rp(\overline{X})$ for any of the data aggregates examined. Nitrate had one site, Phoenix, that failed to meet this criterion. However, SO_4, OC, Al, and Ca all had two or more of the six to seven data aggregates fail this criterion, implying that the measurement errors are not averaging out as would be expected in the case of purely random independent errors.

In the case of SO_4, the observed relative differences in the paired means, $O\_rd(\overline{X})$, are all quite low, less than 4% in absolute terms, but not as low as would be expected, given the relative precision observed in the same data aggregates. The surprisingly high $O\_rd(\overline{X})$ values observed for OC, Ca (up to 14% in absolute terms), and Al (up to 41% in absolute terms) likely indicate that there are important errors affecting the accurate quantification of these parameters. The discrepancy between the observed errors in the means, $O\_rd(\overline{X})$, and the expected errors, $E\_rp(\overline{X})$, does not necessarily imply that the characteristic relative precisions are too low. The discrepancy could also result from a lack of total independence in the uncertainties of individual measurements. For example, a flow bias in one of the collocated samplers related to an annual calibration would be shared among all measurements from that module for the year and therefore would not "cancel out". The model for the expected uncertainty in the means, E_rp($\overline{X}$), is based upon the assumption of independent and identically distributed (i.i.d) random errors and therefore is an invalid model for the case where there are nonrandom or dependent measurement errors, and E_rp($\overline{X}$) will underestimate the true uncertainty in the mean value under these conditions. The results suggest that there are dependent nonrandom sampling errors that are large enough to significantly impact the equivalence of measurements from paired IMPROVE samplers.

Given that, for at least a portion of the parameters involved in the RCFM model, the relative precision does not provide a meaningful estimate of the observed uncertainty in mean concentrations using the parametric E_rp($\overline{X}$) model, a better statistic for defining the comparability of the mean concentrations might be the relative difference in the paired mean concentrations, O_rd($\overline{X}$). Thus, comparability will be defined as the relative difference that could be expected to be found between any two mean concentrations when no actual difference exists between the sampled populations. Based on the standard deviation of the O_rd($\overline{X}$) values observed in the site-parameter aggregates, a conservative two-sigma estimate of the comparability would be 2 times the relative precision of the mean concentrations (2*STDEV(O_rd($\overline{X}$)) This would give us estimates of 3–5% for S and SO_4; 7–10% for NO_3, Fe, Si, and Ti; 12–20% for Ca, OC, and EC; and 37% for Al (Table E.16 in section E.8 lists all the exact values). These comparability estimates provide an upper bound for the precision of either measurement but exclude any contribution from possible systematic errors. Mean concentrations utilizing longer averaging periods would be expected to have a higher degree of precision for the unbiased parameters than observed here, so the actual comparability for most or the average concentrations reported in Chapters 2 and 3 is likely better. However, using what is likely an upper bound estimate of the typical comparability of the mean concentrations should aid in not overinterpreting the corresponding spatial and temporal trends.

Table E.7. The observed relative differences in the paired mean concentrations, O_rd($\bar{X}$), and the expected uncertainty for the mean concentrations in the parameter-only aggregate (network), and the parameter-site aggregates of the in-network IMPROVE collocated data. The observed relative difference in paired means is given by O_rd($\bar{X}$)= $\frac{(\bar{X}-\bar{Y})*2}{(\bar{X}+\bar{Y})}$ where $\bar{X}$ and $\bar{Y}$ represent the paired means. The expected relative precision in paired means is given by E_rp($\bar{X}$)= $\frac{rp}{\sqrt{n}}$ where rp is the relative precision for the parameter of interest.

	Network	Network	MEVE1	MEVE1	OLYM1	OLYM1	PMRF1	PMRF1	SAFO1	SAFO1	TRCR1	TRCR1	PHOE1	PHOE1		
	O_rd	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**	**O_rd**	**E_rp**
Al	1.6%	1.0%	-12.9%	2.8%	6.3%	2.2%	41.1%	2.9%	-5.6%	1.8%	12.1%	1.7%	3.9%	3.2%		
Ca	-5.0%	0.7%	-14.1%	3.7%	-5.7%	1.1%	1.3%	1.5%	-8.9%	1.3%	-6.1%	1.4%	0.8%	2.7%		
Fe	-1.6%	0.7%	-10.8%	3.1%	-4.7%	1.6%	2.3%	1.4%	-3.6%	1.2%	0.1%	2.1%	0.4%	2.6%		
Si	-0.9%	0.8%	-10.2%	2.9%	-3.0%	1.6%	4.1%	2.2%	-2.7%	1.2%	2.2%	2.0%	2.0%	2.8%		
Ti	-1.3%	0.8%	-9.0%	3.0%	-4.3%	1.7%	1.9%	2.4%	-3.9%	1.4%	-1.0%	2.2%	1.0%	2.7%		
S	-0.2%	0.3%	-1.9%	0.8%	-3.3%	0.8%	0.1%	0.9%	2.0%	0.7%	-4.6%	0.9%	-2.7%	1.1%		
	Network	**Network**	**BIBE1**	**BIBE1**	**BLMO1**	**BLMO1**	**FRRE1**	**FRRE1**	**GAMO1**	**GAMO1**	**LAVO1**	**LAVO1**	**PHOE1**	**PHOE1**	**MACA1**	**MACA1**
NO_3	2.1%	0.6%	0.7%	1.2%	0.7%	0.5%	0.4%	0.5%	6.3%	1.4%	3.7%	1.9%	8.2%	1.7%	0.2%	0.6%
SO_4	0.6%	0.2%	1.2%	0.6%	2.5%	0.7%	1.6%	0.2%	3.0%	0.5%	2.8%	0.4%	3.5%	0.4%	-0.7%	0.1%
	Network	**Network**	**EVER1**	**EVER1**	**HEGL1**	**HEGL1**	**HOOV1**	**HOOV1**	**MELA1**	**MELA1**	**SAWE1**	**SAWE1**	**PHOE1**	**PHOE1**	**SENE1**	**SENE1**
OC	-1.6%	0.9%	-7.8%	1.2%	-7.4%	2.2%	-8.8%	2.9%	-5.3%	1.6%	14.5%	3.5%	-0.9%	1.0%	12.3%	2.5%
EC	5.2%	1.0%	-1.6%	2.3%	-6.3%	2.9%	3.4%	3.1%	-0.7%	2.2%	5.8%	2.2%	14.0%	1.5%	-0.4%	3.6%

E.5 STN IN-NETWORK COLLOCATED DATA

In-network collocated data for the STN were not examined in the same manner as the in-network IMPROVE and cross-network IMPROVE and STN collocated data sets. Relative precision estimates were reported in graphical form for the MASS, RAAS, and SASS sampler types for the 2002–2004 time period in the October 2005 STN newsletter [Rice, 2005]. A network average for each parameter was calculated for the purposes of this report by picking the values for each sampler type from the chart. While it is not specified in the report, it is assumed that these estimates do not include values less than three times the reported mdl. The relative precision calculated for each parameter is reported in Table E.8.

Table E.8. Relative Precisions observed in the in-network STN collocated data.

	Al	Ca	Fe	Si	Ti	S	SO_4	NO_3	EC	OC
rp	40%	22%	20%	23%	32%	11%	16%	18%	23%	20%

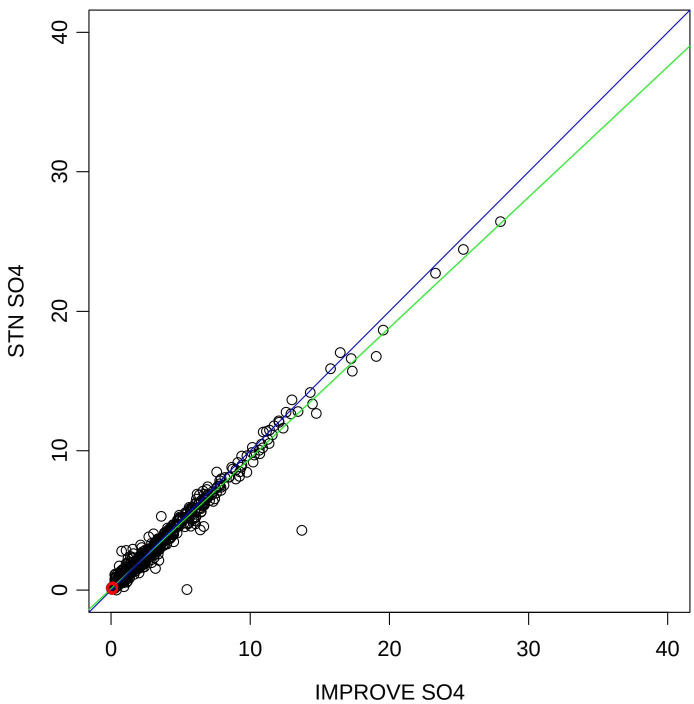
E.6 IMPROVE-STN CROSS-NETWORK COLLOCATED DATA

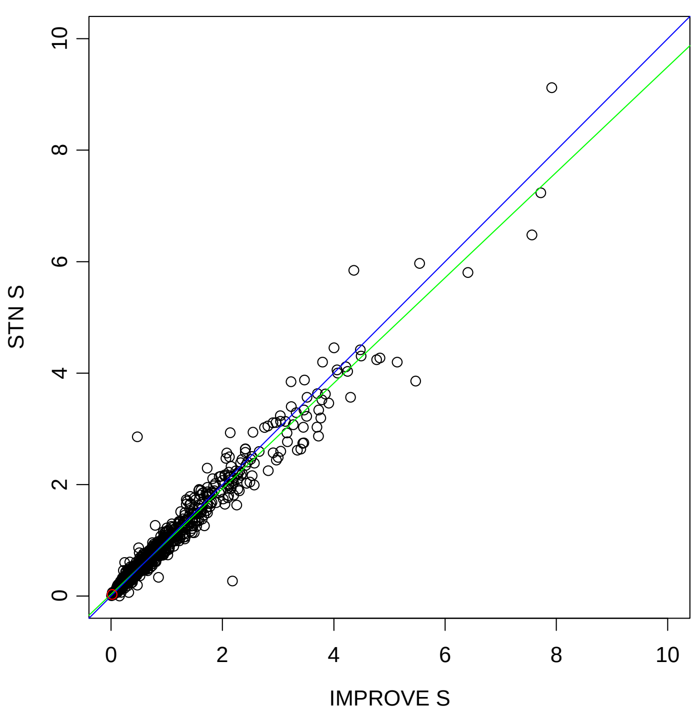
Step 1

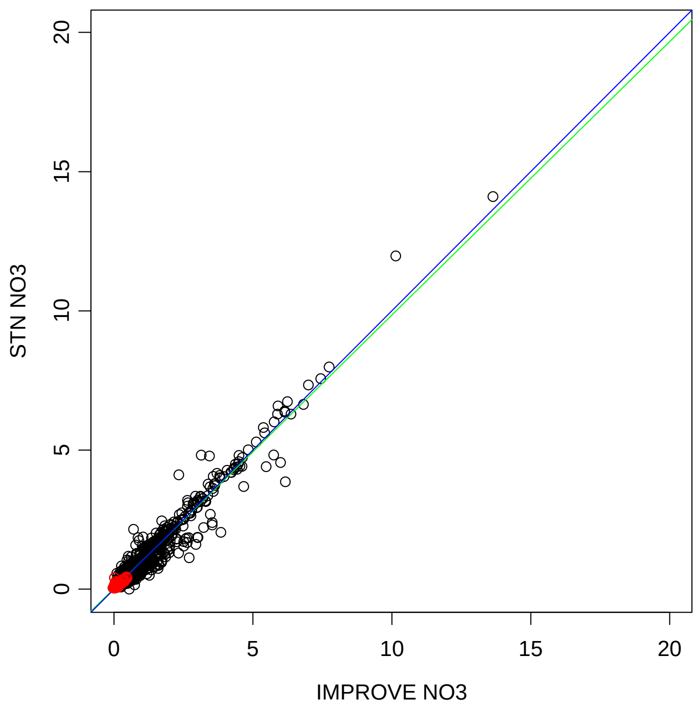
The same procedure performed in section E.4 for the in-network IMPROVE collocated data was followed; the IMPROVE-STN collocated data population was explored to assess if the collocated data appear to follow a 1:1 relationship (Figure E.6, panels a–k). The cross-network collocated data had lower R^2 values for all of the parameters examined (Table E.9), as compared to the in-network IMPROVE collocated data (Table E.2). The R^2 values ranged from 0.98 for SO_4 to 0.19 for Al in the cross-network data, as compared to 1 for SO_4 and 0.75 for Al in the in-network data. The slopes for the parameters typically present at higher concentrations (mean concentration > 0.5 µg/m^3) exhibited OLS slopes in the range of 0.87–1.04, similar to the range for all parameters in the in-network IMPROVE collocated data. The elements typically associated with soil—Al, Fe, Ca, Si, and Ti—all had slopes less than 0.55. The OLS intercepts ranged from 0.00 to 0.11 for all parameters besides the IMPROVE and STN OC fit where the intercept was 1.15. The large bias in the OC comparison between IMPROVE and the standard STN OC values was expected due to the differences in analytical methods and the non-blank-corrected STN data. The correlation and regression results suggest a weaker match to a 1:1 model for the IMPROVE-STN collocated data as compared the IMPROVE-IMPROVE collocated data, with the comparisons between the soil elements and the standard STN OC measurements showing the greatest disagreements.

Table E.9. OLS slope and intercept coefficients, the R^2 value, and the paired mean values for the collocated concentration values are given for each parameter. The summary statistics include all available collocated data from the six IMPROVE-STN collocated samplers (Figure E.1).

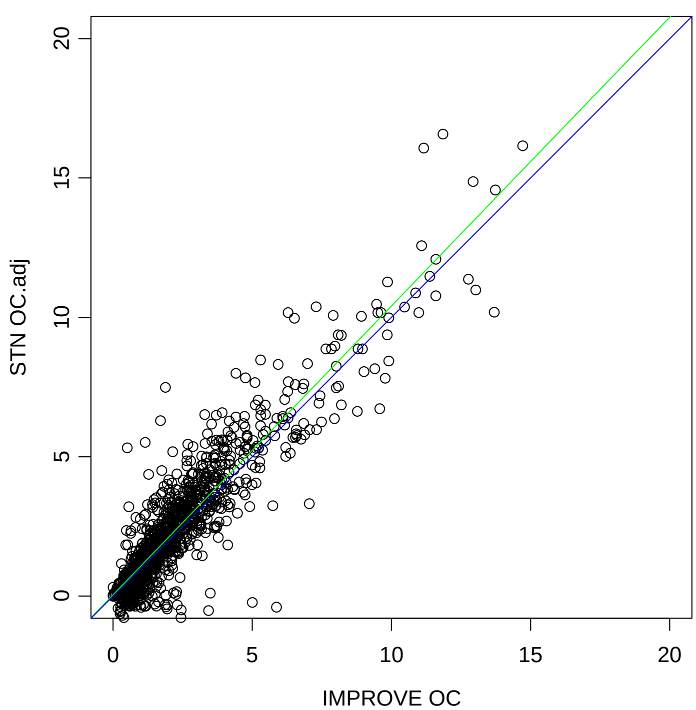
	IMPROVE Mean	STN Mean	Slope	Intercept	R^2
SO_4	2.28	2.25	0.94	0.11	0.98
S	0.76	0.75	0.95	0.03	0.96
NO_3	0.85	0.86	0.98	0.02	0.95
OC.adj	2.30	2.31	1.04	0.04	0.83
Si	0.26	0.15	0.53	0.01	0.77
OC	2.30	3.53	1.03	1.15	0.76
Ca	0.08	0.06	0.51	0.02	0.73
EC	0.58	0.55	0.87	0.05	0.72
Fe	0.11	0.08	0.55	0.02	0.70
Ti	0.01	0.01	0.54	0.00	0.59
Al	0.07	0.04	0.36	0.02	0.19


STN SO4
IMPROVE SO4
0
10
20
30
40

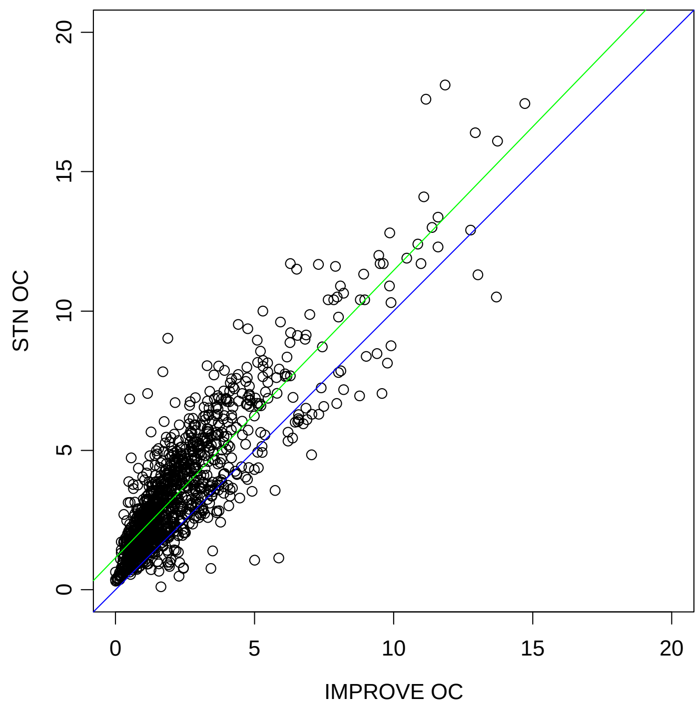

STN S
IMPROVE S
0
2
4
6
8
10



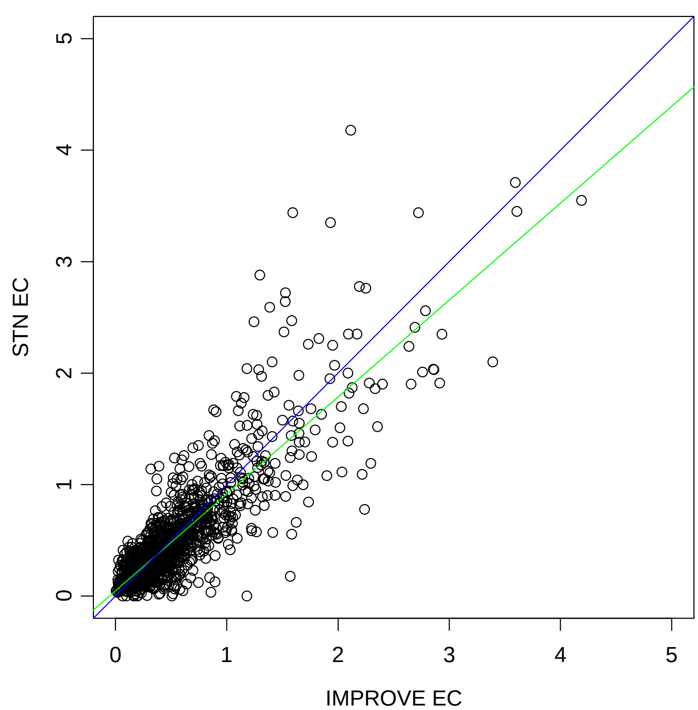
STN NO3
IMPROVE NO3
0
5
10
15
20



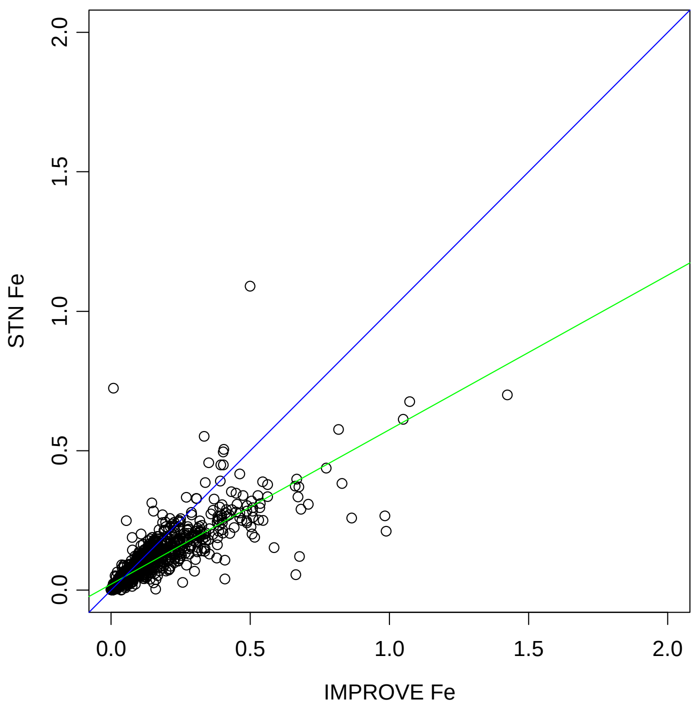
STN OC.adj
IMPROVE OC
0
5
10
15
20

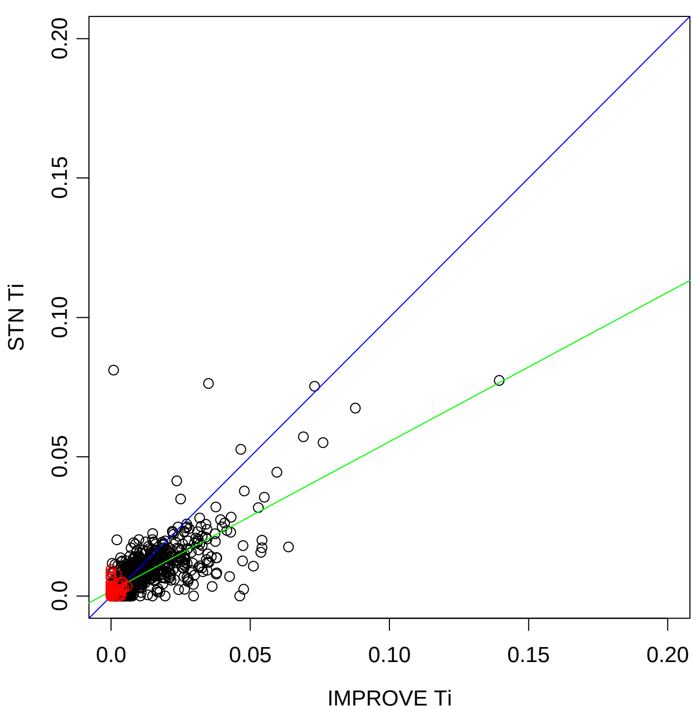


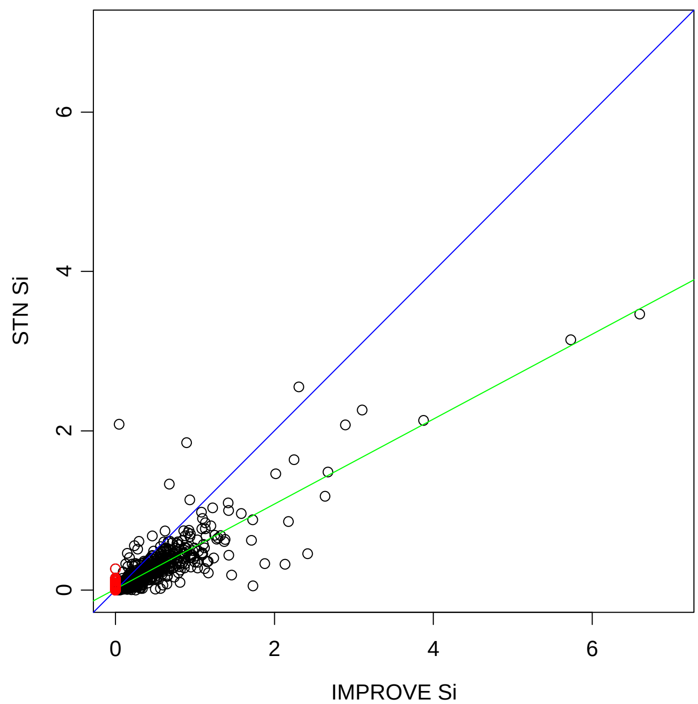
STN OC
IMPROVE OC
0
5
10
15
20

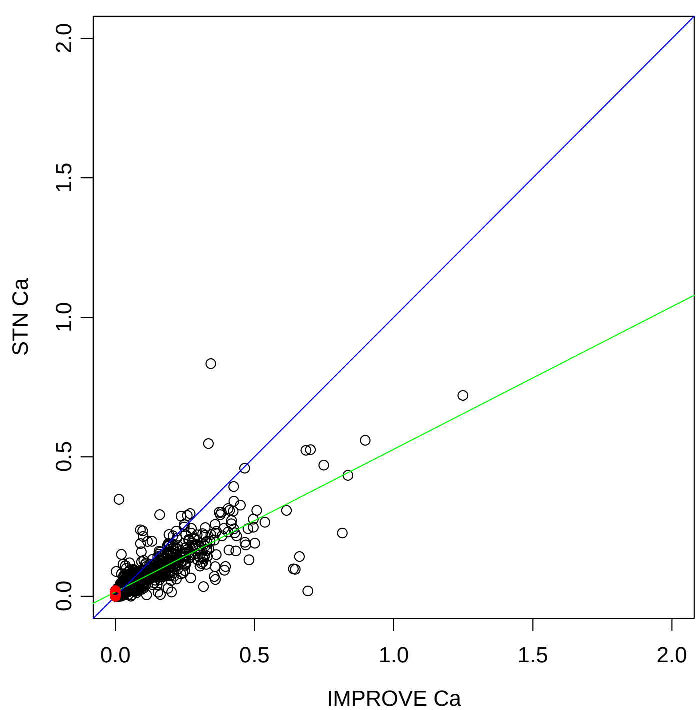


STN EC
IMPROVE EC
0
1
2
3
4
5


STN Fe
IMPROVE Fe
0.0
0.5
1.0
1.5
2.0


STN Ti
IMPROVE Ti
0.0
0.05
0.10
0.15
0.20


STN Si
IMPROVE Si
0
2
4
6


STN Ca
IMPROVE Ca
0.0
0.5
1.0
1.5
2.0

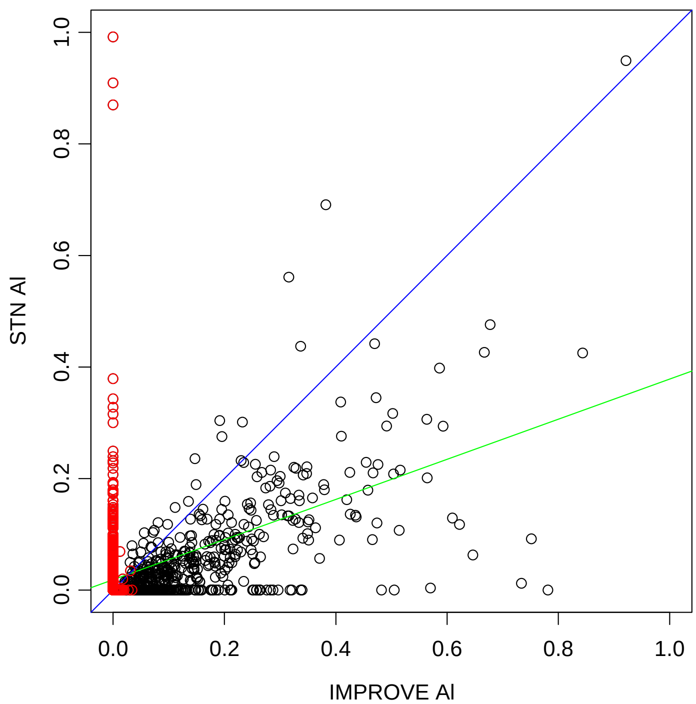


Figure E.6. a–k. Scatter plots of the cross-network collocated data from the IMPROVE and STN networks for the sites shown in Figure E.1. The IMPROVE sampler is plotted along the X axis and the STN sampler is plotted along the Y axis. Both the X and Y axes are in units of μg/m^3. IMPROVE concentration values less than 3*mdl are displayed in red. The 1:1 line is shown in blue and the OLS regression line is shown in green.

Step 2

Again, the same procedure performed in section E.4 for the in-network IMPROVE collocated data was followed; the populations of d_i were explored to address if there were any important relationships between the d_i and concentration or time and to assess if the measurement errors appear to be drawn from a normal distribution. The F-test for the equality of two standard deviations and the T-test with unequal variances were used to compare if the variability and central tendency of the d_i were significantly different in the low and high concentration groups and between 2002 and 2003.

Examining scatter plots of the d_i versus the average concentration of the sample pairs, the pattern of higher variability in the relative differences at lower concentrations is again present for most parameters (Figure E.7, panels a–k). Again, this effect is most apparent in the blank-corrected parameters. Obvious biases are apparent in the soil elements and the non-blank-corrected OC data, with the median d_i value indicating that IMPROVE is consistently reporting higher soil element concentrations and lower OC concentrations. Furthermore, the errors in most parameters are not symmetrically distributed around the 0 or median d_i line, indicating that additional, more subtle biases are affecting all of the parameters. In a comparison of the figures for the blank-corrected and uncorrected OC data, it becomes clear that while the blank correction of the STN OC measurements results in reduced bias between the IMPROVE and STN measurements (a median value closer to 0), on the average it increases the imprecision in the STN measurements (greater range in the relative errors and a much higher standard deviation) particularly at low concentrations (Figure E.7, Table E.10). This suggests that the STN blank correction procedure is not properly taking into account either spatial or temporal shifts in the blank values and thereby is increasing the imprecision in the STN measurements by subtracting off either too much or too little for particular samples.

The F-tests indicated significant differences between the variability in the low and high concentration groups for S, SO_4, NO_3, EC, OC.adj, Al, Fe, and Ti at the 0.05 alpha level (Table E.10). T-tests indicate significant differences between the low and high concentration mean values for every parameter besides Ca.

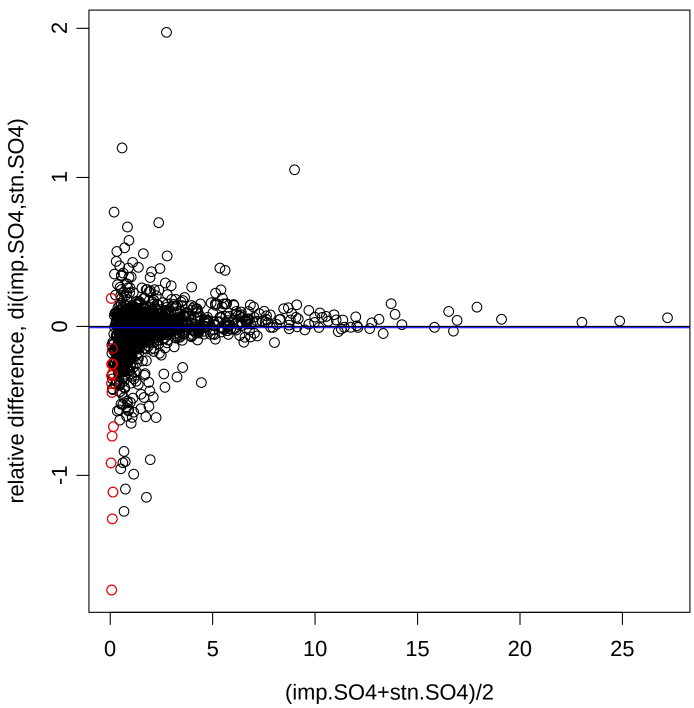
Changes over time in the precision of the measurements are also suggested in the cross-network collocated data for some of the parameters in Figure E.8, panels a–k. Significant differences in mean values from 2002 and 2003 were indicated by the T-tests for all parameters besides NO_3 at the 0.05 significance level (Table E.11). Significant differences in variability were indicated by the F-tests for S, NO_3, OC, OC.adj, and Ti.

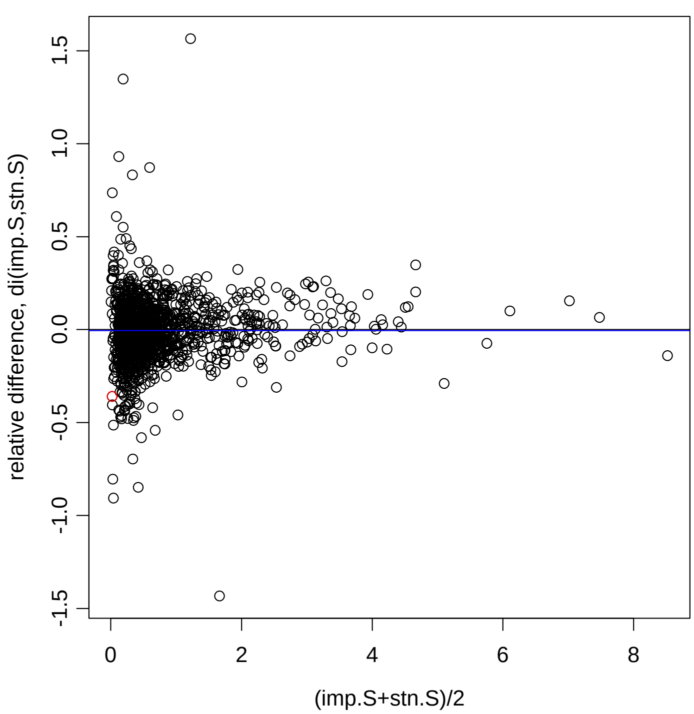
Table E.10. The mean, standard deviation, and number of valid values of the relative differences, d_i, in the low and high concentration groups for the cross-network collocated data. Pairs in bold are significantly different at a 0.05 significance level, based upon T-tests allowing for different variability and F-test for the equality of two standard deviations.

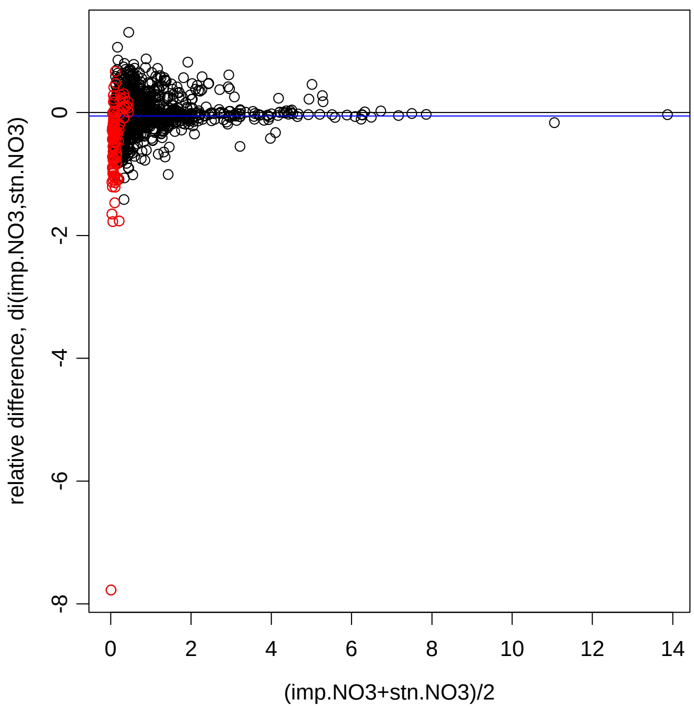
	S		**SO_4**		**NO_3**		**EC**		**OC**		**OC.adj**	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Mean	**-0.035**	**0.017**	**-0.099**	**0.028**	**-0.196**	**0.039**	**-0.027**	**0.157**	**-0.709**	**-0.316**	**0.465**	**-0.012**
Standard Deviation	**0.196**	**0.156**	**0.239**	**0.138**	**0.469**	**0.266**	**0.594**	**0.425**	0.366	0.368	**2.447**	**0.512**
N	623	667	628	658	633	665	575	622	581	617	581	617
	Al		**Ca**		**Fe**		**Si**		**Ti**			
	Low	High	Low	High	Low	High	Low	High	Low	High		
Mean	0.779	0.802	**0.057**	**0.342**	**0.063**	**0.304**	**0.485**	**0.596**	**-0.166**	**0.328**		
Standard Deviation	**0.641**	**0.480**	0.428	0.425	**0.379**	**0.356**	0.431	0.424	**0.661**	**0.500**		
N	137	225	626	642	636	652	539	643	535	629		

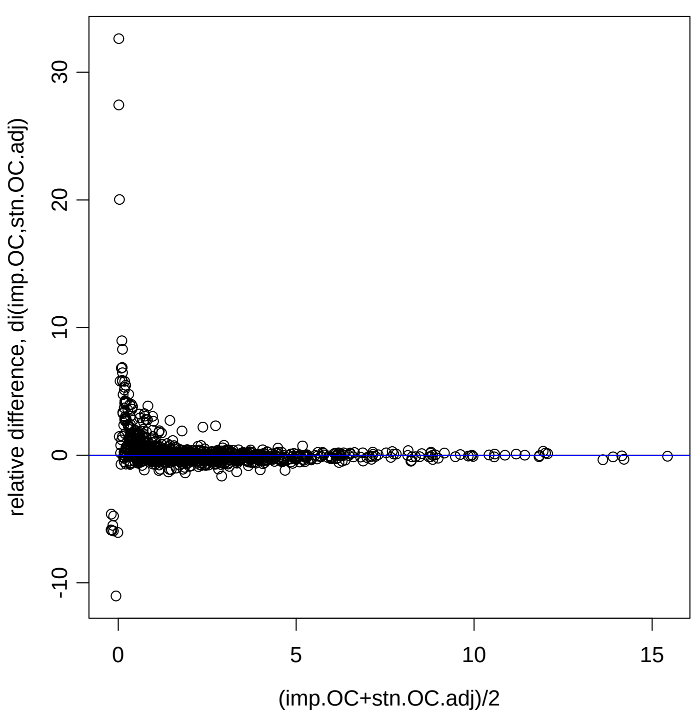
Table E.11. The mean, standard deviation, and number of valid values of the relative differences, d_i, in the first and second half of 2002 and 2003 for the cross-network collocated data. Pairs in bold are significantly different at a 0.05 significance level, based upon T-tests allowing for different variability and F-test for the equality of two standard deviations.

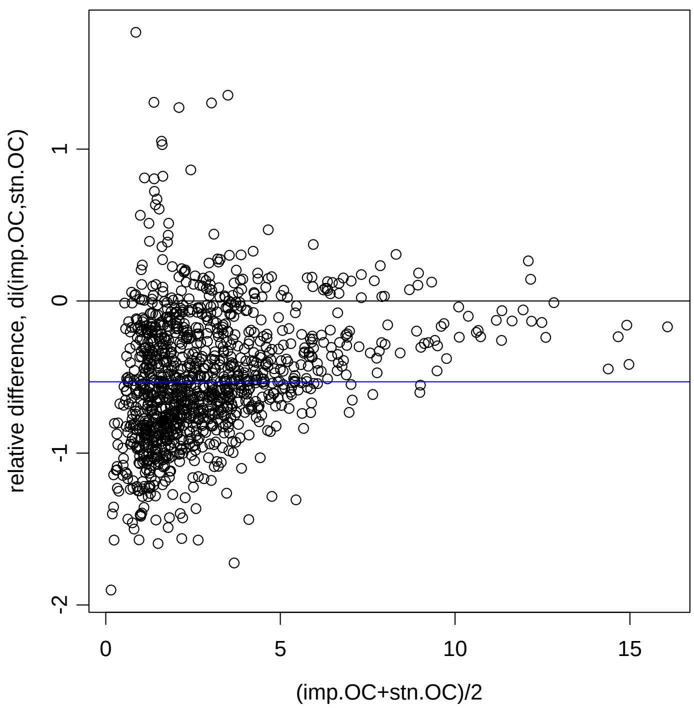
	S		**SO_4**		**NO_3**		**EC**		**OC**		**OC.adj**	
	2002	2003	2002	2003	2002	2003	2002	2003	2002	2003	2002	2003
Mean	**-0.034**	**0.014**	**-0.068**	**0.007**	-0.072	-0.053	**0.003**	**0.138**	**-0.538**	**-0.471**	**0.085**	**0.418**
Standard Deviation	**0.192**	**0.157**	0.205	0.195	**0.332**	**0.296**	0.528	0.507	**0.428**	**0.387**	**1.029**	**2.512**
N	626	532	615	535	627	535	599	464	599	465	599	465
	Al		**Ca**		**Fe**		**Si**		**Ti**			
	2002	2003	2002	2003	2002	2003	2002	2003	2002	2003		
Mean	**0.730**	**0.898**	**0.134**	**0.257**	**0.149**	**0.231**	**0.465**	**0.634**	**0.002**	**0.177**		
Standard Deviation	0.545	0.517	0.453	0.430	0.389	0.372	0.424	0.431	**0.649**	**0.575**		
N	251	101	624	524	627	527	584	481	586	488		

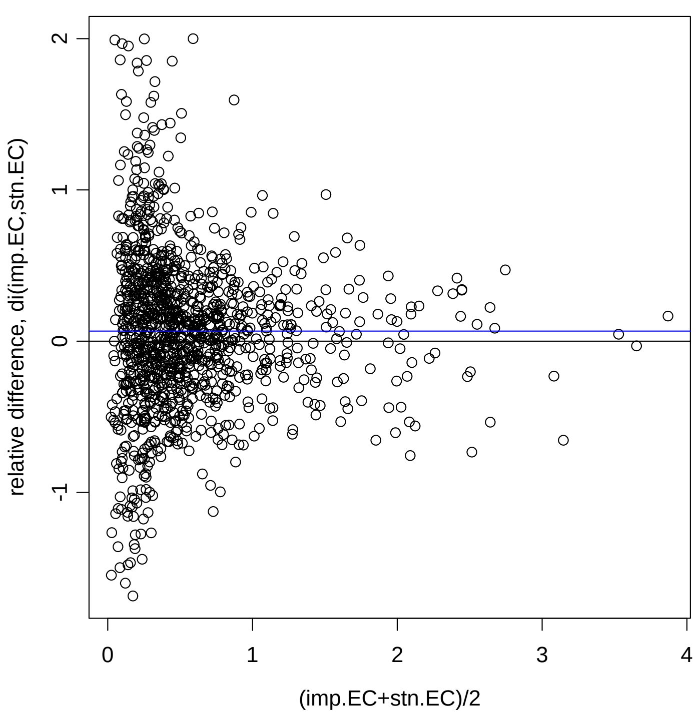

relative difference, di(imp.SO4,stn.SO4)
(imp.SO4+stn.SO4)/2

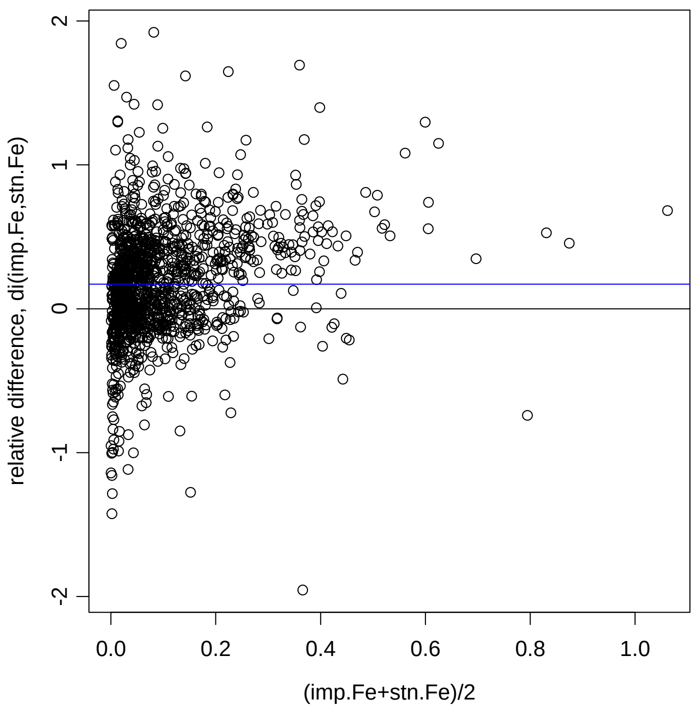

relative difference, di(imp.S,stn.S)
(imp.S+stn.S)/2

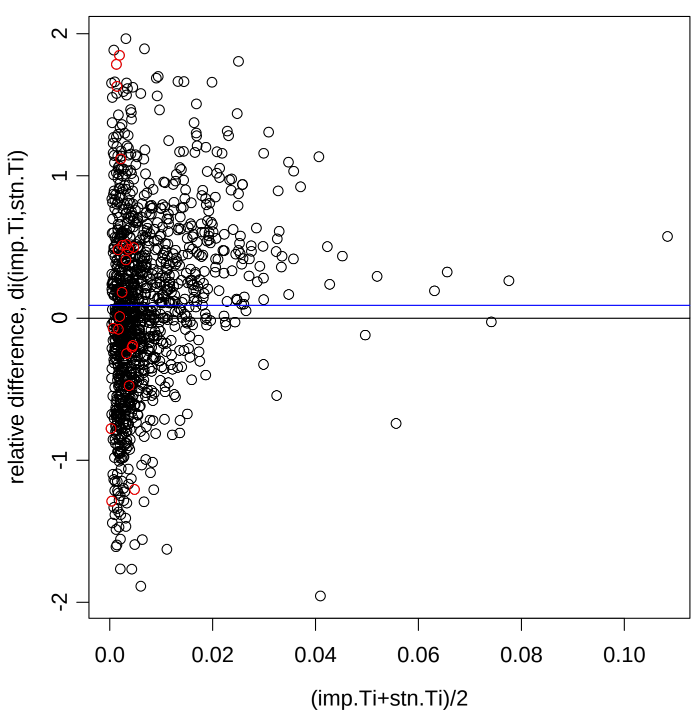

relative difference, di(imp.NO3,stn.NO3)
0
-2
-4
-6
-8
0
2
4
6
8
10
12
14
(imp.NO3+stn.NO3)/2

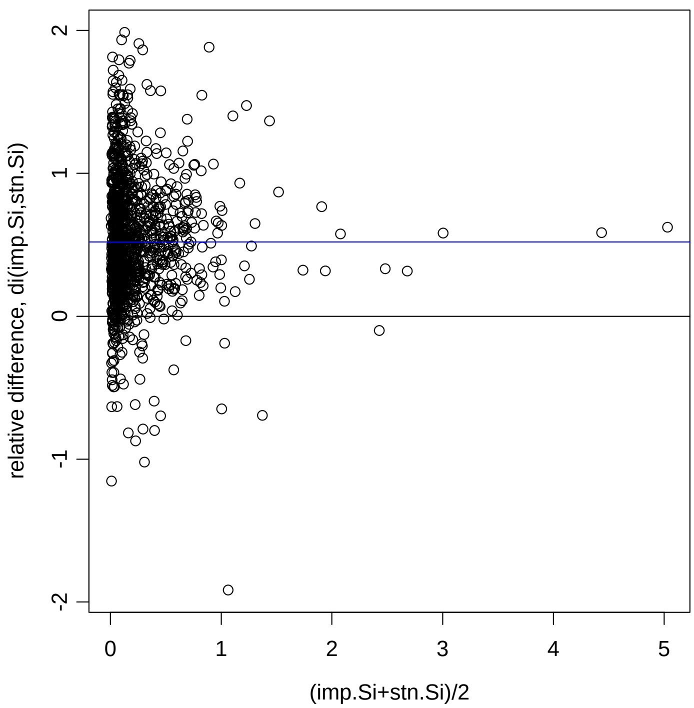

relative difference, di(imp.OC,stn.OC.adj)
30
20
10
0
-10
0
5
10
15
(imp.OC+stn.OC.adj)/2


relative difference, di(imp.OC,stn.OC)
(imp.OC+stn.OC)/2

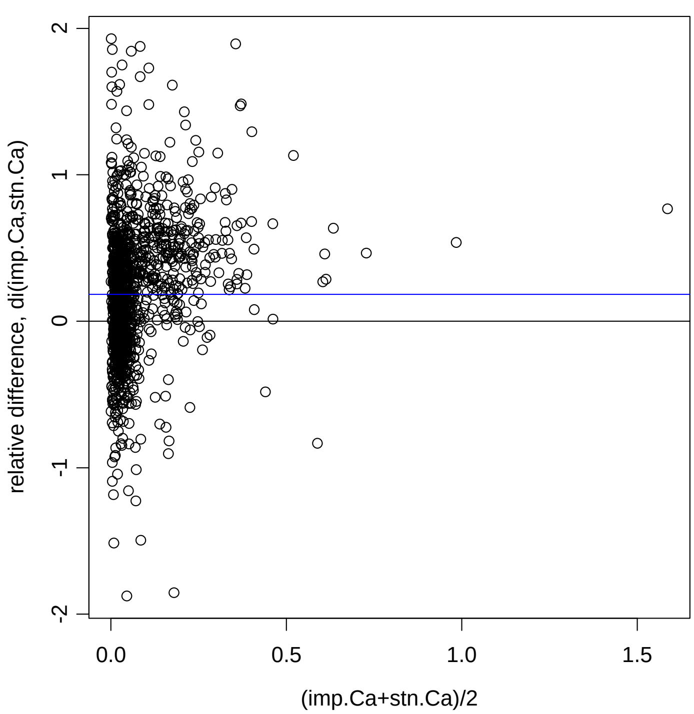

relative difference, di(imp.EC,stn.EC)
(imp.EC+stn.EC)/2


relative difference, di(imp.Fe,stn.Fe)
(imp.Fe+stn.Fe)/2
2
1
0
-1
-2
0.0
0.2
0.4
0.6
0.8
1.0


relative difference, di(imp.Ti,stn.Ti)
(imp.Ti+stn.Ti)/2
2
1
0
-1
-2
0.0
0.02
0.04
0.06
0.08
0.10



relative difference, di(imp.Si,stn.Si)
(imp.Si+stn.Si)/2
-2
-1
0
1
2
0
1
2
3
4
5



relative difference, di(imp.Ca,stn.Ca)
(imp.Ca+stn.Ca)/2
-2
-1
0
1
2
0.0
0.5
1.0
1.5

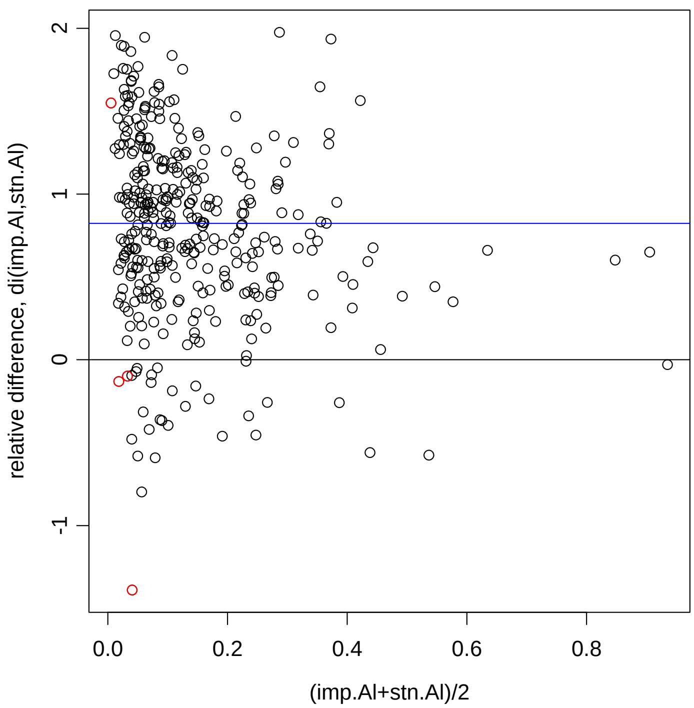
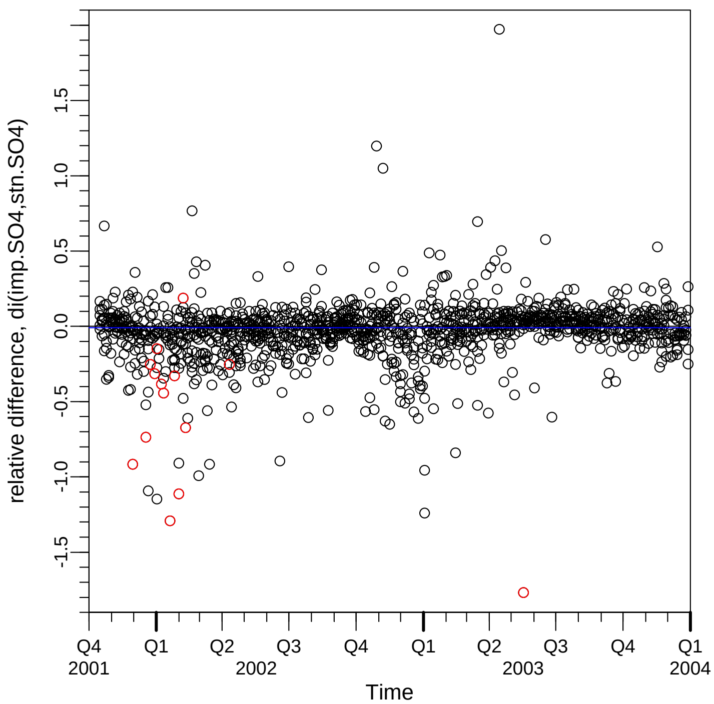
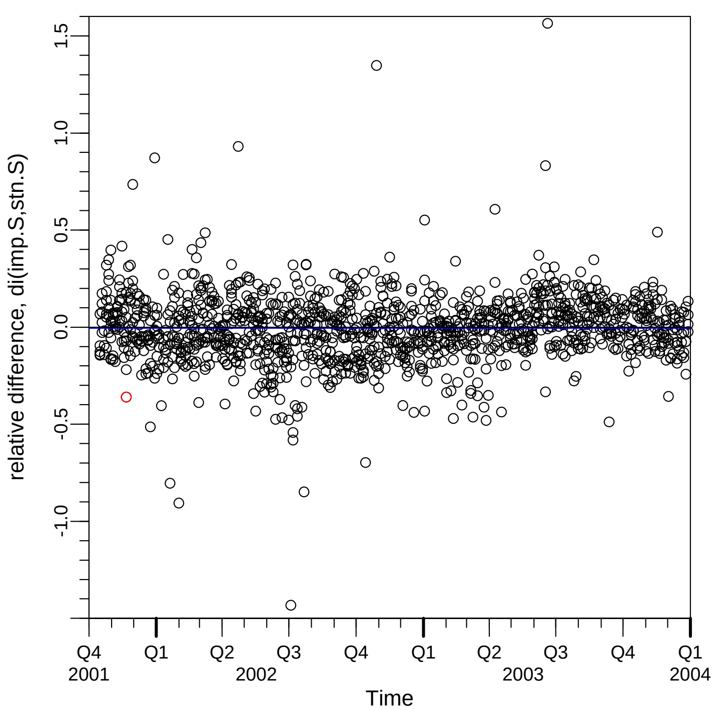
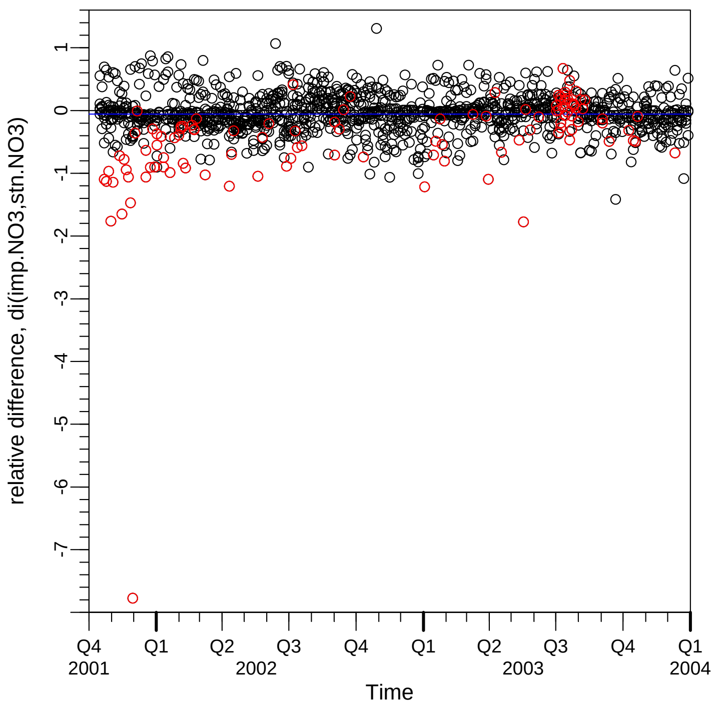
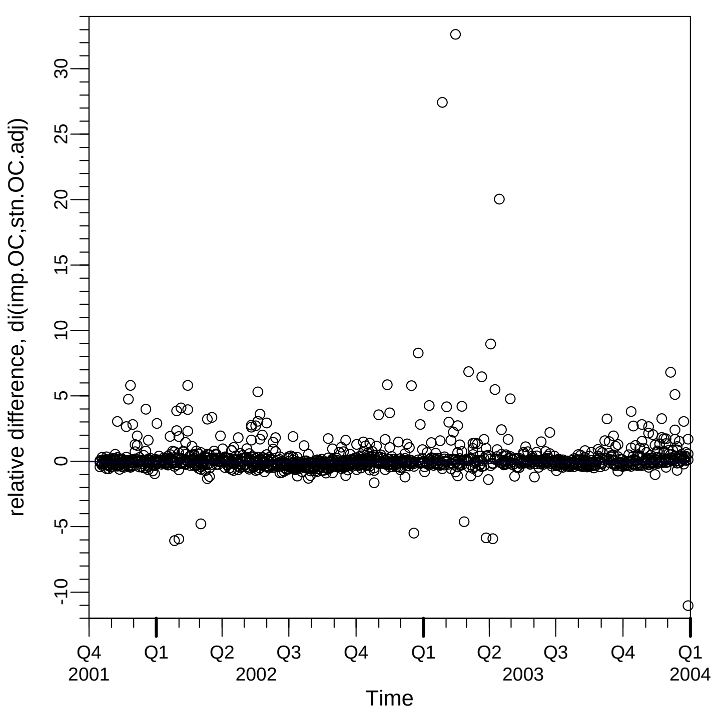


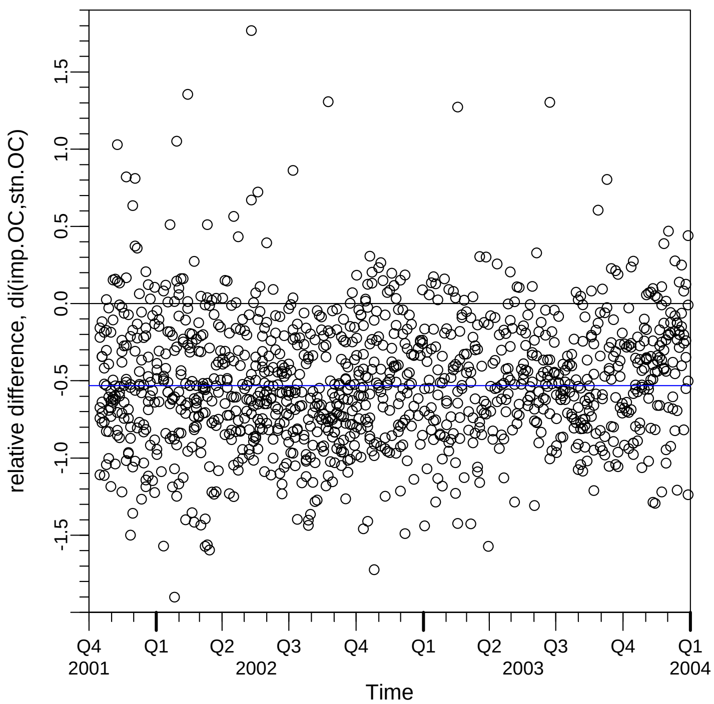
Figure E.7. a–k. Scatter plots of the relative differences of the paired samples, d_i, versus the average concentrations of the sample pairs for the cross-network collocated data from the IMPROVE and STN networks for the sites shown in Figure E.1. The X axis has units of µg/m^3 and the Y axis is unitless. IMPROVE concentration values less than 3*mdl are displayed in red. The blue line is the median d_i value.

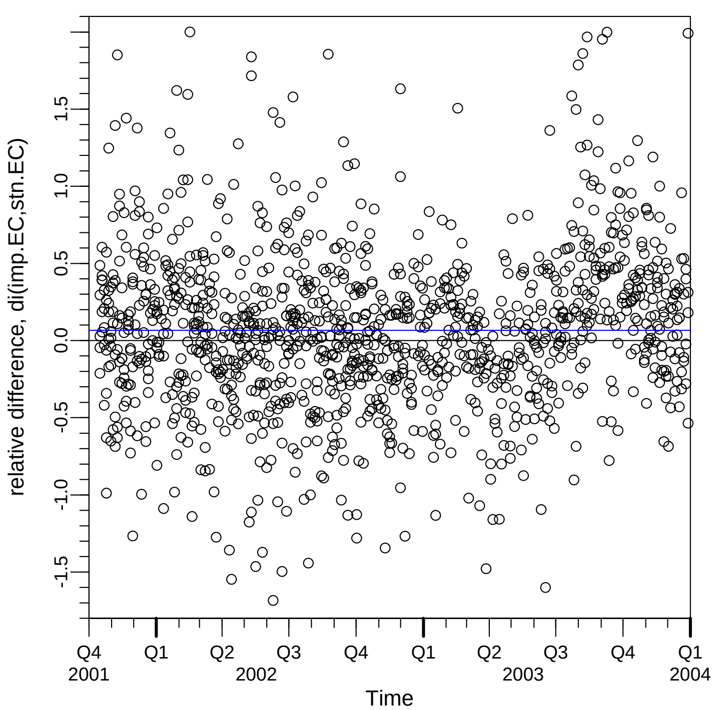

relative difference, di(imp.SO4,stn.SO4)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time

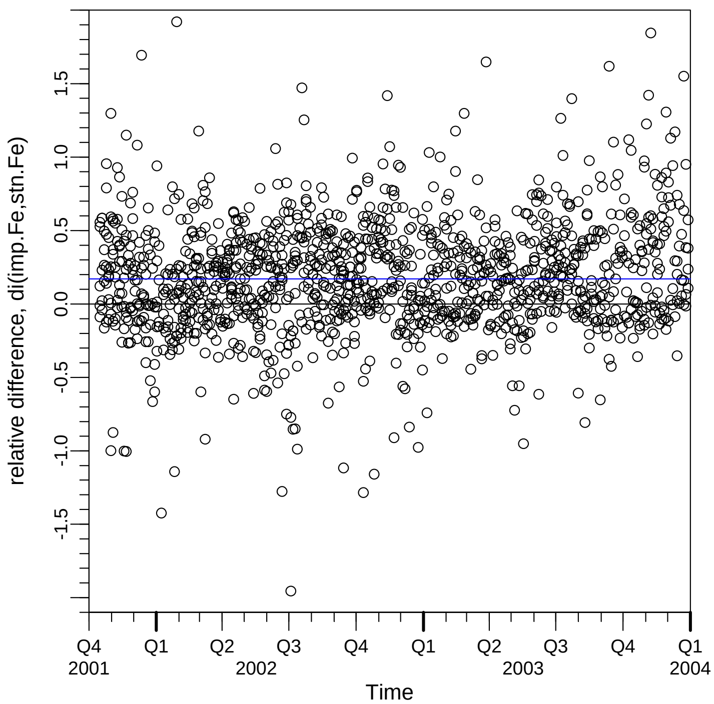

relative difference, di(imp.S,stn.S)
1.5
1.0
0.5
0.0
-0.5
-1.0
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time

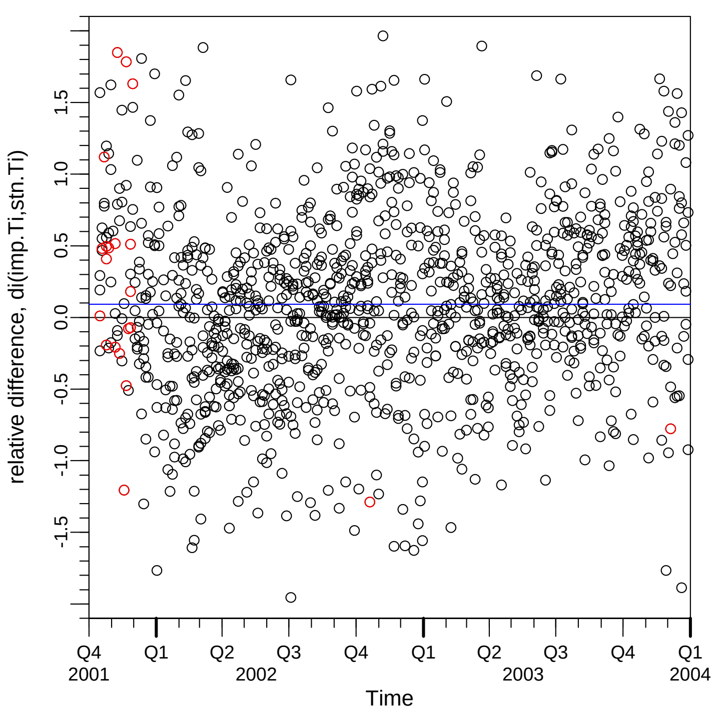

relative difference, di(imp.NO3,stn.NO3)
1
0
-1
-2
-3
-4
-5
-6
-7
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time

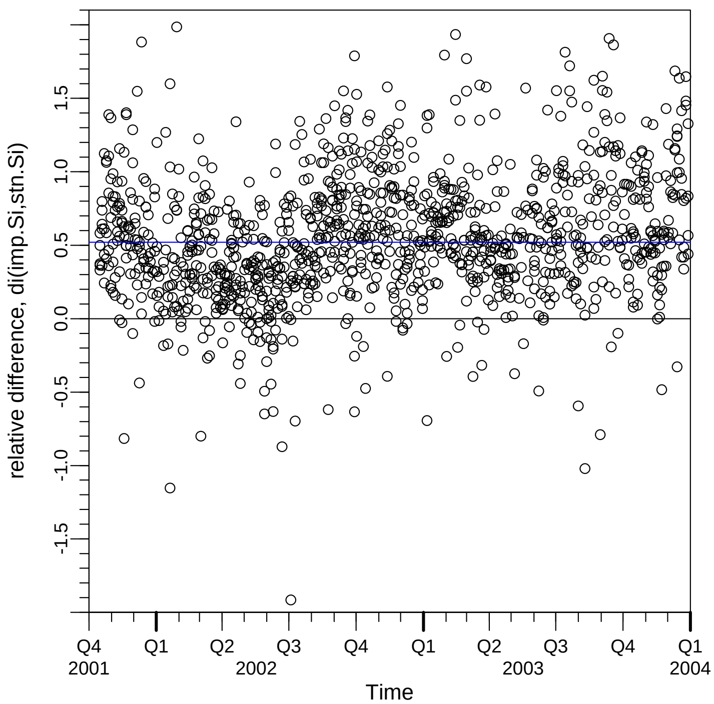

relative difference, di(imp.OC,stn.OC.adj)
30
25
20
15
10
5
0
-5
-10
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time


relative difference, di(imp.OC,stn.OC)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time

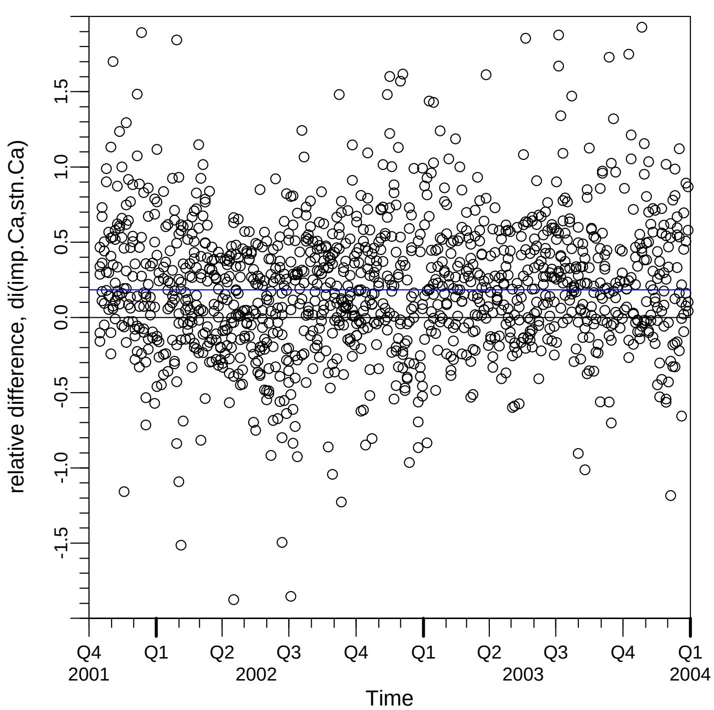

relative difference, di(imp.EC,stn.EC)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Q4
2001
Q1
Q2
2002
Q3
Q4
Q1
Q2
2003
Q3
Q4
Q1
2004
Time


relative difference, di(imp.Fe,stn.Fe)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Q4 2001
Q1
Q2 2002
Q3
Q4
Q1
Q2 2003
Q3
Q4
Q1 2004
Time


relative difference, di(imp.Ti,stn.Ti)
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Q4 2001
Q1
Q2 2002
Q3
Q4
Q1
Q2 2003
Q3
Q4
Q1 2004
Time



relative difference, di(imp.Si,stn.Si)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Q4 2001
Q1
Q2 2002
Q3
Q4
Q1
Q2 2003
Q3
Q4
Q1 2004
Time



relative difference, di(imp.Ca,stn.Ca)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Q4 2001
Q1
Q2 2002
Q3
Q4
Q1
Q2 2003
Q3
Q4
Q1 2004
Time

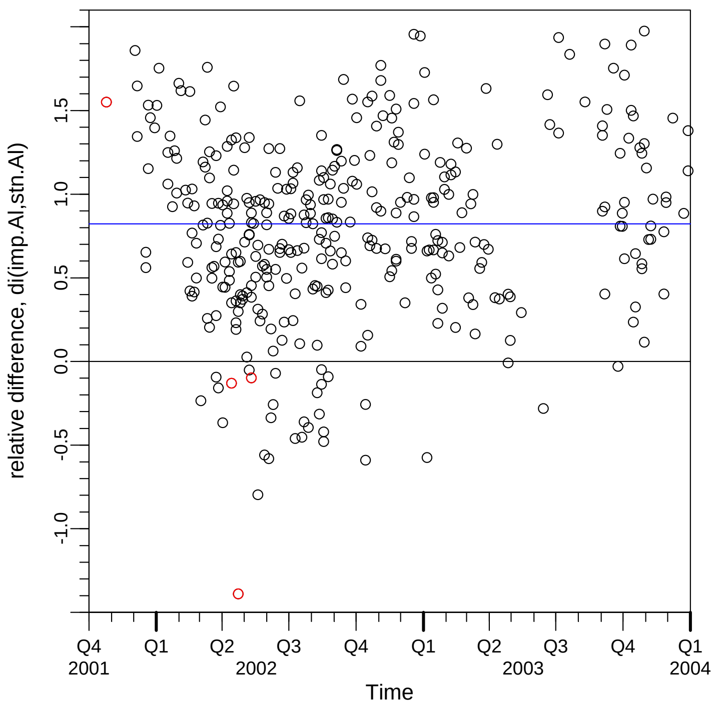
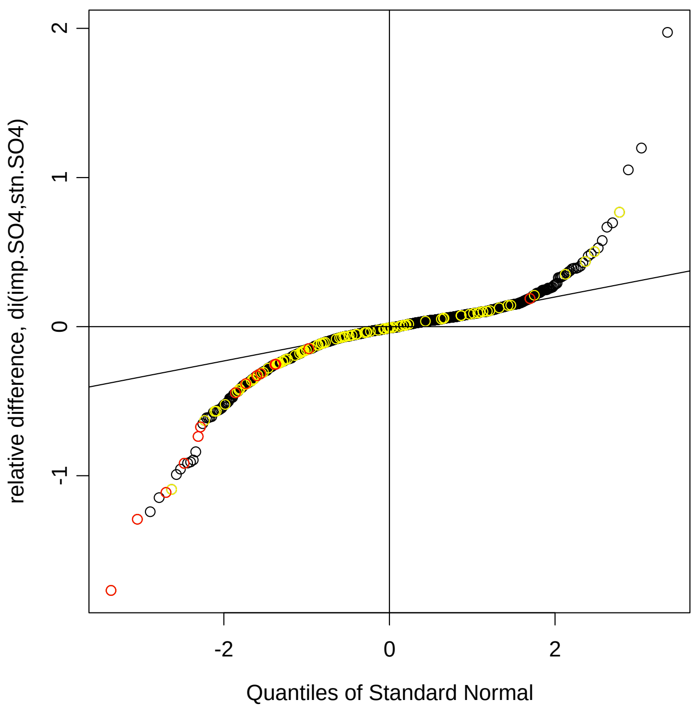
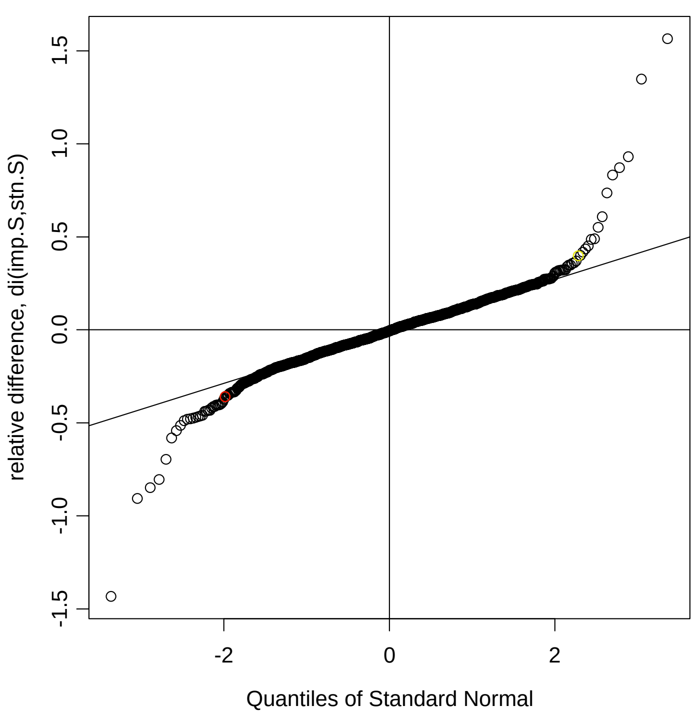


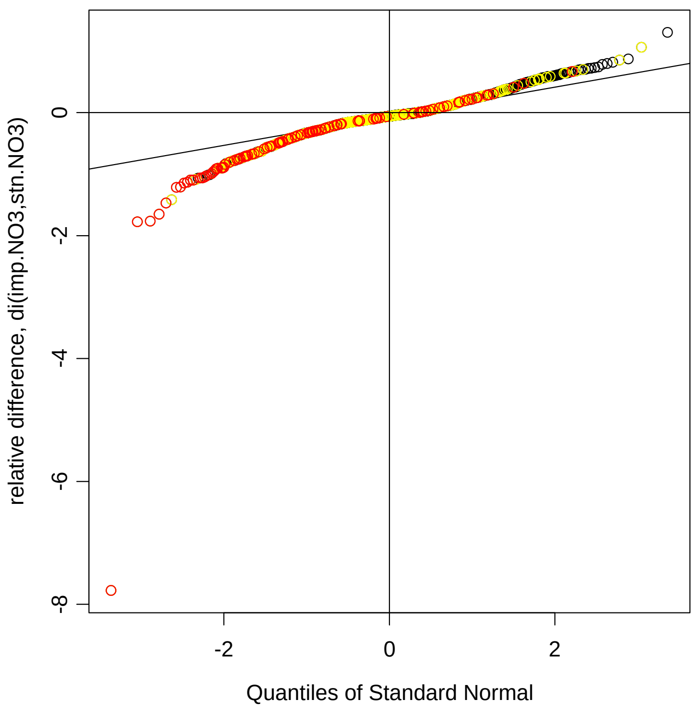
Figure E.8. a–k. Time series of the relative differences of the paired samples, d_i, for the collocated data from the IMPROVE and STN networks for the sites shown in Figure E.1. The X axis has units of time and the Y axis is unitless. IMPROVE concentration values less than 3*mdl are displayed in red. The blue line is the median d_i value.

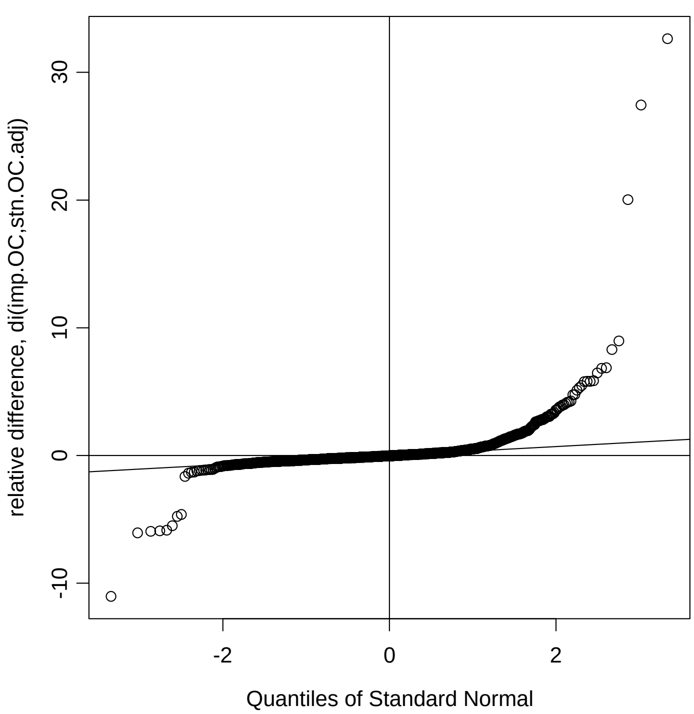
In the case of the cross-network collocated data, the observed measurement errors for all the major parameters again clearly deviate from normality, with the extreme errors larger than would be expected for a single Gaussian distribution (Figure E.9, panels a–k). Nonsymmetry in the length of the tails of the density curves is most apparent for Al, OC, OC.adj, and NO_3. Again, the deviations from the idealized measurement errors are not unexpected in trace chemical analysis, but require special caution on the part of data analysts to ensure that any statistical techniques they use are robust to heteroscedastic and potentially nonnormal measurement errors.

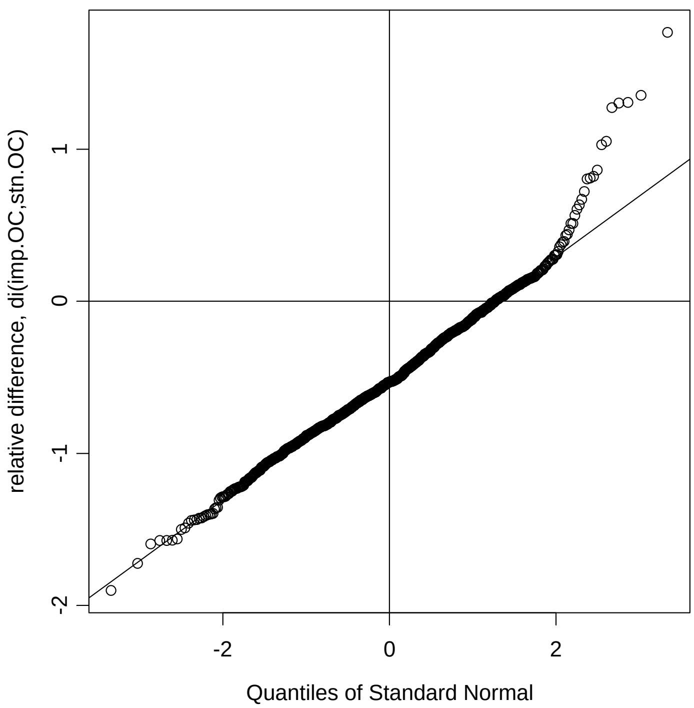
Furthermore, biases in the soil elements and the non-blank-adjusted OC data are immediately noticeable with the central tendency of the population not equal to 0. However, the five soil elements exhibit differing degrees of bias as measured by the distance between the d_i curve and the point(0,0). Biases between the measurement methods are of specific concern to the data analyst since systematic errors by their nature due not "cancel out" and need different attention than do random errors. In the case of significant biases, the data can be dealt with in a variety of ways, depending on the nature of the systematic measurement error and the analyst's requirements. Systematic biases can be corrected if it is known which measurement set is in error. Alternatively, when the biases are systematic but the correct method is unknown, if the analyst is primarily interested in comparing the relative values from the two measurement methods, the data can be calibrated to the method of primary interest to the user. A third option is to account for the bias in the total measurement uncertainty estimate; however, this considerably widens the confidence intervals around the measurements. The biases noted in this analysis are not quantified here nor are they included in the reported comparability estimates in the following sections. Biases will likely be dealt with more fully in future work and are noted here primarily to alert the user to their existence and aid in the interpretation of the comparability estimates developed in this report.

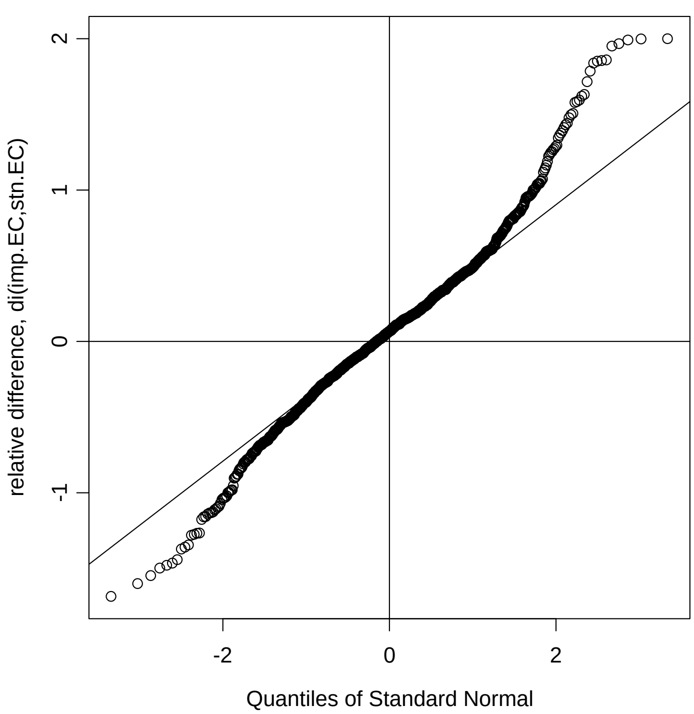

relative difference, di(imp.SO4,stn.SO4)
Quantiles of Standard Normal
-2
0
2
-1
0
1
2

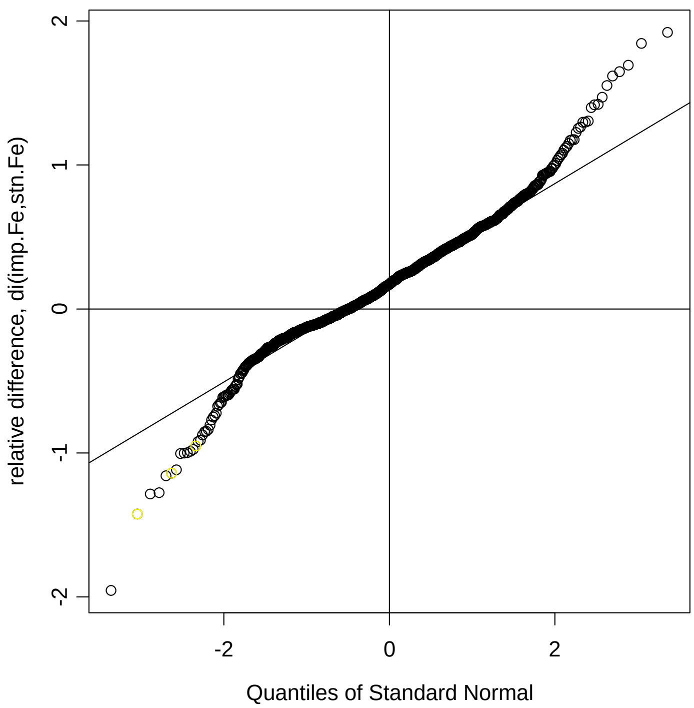

relative difference, di(imp.S,stn.S)
Quantiles of Standard Normal
-2
0
2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

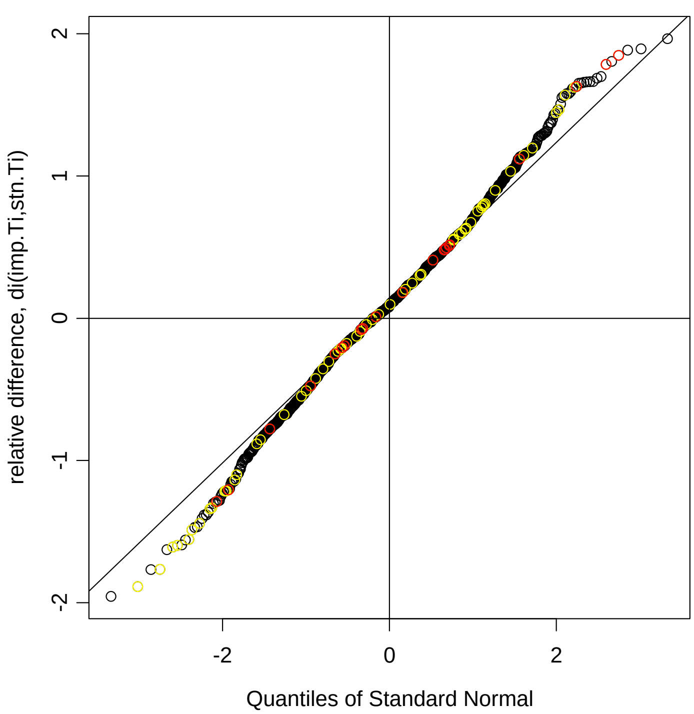

relative difference, di(imp.NO3,stn.NO3)
0
-2
-4
-6
-8
-2
0
2
Quantiles of Standard Normal

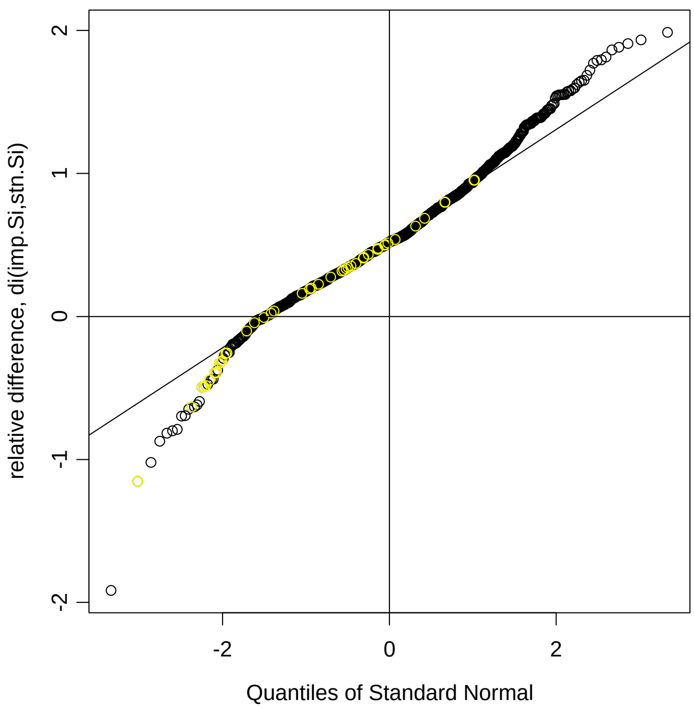

relative difference, di(imp.OC,stn.OC.adj)
30
20
10
0
-10
-2
0
2
Quantiles of Standard Normal

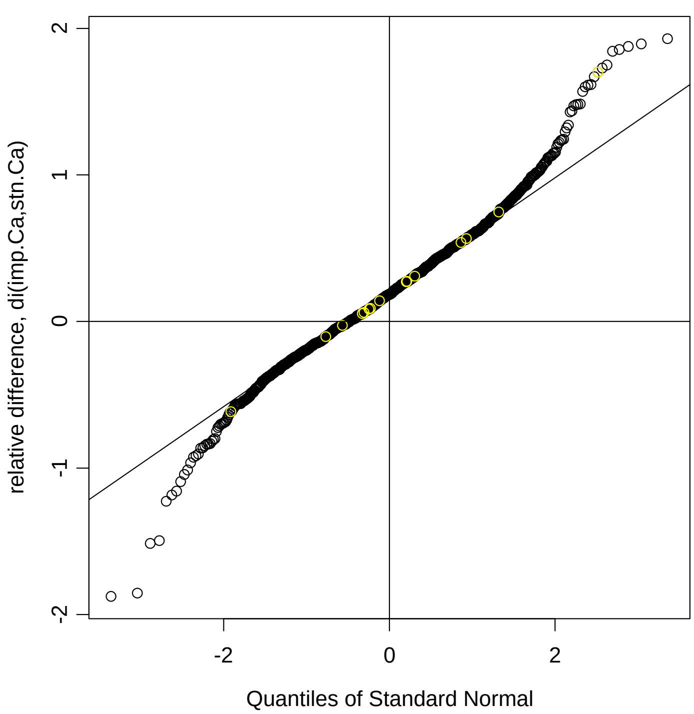

relative difference, di(imp.OC,stn.OC)
1
0
-1
-2
-2
0
2
Quantiles of Standard Normal


relative difference, di(imp.EC,stn.EC)
2
1
0
-1
-2
0
2
Quantiles of Standard Normal


relative difference, di(imp.Fe,stn.Fe)
Quantiles of Standard Normal
-2
-1
0
1
2


relative difference, di(imp.Ti,stn.Ti)
Quantiles of Standard Normal
-2
-1
0
1
2


relative difference, di(imp.Si,stn.Si)
Quantiles of Standard Normal
2
1
0
-1
-2
-2
0
2


relative difference, di(imp.Ca,stn.Ca)
Quantiles of Standard Normal
2
1
0
-1
-2
-2
0
2

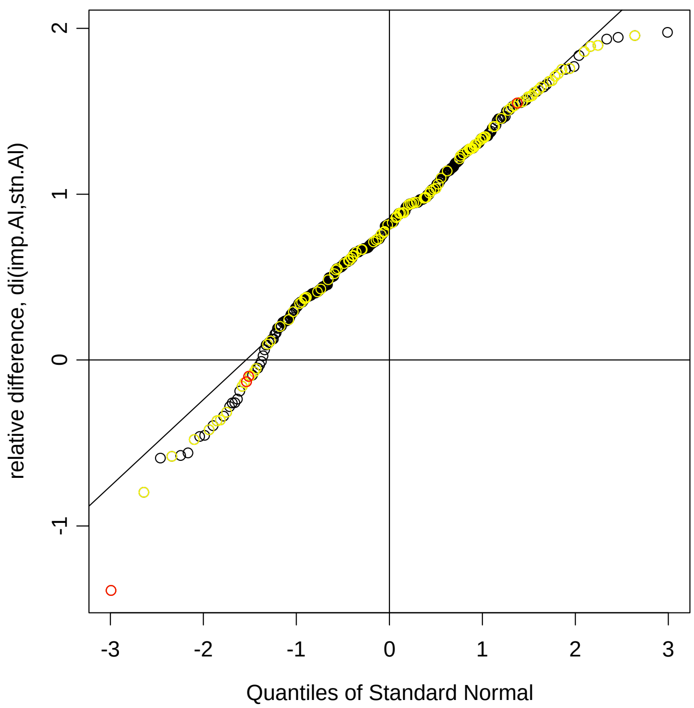


Figure E.9. a–k. Normal probability plots of the relative differences of the paired samples, d_i, for the collocated data from the IMPROVE and STN networks for the sites shown in Figure E.1. The quantiles of the d_i are plotted against the quantiles of the normal distribution. A robust straight line fit is plotted to aid the viewer in interpreting the linearity of the relationship. Values corresponding to concentrations less than 3*mdl and less than 10*mdl are shown in red and yellow, respectively.

Step 3b

The rms relative differences, rms(d), and the relative differences in the corresponding paired averages, O_rd($\overline{X}$), both provide a measure of the combined error observed in the cross-network collocated STN and IMPROVE data (Tables E.12 and E.13). The rms(d) values observed in the cross-network parameter aggregates were between 6 and 64 percentage points higher for S and Al, correspondingly, than those observed in the parameter aggregates in the in-network IMPROVE collocated data (Tables E.5 and E.12, where rms(d)=√2*rp). The O_rd($\overline{X}$) values observed in the cross-network parameter aggregates were within 1 percentage point (in absolute terms) of those observed in the in-network IMPROVE populations for the unbiased parameters S, SO_4, NO_3, and EC. The parameters that exhibited obvious biases in the cross-network comparisons, OC, and the soil elements had O_rd($\overline{X}$) values that were between 25 and 50 percentage points higher in absolute terms than those observed in the in-network IMPROVE data (Table E.13). The blank-corrected OC.adj parameter was unique in the cross-network analysis in that it had a very high rms(d) value, 177%, but a very low O_rd($\overline{X}$) value, -0.4%, again indicating that the blank correction is producing comparable results on the average but not accurately correcting the values on a sample-by-sample basis.

The same definition of comparability and method for deriving estimates of the comparability of multiyear mean concentrations as in section E.4 for the in-network IMPROVE collocated data was applied to the cross-network data. Using this method, the relative precision of the mean values were around 15–20% for Si, SO_4 and S, 33% OC.adj for NO_3, 45–51% for EC and OC, 55–59% for Fe, Ti, Ca and Al, (Table E.16 in section E.8 lists all exact values). The most extreme O_rd($\bar{X}$) value for NO_3 occurs at Puget Sound (PUSO1), which utilized the MASS STN sampler which is subject to significant nitrate losses from the Teflon filter used for particle nitrate quantification. Removing the data for PUSO1, we would estimate the comparability of mean nitrate concentrations for the other sampler types to be around 14%. While these comparability estimates provide an upper bound for the precision of either measurement, they exclude any contribution from the observed systematic errors.

Table E.12. RMS Relative differences observed in the parameter-only aggregate (network) and the parameter-site aggregates of the cross-network IMPROVE and STN collocated data.

	Network	DOSO1	MORA1	PHOE1	PUSO1	TONT1	WASH1
Al	96.2%	118.2%	59.3%	91.1%	82.9%	85.2%	121.6%
Ca	49.2%	44.9%	57.8%	61.9%	39.2%	62.0%	27.9%
Fe	42.8%	35.8%	46.5%	57.2%	37.0%	54.9%	17.1%
Si	69.5%	82.1%	64.8%	68.0%	63.6%	63.8%	72.5%
Ti	63.7%	70.4%	85.6%	68.0%	63.2%	65.4%	41.1%
NO_3	40.4%	35.8%	98.8%	24.7%	35.8%	34.6%	21.7%
SO_4	20.7%	23.3%	30.6%	23.7%	12.3%	22.6%	14.3%
S	17.9%	22.9%	31.3%	13.8%	14.2%	17.0%	11.6%
OC	65.6%	67.7%	69.4%	66.7%	26.3%	92.4%	61.6%
OC.adj	177.1%	121.3%	74.8%	46.8%	29.7%	392.4%	69.1%
EC	52.5%	69.8%	63.3%	32.7%	44.4%	67.8%	41.1%

Table E.13. The observed relative differences in the paired mean concentrations, O_rd($\overline{X}$), in the parameter-only aggregate (network) and the parameter-site aggregates of the in-network IMPROVE collocated data.

	Network	DOSO1	MORA1	PHOE1	PUSO1	TONT1	WASH1
Al	43.6%	-4.0%	-6.7%	49.5%	57.1%	38.7%	51.3%
Ca	35.9%	-11.4%	-3.5%	48.3%	22.8%	47.1%	-9.6%
Fe	30.1%	-17.3%	30.5%	47.5%	34.4%	42.3%	-8.2%
Si	51.6%	40.2%	35.6%	56.1%	50.5%	49.6%	47.3%
Ti	26.8%	-34.4%	-10.4%	49.0%	28.0%	19.2%	-0.8%
NO_3	-1.1%	-16.5%	3.4%	-5.8%	29.1%	-9.9%	-8.2%
SO_4	1.4%	3.3%	-11.5%	-6.7%	7.5%	-10.9%	3.4%
S	1.0%	5.6%	-21.8%	1.3%	2.0%	-5.1%	-0.8%
OC	-42.0%	-51.4%	-37.9%	-50.0%	-0.9%	-79.7%	-48.4%
OC.adj	-0.4%	7.9%	-12.2%	-21.0%	11.5%	21.4%	-12.7%
EC	4.8%	35.7%	42.0%	-1.5%	3.7%	-15.5%	8.5%

Similar to the analysis of the IMPROVE collocated data (section E.4, Step 3b), the average concentration for each sampler, the average absolute difference, and the rms relative difference for the parameter using all concentration values and for the low and high concentration groups are given in Table E.14. The average absolute difference was calculated for each parameter using the formula

$$\text{Average Absolute Difference} = Average(|\,Xi - Yi)\,|)$$

The low concentration subpopulation included all sample pairs where the concentration on the IMPROVE sampler was below the 20th percentile for that sampler; the high concentration subpopulation included all samples above the 20th percentile. These additional estimates of the measurement error are included to highlight the concentration dependence of the rms relative difference estimate. The blank-corrected parameters (NO_3, SO_4, OC, OC.adj, and EC) showed the largest differences in combined measurement error between the low and high concentration populations, with the characteristic relative error in the low concentrations approximately 2–10 times higher than in the high concentrations. In contrast, some of the soil elements actually had a higher rms relative difference in the high concentration population. The absolute errors for Ti, Ca, and Fe were all less than half those for S, whereas the relative errors were between 2 and 3 times higher because of the lower typical concentrations of these soil elements. The SO_4 and EC measurements had similar absolute errors, but EC's relative errors were twice those of SO_4, again due to the differences in typical concentration levels. The OC measurements, both the standard and the blank-corrected STN OC values, had the highest cross-network absolute and relative combined measurement errors. This is not surprising given the differences in analytical and blank correction methodologies.

Table E.14. Comparison of alternate calculations of combined measurement error.

	Average Concentration IMPROVE Sampler μg/m^3	**Average Concentration STN Sampler μg/m^3**	**Average Absolute Difference μg/m^3**	**rms Relative Differences for Whole Population**	**rms Relative Differences for the Lowest 20%**	**rms Relative Differences for the Highest 20%**
Al	0.0667	0.0428	0.1060	96.2%	96.1%	86.7%
Ca	0.0832	0.0578	0.0339	49.2%	54.7%	64.6%
EC	0.5808	0.5538	0.1840	52.5%	70.2%	41.9%
Fe	0.1134	0.0837	0.0393	42.8%	41.8%	55.6%
NO_3	0.8546	0.8638	0.1462	40.4%	71.8%	23.6%
OC	2.3017	3.5267	1.3716	65.6%	98.4%	38.2%
OC.adj	2.3017	2.3107	0.6585	177.1%	391.2%	37.5%
S	0.7585	0.7513	0.0854	17.9%	23.1%	15.9%
Si	0.2608	0.1538	0.1324	69.5%	59.6%	71.4%
SO_4	2.2803	2.2477	0.1934	20.7%	34.1%	17.2%
Ti	0.0080	0.0061	0.0037	63.7%	86.7%	68.1%

E.7 OBSERVED IMPROVE-STN ERRORS VERSUS EXPECTED ERRORS FROM OBSERVED IN-NETWORK RELATIVE PRECISION

Step 4

The final step in this analysis was designed to test if the observed combined errors in the cross-network IMPROVE and STN collocated data, rms(d), were qualitatively comparable to what we would have expected given only the observed relative precisions, rp, in the individual networks. The expected combined errors in the joint IMPROVE and STN data population were estimated by summing the in-network relative precisions using simple propagation of error techniques: $E\_rms(d_i)=\sqrt{rp(X)^2 + rp(Y)^2}$. The filtering out of values less than 3*mdl in the STN in-network relative precision values does limit the conclusions that can be made based on any discrepancies between observed and expected combined measurement errors. The observed combined errors in the parameter data aggregates were around 4 percentage points higher than the expected for S and SO_4; 16 for NO_3, EC, and FE; 20–30 for Ca and Ti; 30–40 for OC and Si; and 50 percentage points higher for Al (Table E.15). Thus cross-network performance was right in line with expectations for S and SO_4, a bit poorer than expected for NO_3 and EC, and fair for the biased soil elements and OC measurements. No data was available on the precision of in-network collocated STN data utilizing blank-corrected data, so no evaluation is possible for OC.adj. The higher than expected observed combined errors may in part reflect the excluded higher uncertainty low concentration values in the in-network STN relative precision estimates.

However the significant biases observed in the cross-network collocated data are clearly a significant contributor to the observed combined measurement errors, rms(d), which are not captured by the in-network collocated data.

Table E.15. The observed rms relative differences, rms(d), in the cross-network IMPROVE and STN collocated data and the expected rms relative differences, E_rms(d_i), given the relative precisions, rp, observed in the in-network IMPROVE and STN collocated data.

	rp STN	rp IMPROVE	E_rms(d)	rms(d)
Al	40%	23.0%	46.2%	96.2%
Ca	22%	17.9%	28.3%	49.2%
Fe	20%	17.9%	26.9%	42.8%
Si	23%	19.1%	29.9%	69.5%
Ti	32%	20.3%	37.9%	63.7%
S	11%	8.2%	13.7%	17.9%
SO_4	16%	4.5%	16.6%	20.7%
NO_3	18%	16.3%	24.3%	40.4%
OC	20%	24.0%	31.2%	65.6%
EC	23%	29.2%	37.2%	52.5%

E.8 CONCLUSIONS

This combined analysis of the in-network and cross-network collocated data populations leads to a number of important conclusions regarding the observed measurement errors and the comparability of mean concentrations within and between IMPROVE and STN:

1) The 2 standard deviation uncertainty estimates for the comparison of 1-year mean concentrations between two IMPROVE sites and of 2-year mean concentrations between an IMPROVE and an STN site, based upon the observed relative differences in paired mean concentrations from the collocated samplers, are listed in Table E.16. The 2 standard deviation estimates ranged from 3 to 37% for in-network and from 15 to 59% for cross-network comparisons. The corresponding 1-sigma uncertainties for cross-network comparisons were better than 25% for all parameters besides Al, Ca, and OC. These uncertainty estimates only measure the precision or random error in the mean concentrations—the significant biases found between certain IMPROVE and STN measurements are not accounted for in these uncertainties.

2) There are nonrandom errors present in the IMPROVE measurements that are identified in the in-network collocated data by larger than expected observed relative differences in the paired mean values, O_rd($\overline{X}$). The expectations were based on the model of random errors whereby independent random errors will "cancel each other out" as the sample size of the average value increases at the rate of $1/\sqrt{n}$ ($\sigma_{mean}=\sigma/\sqrt{n}$). The parameters which had multiple aggregates with absolute value of (O_rd($\overline{X}$)/2)>3*E_rp($\overline{X}$), indicating observed relative errors in the paired means more than 3 standard deviations away from expectations, were SO_4, OC, Al, and Ca. These results suggest that for these parameters nonrandom errors (biases) are present such that the characteristic relative error for the parameter does not provide a meaningful estimate of the errors in the mean value under i.i.d. assumptions. This does not necessarily imply that the characteristic

relative error for the parameter is to low; rather it could be that the assumption of independent errors for every measurement is incorrect. For example, a significant flow bias related to an annual calibration in one of the collocated samplers is not independent for each 24-hour sampling period during that year.

3) Biases in the measurements of the soil elements and OC are readily apparent in the analysis of the cross-network collocated data. In the case of the soil elements, the biases are indicated by low R^2 values (<0.55) in the correlation analysis and median relative errors greater than 0. IMPROVE routinely reports higher values for the soil elements as compared to STN. The biases in OC were in the opposite direction; the median d_i value was less than 0, indicating that IMPROVE was consistently reporting lower OC values than STN. The bias observed in the OC measurements is consistent with only the IMPROVE measurements being blank corrected.

4) There are errors present in the IMPROVE and/or STN measurements that are identified in the IMPROVE-STN collocated data that are not apparent in the IMPROVE-IMPROVE or STN-STN collocated data. These additional errors are identified by larger than expected observed rms relative differences, rms(d). The expectations were based upon simple propagation of error techniques, whereby the expected rms relative difference in the cross-network collocated data is the square root of the sum of the squared relative precision observed in each network's in-network collocated data, $E\_rms(d_i)=\sqrt{rp_X^{\ 2}+rp_Y^{\ 2}}$. The observed rms(d) values observed in the IMPROVE-STN collocated data were15–50 percentage points higher than expectations for all parameters besides S and SO_4. It is possible that the increased error observed in the cross-network collocated data is an artifact of the exclusion of values less than 3*mdl in only the STN in-network collocated data. However, the significant biases observed in IMPROVE-STN collocated data explain at least a portion of the observed "excess" uncertainty. In the case of Si, the observed combined measurement errors of 69% are over twice what would be expected given the relative precisions observed in the in-network collocated data. Analysis of cross-network collocated data identifies previously hidden errors, particularly measurement biases, which for most parameters are of considerable magnitude.

5) The IMPROVE measurement errors are inconsistent with idealized random errors—they are not i.i.d., nor are they Gaussian when the entire population is examined. Thus data analysts must be cautious in their use of any statistical techniques that require an assumption of i.i.d. or normal errors.

   a. There are significant nonrandom errors (biases) present in several of the IMPROVE measurements (see conclusion 2) that additionally are likely shared (dependent) among certain subsets of measurements. Sampling errors seem like the probable source for the biases between the collocated IMPROVE measurements, given that they were all analyzed batch-wise so that the analytical conditions should have been very similar.

   b. The IMPROVE measurement errors are heteroscedastic—they show significant relationships with both time and concentration in terms of central tendency and variability, indicating that the distribution of the errors is not the same for the entire sample population. Heteroscedastic measurement errors are expected when the measurement process spans concentrations from below mdl to those that are well quantified. However, many statistical techniques are highly sensitive to heteroscedastic errors and require the data analyst to either pretreat the data by transforming the data to equalize the errors or to select robust techniques.

c. The IMPROVE measurement errors do not follow a single normal distribution and for many parameters do not follow a single symmetrical distribution. This does not necessarily indicate that the error distribution at any fixed concentration is not Gaussian. For example, the high concentrations could all have errors drawn from a normal distribution with mean value $\mu = 0$ and standard deviation σ_1, $N(0, \sigma_1)$, and the lower concentrations have their errors drawn from a normal distribution with the same mean value and a larger standard deviation, $N(0, \sigma_2)$, where $\sigma_1 < \sigma_2$ and the result of grouping all these samples together would be a nonnormal error distribution.

6) The combined measurement errors observed in the cross-network collocated are additionally inconsistent with idealized i.i.d random errors in that they are heteroscedastic and, as a whole, nonnormal. As no analysis of the in-network STN collocated data was done, it is not possible to determine if this is just a result of IMPROVE having heteroscedastic measurement errors or if STN measurements exhibit the same inconsistencies. The same cautions about applying statistics requiring i.i.d. or normal errors apply to any joint analysis of the IMPROVE and STN data.

Table E.16. Estimated comparability for multiyear mean concentrations between sites for 1) in-network IMPROVE comparisons and 2) cross-network IMPROVE-STN comparisons. Systematic errors between are not included in these estimates.

	In-Network Precision for ~1-Year Averages	Cross-Network Precision for ~2-Year Averages *Does Not Include Bias*
Al	37%	58%
Ca	12%	55%
Fe	10%	55%
Si	10%	15%
Ti	8%	59%
S	5%	20%
SO4	3%	16%
NO3	7%	33%
EC	13%	45%
OC	20%	51%
OC.adj		33%

REFERENCES

Chow, J. C., J. G. Watson, L. W. A. Chen, W. P. Arnott, H. Moosmuller, and K. Fung (2004), Equivalence of elemental carbon by thermal/optical reflectance and transmittance with different temperature protocols, *Environmental Science and Technology 38*, 4414-4422.

EPA, Office of Air Quality Planning and Standards, Emissions, Modeling, and Analysis Division, Monitoring and Quality Assurance Group (1999), Particulate matter (PM2.5) speciation guidance, Final Draft, Edition 1.

Flannagan, J. B, and M. R. Peterson (2002), Analysis of speciation network carbon blank data, U.S. EPA Contract 68-D-98-032, Work Assignment 4-05.

Hollander, M., and D. A.Wolfe (1999), Nonparametric Statistical Methods-Second Edition, John Wiley & Sons, Inc., New York.

Rao, T., N. Frank, A. Rush, and F. Dimmick (2003), Chemical speciation of PM2.5 in urban and rural areas, National Air Quality and Emissions Trends Report, EPA Publication No. EPA 454/R-03-005. http://www.epa.gov/airtrends/aqtrnd03/pdfs/2_chemspecofpm25.pdf

Research Triangle Institute (2002), Analysis of speciation network carbon blank data, Draft Report, August 30.

Rice, J. (2005), Collocated speciation precision data for 2002-2004, PM2.5 Speciation Newsletter, October. http://www.epa.gov/ttnamti1/files/ambient/pm25/spec/spnews4.pdf

Rice, J. A. (1995), Mathematical Statistics and Data Analysis, second edition, Duxbury Press, Belmont, CA.

Taylor, J.K., (1987), Quality Assurance of Chemical Measurements, Lewis Publishers, Inc., Chelsea, MI.

White, W. H., L. L. Ashbaugh, N. P. Hyslop, and C. E. McDade (2005), Estimating measurement uncertainty in an ambient sulfate trend, *Atmospheric Environment, 39*, 6857-6867.