

# UCD IMPROVE Standard Operating Procedure #351

## Data Processing and Validation

*Interagency Monitoring of Protected Visual Environments  
Air Quality Research Center  
University of California, Davis*

*Version 3.0*

Updated By: Xiaolu Zhang Date: 12/18/2017

Approved By: Katrine Gorham Date: 12/20/2017

# IMPROVE SOP #351: DATA PROCESSING AND VALIDATION

## TABLE OF CONTENTS

1.	Purpose And Applicability.....	4
2.	Responsibilities.....	4
2.1	Project Manager.....	4
2.2	Quality Assurance Officer.....	4
2.3	Data Validation Specialist.....	4
3.	Required Equipment And Materials.....	4
4.	Data Processing And Validation Overview.....	6
5.	Data Processing.....	8
5.1	Units.....	8
5.2	Artifacts.....	9
5.3	Volume.....	10
5.4	Concentration, Uncertainty, And Method Detection Limit.....	11
5.4.1	PM <sub>2.5</sub> And PM <sub>10</sub> Mass (A- And D-Modules).....	12
5.4.2	Ions (B-Module).....	13
5.4.3	Carbon (C-Module).....	14
5.4.4	Elements (A-Module).....	16
5.4.5	Laser Absorption (A-Module).....	18
5.5	Equations Of Composite Variables.....	18
5.5.1	Sulfate By XRF (S3) And Ammonium Sulfate (NH <sub>4</sub> SO <sub>4</sub> ).....	19
5.5.2	Ammonium Nitrate (NH <sub>4</sub> NO <sub>3</sub> ).....	19
5.5.3	Soil.....	19
5.5.4	Non-Soil Potassium (KNO <sub>3</sub> ).....	20
5.5.5	Organic Carbon By Mass (OMC).....	20
5.5.6	Reconstructed Mass Using Carbon Measurements (RCMC).....	21
5.5.7	Reconstructed Fine Mass (RCMN).....	21
6.	Data Validation.....	21
6.1	Definition Of Status Flags.....	21
6.2	Level 1 Validation Procedures.....	22
6.2.1	Flow Validation.....	23
6.2.2	Level 1B Checks.....	25
6.3	Level 2 Validation Procedures.....	25
6.3.1	Cross-Module Comparison.....	26
	A-Module Versus C-Module.....	27
	A-Module Versus D-Module.....	27
	Pm <sub>2.5</sub> Reconstructed Mass Versus Gravimetric Mass.....	28
6.3.2	Long-Term Network-Wide Checks.....	29
6.3.3	Final Review.....	32
7.	Data Delivery.....	32
7.1	Submission To CIRA.....	32
7.2	Submission To AQS.....	32
8.	References.....	33

## LIST OF FIGURES

Figure 1. Data processing flow chart .....	6
Figure 2. PM <sub>2.5</sub> sampler flow rate versus the cyclone pressure transducer reading for the IMPROVE-calculated flow rate and the measured (actual) flow rate. ....	24
Figure 3. S3/SO <sub>4</sub> comparison plot for the GRGU1 site showing the 1/21/2016 sample pair as an outlier (green x).....	27
Figure 4. Comparison plot of light absorption coefficient measurements ( $f_{\text{abs}}$ , times 100) from A-Module and elemental carbon (EC) measurements from C-Module at WICA1 site. ....	27
Figure 5. Time series plot of PM <sub>10</sub> and PM <sub>2.5</sub> masses and their ratio at BOWA1 site. ....	28
Figure 6. Time series plot of PM <sub>2.5</sub> gravimetric mass, reconstructed mass without nitrate (RCMC) and reconstructed mass with nitrate (RCMN) and their ratios at LOND1 site.....	29
Figure 7. Scatter plot of sulfur (×3) versus sulfate for the entire IMPROVE network. ....	30
Figure 8. Scatter plot of chlorine versus chloride for the entire IMPROVE network. ....	30
Figure 9. Ratio of PM <sub>2.5</sub> mass (A) over PM <sub>10</sub> mass (D) at ACAD1 site, represented as raw measurements not adjusted for flow rates. Points are individual sample days (pink = Q1, green = Q2, blue = Q3, purple = Q4). Black line is the multi-year monthly mean. Blue line is the locally weighted average (LOESS).....	31
Figure 10. Multi-year monthly 10% percentile (top), median (middle) and 90% percentile (bottom) of organic carbon (OC) concentrations (in ng/m <sup>3</sup> ) for the whole IMPROVE network from 2011 to 2016.....	31

## LIST OF TABLES

Table 1. Units of data delivered to CIRA and AQS. ....	9
Table 2. Fractional uncertainty for the mass.....	13
Table 3. Fractional uncertainty for the ions .....	14
Table 4. Analytical detection limits for the carbon species (from DRI).....	15
Table 5. Fractional uncertainty for the carbon species. ....	16
Table 6. Fractional uncertainty for the elemental species.....	17
Table 7. Fractional uncertainty for the laser absorption data.....	18
Table 8. Status flags and their definitions.....	22
Table 9. Definitions and application criteria of automatic flow flags for PM <sub>2.5</sub> .....	23
Table 10. Definitions and application criteria of automatic flow flags for PM <sub>10</sub> . ....	25

## **1. PURPOSE AND APPLICABILITY**

This standard operating procedure (SOP) provides an overview of the procedures for processing and validating IMPROVE data. Data processing and data validation are performed in parallel.

## **2. RESPONSIBILITIES**

This section describes the responsibilities of those involved in data processing and validation.

### **2.1 PROJECT MANAGER**

The project manager oversees all aspects of the program.

### **2.2 QUALITY ASSURANCE OFFICER**

The quality assurance officer

- devises techniques that improve the efficiency, traceability, and accuracy of the data management;
- characterizes measurement deficiencies;
- quantifies measurement uncertainty;
- analyzes IMPROVE data to assess the influence of specific components of the measurement process and to assess the quantitative ramifications of procedural changes;
- conducts ongoing evaluation of the program's progress in data quality;
- develops and applies approaches for assessing data quality;
- proposes program improvements;
- and, develops automated data quality checks to increase the objectivity and consistency of the data validation process.

### **2.3 DATA VALIDATION SPECIALIST**

The data validation specialist

- receives analyzed data;
- reviews ion and carbon data for errors;
- reviews and evaluates flow data;
- examines flow data for disparities and errors;
- processes species concentrations;
- examines concentration data site by site for consistency;
- notes and resolves inconsistencies in the data;
- and, submits data to project sponsors, Cooperative Institute for Research in the Atmosphere (CIRA) and the EPA Air Quality System (AQS).

## **3. REQUIRED EQUIPMENT AND MATERIALS**

The data processing and validation requires all operational and analytical data be loaded into the SQL IMPROVE database, maintained by UCD. The types of data include the following:

- Basic filter information, such as sample date, site, purpose, and status. These data are recorded during filter preparation and handling and are stored in the filter.Filters table.
- Flow rates. Raw flow readings are acquired from sampler flashcards and stored in the sampler.FlowSourceData table. In addition, handwritten log sheets that contain flow and other sampling information recorded by the operator are stored in filter.Filters. A SQL procedure called sampler.spFilterAverageFlowRates is used to calculate 24-hour average flow rates for each filter based on the raw flow readings or log sheet data. These are stored in the table sampler.AverageFlows.
- Pre- and post-sampling filter mass values. These are acquired in the UCD gravimetric lab and stored in the analysis.Mass table.
- Carbon analysis results. These are acquired from files generated by DRI TOR analysis lab and are stored in analysis.Carbon.
- Ions analysis results. These are acquired from files generated by RTI IC analysis lab and are stored in analysis.Ions.
- Elements analysis results. These are acquired from the UCD XRF analysis lab through a custom ingestion process and are stored in xrf.MassLoadings.
- Optical absorption analysis results. These are acquired from the UCD HIPS analysis lab through a custom ingestion process and are stored in analysis.HIPS.

UCD has developed several custom tools for data processing and validation:

**crocker:** This program (a package in the R programming language) provides functions for processing raw filter weights, mass loadings, and flow rates into concentrations, uncertainties, and method detection limits (MDLs). *crocker* also provides utility functions that are used in the online data validation tools (see Section 6).

**datvalIMPROVE:** This R package provides functions for performing routine validation and QC (see Section 6.3).

**IMPROVE Management Website:** <http://webapp.improve.crocker.ucdavis.edu/>. This web application provides all UCD laboratory staff with viewing access to relevant tables within the SQL database. Functions within the application with relevance to data processing and validation include the following.

- XRF Section: an interface for processing XRF elemental mass loadings, managing processed sets, and applying flags.
- Filter Section: web pages for searching for specific filters, reviewing operational and analytical data associated with a filter, or applying flags and comments.
- Analysis Data Section: pages for importing and viewing carbon, ions, and optical absorption data, as well as exporting final processed and validated concentration data to AQS format for delivery.

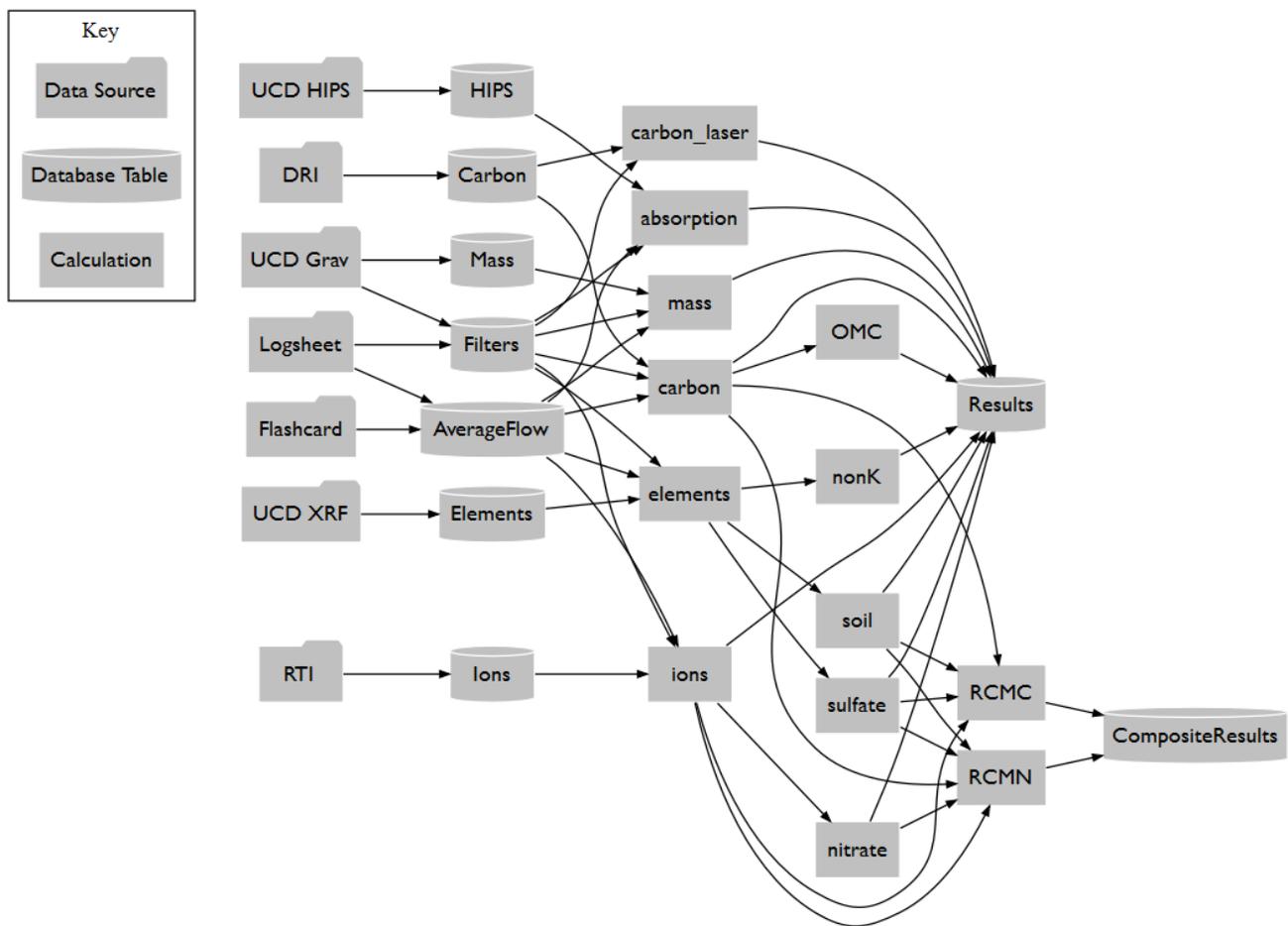
**Flow Graphs:** <http://analysis.crocker.ucdavis.edu:3838/FlowRates/>. This web application provides interactive visualizations of the raw 15-minute flow rates and temperatures in the SQL database.

**IMPROVE Data Site:** <http://analysis.crocker.ucdavis.edu:3838/ImproveData/>. This web application provides interactive visualizations of processed concentrations, uncertainties, and MDLs, plus custom tools for validation as described in Section 6.3.

#### 4. DATA PROCESSING AND VALIDATION OVERVIEW

Data processing for IMPROVE consists of reducing and combining information collected in the field and laboratories to calculate concentrations, uncertainty estimates, and method detection limits (MDLs). Figure 1 shows the flow of data; the process moves from left to right in the figure, and the concentration data are the final products. Calculations are performed within the *crocker* R package.

**Figure 1.** Data processing flow chart.



Data validation is not a linear process, and a significant amount of data validation is performed by the analytical laboratories before the data are delivered to the data validation specialist. The SOPs for the analytical laboratories describe their data validation procedures in detail.

Watson et al. (1995) define a three-level data validation process for environmental measurement studies. The levels are only intended as general guidelines. The IMPROVE data delivered to

CIRA and AQS are considered to be a mixture of Level 1B and Level 2 validated data. The levels are applied to IMPROVE as follows:

**Level 0:** Data at this level are, in essence, raw data, obtained directly from the data-acquiring instruments. These data can be reduced or reformatted, but are unedited and unreviewed, without any adjustments for known biases or problems that might have been identified during preventative maintenance checks or audits. These data may monitor instrument operations on a frequent basis. Averaging times represent the minimum intervals recorded, and these data may need to be aggregated to obtain averages for the sampling periods. Level 0 data have not been edited for instrument downtime, nor have procedural adjustments for baseline shifts, span changes, or known problems been applied. IMPROVE Level 0 data validation includes:

- 15-minute averaged pressure transducer and temperature data downloaded from the sampler flashcards before automated validation tests.
- Sample weight measurements before automated validity tests.
- XRF raw spectra.

**Level 1A:** Data at this level have passed several qualitative reviews for accuracy and completeness. The focus of Level 1A validation is to obtain as complete a data set as possible. IMPROVE Level 1A data validation includes:

- Reviewing operator log sheets to verify operation of the sampler.
- Verifying operator log sheet entries against sampler flashcard data.
- Assigning correct flow and temperature source codes.
- Assigning status flags to invalid or questionable samples to reflect sampler malfunctions, site or laboratory operator errors, or power outages.
- Identifying, investigating, and flagging data that are beyond reasonable bounds or that are unrepresentative of the variable being measured (e.g., flow rate measurements that change significantly over the sampling period).

**Level 1B:** Data at this level have passed additional automated quantitative and qualitative reviews for accuracy and internal consistency. Discrepancies that cannot be resolved are reported to the measurement laboratories for investigation. Data that deviate from consistency objectives are individually examined for errors. Obvious outliers (e.g., -85 °C temperature) are invalidated by applying a status flag. Changes to the data (e.g., swapping dates on consecutive samples) are recorded and documented by applying status flags and providing comments. Level 1B data review is carried out using custom software developed for this purpose. IMPROVE level 1B data validation includes:

- Immediately verifying balance measurements to ensure that
  - the range is within specified limits;
  - and, the post-weight is greater than the pre-weight.
- Examining daily flow rates based on a report that identifies flow rates with significant variations over 24 hours.
- Setting status flags when deviations from nominal operational settings have occurred (e.g., flow rates outside quantitative tolerances).
- Examining the ion and carbon field blank data for evidence of sample swaps.
- Examining individual data points identified as potential sample swaps between two adjacent dates on A-, B-, or C-Module.

- Comparing the analytical data to expectations based on historical data.

**Level 2:** Level 2 data validation takes place after data from various measurement methods have been assembled in the SQL database. Level 2 checks involve comparisons of collocated measurements and internal consistency tests. Data submitted to CIRA and AQS are considered to be validated at Level 1B and Level 2. Additional Level 2 data validation is performed by CIRA.

IMPROVE Level 2 data validation consists of site-by-site and network-wide examination of time series and scatter plot of data, including:

- Comparing sulfur and sulfate concentrations.
- Comparing elemental carbon and light absorption coefficients.
- Examining PM<sub>10</sub> mass and PM<sub>2.5</sub> mass for cases where PM<sub>2.5</sub> is greater than PM<sub>10</sub>.
- Comparing PM<sub>2.5</sub> gravimetric mass and reconstructed mass.

**Level 3:** This level of data review is applied after data delivery and is beyond the scope of data validation performed by UCD. At this level, the data are reconciled with other research findings, such as modeling results or theoretical predictions. Level 3 validation continues for as long as the CIRA and AQS databases are maintained.

## 5. DATA PROCESSING

The following section is a discussion of the equations used to calculate concentrations, and associated uncertainties and MDLs. The flow rate calculations are performed in the SQL database, while other calculations are all performed by R. To calculate values for all measured and derived parameters, the following commands are run in an R environment:

```
[month_data] <- crocker::improve_calculate_all([YYYY], [MM], "production")
```

This command will calculate concentrations, uncertainties, and MDLs for all measured and derived parameters for the year (*[YYYY]*) and month (*[MM]*) and return them (in memory) to the variable *[month\_data]*. The last argument in the command specifies that the calculations will use the “production” database (i.e., the UCD SQL IMPROVE operational database).

Following the processing calculation above, the results are uploaded to the SQL database in preparation for validation using the following command:

```
[post] <- crocker::improve_post_results([month_data], 'production', delete.file = FALSE,
AnalysisQcCode = 1, comment = ['Initial Posting'])
```

This command appends the processed data to the analysis.Results and analysis.CompositeResults table in the production database as an analysis set. It also inserts a record into the analysis.Sets table that provides summary information for this set, including the comment and the AnalysisQcCode. An AnalysisQcCode of 1 is used for valid routine data. The command also creates a text file copy (improveAnalysisResults\_[Date]\_[Time].txt) for archive that is stored in \\CL-SQL\Production\_DB\_Inserts.

### 5.1 UNITS

Table 1 lists the data names, type, and units for all data submitted to CIRA and AQS. For mass, ions, carbon, elements, and *fabs*, the units listed are also used for uncertainty and MDL. NA indicates that the data type is not reported to the corresponding agency.

**Table 1.** Units of data delivered to CIRA and AQS.

Data type	Data names	CIRA units	AQS units
Flow Rate	Flow	L/min	NA
Elapsed Time	ET	hours	NA
Gravimetric mass	PM2.5, PM10	ng/m <sup>3</sup>	µg/m <sup>3</sup>
Ions	Cld, NO <sub>2</sub> , NO <sub>3</sub> , SO <sub>4</sub>	ng/m <sup>3</sup>	µg/m <sup>3</sup>
Carbon	OC1, OC2, OC3, OC4, OC, OPTR, EC1, EC2, EC3, EC, TC, OPTT, OPTR at other wavelength, OPTT at other wavelength	ng/m <sup>3</sup>	µg/m <sup>3</sup>
		ng/m <sup>3</sup>	NA
Carbon_laser	RefF_wavelength, Refl_wavelength, RefM_wavelength, TransF_wavelength, Transl_wavelength, TransM_wavelength	reading	NA
Elements	H, Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, As, Pb, Se, Br, Rb, Sr, Zr	ng/m <sup>3</sup>	µg/m <sup>3</sup>
Light absorption	f <sub>abs</sub>	Mm <sup>-1</sup>	NA
Composite species	OMC, NHNO, NHSO, PM <sub>10</sub> -PM <sub>2.5</sub>	NA	µg/m <sup>3</sup>

## 5.2 ARTIFACTS

An artifact is defined as any increase or decrease of material on the filter that positively or negatively biases the measurement of ambient concentration. Artifact corrections are made on the ion, carbon, and element measurements. Artifact examples:

- (1) Contamination of the filter medium (positive).
- (2) Contamination acquired by contact with the cassettes or in handling (positive).
- (3) Adsorption of gases during collection that are erroneously measured as particles (positive).
- (4) Volatilization of particles during collection and in handling (negative).
- (5) Fall-off of particles during handling after collection (negative).

For the ion measurements, the artifact correction method attempts to account for the first two types of artifacts. The sum of the two contamination artifacts is estimated through the analysis of field blanks. The field blanks are handled as normal filters (loaded into cassettes and cartridges, shipped to and from the field, and left in the sampler for a week) except that no air is drawn through them. The field blanks are collected randomly at all sites on a periodic basis. When there are  $\geq 3$  field blanks in a month, the artifact corrections are calculated as the median loading measured on the field blanks; otherwise, values from the previous month are used. Artifact corrections are subtracted from each ambient concentration for the corresponding month. Prior to June 2002, artifacts corrections were calculated by seasonal quarter instead of by month.

For the carbon measurements, the artifact correction method attempts to account for the first three types of artifacts. Prior to the sample date of 2005, artifacts were estimated using secondary filters (CS status flag). Secondary filters were loaded into the filter cassettes behind the primary filters at select sites only (6-13 sites, depending on the time period). The assumption was that the first filter collects all the particles and absorbs some organic gases while the secondary filter absorbs an equivalent amount of organic gases. The carbon artifact correction values for each carbon fraction were calculated as the median loading measured on the secondary filters each month. Beginning with 2005 data, the artifact correction values are based on the field blanks. From August 2008 to January 2013 field blanks were collected only at sites with secondary

filters, and used both a primary and secondary filter. Analyses of these stacked field blanks showed that each of the individual field blank filters had significantly lower loadings than the single field blanks, as discussed here:

[http://vista.cira.colostate.edu/improve/Activities/Meetings/2012Pres/Dillner\\_OCArtifactAdjustmentIMPROVEOct2012.pdf](http://vista.cira.colostate.edu/improve/Activities/Meetings/2012Pres/Dillner_OCArtifactAdjustmentIMPROVEOct2012.pdf)

Therefore, loadings from the paired field blank filters are multiplied by a factor of 1.42 to adjust the loading to be equivalent to single field blank (FB) loadings. When there are  $\geq 3$  field blanks (CS for data prior to 2005, FB and FP for data 2005 onward) in a month, the artifact corrections are calculated as the median loading measured on the field blanks; otherwise, values from the previous month are used. Artifact corrections are subtracted from each ambient concentration for the corresponding month.

Measurements are not corrected the two negative artifact types (volatilization and fall-off). The measured low-temperature organics may be much less than in the atmosphere because of volatilization of particles during the remainder of the sampling. Volatilization of nitrate and chlorine from nylon is assumed to not be significant. Depending on the environmental conditions, some ammonium nitrate collected on polytetrafluoroethylene (PTFE) filters may volatilize. In those cases the fine mass on the PTFE filter may underestimate the ambient mass concentrations.

For discussion of field blank correction for element measurements, see Section 5.4.4.

### 5.3 VOLUME

The volume is the product of the flow rate and the sample duration. The sample duration is determined using elapsed time (ET), as recorded by the sampler controller on the memory card.

For the PM<sub>2.5</sub> modules, the flow rate is measured using two independent methods. The first method measures the pressure drop across the cyclone using a pressure transducer and employs a measuring orifice equation. The second method measures the pressure immediately before the critical orifice and employs the equation for flow through a critical orifice. If the cyclone transducer measurements are determined to be in error, the critical orifice transducer measurements are used.

Prior to 2016, the 15-minute pressure measurements were averaged over the sample period (nominally 24 hours), and used to calculate the flow rate. Beginning with January 2016 data, the flow rate is calculated for each 15-minute pressure measurement, and used to calculate an elapsed time-weighted average flow rate for the sampling period. If the sampler flow rate drops during sampling (due to filter clogging, for example), the two methods will produce different results (1-3% difference); otherwise they are nearly identical.

The sampler flow rate for A-, B-, and C-Module is calculated using equation 351-1.

$$Q = 10^a M^b * F(elev) * \sqrt{\frac{T + 273.15}{293.15}} \quad (351-1)$$

Q = volumetric flow rate (using site-specific temperature and pressure, not STP)

a, b = calibration coefficients

M = cyclone transducer reading. If the transducer readings are taken from the controller screen, they can be used in equation 351-1 directly. If the transducer readings are taken from the flashcard file, they must be divided by 100.

$F(elev)$  = elevation factor to account for pressure difference between sea level and site.

T = ambient temperature in degrees Celsius at time of sampling.

For D-Module, the sampler flow rate is calculated using equation 351-2.

$$Q = (c + d * G) * F(elev) * \sqrt{\frac{T + 273.15}{293.15}} \quad (351-2)$$

Q = volumetric flow rate

c, d = calibration coefficients

G = critical orifice transducer reading. If the transducer readings are taken from the controller screen, they can be used in equation 351-2 directly. If the transducer readings are taken from the flashcard file, they must be divided by 100.

$F(elev)$  = elevation factor to account for pressure difference between sea level and the site.

T = ambient temperature in degrees Celsius at time of sampling.

Equation 351-2 may be used to calculate the flow rate on A-, B-, or C-Modules if the cyclone transducer readings are invalid.

#### 5.4 CONCENTRATION, UNCERTAINTY, AND METHOD DETECTION LIMIT

The calculations described in this section are performed in R. The mass of material on the filter is equal to the difference between the mass measured on the sample and the mass on the unused filter. The mass on the unused filter is determined from the pre-weight of each individual PTFE filter for the gravimetric analysis, and from the median of field blank loadings for the ion and carbon measurements. The concentration is calculated using equation 351-3.

$$C = \frac{A - B}{V} \quad (351-3)$$

C = ambient concentration (ng/m<sup>3</sup>)

A = mass measured on sample (ng/filter or ng/cm<sup>2</sup>)

B = artifact mass (ng/filter) = pre-weight, monthly median of ion or carbon field blanks (FB). Prior to June 2002, artifacts were calculated by seasonal quarter instead of by month.

V = sample air volume (m<sup>3</sup>) = Q \* Elapsed Time

The uncertainty is reported with each concentration. The general model for the uncertainty is a quadratic sum of two components of uncertainty as shown in Equation 351-4.

$$\sigma(c) = \sqrt{[fC]^2 + \left[\frac{\sigma_a}{V}\right]^2} \quad (351-4)$$

$\sigma_a$  = constant analytical uncertainty. This term results from additive sources of uncertainty, such as those related to background contamination of the filters. For large concentrations, this is small compared to the fractional term.

$f$  = fractional uncertainty. This term results from various sources of proportional uncertainties, such as the analytical calibration and flow rate measurements. The fractional uncertainties ( $f$ ) are determined from the collocated measurements. The calculations are based on a study by Hyslop and White (2008), which determined that a robust statistic best fit the distribution of observed differences between the collocated measurements. Briefly, it is obtained by taking the root mean square of the collocated precisions ( $cp$ ) of the most recent 3 to 5 years of data, depending on the availability. If the count of pairs is less than 60, 0.25 is adopted as  $f$ . The (one sigma) precision of observation pairs of a year is calculated as one-half the difference between the 84<sup>th</sup> and 16<sup>th</sup> percentiles of the signed scaled relative differences ( $srd$ ). Equations are:

$$srd = \frac{(Collo - Routine) / \sqrt{2}}{(Collo + Routine) / 2} \quad (351-5)$$

$$cp = \frac{(84^{th} \text{ percentile\_of\_ } srd) - (16^{th} \text{ percentile\_of\_ } srd)}{2} \quad (351-6)$$

$$f = \sqrt{\frac{1}{n} \sum_{i=1}^n cp_i^2} \quad (351-7)$$

The MDLs are also provided with each concentration in the SQL database. For ions and carbon the MDLs are calculated as 2× the standard deviation of the field blank loading, with analytical uncertainties as floor values. For elements, the MDLs are calculated as 95<sup>th</sup> percentile minus median of field blanks.

#### 5.4.1 PM<sub>2.5</sub> AND PM<sub>10</sub> MASS (A- and D-Modules)

PM<sub>2.5</sub> mass is measured gravimetrically on the polytetrafluoroethylene (PTFE) filter from A-Module. PM<sub>10</sub> mass is measured gravimetrically on the PTFE filter from D-Module. The post- and pre-weights are stored in table analysis.Mass. The constant analytical uncertainty,  $\sigma_a$ , in equation 351-4 is equal to 5  $\mu\text{g}$  for all filters. The mass concentration, uncertainty, and MDL in nanograms per cubic meter are calculated using the following equations:

$$C_{Mass} = 10^6 \text{ ng/mg} * \left( \frac{Postweight - preweight}{V} \right) \quad (351-8)$$

$$\sigma_{Mass} = 1000 \text{ ng/}\mu\text{g} * \sqrt{\left( \left( \frac{5\mu\text{g}}{V} \right)^2 + \left( \frac{C_{Mass} * f}{1000 \text{ ng/}\mu\text{g}} \right)^2 \right)} \quad (351-9)$$

$$mdl_{Mass} = 1000 \text{ ng/}\mu\text{g} * \frac{10\mu\text{g}}{V} \quad (351-10)$$

f = fractional uncertainty (Table 2).

**Table 2.** Fractional uncertainty for the mass

Species	f reported for data 2/28/1995 – 12/31/2006	f reported for data 1/1/2017 -
PM <sub>2.5</sub>	0.03	0.03
PM <sub>10</sub>	0.03	0.07

#### 5.4.2 IONS (B-Module)

Ions are measured by ion chromatography using the nylon filter from B-Module. Ions data are stored as micrograms per filter in analysis.Ions table in the SQL database.

For chloride, nitrate, and sulfate, the concentration, uncertainty, and MDL in nanograms per cubic meter are calculated using the following equations:

$$C_{ion} = 1000 \frac{ng}{\mu g} * \frac{(A_{ion} - B_{ion})}{V_{Bmodule}} \quad (351-11)$$

$$\sigma_{ion} = 1000 \frac{ng}{\mu g} * \frac{\sqrt{(Max(\sigma_{dfb}, mdl_{analytical}))^2 + (f * (A_{ion} - B_{ion}))^2}}{V_{Bmodule}} \quad (351-12)$$

$$mdl = 1000 \frac{ng}{\mu g} * 2 * \frac{Max(\sigma_{dfb}, mdl_{analytical})}{V_{Bmodule}} \quad (351-13)$$

For nitrite, the concentration and MDL equations are the same as above, while the uncertainty equation is affected by the concentration value:

When  $C_{ion} > 0$

$$\sigma_{ion} = 1000 \frac{ng}{\mu g} * \frac{\sqrt{(Max(\sigma_{dfb}, mdl_{analytical}))^2 + (f * (A_{ion} - B_{ion}))^2}}{V_{Bmodule}} \quad (351-14)$$

When  $C_{ion} \leq 0$

$$\sigma_{ion} = 0 \quad (351-15)$$

$B_{ion}$  = median of the field blank measurements when there are  $\geq 3$  field blanks in a month; otherwise, values from the previous month are used.

$mdl_{analytical}$  = 0.03 for chloride, 0.01 for nitrite, 0.05 for nitrate, 0.07 for sulfate, and 0.12 for ammonium. The analytical MDL for each ion is to be used as the overall MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

$\sigma_{dfb}$  = standard deviation of the field blank measurements for the appropriate ion.

f = fractional uncertainty (Table 3).

**Table 3.** Fractional uncertainty for the ions

Species	f reported for data 1/1/2005 – 12/31/2016	f reported for data 1/1/2017 -
Chloride (Cl <sup>-</sup> )	0.08	0.08
Nitrite (NO <sub>2</sub> <sup>-</sup> )	0.22	0.25
Nitrate (NO <sub>3</sub> <sup>-</sup> )	0.04	0.03
Sulfate (SO <sub>4</sub> <sup>=</sup> )	0.02	0.02
Ammonium (NH <sub>4</sub> <sup>+</sup> )	0.02	NA
Sulfur Dioxide (SO <sub>2</sub> )	0.04	NA

### 5.4.3 CARBON (C-Module)

Carbon is measured by thermal optical reflectance (TOR) using the quartz filter from C-Module. Carbon data are stored as micrograms per filter in analysis. Carbon table in the SQL database. For the eight carbon fraction species, the primary source of fractional uncertainty is the separation into temperature ranges. This may be associated with temperature regulation, but it may also be from the inherent variability of the species involved.

The elemental carbon (EC) component is assumed to be all carbon evolved at 550°C and above, after the laser indicates that reflectance has returned to the initial value. The organic carbon (OC) component is assumed to be all carbon evolved at 550°C and below, in a pure helium environment, plus the pyrolyzed organic fraction. The total carbon (TC) is sum of OC and EC.

The concentration, uncertainty, and MDL in nanograms per cubic meter for the carbon species (O1, O2, O3, O4, OP, E1, E2, E3, as well as OC, EC, TC) are calculated using the following equations:

$$C = 1000 \frac{ng}{\mu g} * \frac{(A - B)}{V_{C\ module}} \quad (351-16)$$

$$\sigma = 1000 \frac{ng}{\mu g} * \frac{\sqrt{(\text{Max}(\sigma_{dfb}, t))^2 + (f * (A - B))^2}}{V_{C\ module}} \quad (351-17)$$

$$mdl = 1000 \frac{ng}{\mu g} * 2 * \frac{\text{Max}(\sigma_{dfb}, t)}{V_{C\ module}} \quad (351-18)$$

B = median of the field blank measurements when there are >= 3 field blanks in that month, otherwise the number from the previous month is used.

$\sigma_{dfb}$  = standard deviation of the field blank measurements for the appropriate carbon species.

t = analytical detection limits (Table 4).

f = fractional uncertainty (Table 5).

**Table 4.** Analytical detection limits for the carbon species (from DRI).

Species	t
O1	0.51
O2	0.51
O3	0.51
O4	0.51
OP (OPTR), OPTR at other wavelength	0.15
OPTT , OPTT at other wavelength	0.15
E1	0.15
E2	0.15
E3	0.15
EC	0.15
OC	0.51
TC	0.57

**Table 5.** Fractional uncertainty for the carbon species.

Species	f reported for data 1/1/2005 – 12/31/2016	f reported for data 1/1/2017 -
O1	0.23	0.27
O2	0.15	0.13
O3	0.13	0.13
O4	0.15	0.13
OP (OPTR), OPTR at other wavelength	0.13	0.16
OPTT, OPTT at other wavelength	0.13	0.12
E1	0.10	0.10
E2	0.17	0.18
E3	0.42	0.25
EC	0.12	0.14
OC	0.08	0.09
TC	0.08	0.08

#### 5.4.4 ELEMENTS (A-Module)

Elements are measured using X-ray fluorescence (XRF; PANalytical Epsilon 5) on the PTFE filters from A-Module since 2011. From December 2001 through the end of 2010, all elements, except hydrogen, were analyzed with XRF systems. Hydrogen was analyzed with proton elastic scattering analysis (PESA) until 2010. Prior to December 2001, particle induced x-ray emission (PIXE) was used to analyze the light elements (sodium through manganese) and XRF was used to analyze the heavier elements (iron through lead). The PANalytical XRF instruments report the elements in terms of counts per mV per second, which must be converted into areal densities using element calibration factors (stored in the SQL database). Blank subtraction is performed on the XRF measurements by subtracting the median field blank count from the same instrument that was used to analyze the sample. The field blank correction is specific to each instrument because the counts vary from instrument to instrument. The number of field blanks analyzed on a specific instrument can be quite low in a month, so instead field blanks by month, the last 25 field blanks analyzed are used.

$$\sigma = 1000 \frac{ng}{\mu g} * \sqrt{(0.608 * \text{Max}((\text{blankStat}.\text{Percentile95} - \text{blankStat}.\text{Median}) * \text{p.EValue}, \text{y.DL}))^2 + (f * A_{\text{element}})^2}$$

$A_{\text{element}} = (\text{netCounts} - \text{blankStat}.\text{Median}) * \text{p.EValue}$ , which is the areal density calculated for the element measured by XRF.

$M_{\text{element}} = \text{max}(((\text{blankStat}.\text{Percentile95} - \text{blankStat}.\text{Median}) * \text{p.EValue}), \text{y.DL}[(\text{ug}/\text{cm}^2)]) * 1000$ , which is the areal MDL reported for the element measured by XRF.

$\text{p.EValue} = \text{element calibration factor}$ .

y.DL = static MDL value previously used.

f = fractional uncertainty (Table 6).

blankStat values are specific to a batch of 25 filters analyzed on a specific instrument during the time period that the current sample month was analyzed.

**Table 6.** Fractional uncertainty for the elemental species.

Species	f	
	reported for data 1/1/2005 – 12/31/2016	reported for data 1/1/2017 -
Al	0.09	0.08
As	0.25	0.21
Br	0.10	0.11
Ca	0.06	0.07
Cl	0.14	0.18
Cr	0.22	0.17
Cu	0.12	0.11
Fe	0.06	0.06
K	0.03	0.05
Mg	0.15	0.16
Mn	0.13	0.13
Na	0.14	0.15
Ni	0.16	0.16
P	0.25	0.33
Pb	0.13	0.13
Rb	0.25	0.26
S	0.03	0.03
Se	0.25	0.12
Si	0.10	0.07
Sr	0.16	0.14
Ti	0.11	0.09
V	0.12	0.14
Zn	0.06	0.08
Zr	0.25	0.25

The areal densities, areal uncertainty, and areal MDL (in units of mass/area) are stored in the SQL database and used to calculate concentrations using the following equations:

$$C_{element} = \frac{A_{element} * (Depositarea)}{V} \quad (351-19)$$

$$\sigma_{element} = \frac{U_{element} * (Depositarea)}{V} \quad (351-20)$$

$$mdl_{element} = \frac{M_{element} * (Depositarea)}{V} \quad (351-21)$$

$A_{element}$  = areal concentration of the element measured by XRF.

Deposit area = area of sample deposit on the filter (cm<sup>2</sup>), determined from the filter holder or mask size.

#### 5.4.5 LASER ABSORPTION (A-Module)

Optical absorption is measured by a hybrid integrating plate and sphere (HIPS) system using the PTFE filter from A-Module. The laser absorption measurements are stored in the laser file as reflectance (R) and transmittance (T) values. The filter absorption coefficient ( $f_{abs}$ ) values (Mm) are calculated using the following equations:

$$FABS = \left( 10000 * \text{Log} \left( \max \left( \frac{1025.5 - 1.235 * R}{T}, 0.1 \right) \right) * \frac{(Depositarea)}{V} \right) * \frac{1}{100} \quad (351-22)$$

$$\sigma_{FABS} = \sqrt{\left( \frac{(10000 * 0.014)}{\left( \frac{V}{Depositarea} * 100 \right)} \right)^2 + (f * LRNC)^2} \quad (351-23)$$

$$mdl = 10000 * 0.032 * \left( \frac{Depositarea}{V} \right) * \frac{1}{100} \quad (351-24)$$

f = fractional uncertainty (Table 7).

**Table 7.** Fractional uncertainty for the laser absorption data.

Species	f reported for data 2/28/1995 – 12/31/2006	f reported for data 1/1/2017 -
$f_{abs}$	0.03	0.08

#### 5.5 EQUATIONS OF COMPOSITE VARIABLES

The following composite variables are combinations of the measured concentrations of particles collected on the fine filters. These are used in the Level 2 validation procedures, described in Section 6.3. The concentration is determined along with the uncertainty and MDL. The

composite variable uncertainty calculations assume that the component concentrations are independent and the multiplicative factors have no uncertainty. The independence assumption is not strictly valid for many composites because of common factors, such as volume. However, the effect on the overall uncertainty is too small to warrant the more complicated calculations.

### 5.5.1 SULFATE BY XRF (S3) AND AMMONIUM SULFATE (NHSO)

Sulfur is predominantly present as sulfate in the atmosphere. To compare the sulfur by XRF and the sulfate by ion chromatography, the XRF concentration is multiplied by the ratio of sulfate to sulfur atomic mass ( $96.06/32.06 = 3.0$ ). This composite is labeled S3 in the data validation plots.

The sulfate is generally present as ammonium sulfate,  $(\text{NH}_4)_2\text{SO}_4$ , although it can be present as ammonium bisulfate,  $(\text{NH}_4)\text{HSO}_4$ , sulfuric acid,  $\text{H}_2\text{SO}_4$ , gypsum,  $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ , and, in marine areas, as sodium sulfate,  $\text{Na}_2\text{SO}_4$ . In many cases, the particle will include associated water, this is omitted from the calculation. In order to simplify the calculation it is assumed that all sulfur is present as ammonium sulfate. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$NHSO = 4.125 * S$$

$$S3 = 3 * S \quad (351-25)$$

$$\sigma_{NHSO} = 4.125 * \sigma(S)$$

$$\sigma_{S3} = 3 * \sigma(S) \quad (351-26)$$

$$mdl(NHSO) = 4.125 * mdl(S)$$

$$mdl(S3) = 3 * mdl(S) \quad (351-27)$$

For ammonium bisulfate, sulfuric acid, and sodium sulfate the factors are 3.59, 3.06, and 4.43, respectively. In the first two cases, the actual dry mass associated with sulfate is less than NHSO, and in the third case, more.

### 5.5.2 AMMONIUM NITRATE (NHNO)

This composite is the total dry concentration associated with nitrate, assuming 100% neutralization by ammonium. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$NHNO = 1.29 * \text{NO}_3^- \quad (351-28)$$

$$\sigma_{NHNO} = 1.29 * \sigma(\text{NO}_3^-) \quad (351-29)$$

$$mdl(NHNO) = 1.29 * mdl(\text{NO}_3^-) \quad (351-30)$$

### 5.5.3 SOIL

The soil component consists of the sum of the predominantly soil elements measured by XRF, multiplied by a coefficient to account for oxygen for the normal oxide forms ( $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_2$ ,  $\text{CaO}$ ,  $\text{K}_2\text{O}$ ,  $\text{FeO}$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{TiO}_2$ ), and augmented by a factor to account for other compounds not included in the calculation, such as  $\text{MgO}$ ,  $\text{Na}_2\text{O}$ , water, and  $\text{CO}_2$ . The following assumptions are made:

- Fe is split equally between FeO (oxide factor of 1.29) and Fe<sub>2</sub>O<sub>3</sub> (oxide factor of 1.43), giving an overall Fe oxide factor of 1.36.
- Fine K has a non-soil component from smoke. Based on the K/Fe ratio for average sediment (*Handbook of Chemistry and Physics*), 0.6\*Fe is used as a surrogate for soil K. The oxide factor for K  $\left( K_2O, \frac{39.1 * 2 + 16.0 \text{ g/mol}}{39.1 * 2 \text{ g/mol}} = 1.2 \right)$  is added for a total Fe factor of 0.72\*Fe (0.6\*1.2) for the potassium oxide in soil. This increases the factor for Fe from 1.36 to 2.08.
- The oxide forms of the soil elements account for 86% of average sediment; in order to obtain the total mass associated with soil, the final factors are divided by 0.86 (*Handbook of Chemistry and Physics*). The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$SOIL = 2.2 * \max(Al,0) + 2.49 * \max(Si,0) + 1.63 * \max(Ca,0) + 2.42 * \max(Fe,0) + 1.94 * \max(Ti,0) \quad (351-31)$$

$$\sigma(SOIL) = \sqrt{(2.2 * \max(\sigma(Al),0))^2 + (2.49 * \max(\sigma(Si),0))^2 + (1.63 * \max(\sigma(Ca),0))^2 + (2.42 * \max(\sigma(Fe),0))^2 + (1.94 * \max(\sigma(Ti),0))^2} \quad (351-32)$$

$$mdl(SOIL) = 0 \quad (351-33)$$

The soil variable is calculated for all valid XRF analyses.

#### 5.5.4 NON-SOIL POTASSIUM (KNON)

Non-soil potassium is the measured fine potassium minus the soil potassium estimated from iron. Non-soil potassium is a qualitative tracer of smoke. However, the ratio of potassium/smoke mass may change as the aerosol ages. Particulate smoke potassium may be produced by the transformation of volatilized potassium, and appears to be in a smaller size range than most smoke mass. Close to the smoke source, the particulate potassium may not have time to form. For long-range transport, most other smoke mass may settle out more than potassium mass. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$KNON = (K - 0.6 * FE) \quad (351-34)$$

$$\sigma(KNON) = \sqrt{\sigma^2(K) + [0.6 * \sigma(Fe)]^2} \quad (351-35)$$

$$mdl(KNON) = 0 \quad (351-36)$$

The soil factor of 0.6 may vary slightly with the site; this will produce a small positive or negative offset for baseline values when no smoke is present. Therefore, negative values are retained. KNON is calculated for all valid XRF analyses. If a concentration is less than the MDL, the concentration and uncertainty are assumed to be equal to the MDL.

#### 5.5.5 ORGANIC CARBON BY MASS (OMC)

To determine the total amount of organic mass associated with the organic carbon, the ratio of organic mass to organic carbon is assumed to be 1.8. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$OMC = 1.8 \times OC \quad OMC = 1.8 \times OC = 1.8 \times (O1 + O2 + O3 + O4 + OP) \quad (351-37)$$

$$\sigma_{OMC} = 1.8 \times \sigma_{OC} = 1.8 \times 1000 \frac{ng}{\mu g} * \frac{\sqrt{(\text{Max}(\sigma_{dfb}, t))^2 + (f * (A_{OC} - B_{OC}))^2}}{V_{Cmodule}} \quad (351-38)$$

$$mdl_{OMC} = 1.8 \times mdl_{OC} = 1.8 \times 1000 \frac{ng}{\mu g} * 2 * \frac{\text{Max}(\sigma_{dfb}, t)}{V_{Cmodule}} \quad (351-39)$$

### 5.5.6 RECONSTRUCTED MASS USING CARBON MEASUREMENTS (RCMC)

The reconstructed mass is the sum of sulfate, soil, salt, elemental carbon, and organic mass. The only components not included are water and nitrate. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$RCMC = \max(NHSO, 0) + \max(SOIL, 0) + 1.8 * \max(Cl^-, 0) + \max(LAC, 0) + \max(OMC, 0) \quad (351-40)$$

$$\sigma(RCMC) = \sqrt{(\max(\sigma(NHSO), 0))^2 + (\max(\sigma(SOIL), 0))^2 + [1.8 * \max(\sigma(Cl^-), 0)]^2 + (\max(\sigma(LAC), 0))^2 + (\max(\sigma(OMC), 0))^2} \quad (351-41)$$

$$mdl(RCMC) = 0 \quad (351-42)$$

RCMC is always positive.

RCMC is more relevant at sites where the neutralization of sulfate may be less than 100%, at sites with high nitrate, and at marine sites.

### 5.5.7 RECONSTRUCTED FINE MASS (RCMN)

At sites where ammonium nitrate (NHNO) is present, adding ammonium nitrate to the RCMC can make the reconstructed mass very close to the measured value. The concentrations, uncertainties, and MDLs are calculated using the following equations:

$$RCMN = \max(NHSO, 0) + \max(NHNO, 0) + \max(SOIL, 0) + 1.8 * \max(Cl^-, 0) + \max(LAC, 0) + \max(OMC, 0) \quad (351-43)$$

$$\sigma(RCMN) = \sqrt{(\max(\sigma(NHSO), 0))^2 + (\max(\sigma(NHNO), 0))^2 + (\max(\sigma(SOIL), 0))^2 + [1.8 * \max(\sigma(Cl^-), 0)]^2 + (\max(\sigma(LAC), 0))^2 + (\max(\sigma(OMC), 0))^2} \quad (351-44)$$

$$mdl(RCMN) = 0 \quad (351-45)$$

## 6. DATA VALIDATION

This section discusses the Level 1 and Level 2 validation processes that occur once the data are received from the field and laboratories. Data validation performed at UCD involves assessing the quality, reliability, and integrity of the data.

### 6.1 DEFINITION OF STATUS FLAGS

Status flags are used as standardized abbreviations describing the status of individual sample results, and are assigned during the Level 1 and 2 validation processes (Table 8). Samples associated with “Terminal” flag are invalidated for a variety of reasons, and no concentration values are reported, whereas those associated with “Informational” flag are still valid samples and concentrations are reported. The “Temporary” flags are replaced before final data reporting.

**Table 8.** Status flags and their definitions.

Status Flag	Description	Flag Type	AQS code
BI	Bad Installation of Sample Cartridge or Filter	Terminal	BJ
CG	Sample Flow Rate Out of Spec.	Informational	W
CL	Sample Flow Rate Out of Limits	Terminal	AH
DA	Sample not analyzed	Terminal	AM
DE	Reported value is an estimate	Informational	LJ
EP	Equipment Problem	Terminal	AN
LF	Sample Flow Rate Out of Spec.	Informational	W
NM	Normal	Informational	
NS	No Sample Collected/Late Sample Change	Terminal	AF
OL	Site Off Line	Terminal	AD
PO	Power Outage	Terminal	AV
QD	Questionable Data	Temporary	4
SA	Sampling Anomaly	Informational	1
SO	Still out	Temporary	
SP	Same-day Field Blank/Sample Swap	Informational	
SW	Sampling Dates Swap	Informational	
TU	Incorrect Time (with time shift $\geq$ 6hrs)	Informational	3
UN	Undetermined Weight	Informational	AM
XX	Sample Destroyed, Damaged or Contaminated	Terminal	AJ
PM	Undefined but allowed by SWAP as informational	No longer used	
NR	Not Reanalyzed by DRI	No longer used	
NA	Not Applicable	No longer used	AM
QA	Quality Assurance	No longer used	4
QC	Quality Control	No longer used	
RF	Really High Flow Rate	No longer used	W

## 6.2 LEVEL 1 VALIDATION PROCEDURES

Level 1 validation is conducted throughout the sample handling and analysis processes. Validation for the gravimetric PM<sub>2.5</sub> and PM<sub>10</sub> masses, PM<sub>2.5</sub> elements, ions, and carbon data is performed by the laboratory technicians responsible for the analyses. The following Technical Information (TI) documents are available for mass validation and XRF data validation:

Mass validation: *Sample Handling TI 2510 General Laboratory Procedures*, Section 5.9

XRF validation: *XRF TI 130e Level I Validation*

Level 1 flow rate validation is performed as a four-step process. Additional Level 1B validation checks are performed on data completeness and field blank validity prior to processing the

concentration data. The following sections discuss the flow validation and Level 1B checks in details.

### 6.2.1 FLOW VALIDATION

Level 0 flow data (raw transducer readings from field log sheets and flashcards) are uploaded to the SQL database by the sample handling lab technicians soon after filter samples return to the lab from the IMPROVE sites. Prior to data uploading, as part of the Level 1A validation process, data from log sheets and flashcards are reviewed in parallel and checked for inconsistency resulting from sampling anomaly and/or sampler malfunction. In these cases, the sample status is changed from NM to a terminal or temporary flag, and filter/sample event comments are provided. When flashcard data is missing or not reliable due to equipment problems, the Flow Source Code (FlwSrc) for the affected sample is changed from default (MC) to log sheet (LC/LO) or nominal value (NF) to ensure accurate calculation of the average flow rate. Detailed procedures on flow data ingestion and Level 1A validation can be found in the *Sample Handling TI 251E Entering Log Sheets and Simple Problem Diagnosis*.

Several Level 1B checks on the 15-minute flashcard flow data are performed by running the *flow.check* function in the *datvalIMPROVE R* package. The function returns a report that lists the samples with abnormal flow variability and abnormal sampling temperature, and number of records for further investigation. The criterion for abnormal flow variability is greater than 8% during a 24-hour sampling period, which is due either to equipment/installation problems or to a steady pressure drop from heavy filter loading. Flow plots are carefully reviewed (IMPROVE Flow Graphs; <http://analysis.crocker.ucdavis.edu:3838/FlowRates/>) to determine flow validity. Proper Flow Source Code is assigned if the primary channel (MC) is not reliable.

Next, the 15-minute flashcard data are processed in SQL environment to derive the daily average flow rate and elapsed time (ET). The flow processing code automatically assign non-normal flow status flags to the samples that have flow rate that deviates from the nominal value or ET shorter than 18 hours. Table 9 and 10 list the types of flow flags and the associated criteria for applying them to PM<sub>2.5</sub> and PM<sub>10</sub> samples, respectively.

**Table 9.** Definitions and application criteria of automatic flow flags for PM<sub>2.5</sub>.

Automatic Flow Flag	Definition	Type	Criteria for Application for PM <sub>2.5</sub> Samples
CL	Clogged Filter	Terminal	Flow rate < 15 L/min for more than 6 hours if flashcard data are used Average flow rate < 15 L/min if log sheet values are used
CG	Clogging Filter	Informational	Flow rate < 18 L/min for more than 6 hours if flashcard data used Average flow rate < 18 L/min if log sheet values are used
LF	Low/high flow rate	Informational	Average flow rate < 19.7 L/min or > 24.1 L/min
PO	Power Outage	Terminal	Elapsed time < 1080 minutes (18 hours)

The cyclone characterization test carried out in the summer of 2016 on the IMPROVE PM<sub>2.5</sub> cyclones shows consistent results with the characterization performed by John and

Reischl (1980). The particle size cut of the cyclone at any operating flow rate can be determined from the following equation:

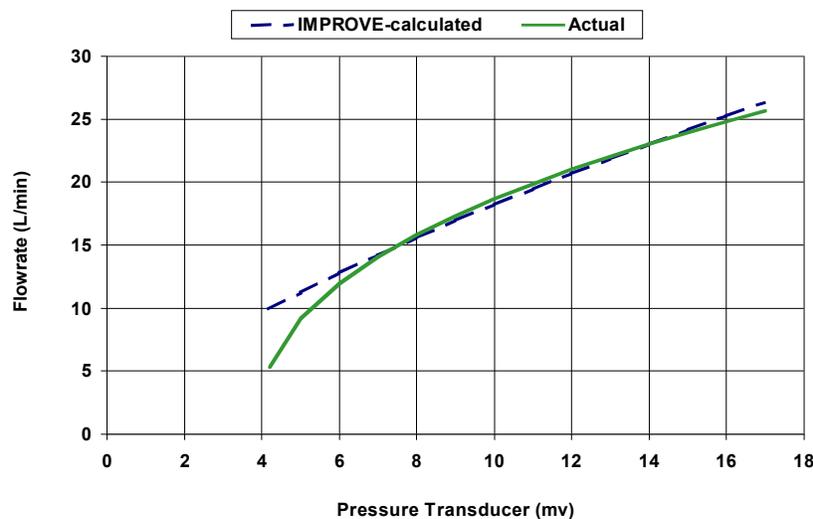
$$D_{50} = 52.5 * Q^{-0.99} \quad (351-46)$$

$D_{50}$  is the 50% cutoff diameter (in  $\mu\text{m}$ ) and  $Q$  is the flow rate (in L/min). Note that at the nominal flow rate of 23 L/min, the 50% cutoff diameter is 2.36  $\mu\text{m}$  rather than 2.5  $\mu\text{m}$ .

The criteria for the CL, CG, and LF flags are determined based on calculation limitations, performance testing, and particle size cut. Figure 2 illustrates a typical relationship between  $\text{PM}_{2.5}$  flow rate and the cyclone pressure transducer measurement. The dashed line shows the calculated flow rate from equation (351-1) whereas the solid line shows the measured flow rate. The response of the actual flow rate to the change in pressure is no longer linear below approximately 15 L/min and therefore the calculated flow rate < 15 L/min is inaccurate. If >24 15-min (6 hours in total) flow rate readings are below 15 L/min, or if the average flow rate is below 15 L/min when log sheet data are used, the sample is flagged as CL and no concentration data are reported. The  $\text{PM}_{2.5}$  cyclone cut point is 3.6  $\mu\text{m}$  at 15 L/min.

The criteria for applying CG and LF flags are based primarily on cut point characterization of the  $\text{PM}_{2.5}$  cyclone. The cut point is 3.0  $\mu\text{m}$ , 2.75  $\mu\text{m}$ , and 2.25  $\mu\text{m}$  at 18 L/min, 19.7 L/min, and 24.1 L/min, respectively. The 2.25 - 2.75  $\mu\text{m}$  range is considered a reasonable range of particle cut points for a data value to be labeled as  $\text{PM}_{2.5}$ .

**Figure 2.**  $\text{PM}_{2.5}$  sampler flow rate versus the cyclone pressure transducer reading for the IMPROVE-calculated flow rate and the measured (actual) flow rate.



A similar set of flags is applied to the  $\text{PM}_{10}$  data (Table 10), but with several differences in the criteria, due principally to the lower flow rate at which the  $\text{PM}_{10}$  sampler operates. The relationship between the  $\text{PM}_{10}$  Sierra cyclone and particle size cut is not well characterized so the criteria are determined somewhat arbitrarily. It is important to note that under circumstance of a failing pump that produces less vacuum, equation (351-2) is no longer true and the calculated flow rates for the  $\text{PM}_{10}$  module are not valid.

**Table 10.** Definitions and application criteria of automatic flow flags for PM<sub>10</sub>.

Validation Flag	Definition	Type	Criteria for Application for PM <sub>10</sub> Samples
CL	Clogged Filter	Terminal	Flow rate < 10 L/min for more than 6 hours if flashcard data are used Average flow rate < 10 L/min if log sheet values are used
CG	Clogging Filter	Informational	Flow rate < 14 L/min for more than 6 hours if flashcard data are used; Average flow rate < 14 L/min if log sheet values are used
LF	Low/high flow rate	Informational	Average flow rate < 15 L/min or > 18 L/min
PO	Power Outage	Terminal	Elapsed time < 1080 minutes (18 hours)

Finally, all samples flagged as terminal (i.e. CL and PO) by the flow processing code are manually reviewed for errors. In cases where valid samples are flagged as invalid (e.g. corrupt flashcard files or faulty transducer readings), the flow source code is changed and average flow rate reprocessed to correct the sample status.

### 6.2.2 LEVEL 1B CHECKS

The analysis data reported by the measurement laboratories are ingested into the SQL database in their corresponding tables (i.e. analysis.Carbon, analysis.CarbonLaser, analysis.HIPS, analysis.Ions, and analysis.Mass). Several checks are performed using the *datvalIMPROVE* package in R, including:

- Data Completeness: the *completeness.check* function returns records with missing analytical data for each module.
- Field Blank Swap: the *ions\_fb.check* and *carbon\_fb.check* functions check for possible swap between same-day field blanks and samples for nylon and quartz filter samples.

Following the checks, concentrations, MDLs, and uncertainties are processed and posted in the analysis.Results table using the *improve\_calculate\_all* function in the *crocker* package. Additional checks on elapsed time and sampling days are performed by running the *etime.check* and *day.count* functions in *datvalIMPROVE*. These checks ensure there are no records with ET greater than 24 hours and no sites with less than 10 or more than 11 sampling days (February is typically an exception).

### 6.3 LEVEL 2 VALIDATION PROCEDURES

Level 2 validation is performed by comparing site-by-site concentration data obtained from different modules as well as by assessing network-wide long-term trends using a variety of computer R scripts and data visualization tools.

### 6.3.1 CROSS-MODULE COMPARISON

#### A-Module vs. B-Module

Quality assurance for the A- and B-Modules consists of comparing the measured concentrations of sulfur and sulfate. Sulfur concentrations are reported through elemental analysis of the PTFE filter from A-Module, while sulfate concentrations are determined by ion chromatography analysis of the nylon filter from B-Module. Discrepancies between A-Module sulfur (times three, S3) and B-Module sulfate (SO4) concentrations are investigated. If analytical error is suspected, a request is sent to the corresponding laboratories for a reanalysis of the sample.

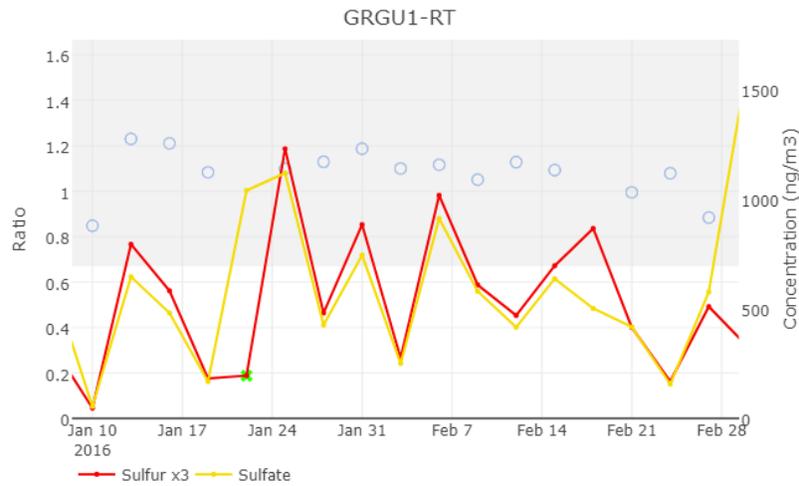
The *swap.check* function in the *datvalIMPROVE* package returns samples marked as “swap” and/or “outlier”. For checking possible sample swaps, successive pairs of data are examined using the algorithm outlined below. In equation (351-47), two indices for each pair of sulfur and sulfate data are calculated using data from the current and the next sampling days (referred to as subscript 1 and 2, respectively).

$$Index1 = \left( \frac{S3_1}{SO4_1} - 1 \right) \times \left( \frac{S3_2}{SO4_2} - 1 \right) \quad Index2 = \left( \frac{S3_1}{SO4_2} - 1 \right) \times \left( \frac{S3_2}{SO4_1} - 1 \right) \quad (351-47)$$

If PM<sub>2.5</sub> sulfur are in the form of sulfate, the S3/SO4 ratio is close to unity. If the samples are not subject to a swap, *Index1* would be close to zero and *Index2* would be large (and may be either positive or negative). The criterion for flagging a pair of samples as swap is when *Index1* < -0.03 and -0.05 < *Index2* < 0.05, which have been set empirically. The criterion for the “outlier” flag is when the S3/SO4 ratio < 0.667 or > 1.8.

The S3/SO4 plot on the IMPROVE Data Site is used to further investigate samples flagged as swap and/or outlier. Figure 3 shows an example of an outlier pair at the GRGU1 site on 1/27/2016. On that day, the sulfate concentration is 1041.06 ng/m<sup>3</sup> while the S3 is 195.51 ng/m<sup>3</sup>, yielding a S3/SO4 ratio of 0.19, well below the acceptable range. In cases like this, the flow rate and elapsed time are first examined to make sure the correct flow source code is assigned. If analytical error is suspected, the XRF and/or IC laboratories would perform a reanalysis. If the reanalysis results resolve the issue, the sample mass loadings would be updated in the SQL database and the concentration data reprocessed. If the reanalysis results are the same as the original analysis, the samples may be flagged as terminal with XX (Sample Destroyed, Damaged or Contaminated) status.

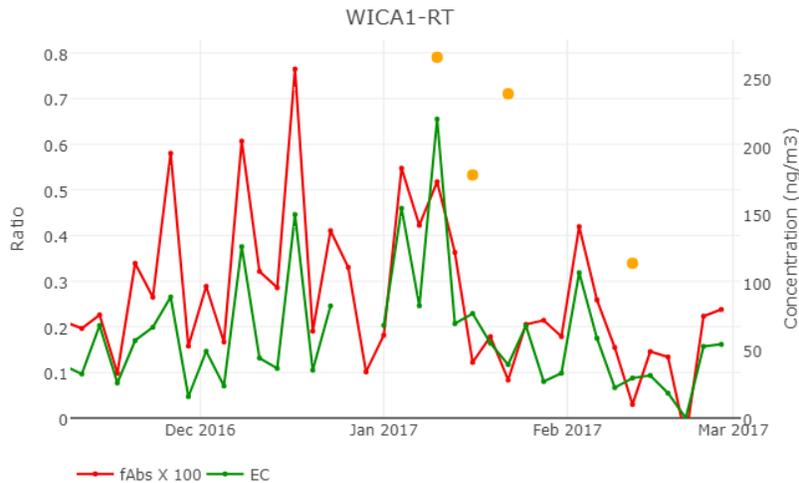
**Figure 3.** S3/SO4 comparison plot for the GRGU1 site showing the 1/21/2016 sample pair as an outlier (green x).



### A-Module versus C-Module

The light absorption coefficient ( $f_{abs}$ ) at 635 nm measurement from A-Module by HIPS is qualitatively compared with the C-Module elemental carbon (EC) concentration determined by the DRI Thermal-Optical OC/EC analyzer. Visual inspection on the data is performed to identify outliers using the  $f_{abs}/EC$  plot on the IMPROVE Data site. The  $f_{abs}$  and EC are typically correlated to some extent (Figure 4). If analytical error in either measurement is suspected, other measurement data from the same module would be examined together to determine validity of the sample.

**Figure 4.** Comparison plot of light absorption coefficient measurements ( $f_{abs}$ , times 100) from A-Module and elemental carbon (EC) measurements from C-Module at WICA1 site.



### A-Module versus D-Module

A-module  $PM_{2.5}$  mass and D-module  $PM_{10}$  mass are reviewed and compared (Figure 5). The *mf\_mt.check* function in the *datvalIMPROVE* package returns a list of samples flagged as mass outliers if any of the following criteria are met:

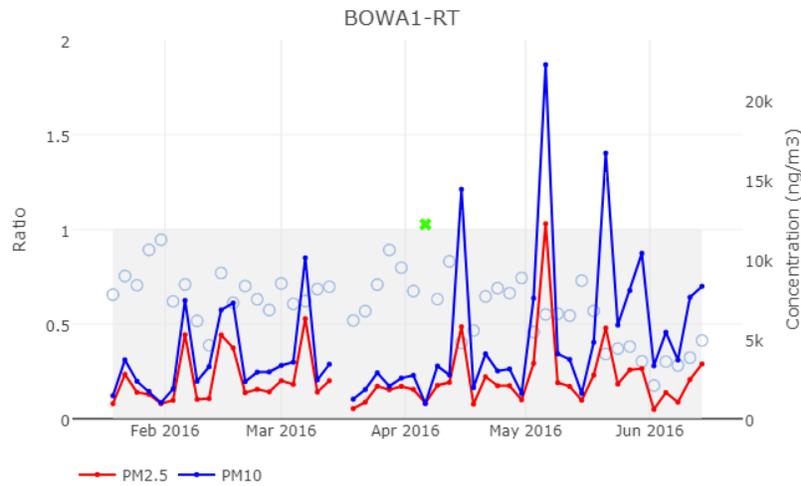
- $PM_{2.5}$  or  $PM_{10}$  mass concentration is negative.
- $PM_{2.5}$  mass is greater than  $PM_{10}$  mass and  $Z\_score > 1$ .
- $PM_{10}$  mass is abnormally high and  $Z\_score > -43$  (the number 43 is set empirically).

$Z\_score$  is calculated using equation (351-48)

$$Z\_score = 1.41 \times \frac{PM_{2.5} - PM_{10}}{\sqrt{(unc_{PM_{2.5}})^2 + (unc_{PM_{10}})^2}} \quad (351-48)$$

For samples that are flagged for one of the above cases, the filter is reweighed to confirm the post-weight; the pre-weight cannot be re-determined after sampling. Samples with invalid mass concentrations are flagged as “UN” (Undetermined Weight).

**Figure 5.** Time series plot of  $PM_{10}$  and  $PM_{2.5}$  masses and their ratio at BOWA1 site.



### PM<sub>2.5</sub> reconstructed mass versus gravimetric mass

The  $PM_{2.5}$  reconstructed masses, RCMC and RCMN, are calculated by equations 351-40 and 351-43, respectively. RCMC and RCMN are compared to the gravimetric mass (MF) as a check of all measured components from A-, B-, and C-Modules (Figure 6). The *mf\_rmc.check* function in the *datvalIMPROVE* package returns a list of samples flagged as outliers if any of the following criteria are met:

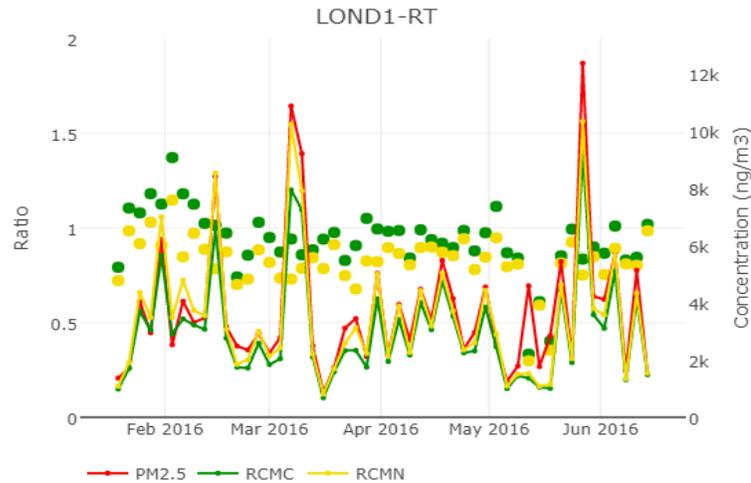
- RCMC is higher than two times of MF, and the RCMC  $Z\_score > 3$ ; the number 3 is set empirically. These samples are accompanied with a comment “MF << RCMC”.
- The RCMN  $Z\_score < -22$ ; the number 22 is set empirically. These samples are accompanied with a comment “MF >> RCMN”.

$Z$  scores are calculated as follows:

$$RCMC\_Z\_score = 1.41 \times \frac{RCMC - PM_{2.5}}{\sqrt{(unc_{PM_{2.5}})^2 + (unc_{RCMC})^2}} \quad (351-49)$$

$$RCMC\_Z\_score = 1.41 \times \frac{RCMC - PM_{2.5}}{\sqrt{(unc_{PM_{2.5}})^2 + (unc_{RCMC})^2}} \quad (351-50)$$

**Figure 6.** Time series plot of PM<sub>2.5</sub> gravimetric mass, reconstructed mass without nitrate (RCMC) and reconstructed mass with nitrate (RCMN) and their ratios at LOND1 site.

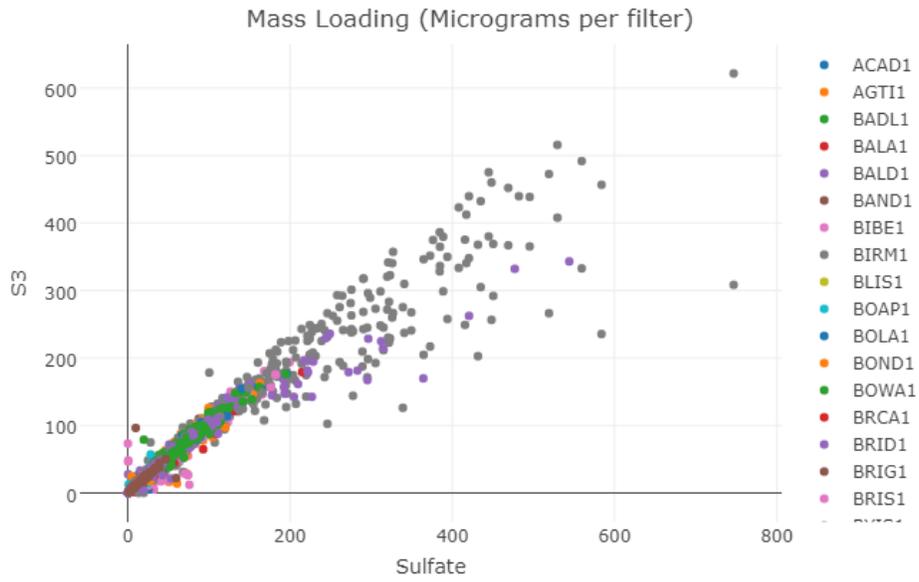


### 6.3.2 LONG-TERM NETWORK-WIDE CHECKS

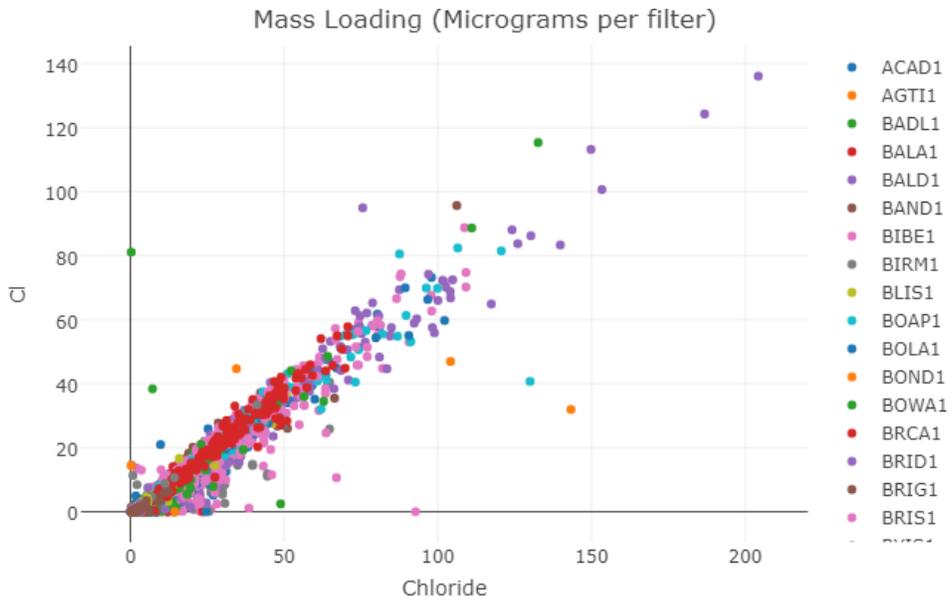
Several data visualization tools and control plots are developed for long-term network-wide checks in addition to the site-by-site monthly data evaluation. These checks help reveal the long-term trends and seasonal patterns if any, as well as any network-wide problems. Below are examples of the tools and plots routinely used and reviewed:

- Scatter plot of S3 versus SO4 mass loadings for the whole network (Figure 7). This plot is accessible from the IMPROVE Data site, "Early Review" tab.
- Scatter plot of chlorine versus chloride mass loadings for the whole network (Figure 8). This plot is accessible from the IMPROVE Data site, "Early Review" tab.
- Time series plot of the A-to-D mass loading ratio showing the long-term trend and historical data at a given site (Figure 9). This tool is accessible from the IMPROVE Data site, "Mass Review" tab.
- Monthly median, 90%, and 10% percentiles of the concentration data for all reported species. Figure 10 shows an example time-series plot for OC concentrations between 2011 and 2016. These plots are generated in R.

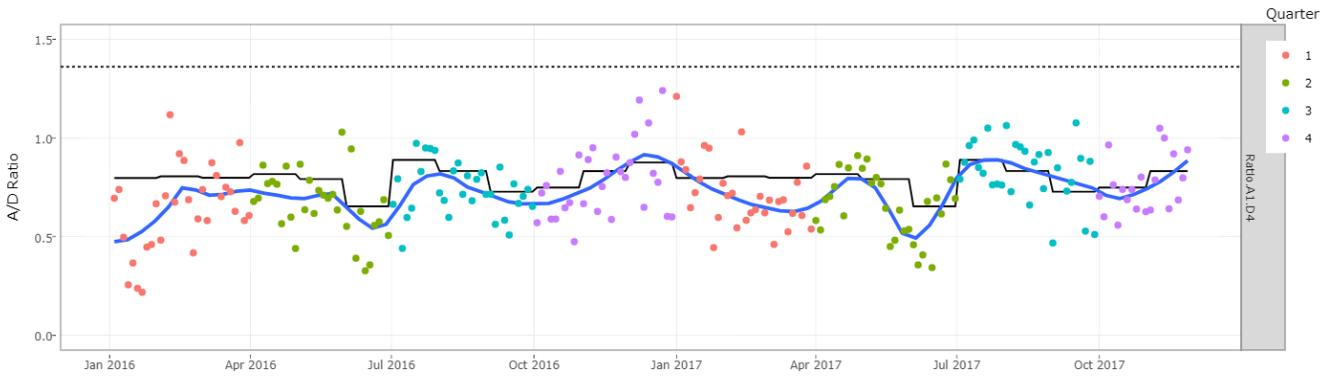
**Figure 7.** Scatter plot of sulfur ( $\times 3$ ) versus sulfate for the entire IMPROVE network.



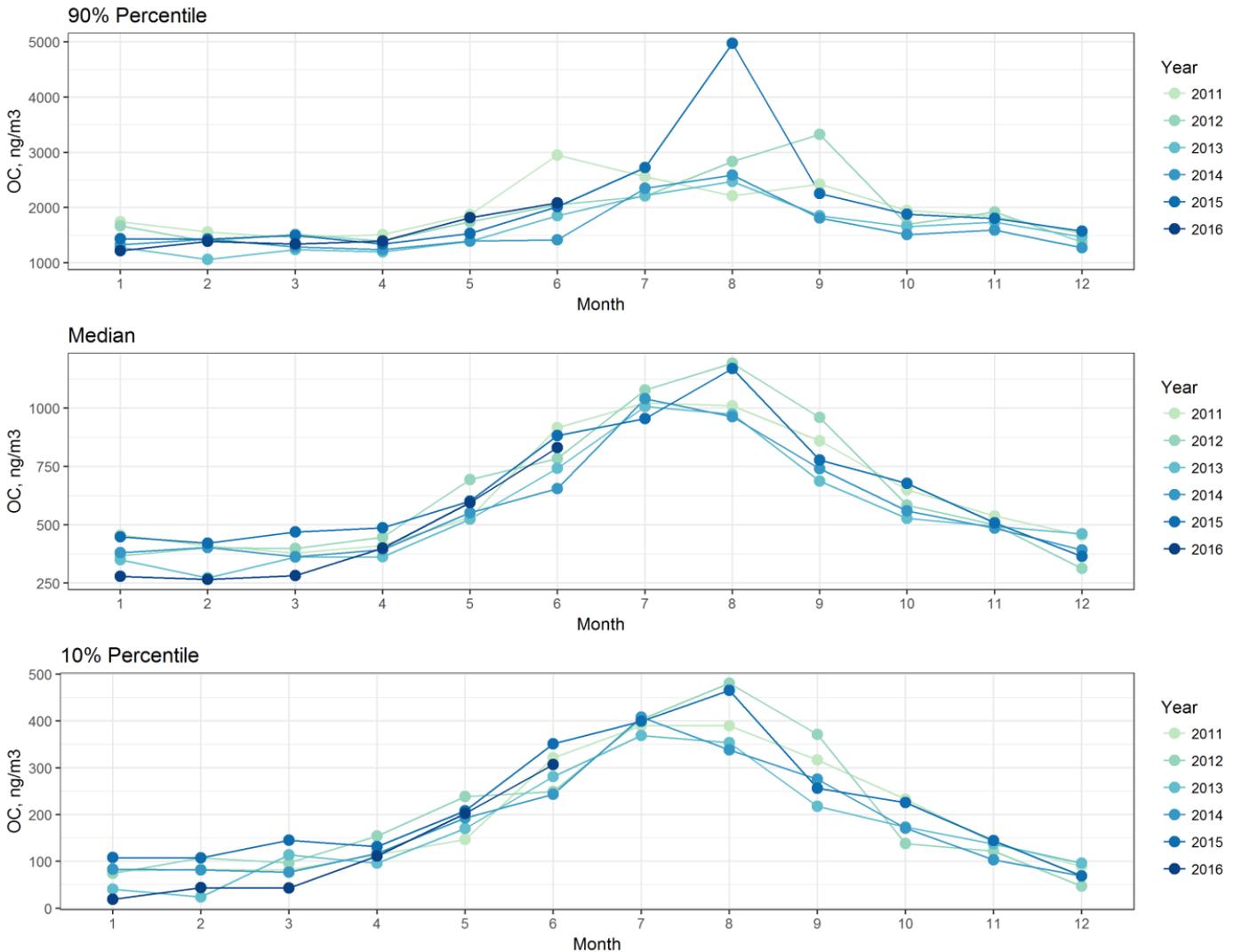
**Figure 8.** Scatter plot of chlorine versus chloride for the entire IMPROVE network.



**Figure 9.** Ratio of PM<sub>2.5</sub> mass (A) over PM<sub>10</sub> mass (D) at ACAD1 site, represented as raw measurements not adjusted for flow rates. Points are individual sample days (pink = Q1, green = Q2, blue = Q3, purple = Q4). Black line is the multi-year monthly mean. Blue line is the locally weighted average (LOESS).



**Figure 10.** Multi-year monthly 10% percentile (top), median (middle) and 90% percentile (bottom) of organic carbon (OC) concentrations (in ng/m<sup>3</sup>) for the whole IMPROVE network from 2011 to 2016.



### 6.3.3 FINAL REVIEW

There are a few final checks before submission of the data delivery file:

- The *QD.check* function in *datvalIMPROVE* returns a list of the QD samples for a given time period. The QD status is normally assigned by the sample handling lab technicians during initial inspection of the physical samples and the raw flow rate data. Every case is investigated, and resolved by either changing the status back to NM or replacing the QD with appropriate terminal or informational flags. There should be no records with QD in the status field in the delivery files.
- The *ObjCode.check* function in *datvalIMPROVE* performs a check on the ObjectiveCode field in the data file. This field should only contain RT (routine) or CL (collocated).
- The *ValidSta\_NullData* function in *datvalIMPROVE* checks if there are valid data (instead of -999) for all valid samples.

## 7. DATA DELIVERY

After Level 2 data validation is complete, the data files are submitted to Cooperative Institute for Research in the Atmosphere (CIRA) and the EPA Air Quality System (AQS) databases.

### 7.1 SUBMISSION TO CIRA

FED export files are created using the *improve\_export\_fed* and *improve\_export\_wide* functions in the *crocker* package, in which the year, month, and server for both functions are entered. The functions create “skinny” and “wide” versions of the dataset, and both are submitted. The files are saved under *U:\IMPROVE\FED Export*, named ‘IMPROVE\_Data\_YYYY\_MM\_server’ and ‘IMPROVE\_WideData\_YYYY\_MM\_server’ (e.g., ‘IMPROVE\_Data\_2017\_02\_production’), respectively. These files are compressed into a zip folder and are emailed to the CIRA correspondent(s) as an attachment.

### 7.2 SUBMISSION TO AQS

Data files are prepared and delivered to AQS following these steps:

1. Create the AQS delivery files using the IMPROVE Management Site, ‘Analysis Data’ tab. Choose AQS for the ‘Format’ and fill in the Year, Start Month and End Month.
2. Click continue to automatically generate the file. Save the file to the UCD U Drive: *U:\IMPROVE\AQS\AQS Export*.
3. Open a web browser and navigate to the EPA Exchange Network Services website, <https://enservices.epa.gov/login.aspx>. Use credentials to login.
4. From the home screen, search for the AQS submission form by clicking on the “Go” button of the Exchange Network Services bar. Type “AQS” into the search bar. The search results will show all available processes associated with the AQS system. Choose the service that has “AQS Submit” specified in the “Service Name” field. This will take the analyst to the AQS submission form.
5. Fill out the submission form, specifying email address, AQS user ID, screening group (IMPROVE), the file type (FLAT), the final processing step (POST), and whether or not

to stop on errors (YES). Use the “Choose File” button to select the file generated from the previous step. Press the “SEND DATA” button to submit the form.

6. Monitor progress of the data submission through the same web portal.

## **8. REFERENCES**

Hyslop, N.P. and White, W.H. (2008) Estimating Precision Using Duplicate Measurements. *J. Air & Waste Manage. Assoc.* **59**:1032–1039.DOI:10.3155/1047-3289.59.9.1032.

John, W. and Reischl, G.P. (1980) A Cyclone for Size-Selective Sampling of Ambient Air, *J. Air Pollut. Control Assoc.*, 30 (8), 872-876.

Watson, J.G.; Liroy, P.J.; Mueller, P.K. (1995). The measurement process: Precision, accuracy, and validity. In *Air Sampling Instruments for Evaluation of Atmospheric Contaminants*, 8th; Cohen, B. S., Hering, S. V., Eds.; American Conference of Governmental Industrial Hygienists: Cincinnati, OH, 187-194.