UCD IMPROVE Standard Operating Procedure #351

Data Processing and Validation

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

June 15, 2020
Version 5.0

Updated By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 7/2/2020

Approved By: DocuSigned by: Katrina Gorham 1B2408CC2DFE4FF... Date: 7/2/2020

UCDAVIS
AIR QUALITY RESEARCH CENTER

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. PURPOSE AND APPLICABILITY

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

2. RESPONSIBILITIES

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

2.1 Data & Reporting Group Manager

The Data & Reporting Group Manager oversees all aspects of data ingestion, processing, validation, and reporting.

2.2 LEAD Quality Assurance Officer

The lead quality assurance officer:

- devises techniques that improve the efficiency, traceability, and accuracy of the data management;
- develops validation criteria, automated and manual checks, and visualization tools for assessing data quality and consistency;
- reviews method detection limit (MDL) and uncertainty;
- identifies sampling or measurement deficiencies and proposes solutions/improvements;
- critically evaluates the data using knowledge of air quality and atmospheric chemistry to better understand trends and biases in the data at program level scale.

2.3 Quality Assurance Officer

The quality assurance officer:

- receives and ingests the analytical data to the University of California, Davis (UCD) IMPROVE database;
- reviews operational and analytical data for errors or incompleteness;
- processes species concentrations and posts monthly dataset to the UCD IMPROVE database;
- performs automated and manual validation checks on concentration data and determines the validity of samples;
- analyzes time-series and spatial trends in network data to assess data consistency due to sampling, measurement or procedural changes;
- identifies sampling or measurement deficiencies and proposes solutions/improvements;
- communicates with laboratories regarding analytical issues and/or reanalysis requests;
- submits Level 2 validated data to project sponsors, Cooperative Institute for Research in the Atmosphere (CIRA), the EPA Air Quality System (AQS), and UCD CSN & IMPROVE Archive (CIA) databases.

3. REQUIRED EQUIPMENT AND MATERIALS

The data processing and validation requires all operational and analytical data be loaded into the UCD IMPROVE database (Improve_2.1). The types of data include:

- 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 a raw flow readings are either acquired from sampler flashcards and stored in the *sampler.FlowSourceData* table (for V2 controllers) or uploaded daily by the controller and stored in the sampler.FlowSourceDataV2 table (for V4 controllers). In addition, handwritten log sheets that contain flow readings and other sampling information recorded by the operator are stored in the *filter.Filters* and *filter.SampleCartridges* tables.
- Average flow rates (24-hour average) are calculated using a SQL procedure called *sampler.spFilterAverageFlowRates* for each filter based on the raw flow readings or log sheet data. These are stored in the *sampler.AverageFlows* table.
- Pre- and post-sampling filter mass values are acquired in the UCD Sample Handling Laboratory and stored in the *analysis.Mass* table.
- Carbon analysis results are acquired from files generated by Desert Research Institute (DRI; Reno, NV) TOR Laboratory and are stored in the *analysis.Carbon*, *analysis.CarbonLaser*, and *analysis.CarbonRun* tables.
- Ions analysis results are acquired from files generated by Research Triangle Institute (RTI; Research Triangle Park, NC) IC Laboratory and are stored in the *analysis.Ions* table.
- Elements analysis results are acquired from the UCD XRF Laboratory through a custom ingestion process and are stored in two tables in the database: *xrf.SampleAnalysis* and *xrf.DeviceCounts*. These are the main tables with mass loading results, reported as raw areal densities from the XRF instruments (ug/cm2). The *DeviceCounts* table contains the XRF results for each element and the *SampleAnalysis* table contains information about the filter analyzed, the instrument used for analysis, and the date and time of analysis.
- Optical absorption analysis results are acquired from the UCD Hybrid Integrating Plate/Sphere (HIPS) Laboratory through a custom ingestion process and are stored in the *hips.SampleAnalysis* table.

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 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 quality control (QC) (see Section 6.3).

IMPROVE Management Website (https://improve.aqrc.ucdavis.edu/): This web application provides all UCD laboratory staff with viewing access to relevant tables within the UCD IMPROVE database. Functions within the application relevant to data processing and validation include:

- The Filter Section (https://improve.aqrc.ucdavis.edu/Filters) consists of web pages for searching for specific filters, reviewing operational and analytical data associated with a filter, or applying flags and comments.
- The XRF Section (https://improve.aqrc.ucdavis.edu/Xrf/Home) is an interface for processing XRF elemental mass loadings, managing processed sets, and applying flags.
- The Analysis Data Section (https://improve.aqrc.ucdavis.edu/AnalysisData/Home) consists of web 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.
- The Operations Section (https://improve.aqrc.ucdavis.edu/Operations/Home) is a live display of the sampler status for the sites equipped with the V4 controllers.

Flow Graphs (http://analysis.crocker.ucdavis.edu:3838/FlowRates/): This web application provides interactive visualizations of the raw 15-minute flow rates and temperatures as well as the processed 24-hr average flow rate in the UCD IMPROVE 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 INGESTION

Prior to data processing and validation, data are ingested for each of the analysis pathways: (1) carbon results from DRI, (2) ions results from RTI, and (3) elemental and optical absorption results from UCD.

4.1 Carbon Results

Carbon analysis results are sent from DRI to UCD via email in .xml format, including three files:

1. CarbonData.xml
2. CarbonInformation.xml
3. CarbonLaser.xml

All three files are ingested using the UCD IMPROVE Management website. Figure 1 shows a screenshot of the carbon data upload page, which is accessed via the Analysis Data Section as described in Section 3, selecting the **Carbons** tab, and clicking the **Ingest Data** button. To ingest the files from the data upload page, select the relevant files, create a name for the import batch under *Batch Label*, and click **Submit**. *CarbonInformation*, *CarbonLaser*, and *CarbonData* are ingested simultaneously, and an automated validity check is performed (Table 1). Results from the validity check will indicate upload failures. The Quality Assurance Officer will review the upload results and notify the Lead Quality Assurance Officer if there are upload failures from

validation errors. After ingest, the source files are stored on the file server at U:\IMPROVE\RawDataReceived\Carbon DRI\Imported, within a folder which is named in accordance with the sample period covered by the source files. After successfully ingesting the results, details of the ingest file are recorded in a logfile located at U:\IMPROVE\RawDataReceived\Carbon DRI\Carbon_Ingest_log.xlsx.

Figure 1. Carbon analysis results upload page.

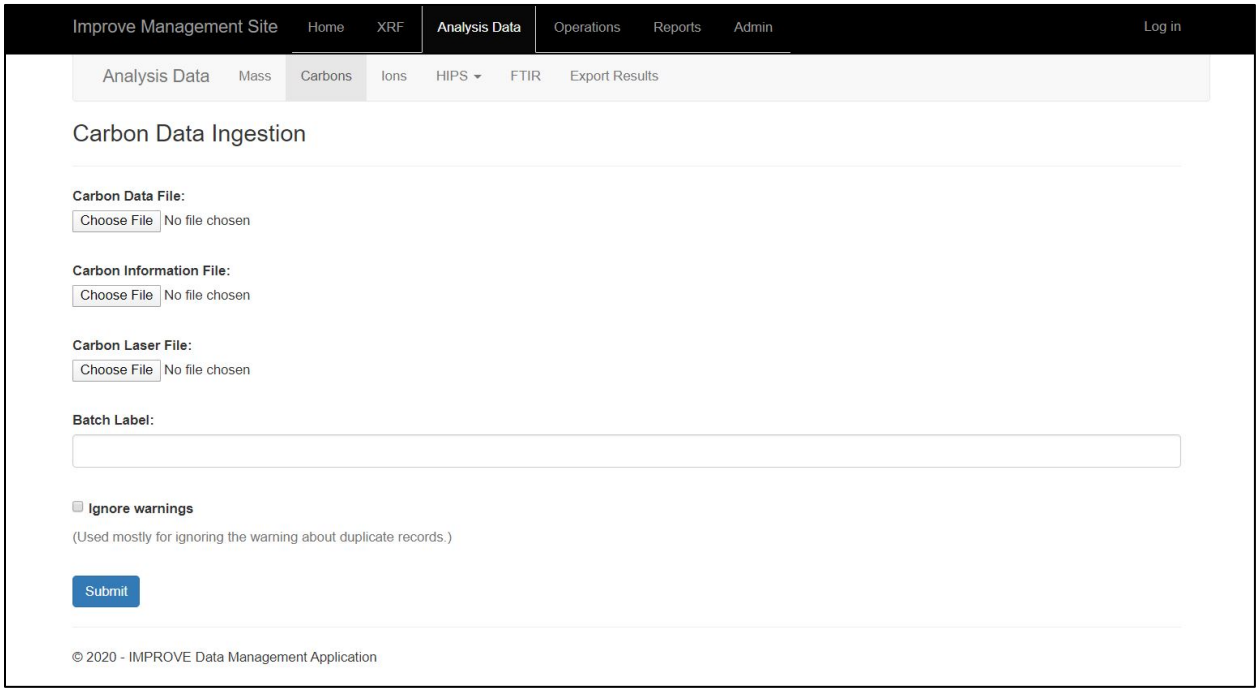


Table 1. Automated validity checks performed during carbon data upload.

Check	**Action**
Basic schema validation on xml files	Error
No filter found for record	Warning
Filter.Module doesn't match record Site field	Warning
Record is marked as re-analysis	Warning
CarbonLaser file has records missing wavelength	Warning
Found more parameter records than expected for an analysis	Warning
Parameter missing for an analysis	Warning
Comment from DRI on analysis	Note
Parameter/record already recorded in database	Warning

Incomplete analysis record (missing entries in either Carbon/CarbonLaser/CarbonInfo file)	Warning

4.2 Ions Results

Ions analysis results are sent as one file from RTI to UCD via email in .csv format. The naming convention of the ions data includes the year followed by the ions data set number (e.g. '2020 1 2 3 data export to UCD').

The ion analysis records are ingested using the UCD IMPROVE Management website. Figure 2 shows a screenshot of the ions data upload page, which is accessed via the Analysis Data Section as described in Section 3, selecting the Ions tab, and clicking the **Upload Data** button. To ingest a file from the data upload page, select the relevant file and click **submit**. An automated validity check is performed, and results from the validity check will indicate if there are upload failures (Table 2). The Quality Assurance Officer will review the upload results and notify the Lead Quality Assurance Officer if there are upload failures from validation errors. After ingest, the source files are stored on the file server at U:\IMPROVE\RawDataReceived\Ions RTI\Ingested. After successful ingest, the relevant details from the file are recorded in a log file located at U:\IMPROVE\RawDataReceived\Ions RTI\Ions_DataIngest_Log.xlsx.

Figure 2. Ions analysis results upload page.

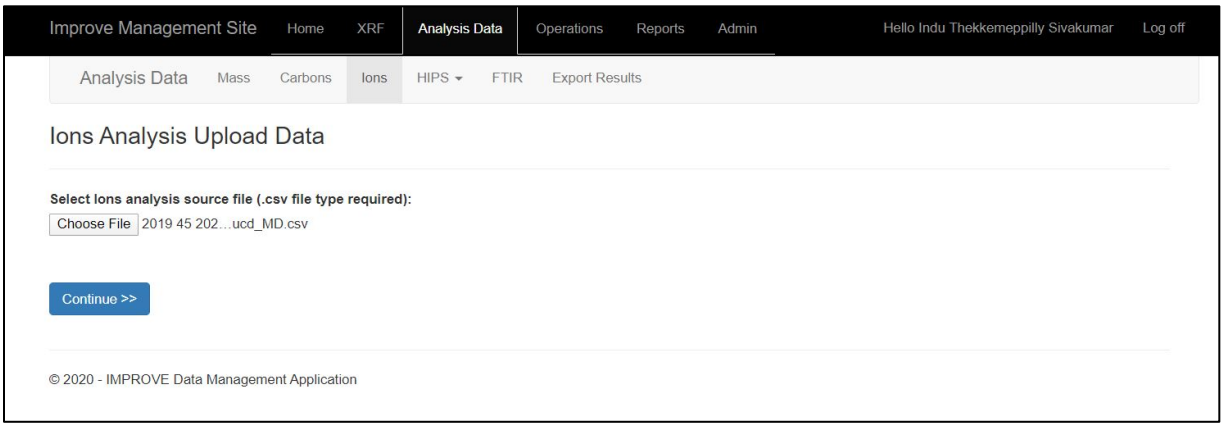


Table 2. Automated validity checks performed during the ions data upload.

Check	**Action**
Basic schema validation on csvfiles	Error message
No filter is found for record	Error message
Data already exists for filter record	Warning message
Parameter missing for a filter	None
Parameter already recorded in database	None

4.3 Element and Optical Absorption Results

Elemental analysis is performed at UCD. The PANalytical XRF software generates results files, which are automatically ingested. The results files are transmitted to a directory on the PANalytical XRF PC (C:\PANalytical\Transmission), and a Windows service (internally named *XRF Data Transfer*) monitors a transmission directory, checking hourly for new files. The XRF results files are standard text files with the extension *.qan*. The file name includes XRF analysis dates and times in the format *YYYYMMDDHHMMSS.qan*. The results files and contents are automatically parsed and ingested into tables in the UCD IMPROVE database.

Optical absorption analysis is performed at UCD. The HIPS instrument generates results in an Excel template file. The data are then filtered to exclude results not related to samples (e.g. hole measurements), and are copied/pasted into a spreadsheet formatted for upload into the database and saved as a comma separated value (csv) file. The csv file of results are uploaded by a laboratory technician into the UCD IMPROVE database through the IMPROVE Management Website (see Section 3).

4.4 Re-ingesting

If errors are identified in the source files from DRI or RTI that cause the import to fail, or if results are updated as part of the validation and reanalysis process, new files must be requested and provided for ingestion. Upload the new files using the process described in Sections 4.1 and 4.2.

For carbon, whether the files contain new batches of data or reanalysis results, take care to ingest with the *ignore warnings* box unchecked. Scrutinize the messages and warnings to check for errors and take note of further actions that may be required after the data is ingested (e.g. changing analysis QC codes). The import process indicates if there are matching existing records; existing records are not updated, only new records are added (including cases with different analysis results from the sample filter). Once the messages have been reviewed and addressed, re-run the ingest process with the *ignore warnings* box checked.

For ions, the data are ingested without any changes to the original process; the QC code is updated in the UCD IMPROVE database.

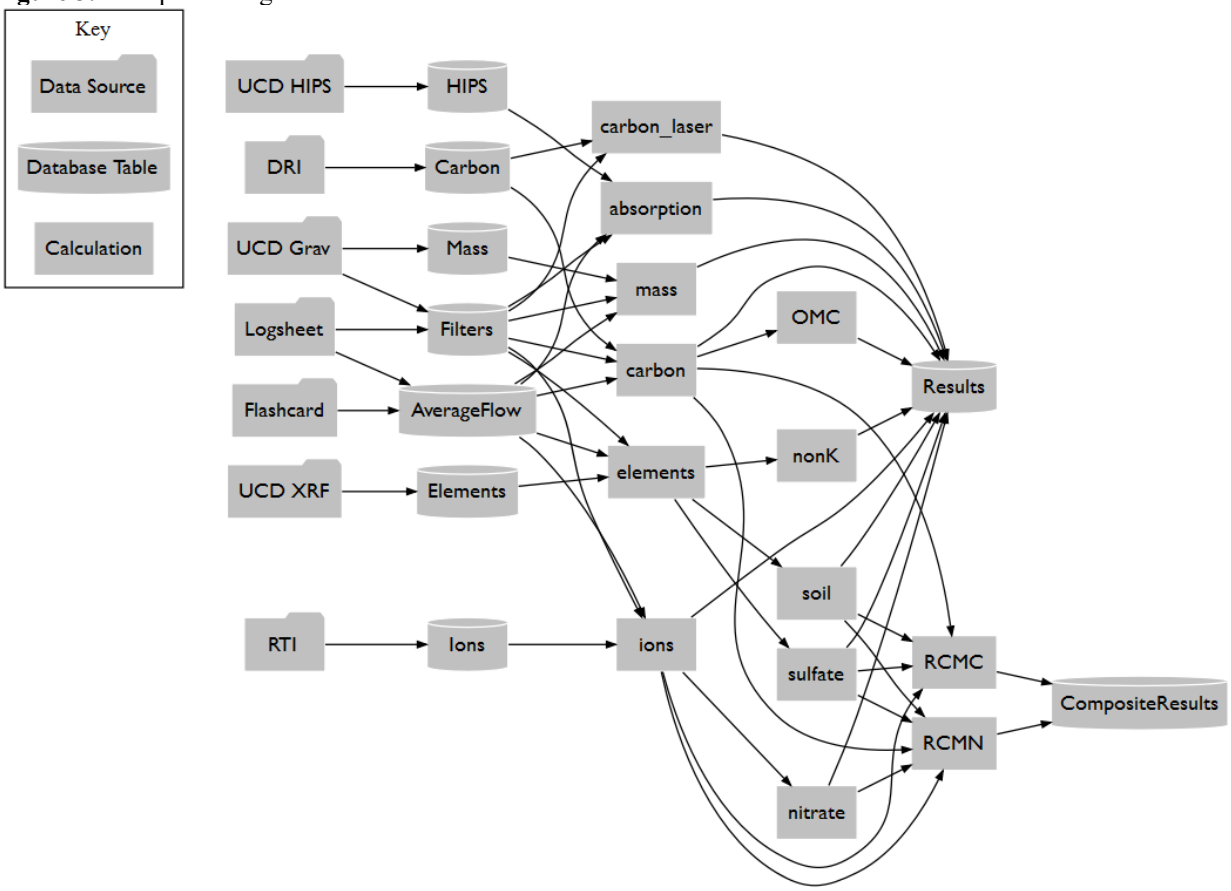
4.5 Issue Tracking

Software bugs and data management issues are tracked through JIRA tracking software. All users have access to the internal UCD JIRA website and can submit, track, and comment on issues.

5. DATA PROCESSING

Data processing for IMPROVE consists of reducing and combining data from the sampling and analytical laboratories to calculate concentrations, uncertainty estimates, and method detection limits (MDLs). Figure 3 shows a flow chart for the IMPROVE data processing.

Figure 3. Data processing flow chart.



Calculation of concentrations and associated uncertainties and MDLs are performed within the *crocker* R package, while flow rate calculations are performed in the UCD IMPROVE database. To calculate values for all measured and derived parameters, the following command is run in an R environment:

[month_data] <- crocker::improve_calculate_and_post ([YYYY], [MM], server, AnalysisQcCode = 1, comment = "", replacingId = NULL, replacingQcCode = NULL)

This command will calculate concentrations, uncertainties, and MDLs for all measured and derived parameters for the year (*[YYYY]*) and month (*[MM]*) and upload the results to the UCD IMPROVE database (*server*, e.g. "production") specified in the command in preparation for validation. The processed concentration data are appended to the analysis.Results and analysis.CompositeResults table in the UCD IMPROVE database (Improve_2.1). A record that contains summary information for the data set, including the comment and the AnalysisQcCode, is inserted into the analysis.Sets table. An AnalysisQcCode of 1 is used for valid routine data.

5.1 Units

Table 3 lists the data types, parameters, and units for all data delivered to the CIRA, AQS, and UCD CIA databases (see Section 7). For mass, ions, carbon, elements, and light absorption, the

units listed are also used for uncertainty and MDL. NA indicates that the data type is not reported to the corresponding database.

Table 3. Units for data delivered to the CIRA, AQS and UCD CIA databases.

Data type	Parameter	CIRA unit	AQS unit	UCD CIA unit
Flow Rate	Flow	L/min	NA	NA
Elapsed Time	ET	min	NA	NA
Gravimetric mass	PM2.5, PM10	ng/m^3	$\mu g/m^3$	$\mu g/m^3$
Ions	Cld, NO_2, NO_3, SO_4	ng/m^3	$\mu g/m^3$	$\mu g/m^3$
Carbon	OC1, OC2, OC3, OC4, OC, OPTR, EC1, EC2, EC3, EC,	ng/m^3	$\mu g/m^3$	$\mu g/m^3$
	TC, OPTT, OPTR at other wavelength, OPTT at other wavelength	ng/m^3	NA	NA
Carbon_laser	RefF_wavelength, Refl_wavelength, RefM_wavelength, TransF_wavelength, Transl_wavelength, TransM_wavelength	reading	NA	NA
Elements	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^3	$\mu g/m^3$	$\mu g/m^3$
Light absorption	*f*abs	Mm^{-1}	NA	NA
Composite species	OMC, NHNO, NHSO, PM_{10}-$PM_{2.5}$	NA	$\mu g/m^3$	$\mu g/m^3$

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 applied to the ion, carbon, and element measurements. Artifact examples include:

(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 and is estimated using data from field blanks. 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 ≥ 50 field blanks in a month, the artifact correction is calculated for each species as the median loading measured on the field blanks. Otherwise, values from the previous month(s) are included until at least 50 field blanks are available. Artifact corrections are subtracted from each ambient concentration for the corresponding month.

For the carbon measurements, the artifact correction method attempts to account for the first three types of artifacts and is estimated using data from 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 ≥ 50 field blanks in a month, the artifact correction is calculated for each species as the median loading measured on the field blanks; otherwise, values from the previous month(s) are included until at least 50 field blanks are available. Artifact corrections are subtracted from each ambient concentration for the corresponding month. For further background information and detail regarding past use of stacked filters for artifact correction and subsequent application of a correction factor, see data advisories:

http://vista.cira.colostate.edu/Improve/wp-content/uploads/2016/04/Dillner_OCArtifactAdjustmentIMPROVEOct2012.pdf and http://vista.cira.colostate.edu/improve/Data/QA_QC/Advisory/da0032/da0032_OC_artifact.pdf

Measurements are not corrected for the two negative artifact types (volatilization and fall-off). The measured mass loadings for the higher-volatility organics may be much less than those in the atmosphere because of volatilization of particles during the remainder of the sampling or during transportation. Volatilization of nitrate and chloride from the nylon filters is assumed to be insignificant. Depending on the environmental conditions, some ammonium nitrate collected on polytetrafluoroethylene (PTFE) filters may volatilize. In those cases, fine mass on the PTFE filter may underestimate the ambient $PM_{2.5}$ mass concentrations.

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

5.3 Volume

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

For the $PM_{2.5}$ modules (1A, 2B, and 3C modules), the flow rate is determined from measurement of static pressure across the cyclone using a pressure transducer (referred to as the CYC value). Since the pressure is measured before the filter, a decrease in measured flow rate could correspond with a lightly loaded filter since a smaller volume of air is being sampled. Prior to 2016, the 15-minute pressure measurements were averaged over the whole sampling period (nominally 24 hours) for calculating the average flow rate. Beginning data for samples collected in January 2016, the average flow rate is an elapsed time-weighted average, calculated from the individual 15-minute pressure measurement. The sampler flow rate for 1A, 2B, and 3C modules is calculated using equation 351-1.

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

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 the PM_{10} module (4D module), the flow rate is determined from measurement of absolute pressure downstream of the filters near the critical orifice using a pressure transducer (referred to as the ORI value); the CYC value is not available for the 4D module. Since the pressure is measured after the filter, a decrease in measured flow rate could be indicative of a heavily loaded filter or filter clogging that is restricting the flow. The sampler flow rate is calculated using equation 351-2.

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

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.

The calibration coefficients (a, b, c, and d) in equations (351-1) and (351-2) have historically been site-specific. Staring with data from samples collected January 2018, a set of universal flow constants for the V4 controller cyclone (CYC; equation 351-1) and orifice (ORI; equation 351-2). The constants are reviewed annually and updated as needed; the values are expected to vary minimally from year to year (Table 4).

Table 4. Universal flow constants for the V4 controllers.

Module	Intercept (a, c)*	Slope (b, d)*
$PM_{2.5}$	1.4891	0.3797
PM_{10}	1.320	1.325

* Applied to data from 1/1/2018 onward.

5.4 Concentration, Uncertainty, and Method Detection Limit

The calculations described in this section are performed in R using the R function listed at the beginning of Section 5.

The concentration is calculated using equation 351-3, where 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. For gravimetric analysis, the mass on the unused filter is determined from the pre-weight of individual PTFE filters. For measurement of ions and carbon, the mass on the unused filter is determined from the median of field blank loadings. For calculation of element concentrations, see Section 5.4.4.

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

C = ambient concentration (ng/m^3)

A = mass measured on sample (ng/filter or ng/cm^2)

B = artifact mass (ng/filter or ng/cm^2) = pre-weight or monthly median of ion or carbon field blank mass loading

V = sample air volume (m^3) = Q * Elapsed Time

Q = volumetric flow rate

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.

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

σ_a = analytical uncertainty. This is a constant term from additive sources of uncertainty, such as those related to background contamination of the filters. Analytical uncertainty is determined and reported by the laboratories. For large concentrations, this is small compared to the fractional term.

V = sample air volume (m^3)

C = ambient concentration (ng/m^3)

f = fractional uncertainty. This term results from various sources of proportional uncertainties, such as the analytical calibration and flow rate measurements. Beginning with data from samples collected January 2018, fractional uncertainties (f) are determined using the most recent two years of data from collocated measurements (352-5 and 351-6). If the count of collocated pairs over the two year period is less than 60, a value of 0.25 is adopted as f.

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

$$f = \frac{(84th\ percentile\ of\ srd) - (16th\ percentile\ of\ srd)}{2} \quad (351\text{-}6)$$

The *improve_fracUnc* function is run using the *crocker* R package to calculate and post a new set of fractional uncertainties as well as to replace older sets, when necessary.

improve_fracUnc(startdate, enddate, effectivedate, server = "production", AnalysisQcCode = 1, comment = "", replacingId = NULL, replacingQcCode = NULL)

For example, processing the 2019 concentration data should use the fractional uncertainties (f) calculated from 1/1/2017 through 12/31/2018 data. The function *improve_fracUnc* calculates and directly imports fractional uncertainty into database tables, Improve_2.1.analysis. UncertaintySets and Improve_2.1.analysis.Uncertainties.

improve_fracUnc(startdate = "2017-01-01", enddate = "2018-12-31", effectivedate = "2019-01-01", server = 'production', comment = "New set to be applied beginning with 2019 data")

For further details, refer to the function help file in R.

The MDLs are also reported with each concentration. Beginning with data from samples collected January 2018, MDLs for ions, carbon, and elements are calculated as 95th percentile minus median of field blanks.

5.4.1 $PM_{2.5}$ and PM_{10} Mass (1A and 4D Modules)

$PM_{2.5}$ mass is measured gravimetrically on the PTFE filter from the 1A Module. PM_{10} mass is measured gravimetrically on the PTFE filter from the 4D Module. The pre- and post-weights (as micrograms per filter) are stored in the *analysis.Mass* table in the UCD IMPROVE database.

The constant analytical uncertainty, σ_a, in equation 351-4 is equal to 5 μg for all filters. The mass concentration (C_{Mass}), uncertainty (σ_{Mass}), and MDL (mdl_{Mass}) in nanograms per cubic meter are calculated using the following equations:

$$C_{Mass} = 10^6 \, ng/mg * \left(\frac{Postweight - preweight}{V} \right) \quad (351\text{-}7)$$

$$\sigma_{Mass} = 1000 \, ng/\mu g * \sqrt{\left(\left(\frac{5 \mu g}{V} \right)^2 + \left(\frac{C_{Mass} * f}{1000 \, ng/\mu g} \right)^2 \right)} \quad (351\text{-}8)$$

$$mdl_{Mass} = 1000 \, ng/\mu g * \frac{10 \mu g}{V} \quad (351\text{-}9)$$

Where,

V = A-Module sample air volume (m^3)

postweight = mass of filter after sampling

preweight = mass of filter before sampling

f = fractional uncertainty (Table 5).

Table 5. Fractional uncertainty for the mass.

Species	**f reported for data 2/28/1995 – 12/31/2006**	**f reported for data 1/1/2017 – 12/31/2017**	**f reported for data 1/1/2018 – 12/31/2018**	**f reported for data 1/1/2019 – current**
$PM_{2.5}$	0.03	0.03	0.04	0.04
PM_{10}	0.03	0.07	0.07	0.08

5.4.2 Ions (2B Module)

Ions are measured by ion chromatography using the nylon filter from the 2B Module. Ions data (as micrograms per filter) are stored in the *analysis.Ions* table in the UCD IMPROVE database.

The concentration (C_{ion}), uncertainty (σ_{ion}), and MDL (mdl_{ion}) in nanograms per cubic meter are calculated for the ion species using the following equations; however, for nitrite, when the concentration is less than or equal to zero, uncertainty is reported as zero:

$$C_{ion} = 1000\frac{ng}{\mu g} * \frac{(A_{ion} - B_{ion})}{V_{B\,module}} \quad (351\text{-}10)$$

$$\sigma_{ion} = 1000\frac{ng}{\mu g} * \frac{\sqrt{(0.608 * Max(P95 - B_{ion}, mdl_{analytical}))^2 + (f * (A_{ion} - B_{ion}))^2}}{V_{B\,Module}} \quad (351\text{-}11)$$

$$mdl_{ion} = 1000\frac{ng}{\mu g} * \frac{Max(P95 - B_{ion}, mdl_{analytical})}{V_{B\,Module}} \quad (351\text{-}12)$$

Where,

A_{ion} = ambient mass loading in µg/filter

B_{ion} = median of the field blank mass loading in µg/filter when there are ≥ 50 field blanks in a month; otherwise, values from the previous month are used.

$V_{B\,module}$ = B-Module sample air volume (m^3)

P95 = 95th percentile of field blank measurements in µg/filter

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory (0.03 for chloride, 0.01 for nitrite, 0.05 for nitrate, and 0.07 for sulfate). The analytical MDL is considered the 'floor value' and is used as the reported MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

f = fractional uncertainty (Table 6).

Table 6. Fractional uncertainty for ions.

Species	**f reported for data 1/1/2005 – 12/31/2016**	**f reported for data 1/1/2017 – 12/31/2017**	**f reported for data 1/1/2018 – 12/31/2018**	**f reported for data 1/1/2019 – current**
Chloride (Cl^-)	0.08	0.08	0.08	0.09
Nitrite (NO_2^-)	0.22	0.25	0.25	0.25
Nitrate (NO_3^-)	0.04	0.03	0.04	0.04

Sulfate ($SO_4^=$)	0.02	0.02	0.02	0.03

5.4.3 Carbon (3C Module)

Carbon is measured by thermal optical reflectance (TOR) and thermal optical transmittance (TOT) using the quartz filter from the 3C Module. The seven carbon fractions (OC1-OC4, EC1-EC3) and organic pyrolized carbon (OP) are recorded in micrograms per filter and stored in the analysis.Carbon table in the UCD IMPROVE database. For the carbon fractions, the primary factors that determine the fractional uncertainty are the homogeneity of the sample deposit and the accuracy of the temperature set point in each stage. For OP, the primary factors that determine the fractional uncertainty are the laser signal stability and the accuracy of the split point placement.

The TOR elemental carbon (ECTR) component is assumed to be all carbon evolved at 580°C and above, after the laser indicates that reflectance has returned to the initial value. The TOR organic carbon (OCTR) component is assumed to be all carbon evolved at 580°C and below, in a pure helium environment, plus the OP fraction. The total carbon (TC) is sum of OCTR and ECTR. Only the TOR OC and EC are calculated and reported.

The concentration, uncertainty, and MDL in nanograms per cubic meter for the carbon species (OC1, OC2, OC3, OC4, OPTR, OPTT, EC1, EC2, EC3, as well as OCTR, ECTR, TC) are calculated using the following equations:

$$C = 1000\frac{ng}{\mu g} * \frac{\left(A - B\right)}{V_{C\,module}} \tag{351-13}$$

$$\sigma_{Carbon} = 1000\frac{ng}{\mu g} * \frac{\sqrt{(0.608 * Max\left(P95 - B_{carbon}, mdl_{analytical}\right))^2 + (f * (A_{carbon} - B_{carbon}))^2}}{V_{C\ Module}} \tag{351-14}$$

$$mdl_{Carbon} = 1000\frac{ng}{\mu g} * \frac{Max\left(P95 - B_{carbon}, mdl_{analytical}\right)}{V_{C\ Module}} \tag{351-15}$$

Where,

A_{carbon} = ambient mass loading in µg/filter

B_{carbon} = median of the field blank mass loading in µg/filter when there are ≥ 50 field blanks in that month, otherwise the number from the previous month is used.

$V_{C\ Module}$ = C-Module sample air volume (m^3)

P95 = 95th percentile of field blank measurements in µg/filter

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory (Table 7). The analytical MDL is considered the 'floor value' and is used as the reported MDL in the

event that the median value of the field blanks is lower than the respective analytical MDL.

f = fractional uncertainty (Table 8).

Table 7. Analytical method detection limits (MDL) for the carbon species.

Species	**MDL**
OC1	0.51
OC2	0.51
OC3	0.51
OC4	0.51
OPTR, OPTR at other wavelength	0.15
OPTT, OPTT at other wavelength	0.15
EC1	0.15
EC2	0.15
EC3	0.15
ECTR	0.15
OCTR	0.51
TC	0.57

Table 8. Fractional uncertainty for the carbon species.

Species	**f reported for data 1/1/2005 – 12/31/2016**	**f reported for data 1/1/2017 – 12/31/2017**	**f reported for data 1/1/2018 – 12/31/2018**	**f reported for data 1/1/2019 - current**
OC1	0.23	0.27	0.23	0.24
OC2	0.15	0.13	0.11	0.10
OC3	0.13	0.13	0.13	0.11
OC4	0.15	0.13	0.13	0.14
OPTR, OPTR at other wavelength	0.13	0.16	0.20	0.21
OPTT, OPTT at other wavelength	0.13	0.12	0.14	0.15
EC1	0.10	0.10	0.11	0.11
EC2	0.17	0.18	0.19	0.21
EC3	0.42	0.25	0.25	0.25
ECTR	0.12	0.14	0.14	0.13
OCTR	0.08	0.09	0.08	0.07
TC	0.08	0.08	0.07	0.07

5.4.4 Elements (1A Module)

Elements are measured using X-ray fluorescence (XRF; PANalytical Epsilon 5) using the PTFE filters from the 1A Module.

The PANalytical XRF instruments report the elements in terms of counts per mV per second, which is converted into areal densities using element calibration factors (stored in the UCD IMPROVE 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. Since the number of field blanks analyzed on a specific instrument during a month may not be statistically sufficient, field blank selection is not grouped by month. Rather, the UCD IMPROVE database uses a procedure (*xrf.spCreateFeldBlankSet*) to select the last 35 field blanks analyzed before the determination date, which is the date of the last sample analyzed from a one-month batch. The selected 35 field blanks are used to calculate batch and instrument specific blank correction. Areal uncertainty ($U_{element}$) is calculated as,

$$U_{element} = 1000\frac{ng}{\mu g} * \sqrt{(0.608 * Max\left((P95 - B_e), mdl_{analytical}\right))^2 + (f * (A_e - B_e))^2} \quad (351\text{-}18)$$

A_e = areal density calculated for the element measured by XRF.

B_e = median areal density of the field blank measured by XRF; ≥ 35 field blanks from before the determination date.

P95 = 95th percentile of field blank measured by XRF.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory. The analytical MDL is considered the 'floor value' and is used as the reported MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

f = fractional uncertainty (Table 9).

0.608 = 1 / 1.645; used to estimate the one-sigma uncertainty at zero concentration from the MDL that is set at the 95th percentile, where 1.645 is the critical value for sigma in a one-tailed test for 5 % significance.

Areal densities, areal uncertainty, and areal MDL (in units of mass/area) are calculated during processing of XRF results. The concentration ($C_{element}$), uncertainty ($\sigma_{element}$), and MDL ($mdl_{element}$) in nanograms per cubic meter for the element species are calculated using the following equations:

$$C_{element} = 1000\frac{ng}{\mu g} * \frac{(A_e - B_e) * (Deposit\ area)}{V} \quad (351\text{-}19)$$

$$\sigma_{element} = \frac{(U_e) * (Deposit\ area)}{V} \quad (351\text{-}20)$$

$$mdl_{element} = 1000\frac{ng}{\mu g} * \frac{Max\left((P95 - B_e), mdl_{analytical}\right) * (Deposit\ area)}{V} \quad (351\text{-}21)$$

Where,

A_e = areal density calculated for the element measured by XRF.

B_e = median areal density of the field blank measured by XRF; ≥ 35 field blanks from before the determination date

Deposit area = area of sample deposit on the filter (cm^2), determined from the filter holder or mask size (approximately 20 mm).

U_e = areal uncertainty reported for the element measured by XRF.

P95 = 95^{th} percentile of field blank measured by XRF.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory. The analytical MDL is considered the 'floor value' and is used as the reported MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

V = 1A Module sample air volume (m^3).

Table 9. Fractional uncertainty for the elemental species.

Species	f reported for data 1/1/2005 – 12/31/2016	f reported for data 1/1/2017 – 12/31/2017	f reported for data 1/1/2018 – 12/31/2018	f reported for data 1/1/2019 - current
Al	0.09	0.08	0.08	0.09
As	0.25	0.21	0.25	0.25
Br	0.10	0.11	0.10	0.09
Ca	0.06	0.07	0.06	0.07
Cl	0.14	0.18	0.14	0.14
Cr	0.22	0.17	0.15	0.17
Cu	0.12	0.11	0.13	0.10
Fe	0.06	0.06	0.05	0.06
K	0.03	0.05	0.03	0.04
Mg	0.15	0.16	0.15	0.15
Mn	0.13	0.13	0.14	0.13
Na	0.14	0.15	0.14	0.14
Ni	0.16	0.16	0.13	0.14

P	0.25	0.33	0.27	0.30
Pb	0.13	0.13	0.14	0.15
Rb	0.25	0.25	0.25	0.25
S	0.03	0.03	0.02	0.03
Se	0.25	0.12	0.25	0.25
Si	0.10	0.07	0.06	0.07
Sr	0.16	0.14	0.13	0.14
Ti	0.11	0.09	0.09	0.09
V	0.12	0.14	0.17	0.17
Zn	0.06	0.08	0.08	0.08
Zr	0.25	0.25	0.25	0.25

5.4.5 Laser Absorption (1A Module)

Optical absorption is measured by a hybrid integrating plate and sphere (HIPS) system using the PTFE filter from the 1A Module. The laser absorption measurements are stored as reflectance (R) and transmittance (T) values in *hips.SampleAnalysis* table in the UCD IMPROVE database.

Results from the HIPS measurement are reported as filter absorption coefficient (*f*Abs) in units of Mm^{-1}, calculated from R and T. The concentration (*f*abs), uncertainty (σ_{fAbs}), and MDL (mdl_{fAbs}) are calculated using the following equations:

$$fAbs = 100 * \frac{\tau_{633} * (Deposit\ Area)}{V_{A\ Module}} \quad (351\text{-}22)$$

Where,

$V_{A\ Module}$ = 1A Module sample air volume (m^3)

Deposit area = area of sample deposit on the filter (cm^2), determined from the filter holder or mask size (approximately 20 mm).

$$\tau_{633} = \log\left(Max\left(\frac{intercept + (slope * refelctance)}{transmittance}, 0.1\right)\right)$$

$$\sigma_{fAbs} = 100 * \frac{\sqrt{\left(\frac{1}{1.65} * Max\left(P95, mdl_{analytical}\right)\right)^2 + \left(f_{unitless} * \tau_{633}\right)^2} * (Deposit\ Area)}{V_{A\ Module}} \quad (351\text{-}23)$$

Where,

P95 = 95th percentile of field blank measurements.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory. The analytical MDL is considered the 'floor value' and is used as the reported MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

$V_{A\ Module}$ = 1A Module sample air volume (m^3)

Deposit area = area of sample deposit on the filter (cm^2), determined from the filter holder or mask size (approximately 20 mm).

$$\tau_{633} = \frac{\log\left(Max\left(\frac{intercept + (slope * refelctance)}{transmittance}, 0.1\right)\right)}{}$$

$f_{unitless}$ = unitless fractional uncertainty calculated from fractional uncertainty (Table 10) and nominal sample volume.

$$mdl_{FAbs} = 100 * \frac{Max(P95,\ mdl_{analytical}) * (Deposit\ area)}{V_{A\ Module}} \quad (351\text{-}24)$$

Where,

P95 = 95th percentile of field blank measurements.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory. The analytical MDL is considered the 'floor value' and is used as the reported MDL in the event that the median value of the field blanks is lower than the respective analytical MDL.

$V_{A\ Module}$ = 1A Module sample air volume (m^3)

Deposit area = area of sample deposit on the filter (cm^2), determined from the filter holder or mask size (approximately 20 mm).

Table 10. 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 – 12/31/2017**	**f reported for data 1/1/2018 – 12/31/2018**	**f reported for data 1/1/2019 - current**
*f*abs	0.03	0.08	0.06	0.05

5.5 Equations of Composite Variables

The following composite variables are combinations of the measured concentrations, and are used in the Level 2 validation procedures described in Section 6.3. For the composite variables, concentration is determined along with the uncertainty and MDL. The 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 more complicated calculations.

5.5.1 Sulfate (3× sulfur from XRF) 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, $(NH_4)_2SO_4$, although it can be present as ammonium bisulfate, $(NH_4)HSO_4$, sulfuric acid, H_2SO_4, gypsum, $CaSO_4 \cdot 2H_2O$, and, in marine areas, as sodium sulfate, Na_2SO_4. In many cases, the particle will include associated water, this is omitted from the calculation. In order to simplify the calculation, all sulfur is assumed to be present as ammonium sulfate. The concentrations (NHSO and S3), uncertainties (σ_{NHSO} and σ_{S3}), and MDLs (mdl_{NHSO} and mdl_{S3}) for ammonium sulfate (NHSO) and sulfate calculated from XRF sulfur (S3) are calculated using the following equations:

$$NHSO = 4.125 * S$$

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

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

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

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

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

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 (NHNO), uncertainties σ_{NHNO}), and MDLs (mdl_{NHNO}) are calculated using the following equations:

$$NHNO = 1.29 * NO_3^- \quad (351\text{-}28)$$

$$\sigma_{NHNO} = 1.29 * \sigma(NO_3^-) \quad (351\text{-}29)$$

$$mdl(NHNO) = 1.29 * mdl(NO_3^-) \quad (351\text{-}30)$$

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 (Al_2O_3, SiO_2, CaO, K_2O, FeO, Fe_2O_3, TiO_2), and augmented by a factor to account for other compounds not included in the calculation, such as MgO, Na_2O, water, and CO_2. The following assumptions are made:

- Fe is split equally between FeO (oxide factor of 1.29) and Fe_2O_3 (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 $\left(\mathrm{K_2O}, \frac{39.1*2+16.0\ \mathrm{g/mol}}{39.1*2\ \mathrm{g/mol}} = 1.2 \right)$ 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:

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

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

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

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:

$$\mathrm{KNON} = (\mathrm{K} - 0.6*\mathrm{Fe}) \quad (351\text{-}34)$$

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

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

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:

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

$$\sigma_{OMC} = 1.8 \times \sigma_{OC} \quad (351\text{-}38)$$

See equation of 351-14 for σ_{OC}.

$$mdl_{OMC} = 1.8 \times mdl_{OC} \quad (351\text{-}39)$$

See equation 351-15 for mdl_{OC}.

5.5.6 Black Carbon

Black carbon is estimated from the initial and final laser readings from the 3C Module quartz filter analysis. For cross-module validation, black carbon is compared to light absorption coefficient (f_{abs}) measured by HIPS from the 1A Module PTFE filter.

$$BC = \frac{ln\,(transfinal - transinitial)}{MAC} \quad (351\text{-}40)$$

TransFinal = Final laser transmittance value of the sample

TransInitial = Initial laser transmittance value of the sample

MAC = Black carbon mass absorption cross-section and it is a constant of 23 m^2/g at 632.8 nm wavelength.

5.5.7 Reconstructed Mass Using Carbon Measurements (RCMC)

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 and uncertainties are calculated using the following equations; negative values are substituted with zero. RCMC concentration is always positive. Uncertainty is calculated as the combination of the individual uncertainties. The MDL for RCMC is zero.

$$RCMC = NHSO + Soil + 1.8 \times Chloride + ECTR + OMC \quad (351\text{-}41)$$

Where,

NHSO = ammonium sulfate concentration

Soil = soil concentration

Chloride = chloride concentration as measured by IC

ECTR = elemental carbon concentration by TOR

OMC = concentration of organic mass by carbon

$$\sigma_{RCMC} = \sqrt{\sigma_{NHSO}^2 + \sigma_{Soil}^2 + (1.8\sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2} \quad (351\text{-}42)$$

$$\text{mdl}_{RCMC} = 0 \quad (351\text{-}43)$$

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.

5.5.8 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 and uncertainties are calculated using the following equations; negative values are substituted with zero. RCMN concentration is always positive. Uncertainty is calculated as the combination of the individual uncertainties. The MDL for RCMN is zero.

$$RCMN = NHSO + NHNO + Soil + 1.8 \times Chloride + ECTR + OMC \quad (351\text{-}44)$$

Where,

NHSO = ammonium sulfate concentration

NHNO = ammonium nitrate concentration

Soil = soil concentration

Chloride = chloride concentration as measured by IC

ECTR = elemental carbon concentration by TOR

OMC = concentration of organic mass by carbon

$$\sigma_{RCMN} = \sqrt{\sigma_{NHSO}^2 + \sigma_{NHSO}^2 + \sigma_{Soil}^2 + (1.8\sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2} \quad (351\text{-}45)$$

$$\text{mdl}_{RCMN} = 0 \quad (351\text{-}46)$$

6. DATA VALIDATION

Data validation performed at UCD involves assessing the quality, reliability, and integrity of the data. 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 databases 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 includes:

- Raw pressure transducer and temperature data from the sampler flashcards or the V4 controllers before automated validity tests.
- Filter 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:

- Verifying filter weight measurements to ensure that
    - the range is within specified limits;
    - 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.
- 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 UCD IMPROVE database. Level 2 validation involves cross-module comparisons of various species. Data submitted to CIRA and AQS databases are considered to be validated at Level 1B and Level 2. Additional Level 2 data validation is performed at 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, black carbon, and light absorption coefficients.
- Examining PM_{10} mass and $PM_{2.5}$ mass for cases where $PM_{2.5}$ is greater than PM_{10} and where $PM_{2.5}$ and/or PM_{10} are zero or negative.
- Comparing $PM_{2.5}$ gravimetric mass and reconstructed mass.
- Comparing organic carbon and elemental carbon.

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.

Data validation is not a linear process, and a significant amount of data validation (including Level 0) is performed by the analytical laboratories before the data are delivered to the quality assurance officer. The SOPs for the analytical laboratories describe their data validation procedures in detail. The following sections discuss the Level 1 and Level 2 validation processes that occur once the data are received from the field and laboratories.

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 11). Samples associated with "Terminal" flag are invalidated for a variety of reasons, and no concentration, uncertainty, or MDL values are reported, whereas those associated with "Informational" flag are still valid samples and concentrations, uncertainties, and MDLs are reported. The "Temporary" flags are assigned for a variety of reasons to aid data validation; they are replaced before final data reporting.

Table 11. 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 >= 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
PC	Possible Contamination	No longer used	4

6.2 Level 1 Validation Procedures

Level 1 validation is conducted throughout the sample handling and analysis processes. Validation for the gravimetric $PM_{2.5}$ and PM_{10} masses, $PM_{2.5}$ elements, optical absorption, ions, and carbon data is conducted by the laboratory technicians performing the analyses. The following Technical Information (TI) documents are available for mass validation and XRF data validation:

Mass validation: *Sample Handling TI 251O General Laboratory Procedures*, Section 5.9

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

HIPS validation: *HIPS SOP 276 Optical Absorption Analysis of $PM_{2.5}$ Samples*

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 detail.

6.2.1 Flow Validation

Flow data from the V4 controllers is automatically transmitted daily to the UCD IMPROVE database for near real-time review by the Sample Handling Laboratory (SHL) and Field Group. Field log sheets and flashcards (with raw pressure transducer readings) are also available as backup flow data and are shipped with the physical sampled filters from the field sites to the UCD SHL. Flow data from the V2 controllers is received by the SHL by flashcard and log sheet; only one IMPROVE site has the older V2 controller (BYIS). As part of the Level 1A validation process, flow data are reviewed 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 automatically transmitted flow data are not available, the flashcard, log sheet, or nominal value can be used instead. 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 raw flow data are performed by running the *flow.check* function (for both the V2 and V4 controller data) in the *datvalIMPROVE* R package. To perform these checks, open an R environment (such as RStudio) and run the following command:

[month_flow] <- datvalIMPROVE::flow.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], site = ['%'], list_all = ['FALSE'], server = 'production')

When *list_all* is set to FALSE, the function returns a report that lists all the samples during the date period specified with abnormal flow variability, abnormal sampling temperature, and number of records for further investigation. If the *list_all* argument is set to TRUE, only the sample events with relative standard deviation out of range will be returned. The analyst can perform the checks for all active sites in the network by setting *site = '%'* or just for a particular site by specifying the site name. Several criteria are checked:

- Abnormal flow variability: > 8 % during a 24-hour sampling period; can be caused by equipment installation problems or steady pressure drop from heavily loaded filter.
- Abnormal sampling temperature: relative standard deviation of temperature < 0.01 % or > 10 %; average temperature < 20 °C or > 40 °C.
- Abnormal number of records: number of 15-minute flow readings is < 72 rows (equivalent to 18 hours of run time) or > 104 rows (equivalent to 26 hours of run time).

Additional criteria implemented for the V4 controller include:

- The 15-minute raw pressure readings that are out of range (CYC pressure < -1.25 or > 1.25; ORI pressure < 0 or > 15) are registered as NULL and excluded from the 24-hour average flow calculation.
- The 15-minute raw cyclone pressure readings that are slightly below 0 (-1.25 ≤ CYC pressure ≤ 0) are treated as 0 in the 24-hour average flow calculation.

6.2.1.1 Flow Validation Report

The flow validation report is generated as an Excel spreadsheet and is populated using the data returned from running the *flow.check* function as described above. The spreadsheet has several tabs as described below:

- V2 Controller Flow Review: This sheet is populated with flow data from sites still using the V2 controller (e.g. BYIS1). Generate this data by running the following command in R:

  View([month_flow]$OldController)

  Copy/paste information into the spreadsheet and color code the modules (A = red, B = Yellow, C = Green, and D = Blue). Three asterisks (***) are used to indicate data issues.

- V4 Controller Flow Review: This sheet is populated using flow data from sites using the V4 controller. Generate this data by running the following command in R:

  View([month_flow]$NewController$MainCheck)

  Copy/paste information into the spreadsheet and color code the modules (A = red, B = Yellow, C = Green, and D = Blue). Three asterisks (***) are used to indicate data issues.

- V4 Controller Solenoid Check: This sheet is populated with flow source records for cases where the open solenoid position is not equal to the cartridge position. Generate this data by running the following command in R:

  View([month_flow]$NewController$SolenoidCheck)

- CG, CL, LF, PO: These sheets contain lists of samples where the flow status is flagged as CG, CL, LF, or PO and require confirmation of appropriate flagging (see Tables 12 and 13). Generate this data by running the following command in R:

  [month_flowflag] <- datvalIMPROVE::flow.status(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], flowflag = [('CG', 'CL', 'LF', 'PO')], server = 'production')

  To generate a list with only one of the flow flags, set the *flowflag* argument to equal one of the four flags. Copy/paste the results to the appropriately labelled sheet within flow validation report.

- No Flow Data: This sheet contains a list of samples that are not in alignment with average flow rates. Generate this data by running the following command in R:

  [month_missing] <- datvalIMPROVE::flow.completeness(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], server = 'production')

To further investigate the data returned from the flow checks and to validate flow data, flow plots are carefully reviewed (IMPROVE Flow Graphs; http://analysis.crocker.ucdavis.edu:3838/FlowRates/). The Flow Source Code is assigned if the primary source (MC; automatically transmitted flow data or flash card) is not reliable. Guidelines for validating flow data include:

- Review the flow charts to identify unstable flow readings. Evaluate to determine if there is an absence of pattern or if the flow is changing gradually during the sampling day. No pattern indicates a potential issue requiring further investigation. Gradual change throughout the sampling period may be caused by heavy loading.
- If automatically transmitted flow data and flashcard data are not available or reliable, use log sheet data which can be retrieved from the Filters page of IMPROVE Management Site.
- The Flow Source Code or Filter Status Code can be updated as needed from the Filters page of IMPROVE Management Site.
- Utilize the Average Flow Plot in the Flow Graphing App to further evaluate flow data.

- Utilize the Early Review page in the IMPROVE Data App to view site-by-site analysis data which can be used to help evaluate flow issues.

Next, flow data are processed in SQL to derive the daily average flow rate and elapsed time (ET). The flow processing code automatically assigns non-normal flow status flags to the samples that have flow rates that deviate from the nominal values. Table 12 and 13 list the types of flow flags and the associated criteria for applying them to $PM_{2.5}$ and PM_{10} samples, respectively.

The SQL query below is used to process the flows using stored procedure; the first six lines state the variables to process flows with. In general, the start and end dates are declared to cover the month(s) of data being processed, and the sampler name is left blank to process flow data for the entire network. The flow processing can be performed on a single site, date, or even filter ID by declaring the appropriate values.

```
DECLARE @RC int
DECLARE @iStartDate datetime = 'mm/dd/yyyy'
DECLARE @iEndDate datetime = 'mm/dd/yyyy'
DECLARE @iSamplerName NVARCHAR(50) = NULL
DECLARE @iFilterId BIGINT = NULL
DECLARE @Debug bit = 1

EXECUTE @RC = [Improve_2.1].[sampler].[spFilterAverageFlowRates]
@iStartDate
,@iEndDate
,@iSamplerName
,@iFilterId
,@Debug
GO
```

Table 12. Definitions and application criteria of automatic flow flags for $PM_{2.5}$.

Automatic Flow Flag	**Definition**	**Type**	**Criteria for Application for $PM_{2.5}$ 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)
EP	Equipment Problem	Terminal	Elapsed time > 1800 minutes (30 hours) or is missing
SA	Sampling Anomaly	Informational	Elapsed time between 1080 minutes (18 hours) - 1380 minutes (23 hours) or 1500 minutes (25 hours) – 1800 minutes (30 hours)

The 2016 IMPROVE $PM_{2.5}$ cyclone characterization test yielded results consistent 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:

$$D_{50} = 52.5 * Q^{-0.99} \quad (351\text{-}47)$$

Where,

D_{50} = 50 % cutoff diameter (in µm)

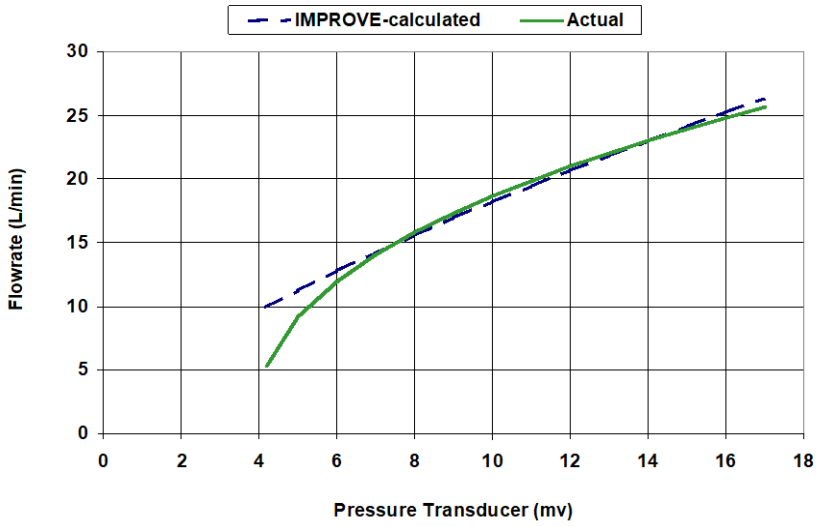
Q = flow rate (in L/min)

Note that at the nominal flow rate of 23 L/min, the 50 % cutoff diameter is 2.36 µm rather than 2.5 µm.

The criteria for the CL, CG, and LF flags are determined based on calculation limitations, performance testing, and particle size cut. Figure 4 illustrates a typical relationship between $PM_{2.5}$ flow rate and the cyclone pressure transducer measurement. The dashed line shows the calculated flow rate 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-minute (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 $PM_{2.5}$ cyclone cut point is 3.6 µm at 15 L/min.

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

Figure 4. $PM_{2.5}$ 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 PM_{10} data (Table 13), but with several differences in the criteria, due principally to the lower flow rate at which the PM_{10} sampler operates. The

relationship between the 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 PM_{10} module are not valid.

Table 13. Definitions and application criteria of automatic flow flags for PM_{10}.

Validation Flag	Definition	Type	Criteria for Application for PM_{10} 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)
EP	Equipment Problem	Terminal	Elapsed time > 1800 minutes (30 hours) or is missing
SA	Sampling Anomaly	Informational	Elapsed time between 1080 minutes (18 hours) - 1380 minutes (23 hours) or 1500 minutes (25 hours) – 1800 minutes (30 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 UCD IMPROVE database to their corresponding tables (e.g., analysis.Carbon, analysis.CarbonLaser, hips.SampleAnalysis,analysis.Ions, and analysis.Mass), as described in Section 4. 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. To perform these checks, run the following command in the R environment:

  [month_year_check] <- datvalIMPROVE::completeness.check(*startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], module_type = ["module"], data_type = ["analysis type"]*, *server = "production")*

  This command will perform the completeness check for data within the date range (*startdate* to *enddate*), for the specific module (*["module"]* can be A, B, C, or D),

and data type (*["analysis type"]* can be xrf, Mass, hips, Ions, or Carbon). The last argument in the command specifies that the calculations will use the production database (i.e. the IMPROVE operational database).

If any analyses are missing, confirm that data are missing and contact the appropriate analysis lab to confirm the status of the results.

- 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. To perform these checks, run the following command in the R environment:

  [month_year_ion_check] <- datvalIMPROVE::ions_fb.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], by = ["ions species"])

  [month_year_carbon_check] <- datvalIMPROVE::carbon_fb.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], by = ["carbon species"])

  This command will perform the checks for data within the date range (*startdate* to *enddate*), and will provide a 'Yes' or 'No' response to indicate if the field blank mass loading of the specified species (*["ions species"]*, e.g. "Sulfate" or *["carbon species"]*, e.g. "ECTR") is higher than the associated sample mass loading.

  Review the results to determine if there are sample and/or the field blank issues. The field blank may have been used as a sample and have similar mass loadings to the sample, and/or the sample may have been used as a field blank and have mass loadings lower than expected. However, the sample should also be investigated for issues independent of a swap. In some instances the sample may have actual low concentrations similar to the field blank. Field blank contamination is also possible, for example zinc contamination from XRF analysis or chloride contamination, in which case only certain field blank species would be elevated relative to the sample.

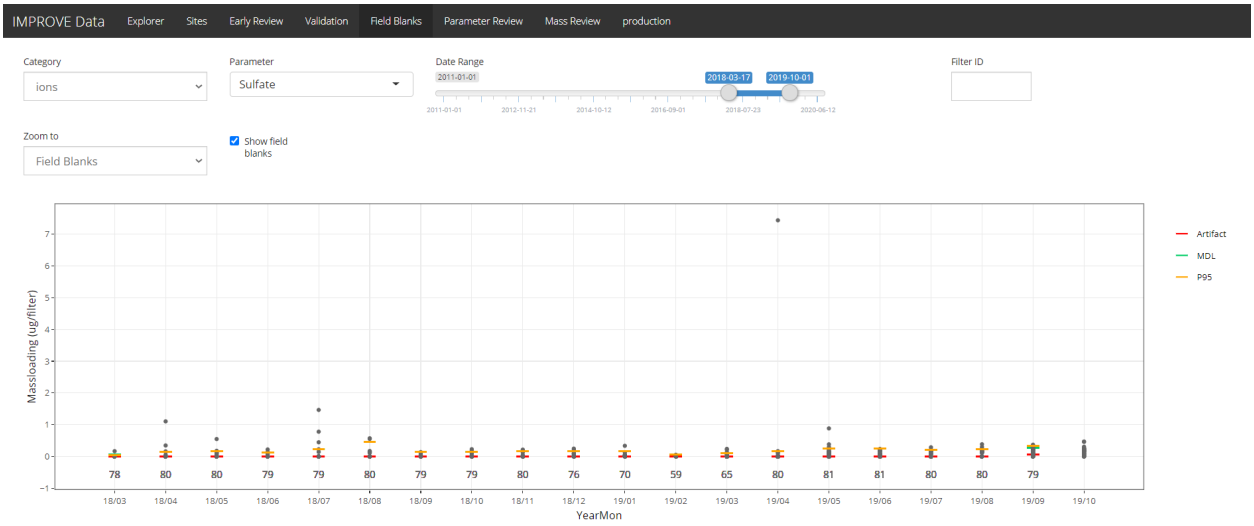
- Evaluate Field Blanks: Typically, for ions, sulfate is the primary species used for sample versus field blank comparison (followed by nitrate and then chloride). For carbon, ECTR, OCTR, OPTR, and TCTC are the primary species used for field blank comparison.

  For all analysis types (ions, carbon, elements, and mass), field blank data across the network can be compared using the Field Blanks tab in the IMPROVE Data website (Figure 5). The mass loading of a specified parameter should be compared to field blank data from the same month as well as to the network history for both high and low cases (although the latter are rare). From the Field Blanks tab, if a point is selected, the mass loadings for all species measured on the field blank and sample filters are displayed for comparison. Plots on the Validation tab should also be reviewed to determine if a sample value is unusually low.

  Artifact and MDL values are calculated using field blank results, and are expected to vary month-to-month; they are calculated for the entire network and can be impacted by shifts in field blank concentrations. As such, the artifact, MDL, and field blank 95^{th} percentile values are reviewed to identify processing issues as well as evaluate the results to determine if any field blank high mass loading cases are causing

unexpected impacts. The artifact and MDL calculation methods are meant to be robust against occasional field blank outliers.

Figure 5. Screen shot of the IMPROVE Data website Field Blanks tab.



* For elements, carbon, and ions, since there's a minimum requirement on field blank count, the statistics (artifact, mdl, P95) are not necessarily calculated from the field blank individuals collected/shown in that month only. The field blanks from adjacent months may have been used too.

Following the checks, concentrations, MDLs, and uncertainties are processed and posted in the analysis.Results table using the *improve_calculate_and_post* function in the *crocker* package. To perform the processing, run the following command in the R environment:

[month_data] <- crocker::improve_calculate_and_post([YYYY], [MM], 'production', AnalysisQcCode = 1, comment = ['Initial Posting'], replacingId = NULL, replacingQcCode = NULL)

This command calculates concentrations, uncertainties, and MDLs for all measured and derived parameters for the year *([YYYY])* and month *([MM])*, using all data from the production database, and appends the processed data to the analysis.Results or analysis.CompositeResults table in the UCD IMPROVE production database as an analysis set. It also inserts a records into the analysis.ResultsSets table that provides summary information for this set, including the *comment* and *AnalysisQcCode*. Rountine data uses *AnalysisQcCode = 1*. During Level 2 validation, the data may be modified and *improve_calculate_and_post* is run again and a new complete data set is posted to the database. When data is re-run/posted, the following actions need to be taken for version control and data integrity:

- Add comment to describe the new dataset;
- Change the analysis QC code of the previously posted dataset(s) by including the data set ID of the previous posting (*replacingId*) and the analysis QC code (*replacingQcCode*) that should be associated with that data set.

The following additional checks are performed:

- Elapsed Time and Sampling Days: Checks are performed by running the *etime.check* and *daycount* 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). To perform these checks, run the following command in the R environment:

[month_time] <- datvalIMPROVE::etime.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production")

[month_days] <- datvalIMPROVE::daycount(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production")

- Questionable Data (QD): To guide the Level 2 validation, a list of filters with the QD flag (QD – questionable data) is generated. To generate the list, run the following command in the R environment:

  [month_QD] <- datvalIMPROVE::QD.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production")

  QD status is typically assigned by the sample handling lab technicians during initial inspection of the physical samples and the raw flow rate data. These cases are investigated by reviewing the data in the Validation plots and other tools, such as comparison of results with neighboring sites. QD flags are resolved and removed by requesting further analysis and/or changing the status back to NM or assigning appropriate terminal or informational flags. There should be no records with QD in the status field in the delivery files.

- Concentration Range: The *Validsta_BadData* function in *datvalIMPROVE* uses a set of criteria listed in the R code to generate a list of results for cases where a valid sample has concentration data outside of defined normal ranges. To generate the list, run the following command in the R environment:

  [month_ValidSta] <- datvalIMPROVE::ValidSta_BadData (startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'])

  The results are reviewed using techniques described in Section 6.3 to investigate potential analysis issues, variations in uncertainty/MDL, and historical and nearby site comparisons. Reanalysis is requested when necessary/possible.

- Object Code: 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). To perform this check, run the following command in the R environment:

  [month_Obj] <- datvalIMPROVE::ObjCode.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production")

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 R scripts and data visualization tools.

6.3.1 Cross-Module Comparison

6.3.1.1 1A Module versus 2B Module

Quality assurance for the 1A and 2B Modules consists of comparing the measured concentrations of sulfur and sulfate. Sulfur concentrations are reported through elemental analysis of the PTFE filter from the 1A Module, while sulfate concentrations are determined by ion chromatography analysis of the nylon filter from the 2B Module. Discrepancies between 1A Module sulfur (times three, S3) and 2B 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". To perform this check, run the following command in the R environment:

[month_swap] <- datvalIMPROVE::swap.check(*startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production", type =["swap or outlier"])*

The *type* argument specifies the records that should be shown in the output and can be "swap", "outlier", "swap and outlier", "swap or outlier", and "all".

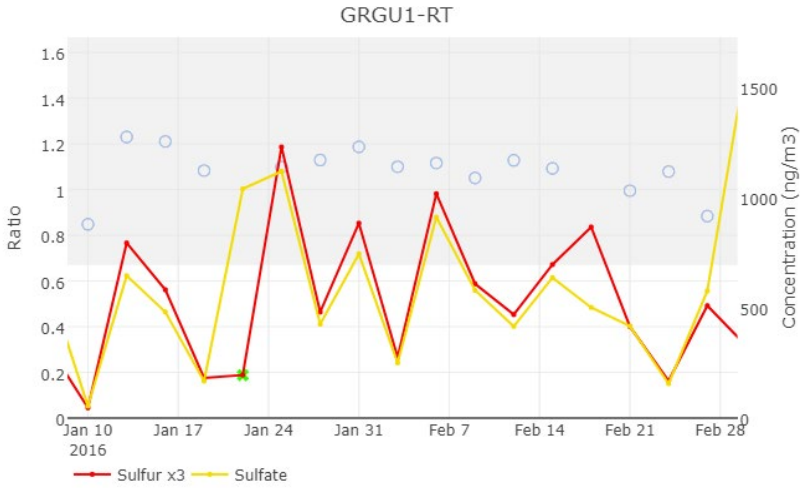
For checking possible sample swaps, successive pairs of data are examined using the algorithm outlined below. In equation (351-47), two indices for each pair of sulfur and sulfate data are calculated using data from the current and the next sampling days (referred to as subscript 1 and 2, respectively).

$$Index1 = \left(\frac{S3_1}{SO4_1} - 1\right) \times \left(\frac{S3_2}{SO4_2} - 1\right) \qquad Index2 = \left(\frac{S3_1}{SO4_2} - 1\right) \times \left(\frac{S3_2}{SO4_1} - 1\right) \qquad (351\text{-}48)$$

If $PM_{2.5}$ 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 plots in the Early Review and Validation tabs on the IMPROVE Data Site (analysis.crocker.ucdavis.edu:3838/ImproveData/) are used to further investigate samples flagged as swap and/or outlier. Figure 6 shows an example of an outlier pair at the GRGU1 site on 1/21/2016. On that day, the sulfate concentration is 1041.06 ng/m^3 while the S3 is 195.51 ng/m^3, 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 perform a reanalysis. If the reanalysis results resolve the issue, the sample mass loadings are updated in the UCD IMPROVE 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 6. S3/SO4 comparison plot for the GRGU1 site showing the 1/21/2016 sample pair as an outlier (green x).



Similar to the sulfur and sulfate comparison, chlorine (from XRF analysis of the Module 1A filter) and chloride (from IC analysis of the Module 2B filter) concentrations can also be compared, and can be used as supporting evidence for issues identified during the sulfur and sulfate comparison. It may also be possible to identify chloride contamination by comparing chlorine to chloride.

When reanalysis yields changes to results, further action is required:

- For elements from the 1A filter, the analysis laboratory will assign the appropriate analysis QC code to each of the result sets so that only one set is marked as valid. The updated results can be viewed in the Early Review S/SO4 plot to confirm that the issue(s) have been resolved. Appropriate comments should be added to the affected filter(s) to indicate that reanalysis was performed, briefly explaining the reasoning, and state which set of results (original or reanalysis) are reported.
- For ions from the 2B filter, the analysis lab sends updated data files, which must be ingested following the steps outlined in Section 4.2. A list should be generated of filter IDs for which additional results have been ingested into the database. The comments from the analysis lab are reviewed to determine which set of analysis results to report, and the analysis QC code in the [IMPROVE_2.1].[analysis].[Ions] production database table must be changed accordingly. For example, if the analysis lab indicates that the reanalysis results should be reported, the invalid analysis QC Code (= 0) should be assigned to the original results and the valid analysis QC Code (= 1) should be assigned the newly ingested reanalysis results. The updated results can be viewed in the Validation plots to confirm that the issue(s) have been resolved. For some cases, the updated results can also be viewed in the Early Review plots to confirm that the issue(s) have been resolved. Appropriate comments should be added to the affected filter(s) to indicate that reanalysis was performed, briefly explaining the reasoning, and state which set of results (original or reanalysis) are reported.

6.3.1.2 1A Module versus 3C Module

The light absorption coefficient (f_{abs}) at 635 nm is measured by HIPS from the 1A Module PTFE filter and is compared qualitatively with the elemental carbon (EC) concentration measured by TOR from the 3C Module quartz filter as well as with the black carbon (BC) concentrations estimated from the initial and final laser readings from the 3C Module quartz filter analysis. Visual inspection of the data is performed to identify outliers using the fAbs, BC, and EC time series plot on the Validation page of the IMPROVE data website. Black carbon and f_{abs} are both optical measurements and are expected to compare well, whereas f_{abs} and EC are determined by different methods and may not be consistently comparable (Figure 7). If analytical error in either measurement is suspected, other measurement data from the same module is examined to determine validity of the sample.

The relationship between EC, BC, and f_{abs} is used to evaluate the carbon and HIPS results and select samples for carbon reanalysis. However, the relationships between these parameters vary across sites and seasons, making quantitative criteria ineffective for identification of outliers. As such, site-specific historical results and results from nearby sites are used to provide insight into anomalous samples. Issues identified during comparison of EC, BC, and f_{abs} results can be further investigated using qualitative checks and criteria to evaluate 3C Module carbon results (OC, EC, and TC) independently of f_{abs} (Table 14).

Table 14. Qualitative checks and criteria for carbon (OC, EC, and TC) validation.

Analytical Issue	Considerations
OC/EC split point	Evaluate and compare OC, EC, and TC values.
Laser response	Evaluate EC 808 nm versus EC 635 nm (ECTR); dissimilar results indicate a laser issue.
Laser issue	Consider EC 635 nm (ECTR) versus all other EC wavelengths; if only EC 635 nm is zero, the issue is likely specific to the 635 nm laser.

In addition, the following points should also be considered:

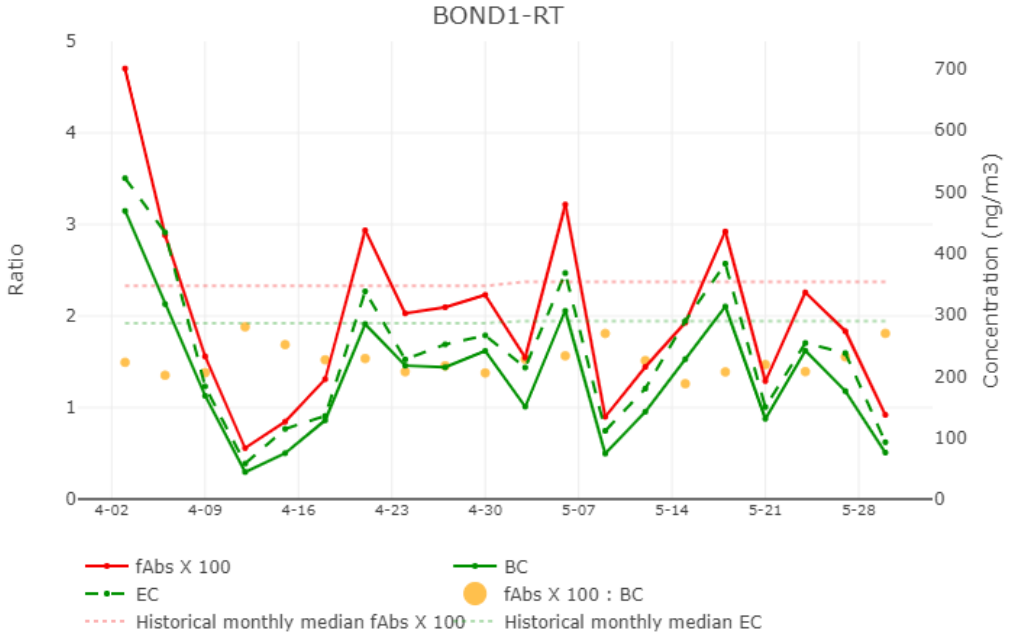
- Consider the trend of ECTR relative to f_{abs} and BC. If ECTR is low, investigate to determine if it is anomalous or if there have been other occurrences in recent months/years.
- Evaluate $PM_{2.5}$ relative to RCMN. If ECTR is unexpectedly high/low, then re-evaluate OCTR and ECTR. If OMC is unexpectedly high/low, then re-evaluate OCTR and ECTR.
- Compare ECTR and OCTR to nearby sites.
- Evaluate the OCTR/ECTR ratio at the site relative to recent days/months/years.
- Investigate ECTR values that are negative or zero. If values are negative, evaluate the original mass loading relative to the artifact correction. If the value is 0.00 but ECTT has a value, there may be a split point issue.
- Compare ECTR results at different wavelengths. For some sources, ECTR 635 nm should be close to ECTR 808 nm. For sources that emit brown carbon (e.g. fire), ECTR 405 nm is larger than ECTR 635 nm. If ECTR = 0 at 635 nm but ECTR at all other wavelengths are non-zero, there is likely an issue with the 635 nm laser.

- Inspect TC replicate and/or reanalysis results. If different is > 10 %, request a third analysis. The maximum number of punches available for a quartz filter is three; there will be cases where reanalysis is not possible. In such cases, proper documentation regarding filter/ sampling events leading to the use of extra punch should be documented.

When reanalysis yields changes to results, further action is required:

- For f_{abs} from the 1A filter, the analysis laboratory will assign the appropriate analysis QC code to each of the result sets so that only one set is marked as valid. The updated results can be viewed in the Validation plots to confirm that the issue(s) have been resolved. Appropriate comments should be added to the affected filter(s) to indicate that reanalysis was performed, briefly explaining the reasoning, and state which set of results (original or reanalysis) are reported.
- For carbon from the 3C filter, reanalysis results received from the analysis laboratory must be ingested following the steps outlined in Section 4.2. A list should be generated of filter IDs for which additional results have been ingested into the database. The comments from the analysis lab are reviewed to determine which set of analysis results to report, and the analysis QC code in the [IMPROVE_2.1].[analysis].[Carbon] production database table must be changed accordingly. For example, if the analysis lab indicates that the reanalysis results should be reported, the invalid analysis QC Code (= 0) should be assigned to the original results and the valid analysis QC Code (= 1) should be assigned the newly ingested reanalysis results. If the analysis laboratory indicates that the reanalysis results are within replicate criteria or if only one species was affected, the replicate or reanalysis analysis QC code (= 2) should be assigned to the relevant set of results and parameters that were unaffected by the issue. The updated results can be viewed in the Validation plots to confirm that the issue(s) have been resolved. Further, the analyst should review the mass loadings for all sets of analysis results for a given filter. Appropriate comments should be added to the affected filter(s) to indicate that reanalysis was performed, briefly explaining the reasoning, and state which set of results (original or reanalysis) are reported

Figure 7. Comparison plot of light absorption coefficient measurements (f_{abs}, times 100) from 1A Module and elemental carbon (EC) measurements and black carbon measurements from 3C Module at BOND1 site.



6.3.1.3 1A Module versus 4D Module

1A module $PM_{2.5}$ mass and 4D module PM_{10} mass are reviewed and compared (Figure 8). The *mf_mt.check* function in the *datvalIMPROVE* package is run using the following command in the R environment:

[month_PM] <- datvalIMPROVE::mf_mt.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production", problemonly = ["TRUE"])

The check returns a list of samples flagged as mass outliers if the *problemonly* argument is set to 'TRUE' and any of the following criteria are met:

- $PM_{2.5}$ or PM_{10} mass concentration is negative (negative value does not necessarily mean invalid).
- $PM_{2.5}$ 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).

Where the Z-score is calculated using equation (351-48),

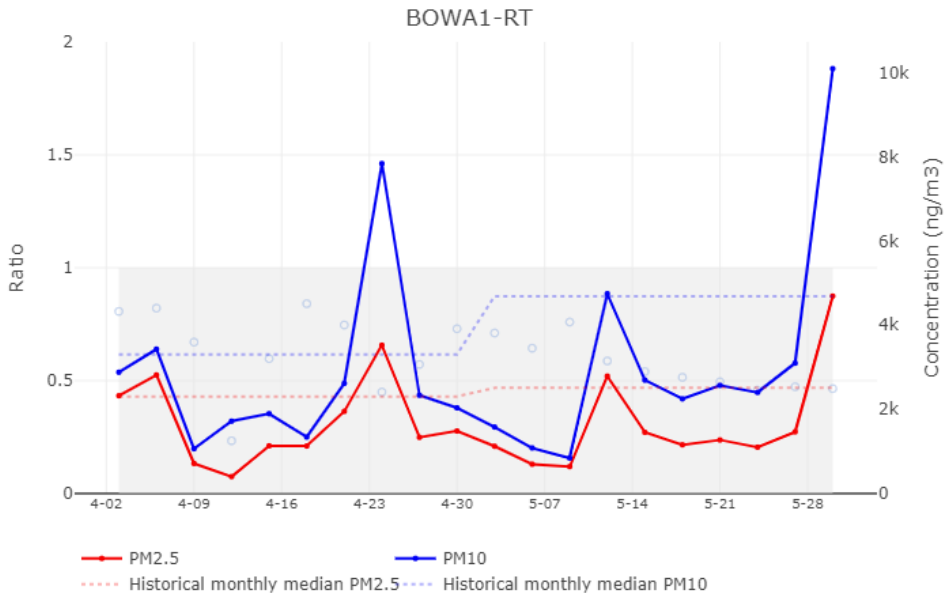
$$Z\_score = 1.41 \times \frac{PM_{2.5} - PM_{10}}{\sqrt{(unc_{PM2.5})^2 + (unc_{PM10})^2}} \quad (351\text{-}49)$$

For samples that are flagged for one of the above cases, further investigation is required to identify the cause:

- Use the mass time-series plot on the Validation page;

- Investigate occurrence of a possible swap ($PM_{2.5}$ to PM_{10} swap, adjacent day swap, etc). If a swap may have occurred request further investigation from the Sample Handling Laboratory, and correct swapped data as needed.
- If the data appear abnormal, request confirmation of the post-weight for the Sample Handling Laboraotry; the pre-weight cannot be re-determined after sampling;
- Samples with invalid mass concentrations are flagged as "UN" (Undetermined Weight).

Figure 8. Time series plot of PM_{10} and $PM_{2.5}$ masses and their ratio at BOWA1 site.



6.3.1.4 $PM_{2.5}$ 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 measured components from the 1A, 2B, and 3C Modules (Figure 9). The *mf_rcm.check* function in the *datvalIMPROVE* package is run using the following command in the R environment:

[month_recon] <- datvalIMPROVE::mf_rcm.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], "production", problemonly = ["TRUE"])

The *mf_rcm.check* returns a list of samples flagged as outliers if the *problemonly* argument is set to 'TRUE' and any of the following criteria are met:

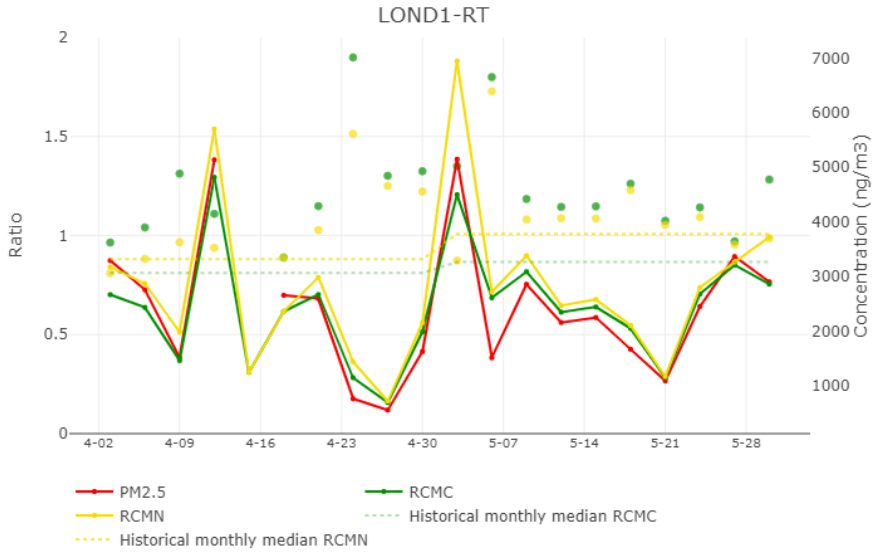
- RCMC is higher than two times MF, and the RCMC Z-score > 3; the number three is set empirically. These samples are accompanied with a comment "MF << RCMC".
- The RCMN Z-score < -22; the number 22 is set empirically. These samples are accompanied with a comment "MF >> RCMN".

Z scores are calculated as follows:

$$RCMC\_Z\_score = 1.41 \times \frac{RCMC - PM_{2.5}}{\sqrt{(unc_{PM2.5})^2 + (unc_{RCMC})^2}} \quad (351\text{-}50)$$

$$RCMN\_Z\_score = 1.41 \times \frac{\mathrm{RCMN} - \mathrm{PM_{2.5}}}{\sqrt{(unc_{PM2.5})^2 + (unc_{RCMN})^2}} \quad (351\text{-}51)$$

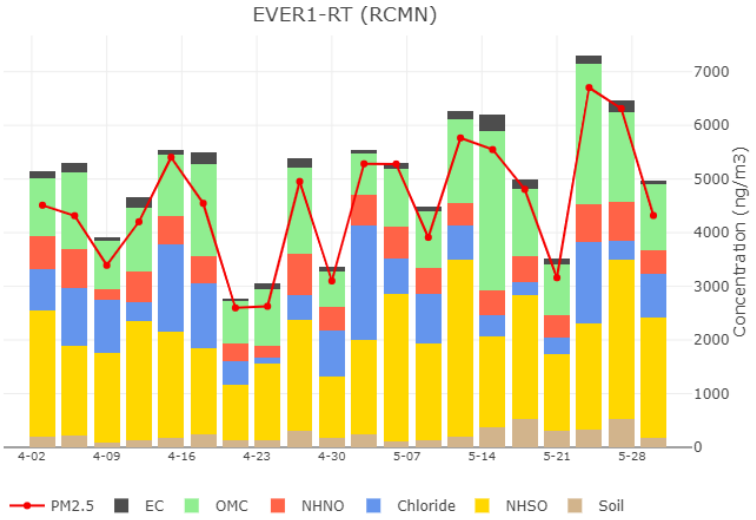
Figure 9. Time series plot of $PM_{2.5}$ gravimetric mass, reconstructed mass without nitrate (RCMC) and reconstructed mass with nitrate (RCMN) and their ratios at LOND1 site.



RCMN is also plotted as a bar plot (Figure 10), along with the $PM_{2.5}$ time series, for comparison of RCMN and $PM_{2.5}$ concentrations and to enable the contributions from the various species to be viewed and evaluated.

6.3.2 Long-Term Network-Wide Checks

Figure 10. Time series for RCMN versus Fine mass at EVER1 site.



Several data visualization tools and control plots are used 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 that are routinely used and reviewed:

- Scatter plot of S3 versus SO4 mass loadings for the whole network (Figure 11). 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 12). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of f_{abs} versus BC (converted from TOR absorption measurements) for the whole network (Figure 13). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Time series plot of the 1A to 4D mass loading ratio showing the long-term trend and historical data at a given site (Figure 14). 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 15 shows an example time-series plot for OC concentrations between 2011 and 2016. These plots are generated in R, and are typically included as part of the IMPROVE Quality Assurance Report.

Figure 11. Scatter plot of sulfur (×3) versus sulfate for the entire IMPROVE network.

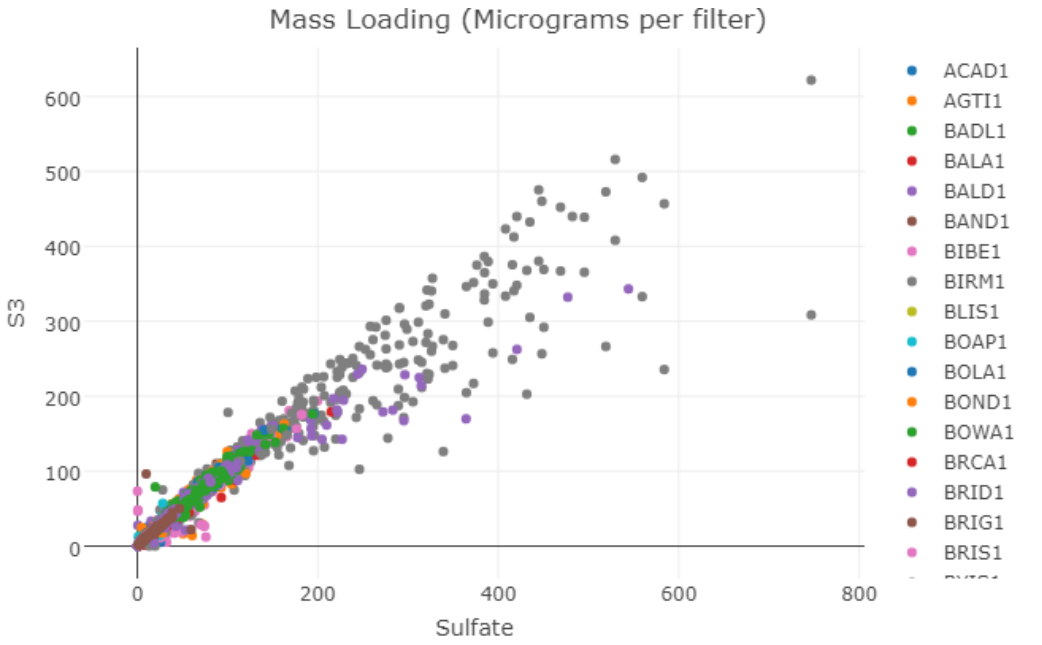


Figure 12. Scatter plot of chlorine versus chloride for the entire IMPROVE network.

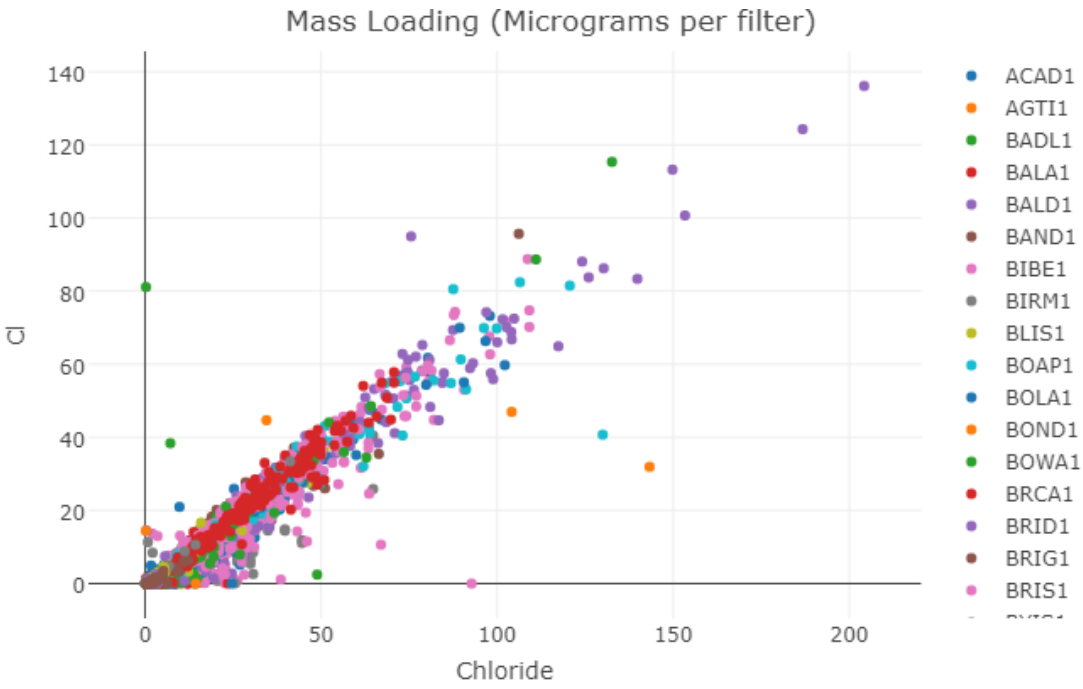


Figure 13. Scatter plot of f_{abs} versus BC (converted from TOR absorption measurement) for the whole network.

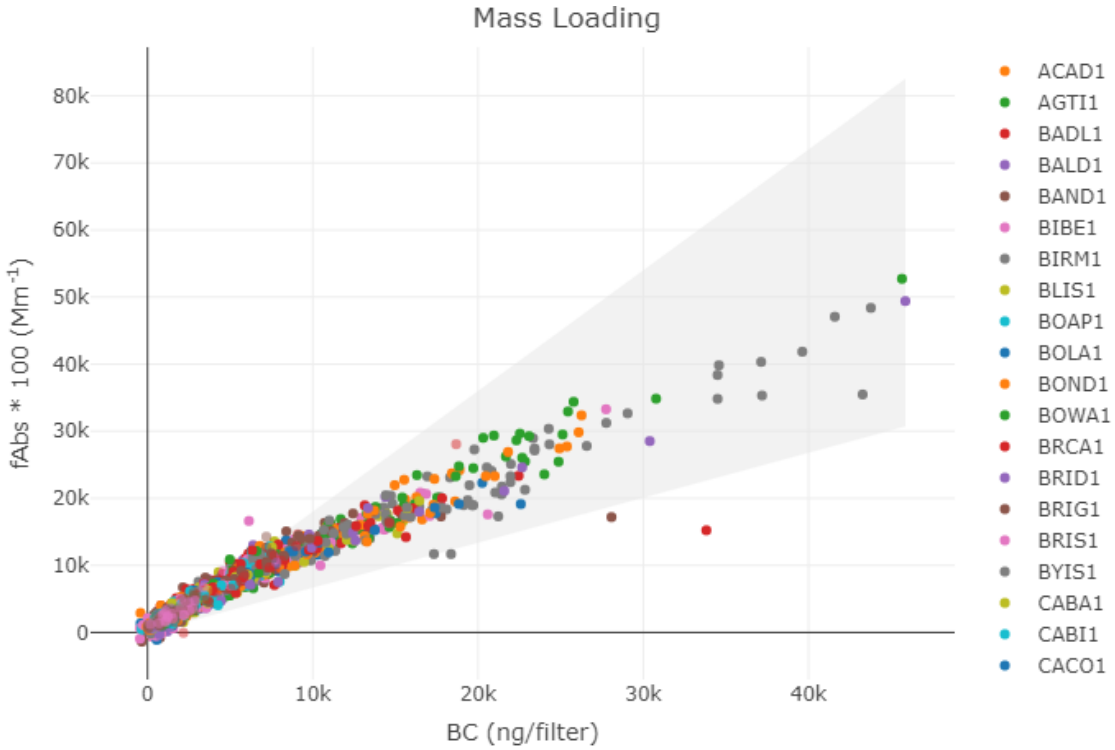


Figure 14. Ratio of $PM_{2.5}$ mass (1A) over PM_{10} mass (4D) at ACAD1 site, represented as raw measurements not adjusted for flow rates. Points are individual sample days (pink = Q1, green = Q2, blue = Q3, purple = Q4). Black line is the multi-year monthly mean. Blue line is the locally weighted average (LOESS).

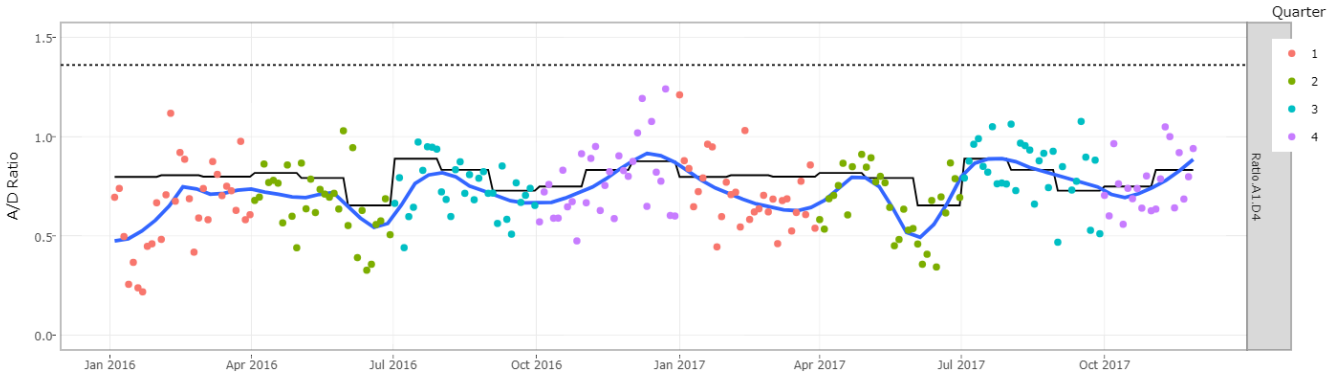
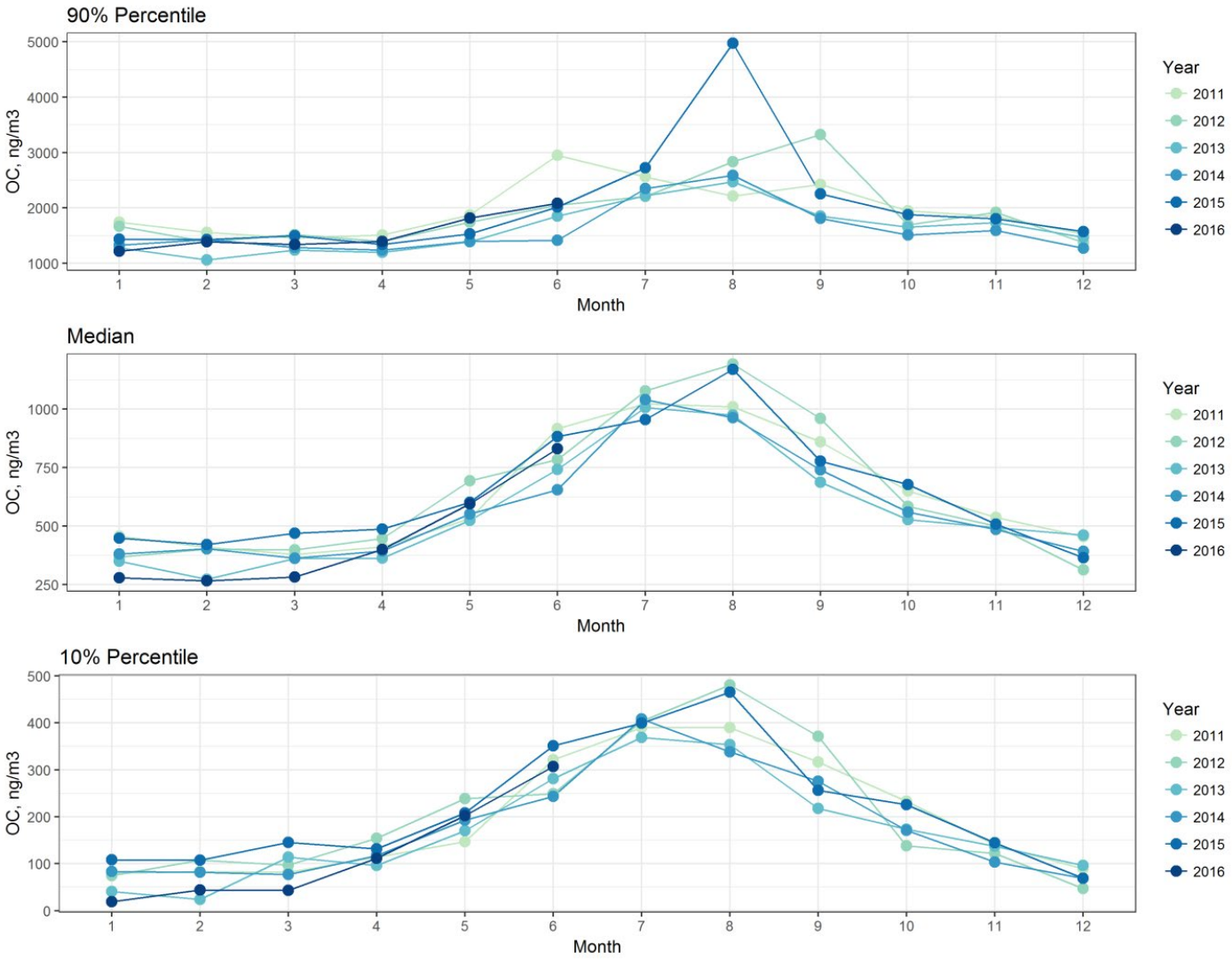


Figure 15. Multi-year monthly 10 % percentile (top), median (middle) and 90 % percentile (bottom) of organic carbon (OC) concentrations (in ng/m^3) for the whole IMPROVE network from 2011 to 2016.



6.3.3 Common Validation Findings

Some validation findings have a tendency to recur periodically, and effort is made to handle and resolve these issues consistently. Some examples of common findings are covered in this section, though those mentioned here are not inclusive of all scenarios or variations.

Swaps

There are several types of swaps in terms of the filter purposes involved and at what point in the process the swap occurred.

- **Field Blank - Sample Swap, Type 1:** Swapped before sampling (all downstream data are swapped, including flow data and all analysis associated with the filters); also referred to as cartridge position swap**:**

Step 1: Use the web app (https://improve.aqrc.ucdavis.edu/Filters/Details) to change the filter purpose. This step applies to both Field Blank and Sample.

Step 2: Also swap all data fields relating to the cartridge position between the relevant filters, including filter position properties (Cartridge Position) and log sheet records.

Step 3: Field data also needs to be swapped, specifically flow data. To do this, confirm whether the controller used at the affected site is new (V4) or old (V2. Update the flow information in [Improve_2.1].[sampler].[FlowSourceDataV2] for the new controller or [Improve_2.1].[sampler].[FlowSourceData] for the old controller. In this case, there is only one filter with flow information; the flow information is assigned to the wrong Filter ID. Update to the correct sample filter ID using the example SQL query below.

UPDATE [Improve_2.1].[sampler].[FlowSourceDataV2]

SET FilterId = Old field blank filter Id

WHERE FilterId = Old Sample filter Id

Step 4: Using the web app given in Step 1, change the filter status to Sample - Field blank swap ('SP').

Step 5: Check the updated information in the Field blank review plot.

Step 6: Add necessary comments to the filters.

Step 7: Reprocess flow following the steps outlined in Section 6.2.

Step 8: Reprocess concentrations following the steps outlined in Section 5.

- **Field Blank – Sample Swap, Type 2:** Swapped after sampling, before analyzing (flow data are fine, analysis data are swapped):

  Step 1: Identify which analysis data are swapped if multiple analysis are performed on that filter. Usually if the filters are swapped at a lab station, all downstream analyses will be swapped too.

  Step 2: Swap the FilterId in the analysis table(s).

  Step 3: Using the web app (https://improve.aqrc.ucdavis.edu/Filters/Details), change the filter status to Sample - Field blank swap ('SP').

  Step 4: Check to ensure that the swap was performed by viewing the data in the Field Blanks and Early Review (if available) tabs. The Early Review tab shows data in mass loading from the analysis table.

  Step 5: Reprocess concentrations following the steps outlined in Section 5.

  Step 6: Review the data in the Field Blanks tab and Validation plots.

  Step 7: Add comments.

- **Sample Date Swap, Type 1:** Swapped before sampling (all downstream data are swapped, including flow data and all analysis associated with the filters); also referred to as cartridge position swap:

  Step 1: Use the web app (https://improve.aqrc.ucdavis.edu/Filters/Details) to change the sample dates. Repeat this step for all affected dates.

  Step 2: Swap all data fields relating to the cartridge position between the relevant filters, including filter position properties (Cartridge Position) and log sheet records.

  Step 3: Check to ensure that the swap was performed by viewing the data in the Early Review tab (if available). The Early Review tab shows data in mass loading from the analysis table.

  Step 4: Field data also needs to be swapped, specifically flow data. To do this, confirm whether the controller used at the affected site is new (V4) or old (V2. Update the flow information in [Improve_2.1].[sampler].[FlowSourceDataV2] for the new controller or [Improve_2.1].[sampler].[FlowSourceData] for the old controller. In this case, the filter IDs need to be assigned to the correct flow data using the example SQL query below.

  *SELECT * FROM [Improve_2.1].[sampler].[FlowSourceDataV2]*

  WHERE FilterId = XXXXXXX and SampleDate ='2019-08-22'

  UPDATE [Improve_2.1].[sampler].[FlowSourceDataV2]

  SET FilterId = YYYYYYY

  WHERE FilterId = XXXXXXX and SampleDate ='2019-08-22'

  *SELECT * FROM [Improve_2.1].[sampler].[FlowSourceDataV2]*

  WHERE FilterId = YYYYYYY and SampleDate = '2019-08-25'

  UPDATE [Improve_2.1].[sampler].[FlowSourceDataV2]

  SET FilterId = XXXXXXX

  WHERE FilterId = YYYYYYY and SampleDate = '2019-08-25'

  Step 5: Using the web app given in step 1, change the filter status to Swapped sample dates ('SW').

  Step 6: Reprocess flow following the steps outlined in Section 6.2.

  Step 7: Reprocess concentrations following the steps outlined in Section 5.

  Step 8: Review the data in the Validation plots.

  Step 9: Add comments.

- **Sample Date Swap, Type 2:** Swapped after sampling, before analyzing (Flow data are fine, analysis data are swapped.)

Step 1: Identify which analysis data are swapped if multiple analyses are performed on that filter. Usually if the filters are swapped at a lab station, all downstream analysis will be swapped too.

Step 2: Swap the FilterId in the analysis table(s).

Step 3: Check to ensure that the swap was performed by viewing the data in the Early Review tab (if available). The Early Review tab shows data in mass loading from the analysis table.

Step 4: Using the web app (https://improve.aqrc.ucdavis.edu/Filters/Details), change the filter status to Swapped sample dates ('SW').

Step 5: Reprocess concentrations following the steps outlined in Section 5.

Step 6: Review the data in the Validation plots.

Step 7: Add comments.

Sampling Anomalies and Questionable Data

Thre are several types of sampling anomalies and questionable data commonly observed during validation. Included here are guidelines for addressing and resolving these issues.

- **Module stack not fully inserted**
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. Has previously occurred for the D-Module stack.
  - Review the data and JIRA notes to determine if this has previously been an issue or if it is a longer-term issue.
  - To determine if the SA flag (SA – samping anomaly) is suitable, review the relevant concentration data and compare with results from other modules. If the cross-module results agreee, apply the SA flag to indicate an operational deviation. If the cross-module results do not agree, consider other actions such as reanalysis or invalidation.
- **Module flow obstruction**
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. Has previously occurred for the B- and D-Modules.
  - Review the data and JIRA notes to determine if this has previously been an issue or if it is a longer-term issue.
  - Notes from previously resolved issues are included here to provide context and framework for handling of future similar cases:
    - D-module flow obstruction example: The SA flag was applied because the impact to the data was not quantifiable and the PM_{10} and $PM_{2.5}$ masses compared relatively well. Some nearby sampling dates had flow rate flagged as low or clogging, but not on all days and a null code was not applied.

        - B-module flow obstruction example: The cross-module comparison ratios were evaluated, and since sulfur and sulfate trended reasonably well together, and there were no outliers, the SA flag was applied rather than invalidating.

- **Possible manifold open / cartridges not seated correctly**
    - Typically flagged QD by the Sample Handling Laboratory with comment applied. For example, a typical comment is: *Module/filter CARTs, possible MANIFOLD open / CART not seated correctly, low FLOW*.
    - Assess the concentration data and compare with other modules. Evaluate the flow and filter statuses.
    - Review JIRA notes to determine if this has previously been an issue or if it is a longer-term issue.
    - Notes from previously resolved issues are included here to provide context and framework for handling of future similar cases:
        - Scenario #1: Comment from Sample Handling Lab indicated, *3C CARTs, possible MANIFOLD open / CART not seated correctly, low FLOW*. The EC and BC data agreed with the fAbs, suggesting that the leak was not severe. The flow rate through the filter was lower than expected and the LF flow status flag was applied. The filter status was kept as NM rather than applying the SA flag. Since LF is a more severe status than SA, the SA flag would not have been reported to end users. If the flow status had been NM, the SA flag would have been applied and reported to the end user.
        - Scenario #2: In some cases, the Sampling Handling Laboratory invalidates filters with the BI terminal flag (BI – bad install) prior to data validation. The Sample Handling Laboratory will invalidate the filter if there was no sample collected, which can be confirmed for 1A and 4D filters when the pre- and post-weight difference is zero. Filters may also be invalidated if the filter deposit is much lighter in appearance relative to the other three filters collected on the same day. If there is uncertainty, the Sample Handling Laboratory applies the QD flag (typical for 2B and 3C filters).

- **Double filter**
    - Typically flagged QD by the Sample Handling Laboratory with comment applied. Most commonly found for 3C filters.
    - To determine if the SA flag (SA – samping anomaly) is suitable, review the relevant concentration data and compare with results from other modules. If the cross-module results agreee, apply the SA flag to indicate an operational deviation. If the cross-module results do not agree, consider other actions such as reanalysis or invalidation

- **Pre-weight unknown**
  - Only applies to 1A and 4D filters, samples and field blanks.
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. For example, a typical comment is: *Module/filter FIL mass difference negative/high, POST weight confirmed, PRE weight unknown*. This can appear as pre- to post-weight difference of zero or negative, high PM_{10}, or $PM_{2.5}$>PM_{10}.
  - Assess the severity of the situation by evaluating the $PM_{2.5}/PM_{10}$ ratio, $PM_{2.5}$ relative to RCMN, and regional mode comparisons.
  - If the pre-weight is unknown, the filter status should have the UN terminal flag (UN – undetermined mass), which invalidates only the mass parameter from the affected filter. If the comment does not mention pre-weight, review the mass data, request re-weigh, and investigate other issues (such barcode assignments in the database).
- **Quartz contamination**
  - This typically applies to 1A and 4D filters only.
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. Quartz contmation occurs on PTFE filters if a screen with quartz deposit is installed. The PTFE and quartz screens are kept apart in the Sample Handling Laboratory, but there is potential for contamination due to human error. White deposit or white specs on the PTFE filter are indications of quartz contamination.
  - Assess the severity of the situation by evaluating the concentration data and compare with results from other modules.
  - If the quartz contamination is deemed to not be significant enough to impact analysis, the filter status should be changed to NM.
- **Insects / large particles**
  - This typically applies to 4D filters.
  - Because of the D-Module sampling design, it is not uncommon to see insects or other large particles such as seeds on the filters. In some cases the Sample Handling Laboratory is able to remove the debris and reweigh the filter. The QD flag and an apporpriate comment are applied to the filter to indicate possible impact to the analysis results.
  - Review the data to determine if the results appear reasonable; if so, change the filter status to NM. Another visual check and/or reanlaysis could be reqeusted if the data appear questionable.
- **Dropped filters**
  - Filters can be dropped at any point during the sampling or analysis process. A comment is typically applied by the laboratory to indicate such. If the filter was dropped in the Sample Hanlding Laboratory, the QD flag is also applied.

- The Sample Handling Laboratory distinguishes between dropping filters on the floor and on the counter, where heavy contamination is assumed for the former.
- Assess the concentration data and compare with other modules. Evaluate ralative to historical data form the site and same day neighboring sites.
- Review the data to determine if the results appear reasonable; if so, change the filter status to NM. Another visual check and/or reanlaysis could be reqeusted if the data appear questionable. The nylon filter from the 2B module will not be available because it was extracted for analysis. Invalidate the filter if the contamination appears to be severe.

- **Wrinkled filter**
  - This is a common occurrence for 3C filters and is observed either at the Sample Handling Laboratory and/or the analysis lab.
  - A wrinkled filter can occur when loading the filter at the lab or in the field. The cartridge may have come lose causing the filter to shift and wrinkle. A wrinkled filter will likely have an uneven/low deposit.
- **Filter blown out / bulging filter**
  - The quartz filters from the 3C module are commonly suspected of being blown out when filter bulging is observed at the Sample Handling Laboratory and/or the analysis lab; 37 mm nylon filters from the 2B module are also sometimes observed to have crinkled edges.
  - For 25 mm quartz filters from the 3C module, it is possible to “suck out” part of the filter when (aggressively) taking off the red caps. While installed in the modules, the edges of the quartz filters are compressed between the screen and a flat lip on the cassette bottom, which weakens the outer edges; the edges will be relatively rough. Bulging filters can also suggest airflow in the wrong direction and can occur if quartz filters are loaded without screens or loaded upside down; for these cases there will be little or no sample deposit.
  - For 37 mm nylon filters from the 2B module, it is possible to crinkle the edges of the filter while loading. For these cases, the filter looks similar to a bulged filter but usually folds flat during sampling. Filter cassettes must be assembled with a press to ensure even pressure.
  - Reivew all data – including the flow data – to determine if and when the filter was disfigured. Flow issues may result in application of flow-related informational or terminal flags (see criteria in Table 12 and Table 13), and may explain concentration discrepancies such as poor sulfur to sulfate agreement. If the flow status is normal and the data appear reasonable, the filter status should be changed to NM.
- **Holes**
  - Holes can be observed for any filter type and range from pin holes to larger holes that destroy the filter. Holes can be introduced at various points during the

sampling and analysis process; filters are flagged QD, invalidated, and/or have comments applied.

- Analysis can be impacted by a hole of any size, and the extent of impact varies by analysis type. As such, all analysis results should be reviewed independently (for example, HIPS analysis may be impacted even though mass analysis is not). If concentration results are suspect, a visual check and reanalysis should be requested, if available. The nylon filter from the 2B module will not be available because it was extracted for analysis. Review the flow data to evaluate potential sampling issues. If the results are determined to have been impacted by the hole, invalidate the filter; if the results are reasonable, change the status to NM.

6.3.4 Final Review

Several final checks are performed before submission of data delivery files to the CIRA, AQS, and UCD CIA databases:

- The *QD.check* function in *datvalIMPROVE* is run again after validation is complete to confirm that there are no remaining records with QD status. No records with QD in the status field should exist in the delivery files.
- The *ObjCode.check* function in *datvalIMPROVE* is run again after validation is complete to confirm that only RT (routine) or CL (collocated) objective codes exist in the data file.
- The *ValidSta_NullData* function in *datvalIMPROVE* checks to determine if there are cases where no value (-999) is reported but the filter is marked as valid. Perform this check using the following command in the R environment:

  [month_ValidNull] <- datvalIMPROVE::ValidSta_NullData(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’], “production”)

  Confirm application of a terminal flag or locate the missing analysis results and follow the steps to reprocess the data for delivery.
- The *MDL_UNC* function in *datvalIMPROVE* checks to determine if calculated MDLs or uncertainties have negative values. To obtain a list of records that meet this criteria, run the following command in the R environment:

  [month_mdl_uncl] <- datvalIMPROVE::MDL_UNC(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’])

  Review records to determine why the uncertainty or MDL is negative and resolve as needed.
- The *sitecount* function in *datvalIMPROVE* is used to determine the site count for a specific FED delivery file. Perform this check using the following command in the R environment:

  [month_site] <- datvalIMPROVE::sitecount(filepath = [‘filepath.csv’])

The *filepath* argument is a character string containing the file path and file name of the wide-format file for delivery to FED, where the file itself is a .csv file format.

Many of the functions described in Section 6.2 and 6.3 can be performed simultaneously using the *datvalIMPROVE::improve_validate* function. This function should be run at the beginning of initial validation as well as prior to delivery. Perform this check using the following command in the R environment, and evaluate the output from the checks described below:

[month_output] <- datvalIMPROVE::improve_validate(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'])

- **output$flow_completeness –** flow.completeness
- **output$flow_status -** flow.status
- **output$etime -** etime.check
- **output$day_count –** daycount
- **output$objective_code –** ObjCode.check
- **output$mass** - mf_mt.check
- **output$rcm -** mf_rcm.check
- **output$swap -** swap.check
- **output$QD -** QD.check

Below are the final checks performed within the *improve_validate* function:

- **output$validsta_null -** ValidSta_NullData
- **output$validsta_bad -** ValidSta_BadData
- **output$mdl_unc -** MDL_UNC

7. DATA DELIVERY

After Level 2 data validation is complete, the data files are submitted to CIRA, AQS, and UCD CIA databases.

7.1 Submission to CIRA

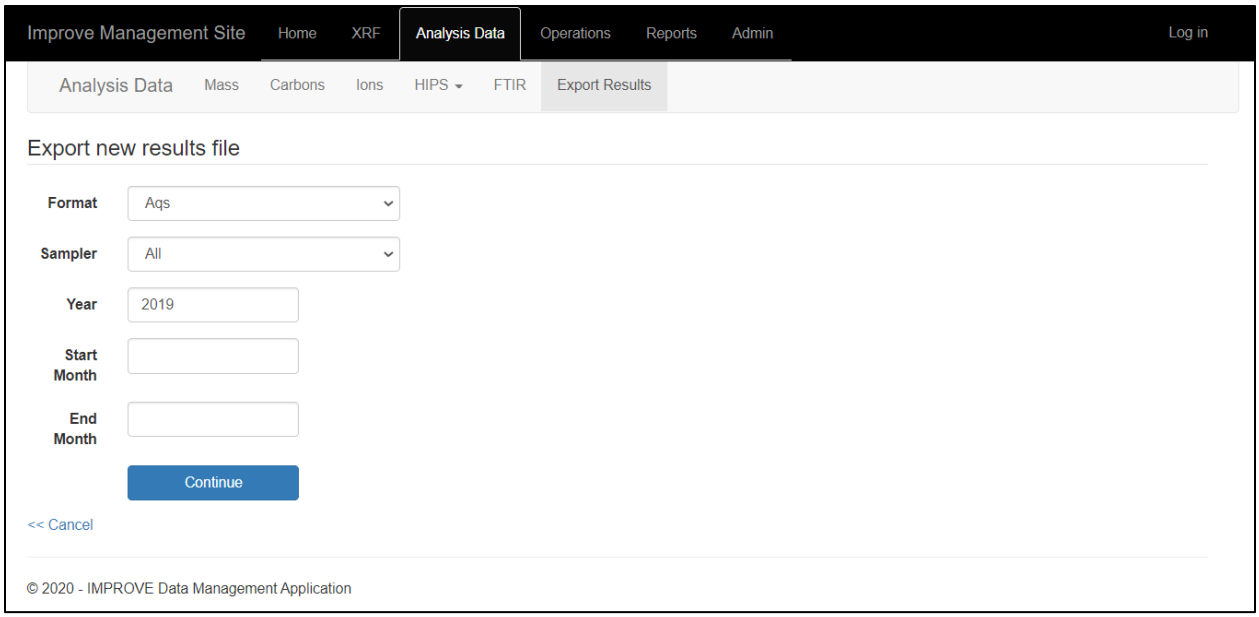
FED export files are created using the *improve_export_fed* and *improve_export_wide* functions in the *crocker* package, in which the year, month, and server for both functions are entered. The functions create "skinny" and "wide" versions of the dataset, and both are submitted. The files are saved 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:

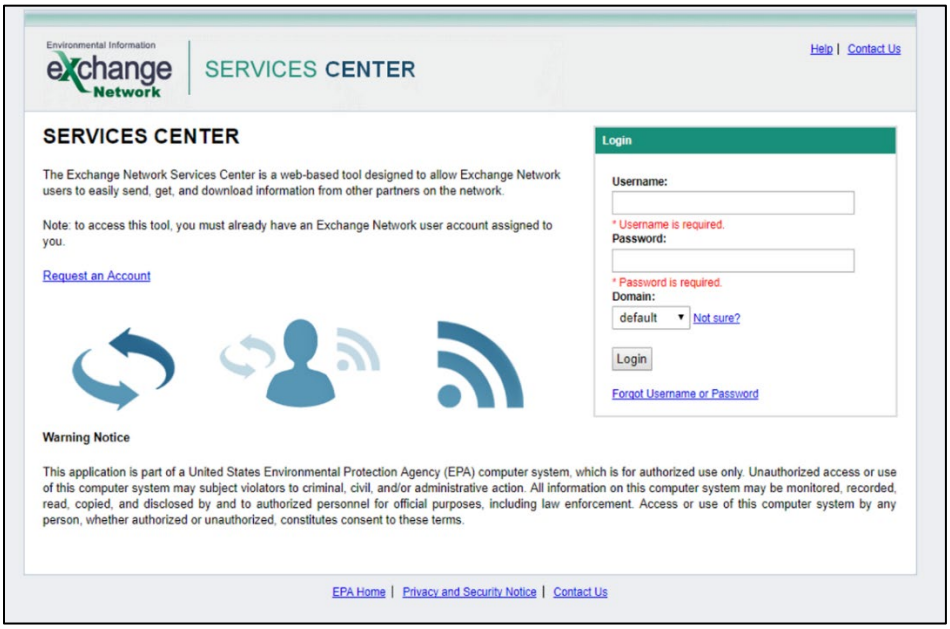
1. Create the AQS delivery files using the **Analysis Data** tab in the IMPROVE Management Site; select the **Export Results** option (Figure 16). Choose **AQS** for the **Format** and fill in the Year, Start Month, and End Month. The AQS delivery file can also be generated for a specific site by selecting the relevant site name from the **Sampler** list.

Figure 16. Screen where the AQS file for delivery can be generated from the Improve Management Site.



2. Click continue to automatically generate the file. Save the file to the UCD U Drive at *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 (Figure 17). Use credentials to login.

Figure 17. Login screen for the EPA's Exchange Network Services Center website.

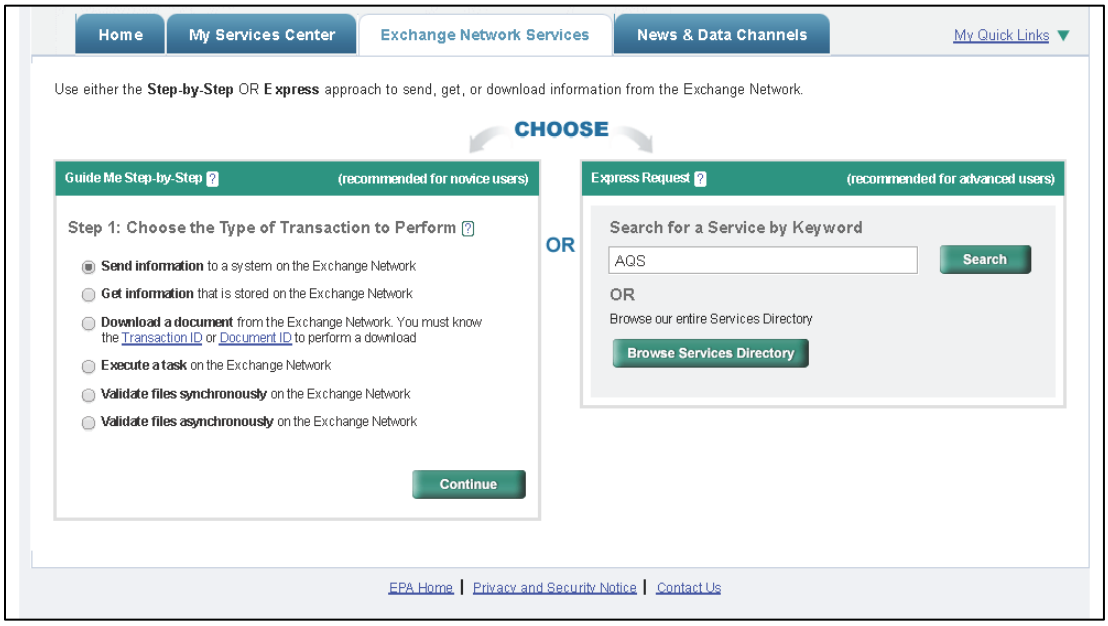


Following login, the home screen is accessed (Figure 18). For efficiency, add the AQS service to the home screen **My Quick Links** bar; however, it is also possible to search for the AQS submission form. To search, use the **Go** button of the **Exchange Network Services** bar.

Figure 18. Home screen of the Exchange Network Services Center website.

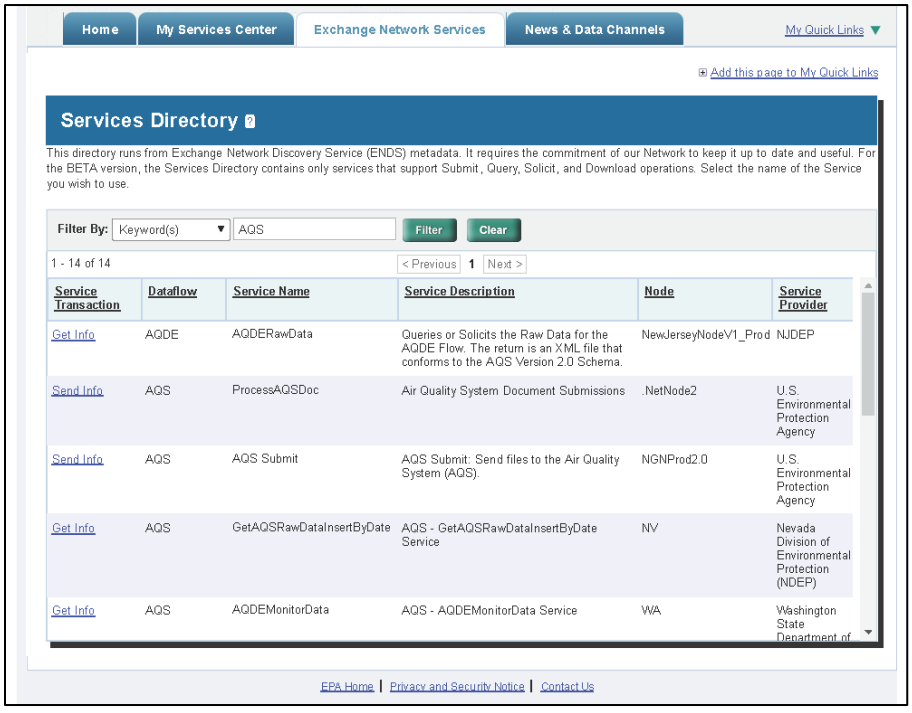
Next, the option for a step-by-step guide and a search bar presented (Figure 19); type **AQS** into the search bar.

Figure 19. Type *AQS* into the search bar.



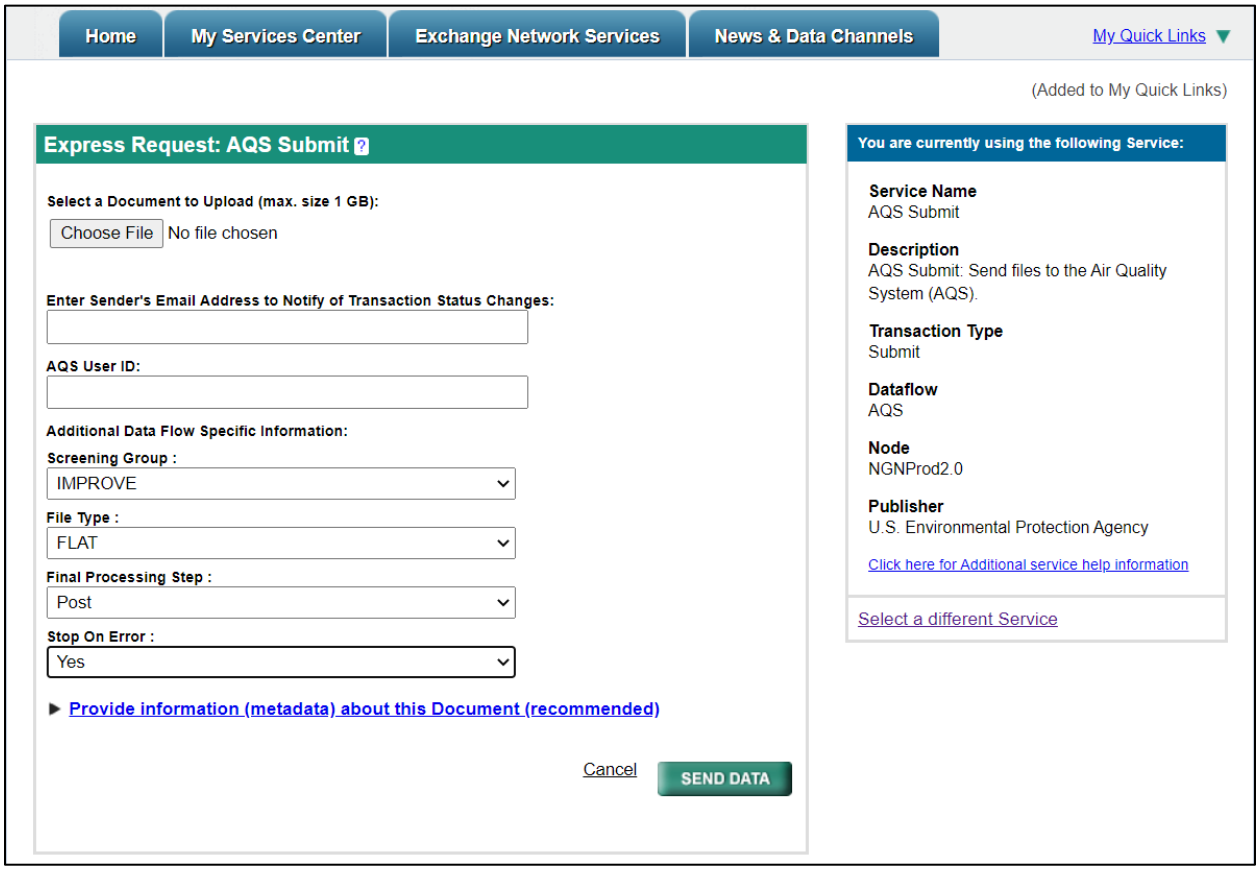
The search results will show all available processes associated with the AQS system (Figure 20). To access the AQS submission form, choose the service that has **AQS Submit** specified in the **Service Name** field (usually the third option listed).

Figure 20. Select the service named *AQS Submit*.



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 to stop on errors (YES). See Figure 21 for an example. Use the **Choose File** button to select the file generated from the previous step. Press the **SEND DATA** button to submit the form. Monitor progress of the data submission through the web portal.

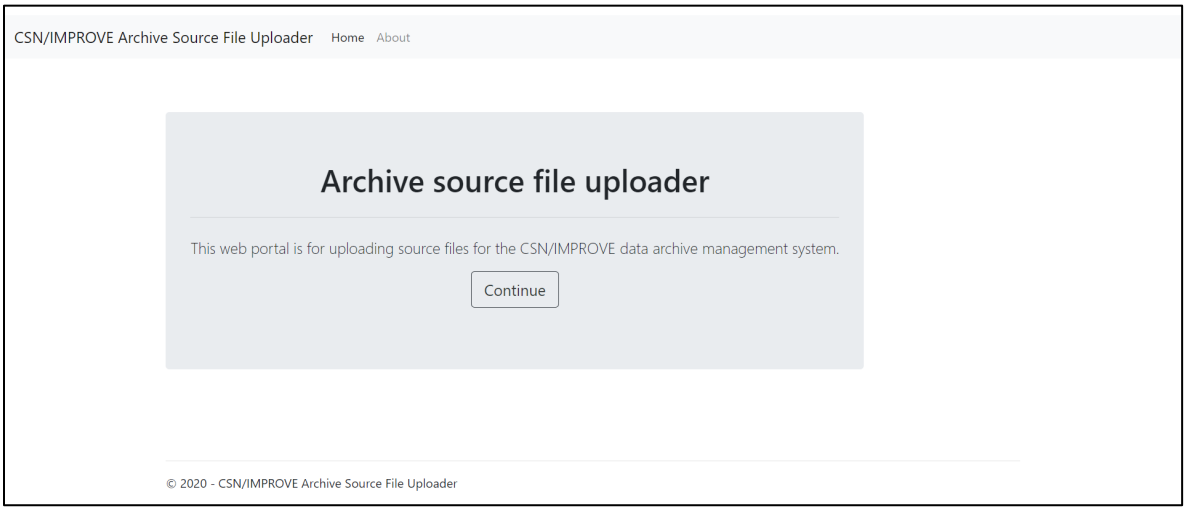
Figure 21. AQS data submission form.



7.3 Submission to UCD CIA

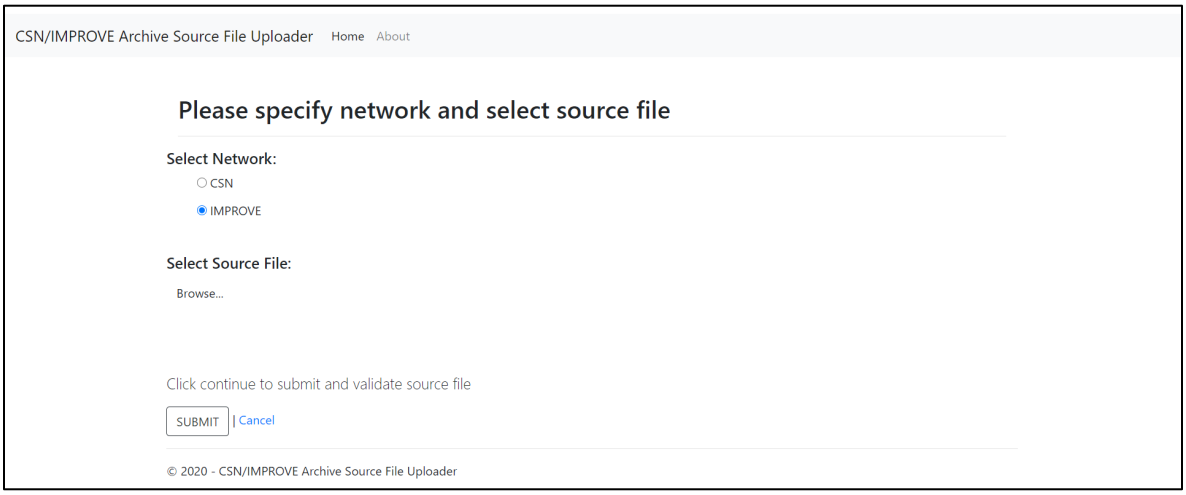
1. Open a web browser and navigate to the UCD CIA submission website, https://cia-uploadportal.azurewebsites.net/ (Figure 22).

Figure 22. UCD CIA submission website home page.



2. Click the **Continue** button in the center of the page.
3. Specify the network of choice that you will be delivering the data for, which in this case is **IMPROVE**. See Figure 23 for an example.

Figure 23. UCD CIA data submission details page.



4. Click **Browse** and select the file generated/submitted successfully to AQS.
5. Once the file is selected, click **Submit**; the next page will indicate if the submission was successful.

8. QUARTERLY FIELD STATUS REPORT

A field status report is generated quarterly to report on the status of all samples collected across the network for the previous quarter. Site status is evaluated relative to the regional haze rule criteria. The following information outlines the steps to generate the report and the checks to perform before delivery.

1. Create the report spreadsheet:
   - For the first quarter of a new year, save a copy of the template report under another name, with the format of *IMPROVE Status Report YYYY Q#.* The template report is located at U:\IMPROVE\Status Reports\Status_Report_Template.xlsx.
   - For the second, third, or fourth quarter, find the last report and save it under a name indicating the relevant quarter number. Previous reports are located at U:\IMPROVE\Status Reports\Reports
   - In the report there are four tabs:
     - Site Status Report
     - Status Flag Table
     - Flag Definitions: available from the database,

       *SELECT **

       FROM [Improve_2.1].[filter].[Statuses]
     - Sampler Locations: Determine if any sites are new, re-started, or have stopped during the relevant quarter by reviewing the date information in the *[Improve_2.1].[sampler].[Samplers]* table in the production database.
2. Populate the report:
   - From the IMPROVE Status page (http://analysis.crocker.ucdavis.edu:3838/ImproveStatus/), access the Network Status and Network Timeline tabs
   - The Network Status tab provides a count of the different statuses used per site, the total number of terminal statuses, which quarter they occur in, the percent complete by quarter, and the number of consecutive invalid statuses.
     - Change the Year and Ending Quarter fields to align with the reporting period.
     - Copy/paste the full table into an untitled spreadsheet.
     - Compare the columns in the Site Status Report tab to the content of the untitled spreadsheet; add new columns to the Site Status Report tab as needed.
     - Confirm that included flags are allowed (for example, the RF flag is no longer used). Investigate cases where unallowed flags are applied; work with the Sample Handling Laboratory to resolve.
     - Add the flag, definition, and result to the Flag Definitions tab of the report spreadsheet if not already listed.
     - Copy/paste content from the untitled spreadsheet to the Site Status Report tab.
     - Color the relevant fields:
       - Percent Complete by Quarter:

         < 75 %, yellow
         < 50 %, red
       - Consecutive Terminal Samples

         > 7, yellow

> 10, red

- Annual Completeness

  < 75 %, red
  Only color for the fourth quarter (Q4) report. Report uncolored values for the first, second, and third quarter reports.

  - Check formatting for consistency, including font style and type, coloring and shading.
  - Consecutive Terminal Samples may require merged cells in order to report a single number per row. Select all the cells to be merged; navigate to the MS Office Home tab and select Merge Across (drop down menu by Merge and Center); click okay for each row for the cells highlighted.
  - Update the date and quarter details at the top of the Site Status Report tab.

- The Network Timeline documents the most severe filter status for each site and date. For this report, exclude the flow status; flow validation often results in changes to status and is performed after this report is generated.
  - Change the Year to be relevant.
  - Do a search for NF statuses. For NF statuses, process the flows using the SQL execution code. If the NF statuses are for the most recent period, run the following code in SQL, changing the Start Date and End Date fields; if successful, a date/time of completion will show in the window:

```
DECLARE @RC int
DECLARE @iStartDate datetime = 'mm/dd/yyyy'
DECLARE @iEndDate datetime = 'mm/dd/yyyy'
DECLARE @iSamplerName NVARCHAR(50) = NULL
DECLARE @iFilterId BIGINT = NULL
DECLARE @Debug bit = 1
EXECUTE @RC =
[Improve_2.1].[sampler].[spFilterAverageFlowRates]
@iStartDate
,@iEndDate
,@iSamplerName
,@iFilterId
,@Debug
GO
```

  If the NF statuses are for a small set of filters/sites/dates, confirm why this is the case and edit the above code above to run on the specified filter, date range, and/or site.
  If the execution code fails, evaluate the warning message and work with the Software & Analysis Group and/or Sample Handling Laboratory to identify the issue and resolve.

- Do a search for no statuses For sites with no statuses, determine if it is a new site or if there is a reason such as paused shipments or the site temporarily offline.

  If the site is new, there may be blank records prior to the start date; if so, leave as-is but make sure the site is not falsely reported as failing the Regional Haze Rule criteria.

  If shipments are paused, work with the Software & Analysis Group and Sample Handling Laboratory; records may need to be added and/or the OL status may need to be manually inserted.

- If no NF or blank statuses, copy the relevant data and paste it into an untitled spreadsheet.
- Compare the site list from the untitled spreadsheet with the sites listed in the Sampler Locations and Status Flag Table tabs. Update site and date information as needed.
- Copy/paste content from the untitled spreadsheet to the Status Flag Table tab. Make sure the full table is copied over, not just the most recent quarter; some filters may have changed status since the last report was generated
- Color the relevant fields:
  - QD flags, yellow
  - Null/terminal flags, red

3. Perform checks prior to delivery:
   - Verify that the color coding is correctly assigned.
   - Status Flag Table tab: Look for blocks of red (invalid) and SO flags. Investigate using JIRA and/or follow up with the Sample Handling Laboratory.
   - Status Flag Table tab: Spot check to ensure that the number of terminal flags is corresponding to those reported in the Site Status Report tab.
   - Status Flag Table tab: Confirm that the sites listed are also shown in the Site Status Report tab and the Sampler Locations tab.
   - Flag Definitions tab: Confirm that the formatting and color coding is correct.
   - Sampler Locations tab: Confirm that new sites have been added.
4. Send to the Data & Reporting Group Manager for review.

9. 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.; Lioy, 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.