DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

UCD IMPROVE Technical Information #351A

Data Ingest

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

October 3, 2022
Version 1.0

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 10/7/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 10/7/2022

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 10/7/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
03/14/22	SRS	All	Previously anthologized version separated into individual TIs.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide information on handling electronic laboratory records from samples collected in the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. This document is intended to guide users on the receiving and validating of IMPROVE laboratory records and ingestion to the University of California, Davis (UCD), IMPROVE database. These include ion analysis results from RTI International (RTI), carbon analysis results from Desert Research Institute (DRI), and gravimetric mass, elemental, and filter absorption analysis results from UCD.

2. SUMMARY OF THE METHOD

Ion analysis results from RTI and carbon analysis results from DRI are received in data files, typically delivered as .csv files and XML files, respectively. The files are ingested into the UCD IMPROVE database using the UCD IMPROVE Data Management website. Gravimetric mass, elemental, and filter absorption analysis results from UCD are automatically ingested.

3. DEFINITIONS

- **AQRC:** Air Quality Research Center.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **DRI:** Desert Research Institute.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.
- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (μm) and smaller. PM_{10} is particulate matter with diameters 10 μm or smaller.
- **RTI:** Research Triangle Institute, International.
- **SQL:** database management system used by AQRC.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon. Also referred to as TOR (Thermal Optical Reflectance) and TOT (Thermal Optical Transmittance).
- **UCD:** University of CA—Davis.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

- **Extensible Markup Language (XML):** a markup language defining a set of rules for encoding documents in a particular format; used for IMPROVE carbon files.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data ingest are described in the associated *UCD IMPROVE SOP #351: Data Processing & Validation*.

9. PROCEDURAL STEPS

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.

9.1 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

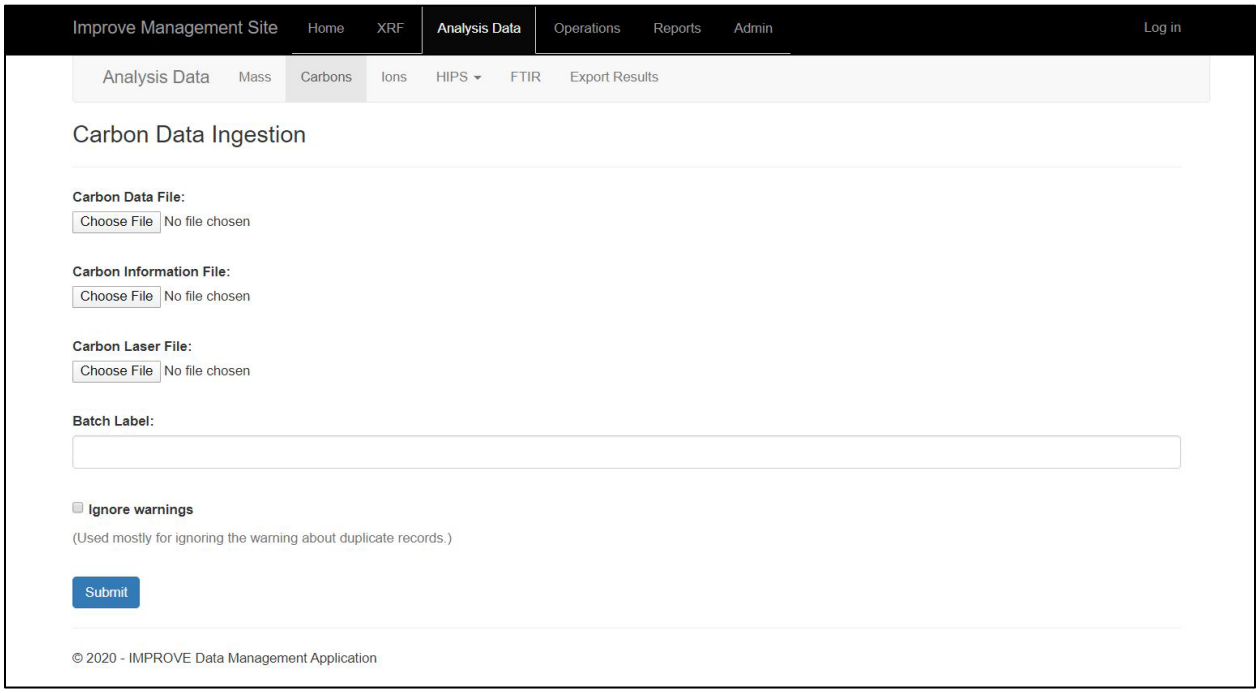
All three files are included in a zip folder which will be saved to U:\IMPROVE\RawDataReceived\Carbon DRI\To be Imported. The files will be extracted by right clicking on zipped folder, selecting 7-Zip and then ‘Extract Here’. A

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

new folder with the same name as the zipped folder will be created in the same location on the U Drive and contain all three files as named above.

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 8, 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**. The suggested batch label is the filename from DRI (First sample date - Last Sample date). *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 the same as the *Batch Label* created for ingest. After successfully ingesting the results, the Quality Assurance Officer will save a copy of the Carbon Data Ingestion summary page by printing the page to a PDF and saving to U:\IMPROVE\RawDataReceived\Carbon DRI\IngestRecord\. Further details of the ingest file are recorded in a log file located at U:\IMPROVE\RawDataReceived\Carbon DRI\Carbon_Ingest_log.xlsx.

Figure 1. Carbon analysis results upload page.



DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

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
Carbon Laser 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/Carbon Laser/Carbon Info file)	Warning

9.2 Ion 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 file is saved to the file server at U:\IMPROVE\RawDataReceived\Ions RTI\To Be Ingested.

The ion analysis records are ingested using the UCD IMPROVE Management website. Figure 2 shows a screenshot of the ions data upload page, accessed via the Analysis Data Section as described in section 8, selecting the Ions tab, and click the **Upload Data** button. To ingest a file from the data upload page, select the relevant file and click **Continue**. An automated validity check is performed, and the validity check results 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. An ingest page after passing validation is shown in Figure 3.If the file is ready for ingestion, click **Submit**. After ingest, the source files are stored on the file server at U:\IMPROVE\RawDataReceived\Ions RTI\Ingested. After successfully ingesting the results, the Quality Assurance Officer will save a copy of the Ions Data Ingestion Status summary page by printing the page to a PDF and saving to U:\IMPROVE\RawDataReceived\Ions RTI\Ingest_record\. Further details of the ingest file are recorded in a log file located at U:\IMPROVE\RawDataReceived\Ions RTI\Ions_DataIngest_Log.xlsx.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 2. Ions analysis results upload page.

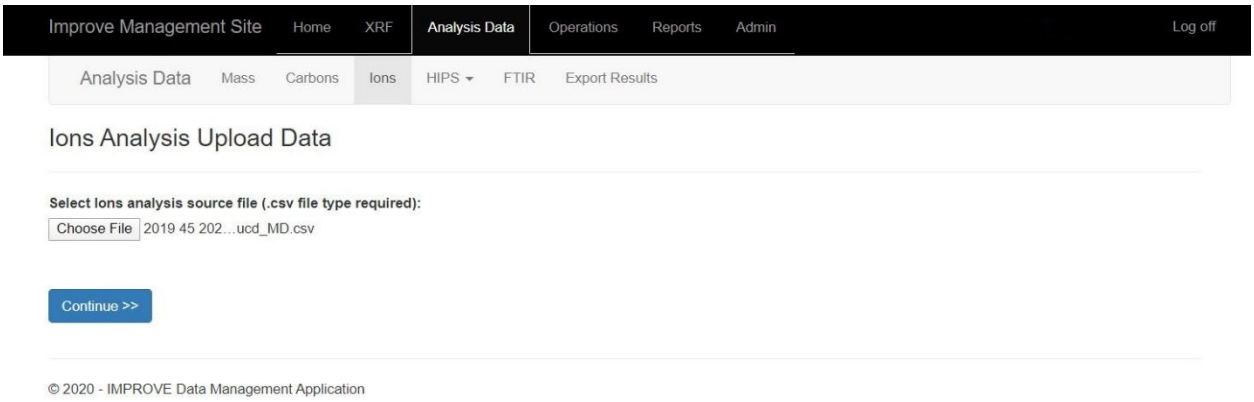
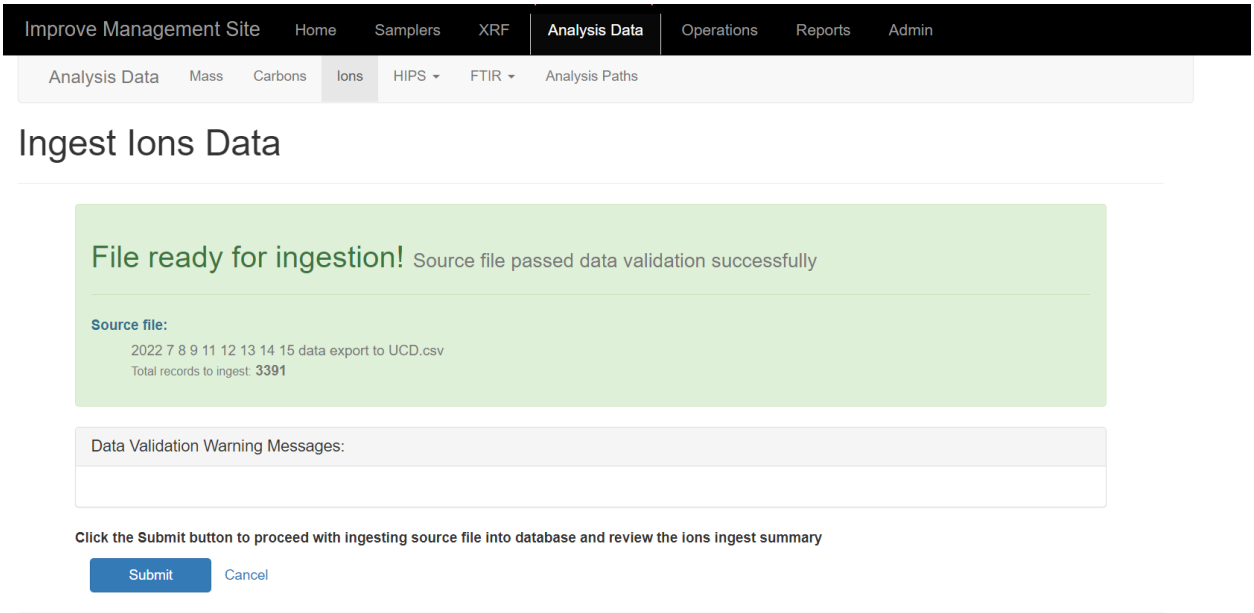


Table 2. Automated validity checks performed during the ions data upload.

Check	**Action**
Basic schema validation on csv files	Error message
No filter is found for record	Error message
Data already exists for filter record	Warning message
Parameter missing for a filter e.g., parameter has null value or parameter column isn't present at all	The ingest process will abort with an error message
Parameter already recorded in database	Warning

Figure 3. Ions ingest page after passing validation



9.3 Element and Optical Absorption Results

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Elemental analysis is performed at UCD. The PANalytical XRF software generates results files and automatically transmits them to a directory on an AQMT file server. A service on the server (internally named *XRF Data Transfer*) monitors the transmission directory, checking every five minutes 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 which are then verified by the operator to be complete and then written to the database. The data are then available on the UCD IMPROVE database.

9.4 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 9.1 and 9.2.

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, if existing records are not updated, or if 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 carbon, if the reanalysis results are used, the analysis QC code can be updated using the tool available at https://improve.aqrc.ucdavis.edu/AnalysisData/Carbons/CarbonsQcReview.

For ions, the data are ingested without any changes to the original process; the QC code is updated using the tool available at https://improve.aqrc.ucdavis.edu/AnalysisData/Ions/IonsQcReview.

9.5 Issue Tracking

Software bugs and data management issues are tracked through JIRA tracking software. All users who have access to the internal UCD JIRA website can submit, track, and comment on issues. Users requesting new tools, modifications to existing tools, or to report bugs specific to the IMPROVE data should add JIRA tickets to the IMPROVE Data Management Software project at https://improve.atlassian.net/jira/software/c/projects/IMPSW/issues/

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

10. QUALITY ASSURANCE AND QUALITY CONTROL

10.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

10.2 Bug Reporting

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

10.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this SOP.

11. REFERENCES

Not applicable.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Section 508 Compliant ☐ Yes ☑ No

UCD IMPROVE Technical Information #351B

Data Processing

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

November 10, 2022
Version 1.1

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 11/10/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 11/10/2022

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 11/10/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

Electronic documents are official. Paper copies are for reference only.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

DOCUMENT HISTORY

Revision	Release Date	Initials	Section/s Modified	Brief Description of Modifications
1.0	03/14/22	SRS	All	Previously anthologized version separated into individual TIs.
1.1	11/10/22	DEY, ITS	9	Corrected flow processing examples. Corrected Pre- and post-weight units.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

TABLE OF CONTENTS

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

LIST OF FIGURES

LIST OF TABLES

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide information on the steps for processing the sampling and analytical data from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. The raw operational information from field sampling is combined with laboratory analytical results to generate concentrations, uncertainties, and method detection limits.

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst will use functions in the *crocker* software package to calculate final results and post them to the UCD IMPROVE database. The analyst will also review any output messages for errors.

3. DEFINITIONS

- **AQRC:** Air Quality Research Center.
- **AQS:** EPA's Air Quality System database.
- **CSN and IMPROVE Archive (CIA) Database:** A database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.
- **Chemical Speciation Network (CSN)**: EPA's $PM_{2.5}$ sampling network, with sites located principally in urban areas.
- **CIRA:** Cooperative Institute for Research in the Atmosphere.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA)
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.
- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **NPS:** National Park Service.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (µm) and smaller. PM_{10} is particulate matter with diameters 10 µm or smaller.
- **SQL:** database management system used by AQRC.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon. Also referred to as TOR (Thermal Optical Reflectance) and TOT (Thermal Optical Transmittance).
- **UCD:** University of CA—Davis.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

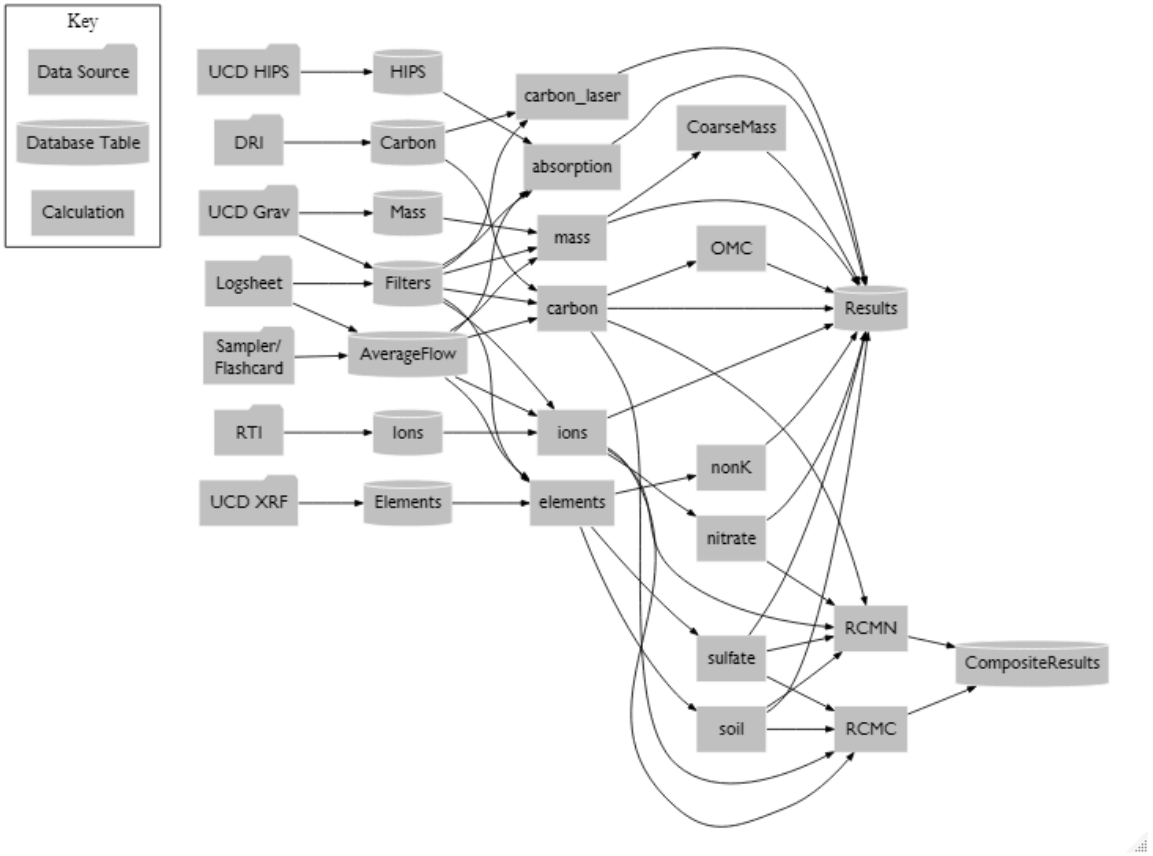
The hardware and software used for IMPROVE data ingest are described in the associated *UCD IMPROVE SOP #351: Data Processing & Validation.*

9. PROCEDURAL STEPS

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 1 shows a flow chart for the IMPROVE data processing.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Figure 1. Data processing flow chart.



Calculation of concentrations and associated uncertainties and MDLs are performed within the *crocker* R package. The flow rate calculations can be performed in the *crocker* R package or in the UCD IMPROVE database. Flow rate calculations are performed before calculating concentrations to ensure the most up-to-date flow data are used.

Flow data are processed using a stored SQL procedure 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 with flow rates that deviate from the nominal values.

The stored SQL procedure is shown below. The first six lines of the SQL query below 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
```

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF


```
DECLARE @iFilterId BIGINT = NULL
DECLARE @Debug bit = 1

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

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.

The flow data can also be processed in the R environment by the *improve_process_flow* function from the *crocker* package. This function uses the main stored procedure (spFilterAverageFlowRates; detailed above) in the IMPROVE database for processing 24-hr average flow rates. To perform flow data processing in the R environment, open an R environment (such as RStudio) and run the following command:

```
[flow] <- crocker::improve_process_flow(start_date = ['YYYY-MM-DD'],
end_date =  = ['YYYY-MM-DD'], sampler =[NULL], filter_id =[NULL], server
= "production")
```

The start date, end date, and server information are required fields like the SQL procedure. Similar to the SQL procedure, the flow processing function can be performed on a single site, date, or even filter ID by declaring the appropriate values.

For processing flow for the entire network, the following code can be used:

```
[flow] <- crocker::improve_process_flow(start_date = ['YYYY-MM-DD'],
end_date =  = ['YYYY-MM-DD'], server = "production")
```

If successful, the code will return a value of zero. If another value is returned, try to run the code directly in SQL in debugging mode to investigate the error messages and inform the software group.

To calculate a single month of concentration, uncertainty, and MDL for all IMPROVE parameters, including reported and non-reported diagnostic parameters, the following command is run the in R environment:

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

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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 (when *server* = 'production'), and appends the processed data to the analysis.Results or analysis.CompositeResults table in the UCD IMPROVE production database (Improve_2.1) 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*. Routine data uses *AnalysisQcCode = 1*.

The *skip* argument gives the flexibility to skip one or more categories of analysis in processing. All analyses except those listed in the *skip* argument will be posted to the database. Any parameters that are derived from the skipped parameters are also skipped. A typical command using the *skip* argument can be as follows:

> *November_data <- crocker::improve_calculate_and_post(2020, 11, server = "production ", skip = "optical")*

If the user wants to skip multiple analysis categories, the categories can be grouped in the *skip* argument, as shown below.

> *November_data <- crocker::improve_calculate_and_post(2020, 11, server = "production", skip = c("optical", "elements"))*

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.

9.1 Units

Table 1 lists the data types, parameters, and units for all data delivered to the CIRA, AQS, and UCD CIA databases (see *UCD IMPROVE TI #351D: Data Delivery*). 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.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Table 1. 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	fAbs	Mm^{-1}	NA	NA
Composite species	OMC, NHNO, NHSO, PM_{10}-$PM_{2.5}$, Soil	NA	$\mu g/m^3$	$\mu g/m^3$

9.2 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 ions, 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

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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 9.4.4.

In the rare situations where not enough field blanks from a given lot are available according to standard practices, the analyst will take one of the following three actions:

- Perform normal data processing for the lot with fewer field blanks than are typically required;
- Use medians of previous field blank statistics if the lot has previously been used. The field blank statistics for the current lot will include medians of the median, 95th percentile ,standard deviation, and field blank count; or
- Use medians of previous field blank statistics for all lots from the same manufacturer if the lot has not previously been used. The field blank statistics for the current lot will include medians of the 95th percentile, standard deviation, and field blank count. The median will be set to zero to negate blank correction.

In all of the above cases, the data will be reviewed to determine if the resulting statistics are appropriate or if another approach should be used. The analyst will add comments to the field blank sets tables in the database noting the deviation from standard procedures.

9.3 Volume

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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 **Error! Reference source not found.** 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.

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

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 **Error! Reference source not found.**

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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 2).

Table 2. 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.

9.4 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 9.

The concentration is calculated using equation 351-3, where the mass of material on the filter is equal to the difference between the mass measured on the sample and the mass on the unused filter. 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 9.4.4.

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

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

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

The uncertainty and MDLs are reported with each concentration. Uncertainties and MDLs for ions, carbon, and elements are calculated using the following equations:

$$\sigma_{(C)} = 1000\frac{ng}{\mu g} * \frac{\sqrt{(0.608*Max(P95-B, mdl_{analytical}))^2+(f*(A-B))^2}}{V} \quad (351\text{-}4)$$

$$mdl_{(C)} = 1000\frac{ng}{\mu g} * \frac{Max(P95-B, mdl_{analytical})}{V} \quad (351\text{-}5)$$

Where,

V = Module sample air volume (m^3)

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

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

$mdl_{analytical}$ = 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. This term results from various sources of proportional uncertainties, such as 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 (351-6 and 351-7). If the count of collocated pairs over the two-year period is less than 60, a value of 0.25 is adopted as f.

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.

$$srd = \frac{(Collo - Routine)/\sqrt{2}}{(Collo + Routine)/2} \quad (351\text{-}6)$$

$$f = \frac{(84th\ percentile\ of\ srd) - (16th\ percentile\ of\ srd)}{2} \quad (351\text{-}7)$$

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. The date range specified must be for a two-year period prior to the current year of data to be processed. The function can also be used for other purposes where the user can specify any time period of interest.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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.

9.4.1 $PM_{2.5}$ 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 milligrams per filter) are stored in the *grav.SampleAnalysis* 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:

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

$$\sigma_{Mass} = 1000\,\frac{ng}{\mu g} * \frac{\sqrt{(0.608 * Max(P95, mdl_{analytical}))^2 + (f * (postweight - preweight))^2}}{V} \quad (351\text{-}9)$$

$$mdl_{Mass} = 1000\,\frac{ng}{\mu g} * \frac{Max(P95, mdl_{analytical})}{V} \quad (351\text{-}10)$$

Where,

postweight = mass of filter after sampling

preweight = mass of filter before sampling

V = A-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 (10 µg/filter for $PM_{2.5}$ and PM_{10}). The analytical MDL is considered the 'floor value' and is used as the

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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

f = fractional uncertainty (Table 3).

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.

Table 3. Fractional uncertainty for the mass.

Species	**f reported for data**					
	2/28/1995 – 12/31/2016	**2017**	**2018**	**2019**	**2020**	**2021**
$PM_{2.5}$	0.03	0.03	0.04	0.04	0.04	0.05
PM_{10}	0.03	0.07	0.07	0.08	0.07	0.05

9.4.2 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 *ions.MassLoadings* 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:

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

$$\sigma_{ion} = 1000 \frac{ng}{\mu g} * \frac{\sqrt{(0.608 * Max(P95 - B_{ion}, mdl_{analytical}))^2 + (f * (A_{ion} - B_{ion}))^2}}{V_{B\ Module}} \quad (351\text{-}12)$$

$$mdl_{ion} = 1000 \frac{ng}{\mu g} * \frac{Max(P95 - B_{ion}, mdl_{analytical})}{V_{B\ Module}} \quad (351\text{-}13)$$

Where,

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

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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 in µg/filter reported from the analytical laboratory (Table 4). 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 5).

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.

Table 4. Analytical method detection limits (MDL) in µg/filter for the ions species.

Species	**Analytical MDLs used for data 1/1/2006 – 12/31/2019**	**Analytical MDLs used for data 1/1/2020 – 12/31/2020**	**Analytical MDLs used for data 1/1/2021 – current**
Chloride (Cl^-)	0.03	0.1	0.1
Nitrite (NO_2^-)	0.01	0.2	0.2
Nitrate (NO_3^-)	0.05	0.16	0.16
Sulfate ($SO_4^=$)	0.07	0.22	0.22

Table 5. Fractional uncertainty for ions.

Species	**f reported for data**					
	1/1/2005 – 12/31/2016	**2017**	**2018**	**2019**	**2020**	**2021**
Chloride (Cl^-)	0.08	0.08	0.08	0.09	0.10	0.09
Nitrite (NO_2^-)	0.22	0.25	0.25	0.25	0.25	0.25
Nitrate (NO_3^-)	0.04	0.03	0.04	0.04	0.04	0.04
Sulfate ($SO_4^=$)	0.02	0.02	0.02	0.03	0.02	0.01

9.4.3 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 pyrolyzed carbon (OP) are recorded in

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

micrograms per filter and stored in the *dricarbon.MassLoadings* 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:

$$C = 1000\frac{ng}{\mu g} * \frac{(A_{carbon} - B_{carbon})}{V_{Cmodule}} \quad (351\text{-}14)$$

$$\sigma_{Carbon} = 1000\frac{ng}{\mu g} * \frac{\sqrt{(0.608 * Max(P95 - B_{carbon}, mdl_{analytical}))^2 + (f * (A_{carbon} - B_{carbon}))^2}}{V_{C\,Module}} \quad (351\text{-}15)$$

$$mdl_{Carbon} = 1000\frac{ng}{\mu g} * \frac{Max(P95 - B_{carbon}, mdl_{analytical})}{V_{C\,Module}} \quad (351\text{-}16)$$

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 in µg/filter reported from the analytical laboratory (Table 6). 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 7).

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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.

Table 6. Analytical method detection limits (MDL) in µg/filter for the carbon species.

Species	Analytical MDLs used for data 1/1/2006 – 12/31/2019	Analytical MDLs used for data 1/1/2020 – 12/31/2020	Analytical MDLs used for data 1/1/2021 – current
OC1	0.51	0.03	0.03
OC2	0.51	0.06	0.04
OC3	0.51	0.18	0.16
OC4	0.51	0.12	0.12
OPTR	0.15	0.12	0.14
OPTR at 405 nm	0.15	0.03	0.05
OPTR at 445 nm	0.15	0.06	0.11
OPTR at 532 nm	0.15	0.08	0.15
OPTR at 780 nm	0.15	0.08	0.12
OPTR at 808 nm	0.15	0.06	0.06
OPTR at 980 nm	0.15	0.12	0.04
OPTT	0.15	0.22	0.22
OPTT at 405 nm	0.15	0.18	0.26
OPTT at 445 nm	0.15	0.21	0.22
OPTT at 532 nm	0.15	0.19	0.24
OPTT at 780 nm	0.15	0.2	0.19
OPTT at 808 nm	0.15	0.19	0.2
OPTT at 980 nm	0.15	0.15	0.21
EC1	0.15	0.07	0.06
EC2	0.15	0.22	0.27
EC3	0.15	0.01	0.01
ECTR	0.15	0.23	0.27
OCTR	0.51	0.31	0.33
TC	0.57	0.43	0.45

* Prior to 2017, data for OP at different wavelengths were not reported.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Table 7. Fractional uncertainty for the carbon species.

Species	f reported for data					
	1/1/2005 – 12/31/2016*	**2017**	**2018**	**2019**	**2020**	**2021**
OC1	0.23	0.27	0.23	0.24	0.21	0.17
OC2	0.15	0.13	0.11	0.10	0.09	0.09
OC3	0.13	0.13	0.13	0.11	0.09	0.09
OC4	0.15	0.13	0.13	0.14	0.16	0.16
OPTR	0.13	0.16	0.20	0.21	0.20	0.19
OPTR at 405 nm	N/A	0.18	0.18	0.19	0.19	0.19
OPTR at 445 nm	N/A	0.17	0.17	0.18	0.18	0.18
OPTR at 532 nm	N/A	0.20	0.21	0.21	0.21	0.2
OPTR at 780 nm	N/A	0.19	0.21	0.22	0.22	0.22
OPTR at 808 nm	N/A	0.19	0.20	0.21	0.22	0.23
OPTR at 980 nm	N/A	0.21	0.23	0.25	0.25	0.24
OPTT	0.13	0.12	0.14	0.15	0.14	0.13
OPTT at 405 nm	N/A	0.13	0.13	0.14	0.13	0.12
OPTT at 445 nm	N/A	0.13	0.13	0.15	0.14	0.13
OPTT at 532 nm	N/A	0.13	0.14	0.15	0.14	0.13
OPTT at 780 nm	N/A	0.13	0.14	0.16	0.14	0.14
OPTT at 808 nm	N/A	0.13	0.15	0.16	0.15	0.14
OPTT at 980 nm	N/A	0.14	0.16	0.17	0.15	0.14
EC1	0.10	0.10	0.11	0.11	0.11	0.12
EC2	0.17	0.18	0.19	0.21	0.22	0.22
EC3	0.42	0.25	0.25	0.25	0.25	0.25
ECTR	0.12	0.14	0.14	0.13	0.13	0.14

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

OCTR	0.08	0.09	0.08	0.07	0.07	0.06
TC	0.08	0.08	0.07	0.07	0.06	0.06

9.4.4 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 filter lot as that of the sample filters. The field blank correction is specific to each filter lot and since the number of field blanks from a filter lot used in a given month may not be statistically sufficient, a minimum of 35 field blanks are required before the median can be calculated. Field blank selection is therefore expanded to include field blanks from previous month(s) until at least 35 field blanks are found. The selected 35 field blanks are used to calculate batch and filter lot-specific blank correction. Areal uncertainty ($U_{element}$) is calculated as,

$$U_{element} = 1000\frac{ng}{\mu g} * \sqrt{(0.608 * Max\left((P95 - B_e), mdl_{analytical}\right))^2 + (f * (A_e - B_e))^2} \quad (351\text{-}17)$$

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 in µg/cm^2 reported from the analytical laboratory (Table 8). 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.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Table 8. Analytical method detection limits (MDL) in µg/cm2 for the elemental species.

Species	**Analytical MDLs used for data 1/1/2006 – 12/31/2019**	**Analytical MDLs used for data 1/1/2020 – 12/31/2020**	**Analytical MDLs used for data 1/1/2021 – current**
Al	0.011	0.011	0.011
As	0.002	0.002	0.002
Br	0.001	0.001	0.001
Ca	0.021	0.003	0.003
Cl	0.002	0.002	0.002
Cr	0.001	0.001	0.001
Cu	0.002	0.001	0.001
Fe	0.012	0.003	0.003
K	0.005	0.001	0.001
Mg	0.021	0.02	0.02
Mn	0.003	0.002	0.002
Na	0.037	0.046	0.046
Ni	0.001	0.001	0.001
P	0.002	0.002	0.002
Pb	0.006	0.003	0.003
Rb	0.002	0.002	0.002
S	0.003	0.001	0.001
Se	0.002	0.001	0.001
Si	0.013	0.005	0.005

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Sr	0.002	0.001	0.001
Ti	0.003	0.001	0.001
V	0.001	0.001	0.001
Zn	0.002	0.002	0.002
Zr	0.012	0.007	0.007

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 – 12/31/2019**	**f reported for data 1/1/2020 – 12/31/2020**	**f reported for data 1/1/2021 - current**
Al	0.09	0.08	0.08	0.09	0.10	0.12
As	0.25	0.21	0.25	0.25	0.25	0.25
Br	0.10	0.11	0.10	0.09	0.09	0.09
Ca	0.06	0.07	0.06	0.07	0.09	0.1
Cl	0.14	0.18	0.14	0.14	0.16	0.18
Cr	0.22	0.17	0.15	0.17	0.16	0.15
Cu	0.12	0.11	0.13	0.10	0.10	0.11
Fe	0.06	0.06	0.05	0.06	0.08	0.09
K	0.03	0.05	0.03	0.04	0.05	0.06
Mg	0.15	0.16	0.15	0.15	0.17	0.2
Mn	0.13	0.13	0.14	0.13	0.13	0.13
Na	0.14	0.15	0.14	0.14	0.15	0.16
Ni	0.16	0.16	0.13	0.14	0.18	0.14
P	0.25	0.33	0.27	0.30	0.30	0.35
Pb	0.13	0.13	0.14	0.15	0.25	0.19
Rb	0.25	0.25	0.25	0.25	0.25	0.25
S	0.03	0.03	0.02	0.03	0.03	0.03

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Se	0.25	0.12	0.25	0.25	0.25	0.25
Si	0.10	0.07	0.06	0.07	0.09	0.1
Sr	0.16	0.14	0.13	0.14	0.14	0.15
Ti	0.11	0.09	0.09	0.09	0.11	0.12
V	0.12	0.14	0.17	0.17	0.12	0.16
Zn	0.06	0.08	0.08	0.08	0.08	0.09
Zr	0.25	0.25	0.25	0.25	0.25	0.25

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:

$$C_{element} = 1000\frac{ng}{\mu g} * \frac{(A_e - B_e) * (Deposit\ area)}{V} \quad (351\text{-}18)$$

$$\sigma_{element} = \frac{(U_e) * (Deposit\ area)}{V} \quad (351\text{-}19)$$

$$mdl_{element} = 1000\frac{ng}{\mu g} * \frac{Max\left((P95 - B_e), mdl_{analytical}\right) * (Deposit\ area)}{V} \quad (351\text{-}20)$$

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).

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

9.4.5 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.Results* table in the UCD IMPROVE database.

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

$$fAbs = 100 * \frac{\tau_{633} * (Deposit\ Area)}{V_{A\ Module}} \quad (351\text{-}21)$$

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).

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

$$\sigma fAbs = 100 * \frac{\sqrt{\left(\frac{1}{1.65} * Max\left(P95, mdl_{analytical}\right)\right)^2 + (f_{unitless} * \tau_{633})^2 * (Deposit\ Area)}}{V_{A\ Module}} \quad (351\text{-}22)$$

Where,

P95 = 95th percentile of field blank measurements.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory (τ_{633} = 0.009, unitless). 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)

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

$$mdl_{fAbs} = 100 * \frac{Max\left(P95, mdl_{analytical}\right) * (Deposit\ Area)}{V_{A\ Module}} \quad (351\text{-}23)$$

Where,

P95 = 95th percentile of field blank measurements.

$mdl_{analytical}$ = analytical MDL reported from the analytical laboratory (τ_{633} = 0.009, unitless). The analytical MDL is considered the 'floor value' and is used as the reported

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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.

	f reported for data					
Species	**2/28/1995 – 12/31/2016**	**2017**	**2018**	**2019**	**2020**	**2021**
fAbs	0.03	0.06	0.06	0.05	0.06	0.07

In the rare situations where not enough field blanks from a given lot are available according to standard practices, see section 9.2 for guidelines on calculating field blank statistics.

9.5 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 *UCD IMPROVE #351C: Data Validation*. 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.

9.5.1 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

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

ammonium sulfate (NHSO) and sulfate calculated from XRF sulfur (S3) are calculated using the following equations:

$$NHSO = 4.125 * S$$

$$S3 = 3 * S \quad (351\text{-}24)$$

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

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

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

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

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.

9.5.2 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:

$$NHNO = 1.29 * NO_3^- \quad (351\text{-}27)$$

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

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

9.5.3 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 $\left(K_2O, \frac{39.1*2+16.0\ g/mol}{39.1*2\ g/mol} = 1.2 \right)$ 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.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- 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:

$$SOIL = 2.2 * \max(Al,0) + 2.49 * \max(Si,0) + 1.63 * \max(Ca,0) + 2.42 * \max(Fe,0) + 1.94 * \max(Ti,0) \quad (351\text{-}30)$$

$$\sigma(SOIL) = \sqrt{(2.2 * \max(\sigma(Al),0))^2 + (2.49 * \max(\sigma(Si),0))^2 + (1.63 * \max(\sigma(Ca),0))^2 + (2.42 * \max(\sigma(Fe),0))^2 + (1.94 * \max(\sigma(Ti),0))^2} \quad (351\text{-}31)$$

$$mdl(SOIL) = 0 \quad (351\text{-}32)$$

The soil variable is calculated for all valid XRF analyses.

9.5.4 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:

$$\text{KNON} = (\text{K} - 0.6*\text{Fe}) \quad (351\text{-}33)$$

$$\sigma(KNON) = \sqrt{\sigma^2(K) + [0.6 * \sigma(Fe)]^2} \quad (351\text{-}34)$$

$$mdl(KNON) = 0 \quad (351\text{-}35)$$

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.

9.5.5 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:

$$OMC = 1.8 \times OC = 1.8 \times (O1 + O2 + O3 + O4 + OP) \quad (351\text{-}36)$$

$$\sigma_{OMC} = 1.8 \times \sigma_{OC} \quad (351\text{-}37)$$

See equation of 351-14 for σ_{OC}.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

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

See equation 351-15 for $mdl_{OC.}$

9.5.6 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 (fAbs) measured by HIPS from the 1A Module PTFE filter.

$$BC = \frac{ln\,(transfinal - transinitial)}{MAC} \quad (351\text{-}39)$$

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.

9.5.7 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.

$$RCMC = NHSO + Soil + 1.8 \times Chloride + ECTR + OMC \quad (351\text{-}40)$$

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

$$\sigma_{RCMC} = \sqrt{\sigma_{NHSO}^2 + \sigma_{Soil}^2 + (1.8\sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2} \quad (351\text{-}41)$$

$$mdl_{RCMC} = 0 \quad (351\text{-}42)$$

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.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

9.5.8 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.

$$RCMN = NHSO + NHNO + Soil + 1.8 \times Chloride + ECTR + OMC \quad (351\text{-}43)$$

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

$$\sigma_{RCMN} = \sqrt{\sigma_{NHSO}^2 + \sigma_{NHSO}^2 + \sigma_{Soil}^2 + (1.8\sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2} \quad (351\text{-}44)$$

$$mdl_{RCMN} = 0 \quad (351\text{-}45)$$

10. DATA PROCESSING CODE

This section describes the flow of data through the data processing code used to calculate concentration, MDL, and uncertainty for all IMPROVE parameters. Figure 1 outlines the flow of data from the sampler and analysis specific database tables to final results. The function improve_process_flow and wrapper function improve_calculate_and_post from the crocker R package are the only functions executed directly by the analyst (see Section 9). The improve_process_flow function processes and posts flow data directly into the database. The improve_calculate_and_post function calls several functions sequentially to calculate first measured and then derived concentrations, as well as uncertainties and MDLs. Source code for these functions is stored in the UCD source repository.

11. QUALITY ASSURANCE AND QUALITY CONTROL

11.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

11.2 Bug Reporting

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

11.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this SOP.

12. REFERENCES

UCD IMPROVE SOP #351: Data Processing and Validation

UCD IMPROVE TI #351A: Data Ingest

UCD IMPROVE TI #351C: Data Validation

Lide, D.R. (Ed.) (2004). *CRC Handbook of Chemistry and Physics* (85th ed.). Boca Raton, FL: CRC Press.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

UCD IMPROVE Technical Information #351C

Data Validation

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

November 14, 2022
Version 1.1

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 11/14/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 11/14/2022

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 11/14/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

DOCUMENT HISTORY

Revision	Release Date	Initials	Section/s Modified	Brief Description of Modifications
1.0	03/14/22	SRS	All	Previously anthologized version separated into TIs.
1.1	11/14/22	DEY, ITS	9.2	Updated status_check function. Minor grammar updates.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

TABLE OF CONTENTS

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

LIST OF FIGURES

LIST OF TABLES

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide information regarding the validation of the analytical data from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. Data from the network are reviewed and validated using a variety of tools. Informational and/or terminal flags (statuses) are applied as appropriate.

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst uses the UCD IMPROVE Data Management website along with custom software in the R language to perform validation. The primary tools for review are summary data tables and comparison figures.

3. DEFINITIONS

- **AQRC:** Air Quality Research Center.
- **AQS:** EPA's Air Quality System database.
- **CSN and IMPROVE Archive (CIA) Database:** A database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.
- **Chemical Speciation Network (CSN)**: EPA's $PM_{2.5}$ sampling network, with sites located principally in urban areas.
- **CIRA:** Cooperative Institute for Research in the Atmosphere.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **datvalIMPROVE:** A custom software package in the R language that contains the data validation code used to collect, compare, and flag the final results.
- **DRI:** Desert Research Institute.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA)
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (μm) and smaller. PM_{10} is particulate matter with diameters 10 μm or smaller.
- **SQL:** database management system used by AQRC.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon. Also referred to as TOR (Thermal Optical Reflectance) and TOT (Thermal Optical Transmittance).
- **UCD:** University of CA—Davis.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data and Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data validation are described in the associated *UCD IMPROVE SOP #351: Data Processing and Validation*.

9. PROCEDURAL STEPS

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

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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 filter readings.
- 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 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.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- 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, carbon, elemental, and mass 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 occurs 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. 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.

9.1 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 13). 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.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Table 1. 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
NF	No Flow	Temporary	
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
QV	Data quality check (for Data Group only)	informational	
SA	Sampling Anomaly	Informational	1
SO	Still out	Temporary	
SP	Same-day Field Blank/Sample Swap	Informational	
SW	Sampling Dates Swap	Informational	
TO	Timing Outside normal bounds	Informational	Y
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	

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

RF	Really High Flow Rate	No longer used	W
PC	Possible Contamination	No longer used	4

9.2 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 data validation and HIPS data validation:

> Mass validation: *UCD IMPROVE TI #251R: General Laboratory Procedures*, section 5.8
>
> HIPS validation: *UCD IMPROVE TI #276C: QA/QC of Analysis of Loaded Filters Using HIPS*

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 before processing the concentration data. Detailed discussion concerning flow validation can be found in *UCD IMPROVE TI 351E: Flow Validation*. The following sections discuss the Level 1B checks in detail.

9.2.1 Level 1B Checks

The analysis data reported by the measurement laboratories are ingested into the UCD IMPROVE database to their corresponding tables (e.g., *dricarbon.MassLoadings*, *dricarbon.SampleAnalysis*, *hips.Results*, *ions.MassLoadings*, and *grav.SampleAnalysis*), as described in section 9 of *UCD IMPROVE TI #351A: Data Ingest.* Once all analysis results for a month are in the database, concentrations, MDLs, and uncertainties are processed and posted in the *analysis.Results* table using the *improve_calculate_and_post* function in the *crocker* package, as described in Section 9 of *UCD IMPROVE TI #351B: Data Processing.*

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

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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*, *elements fb.check*, and *carbon_fb.check* functions check for possible swap between same-day field blanks and samples for nylon, PTFE, 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"], sameday_swap_only = ['FALSE'])

  [month_year_ion_check] <- datvalIMPROVE::elements_fb.check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], by = ["element species"], sameday_swap_only = ['FALSE'])

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

  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 (*["elements species"]*, e.g. "S" or *["carbon species"]*, e.g. "ECTR") is higher than the associated sample mass loading. If *sameday_swap_only* is set to 'FALSE', all records will be returned. To return only the possible same-day swaps, set to 'TRUE'.

  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 from IC analysis, 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 elements sulfur (S) is the primary species (followed by sodium (Na) and then silicon (Si)). For carbon, ECTR, OCTR, OPTR, and TCTC are the primary species used for field blank comparison.

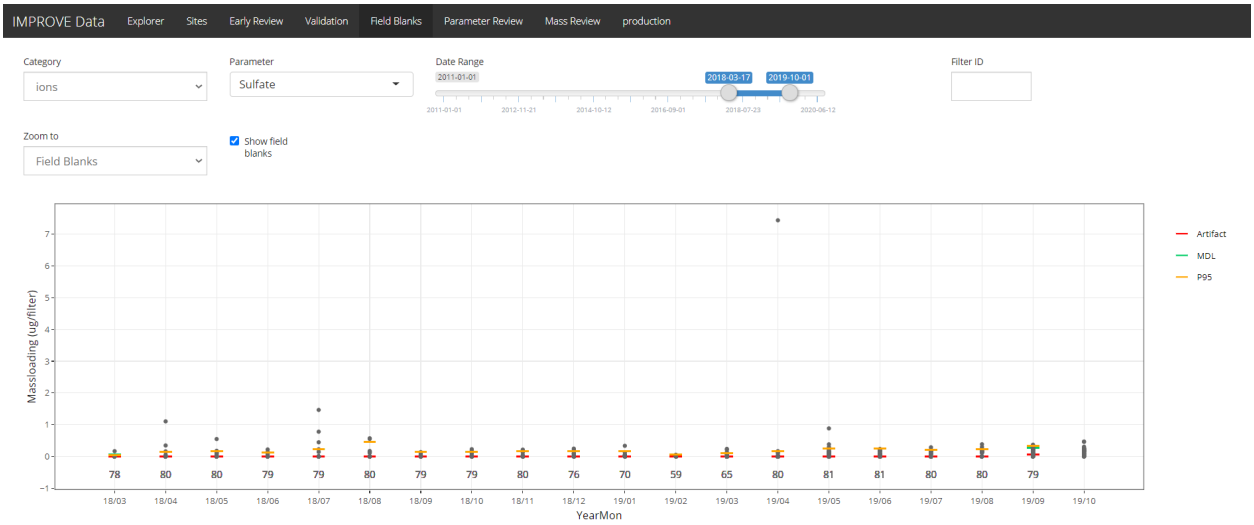
  For all analysis types (ions, carbon, elements, and mass), field blank data across the network can be compared using the Field Blanks tab in the IMPROVE Data website (https://shiny.aqrc.ucdavis.edu/ImproveData/; Figure 1). The mass loading of a specified parameter should be compared to field blank data from the same month as

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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 95th 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 1. 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.

These checks can be performed prior to processing the data (described in IMPROVE TI 251B Data Processing) or after the data have been processed for initial validation. If the checks are performed after data have been processed and issues are identified with field blanks, such as field blank – sample swaps or other field blanks with outlier mass loadings that have been resolved through swapping the data, reanalysis, and/or invalidation, the quality assurance officer should invalidate the appropriate field blank statistic sets in the database so that only correct and valid data are included in the calculations when the data are processed again, typically as part of preparations for data delivery. Table 2 lists the table names where the field blank statistics are saved for each analysis type. An example of the SQL query and update statement is given below. The query and update statement can be modified for each analysis type by selecting the appropriate table name from Table 2.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Table 2. Details of the IMPROVE database table names where field blank statistics are stored for each of the relevant analyses.

Analysis Type	**Field Blank statistics database table**
Mass	[Improve_2.1].[grav].[FBSets]
XRF	[Improve_2.1].[xrf].[FBSets]
HIPS	[Improve_2.1].[hips].[FBSets]
IC	[Improve_2.1].[ions].[FBsets]
TOR	[Improve_2.1].[dricarbon].[FBSets]

To invalidate a field blank set, first query the IMPROVE database using the following SQL query to find the relevant set Id, where '####' represents the four-digit year and '##' the two-digit month.

SELECT Id

FROM [Improve_2.1].[xrf].[FBSets]

WHERE Year= '####' and Month = '##' and AnalysisQcCode = 1

Next, update the AnalysisQccode of the field blank set using the Id from the above query, along with Year, Month using the statement below. A comment explaining the reason for invalidating the set can be added using the update query. If a comment already exists, make sure that it was appended and not overwritten.

UPDATE [Improve_2.1].[xrf].[FBSets]

SET AnalysisQcCode = 0, Comment = XXXX'

WHERE Year= '####' and Month = '##' and Id = '##'

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'], server = "production")

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

- Questionable Data (QD): To guide the Level 2 validation, a list of filters with the QD flag (QD – questionable data) is generated. 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 comparing results with neighboring sites. QD

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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. To generate the list, run the following command in the R environment:

[month_QD] <- datvalIMPROVE::status_check(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’], , status_table = "final", include_FB = TRUE, status = "('QD')", server = “production”)

. The argument *status_table* has two choices depending on which database table is to be queried: "final" (statuses from the final delivery table, which includes filter and analysis status)es and "filter" (statuses from the Filters table, which only considers the filter status). The default is "final". The argument *include_FB* has two choices: include field blank QD filter status (“TRUE”) or not (“FALSE”). The default is TRUE.

- 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’], server = “production”)

  The results are reviewed using techniques described in section 9.3.3 to investigate potential analysis issues, variations in uncertainty/MDL, and historical and nearby site comparisons. Reanalysis is requested when necessary/possible.

- Objective 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’], server = “production”)

Many of the functions described in this section (sections 9.3.2 and 9.3.3 specifically) 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 for initial validation:

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

- **output$flow_completeness –** flow.completeness
- **output$flow_status -** flow.status
- **output$elapsed_time -** etime.check
- **output$day_count –** daycount

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- **output$objective_code –** ObjCode.check
- **output$mass** - mf_mt.check
- **output$rcm -** mf_rcm.check
- **output$swap -** swap.check
- **output$QD -** status_check
- **output$validatsta_bad -** Validsta_BadData

9.3 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.

During Level 2 validation, if the user determines data requires further investigation, the filter status should be changed to 'QV' using the web app. If the initial status is not NM, the user is to add a comment using the web app to note both the initial and new statuses to track and document the changes. Once an initial review of the data is performed, a complete list of filters that need further investigation can be generated by specifying the QV filter status when running the *status_check* function as follows:

> *[month_QV] <- datvalIMPROVE::status_check(startdate = ['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], full_list = TRUE, status = "('QV')", server = "production")*

QV is a status to be used internally by UCD staff only and is not reported in any data delivery. Once investigations are complete, the filter status is updated to the appropriate status using the web app. The *status_check* function for QV is to be rerun before processing the data for delivery to check the dataset does not contain any QV statuses.

9.3.1 Cross-Module Comparison

9.3.1.1 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 an analytical error is suspected, a request is sent to the corresponding laboratories for a reanalysis of the sample.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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'], server ="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 (351C-1), 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).

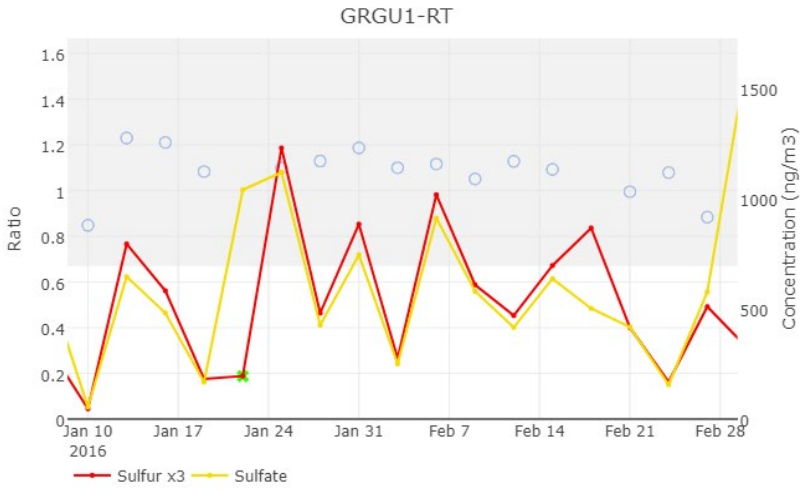
$$Index1 = \left(\frac{S3_1}{SO4_1} - 1\right) \times \left(\frac{S3_2}{SO4_2} - 1\right) \qquad Index2 = \left(\frac{S3_1}{SO4_2} - 1\right) \times \left(\frac{S3_2}{SO4_1} - 1\right) \qquad \text{(351C-1)}$$

If $PM_{2.5}$ sulfur is 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 (https://shiny.aqrc.ucdavis.edu/ImproveData/) are used to further investigate samples flagged as swap and/or outlier. Figure 2 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 an 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 2. S3/SO4 comparison plot for the GRGU1 site showing the 1/21/2016 sample pair as an outlier (green x).

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



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.

After identifying filters that need XRF reanalysis, create a reanalysis list with the first column as 'Filter ID' followed by the relevant columns such as Sampler name, Objective code, Sampling Date, Validation comments, and requested action. This list is then used to generate an XRF reanalysis sheet with Filter barcode and filter purpose information. To generate the XRF reanalysis request sheet from the reweigh list, the following function is used:

datvalIMPROVE::HIPS_XRF_sheet(input path = ['filepath.xlsx'], input_sheet =[NULL], output_path = ['filepath.xlsx'], output_sheet = ['XRF'], server = ['production'])

where *input_path* is the file path and file name of the reanalysis list, and *input_sheet* denotes the relevant sheet within the reanalysis list. If the *input _sheet* is not specified or set to NULL, the function will read the first sheet. The user can specify the name and location of the output file (*output_path*) as well as the sheet name (*output_sheet).* A typical command is shown below:

HIPS_XRF_sheet(input_path = " U:/IMPROVE/Data_Validation/Dec2020_ reanalysis All.xlsx", input_sheet = "XRF",
output_path = "U:/IMPROVE/Data_Validation/XRF_Dec2020_final.xlsx",
output_sheet = "XRF,
server = "production")

The same function can be used to generate a HIPS reanalysis list. The only difference is setting the *output_sheet* to 'HIPS'.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

The reanalysis list should then be sent to the appropriate analysis laboratory.

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 sections 9.2 and 9.4 of UCD IMPROVE TI #351A. 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 is changed using the QC review tool (https://improve.aqrc.ucdavis.edu/AnalysisData/Ions/IonsQcReview). 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 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.

9.3.1.2 1A Module versus 3C Module

The light absorption coefficient (fAbs) 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. Figure 3 shows an example comparison plot of fAbs (times 100), EC, and BC from the BOND1 site. Black carbon and fAbs are both optical measurements and are expected to compare well, whereas fAbs and EC are determined by different methods and may not be consistently comparable. If an analytical error in either measurement is suspected, other measurement data from the same module is examined to determine validity of the sample. If a replicate result is available for carbon,

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

compare the replicates against the original run. If the values do not meet the following criteria by DRI (Table 4), request a reanalysis for the sample.

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

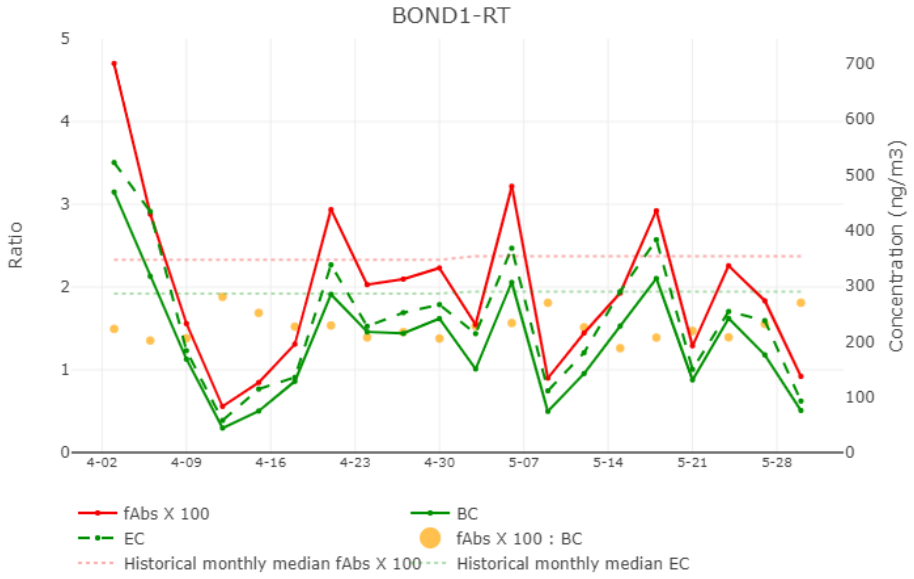


Table 3. Replicate criteria for carbon analysis results.

Range	**Acceptance Levels**
OC or TC < 10 µg/ cm2	< ±1.0 µg/cm2
OC or TC > 10 µg/ cm2	< 10 % of average of the 2 values
EC< 10 µg/ cm3	< ±2.0 µg/cm2
EC > 10 µg/ cm4	< 20 % of average of the 2 values

The relationship between EC, BC, and fAbs 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 the comparison of EC, BC, and fAbs results can be further investigated using qualitative checks and criteria to evaluate 3C Module carbon results (OC, EC, and TC) independently of fAbs (Table 5).

Table 4. 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.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

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 fAbs 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 using the ECTR scatter plot available on the early review tab. 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.

For HIPS reanalysis requests, the following points can be considered. Negative fAbs concentrations are a common observation during data validation. This could be due to various reasons; some frequently found scenarios are listed below.

- Due to issues with filter integrity like holes or tears. The SHL or analysis laboratories may have already added a comment to the filter about such filter integrity observations. If a comment exists but does not mention the location of the hole relative to the analysis area or if no comment has been applied, the data validator should confirm with the analysis laboratories if the hole/tear is in the analysis area and if the result can be considered valid. The data validator should proceed with further actions according to the laboratory's recommendations.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- Due to the nature of the deposit. In some cases, the nature of the deposit results in negative fAbs concentrations. The analysis laboratory may have added a comment noting observations of deposits with reflective nature. In any case, the data validator should reach out to the HIPS analysis laboratory and request an opinion on reanalysis.
- Multiple negative fAbs concentrations near coastal sites. If the data validator observes multiple negative fAbs concentrations at sites which are considered coastal sites, it is likely there was a high sea-salt component to the aerosol. Sea salts are mostly non-absorbing so if sea salts dominate the aerosol composition it is expected that lots of scattering occurs in HIPS and, therefore, negative fAbs results. In such cases, the data validator should review the Cl/Chloride plot and $PM_{2.5}$ vs. RCMN plot (Figures 5 and 6, respectively) on the data validation page of the IMPROVE shiny app to see if the sample composition has high chloride.

Once the HIPS reanalysis list is finalized, use the *datvalIMPROVE* function *HIPS_XRF_sheet* to generate the final HIPS reanalysis request list. The function usage is detailed in section 9.3.1.1. When reanalysis yields changes to results, further action is required:

- For fAbs 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 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 sections 9.1 and 9.4 of the TI #351A. 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].[dricarbon].[SampleAnalysis] production database table must be changed accordingly. This can be done using the QC review tool available at https://improve.aqrc.ucdavis.edu/AnalysisData/Carbons/CarbonsQcReview. 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

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

in the early review 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.

9.3.1.3 1A Module versus 4D Module

1A module $PM_{2.5}$ mass and 4D module PM_{10} mass are reviewed and compared (Figure 4). 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'], server = "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 (351C-2),

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

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 from the Sample Handling Laboratory; the pre-weight cannot be re-determined after sampling;
- Samples with invalid mass concentrations are flagged as "UN" (Undetermined Weight).

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

After identifying filters with a mass discrepancy, create a reweigh list containing the following columns (in the following order); Filter ID, Sampler, Objective Code, Sample Date, Module, Issue Type, Validation comments, and Requested action. This list is then used to generate a reweigh request sheet with various information the weigh lab requires including pre- and post-weight data and information regarding the balance used for weighing. To generate the reweigh request sheet from the reweigh list, the following function is used:

datvalIMPROVE::reweigh_sheet(inputpath = ['filepath.xlsx'], input_sheet =[NULL], output_path = ['filepath.xlsx'], output_sheet = ['Reweigh'], server = ['production'])

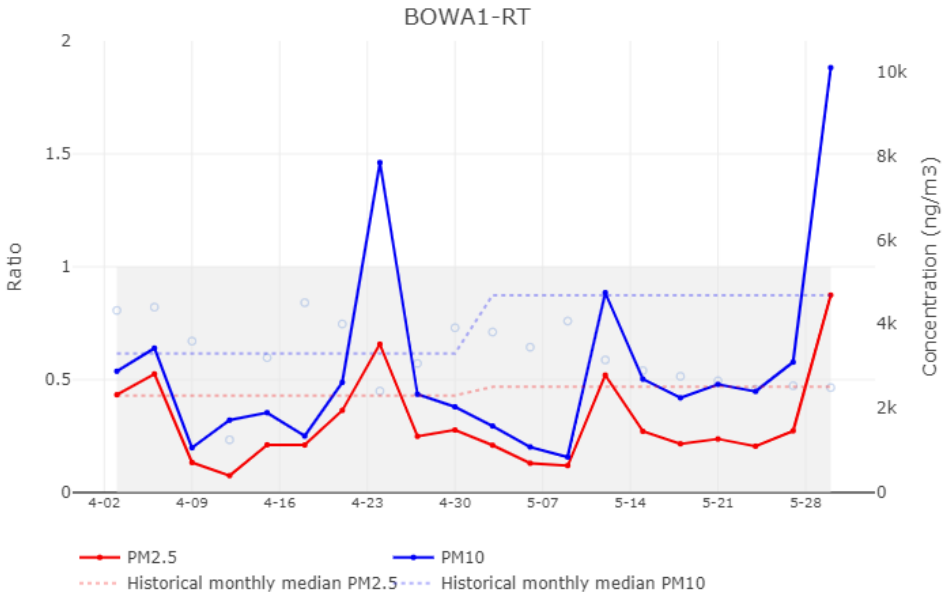
where *inputpath* is the file path and file name of the reweigh list and *input_sheet* denotes the relevant sheet within the reweigh list spreadsheet. The user can specify the name and location of the output file (*output_path*) as well as the sheet name (*output_sheet*), where the default sheet name is "Reweigh" if not specified. A typical command is shown below:

reweigh_sheet(input_path = "C:/IMPROVE_Reweigh_list_Feb2020.xlsx", input_sheet = "ReweighList", output_path = "C:/IMPROVE_Reweigh_list_Feb2020_final.xlsx", output_sheet = "Reweighlist_New", server = "production")

The generated reweigh request sheet is then sent to the weigh lab for cases to be assessed. When reweighing yields changes to results, the validation group reviews the reweigh results along with the weigh lab recommendations before requesting the weigh lab update the results, typically post-weight values, as necessary. Once the data are updated by the weigh lab, the validation group checks the early review plots to confirm the changes are as expected. In cases where results are either still questionable after reweighing or results did not change, due to questionable pre-weights for example, the filter status is updated to UN (Undetermined weight).

Figure 4. Time series plot of PM_{10} and $PM_{2.5}$ masses and their ratio at BOWA1 site.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



9.3.1.4 $PM_{2.5}$ Reconstructed Mass versus Gravimetric Mass

The $PM_{2.5}$ reconstructed masses, RCMC and RCMN, are calculated by equations 351-40 and 351-43 in *UCD IMPROVE TI #351B: Data Processing*. RCMC and RCMN are compared to the gravimetric mass (MF) as a check of measured components from the 1A, 2B, and 3C Modules (Figure 5). 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'], server = "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".

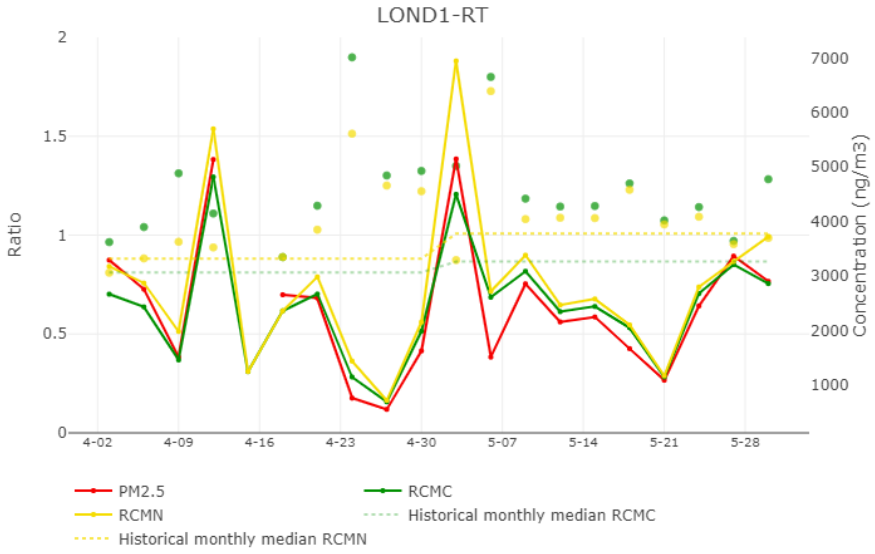
Z scores are calculated as follows:

$$RCMC\_Z\_score = 1.41 \times \frac{RCMC - PM_{2.5}}{\sqrt{(unc_{PM2.5})^2 + (unc_{RCMC})^2}} \qquad (351C\text{-}3)$$

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

$$RCMN\_Z\_score = 1.41 \times \frac{\mathrm{RCMN} - \mathrm{PM}_{2.5}}{\sqrt{(unc_{PM2.5})^2 + (unc_{RCMN})^2}} \quad \text{(351C-4)}$$

Figure 5. Time series plot of RCMC, RCMN, and $PM_{2.5}$ concentrations and their respective ratios at LOND1 site.

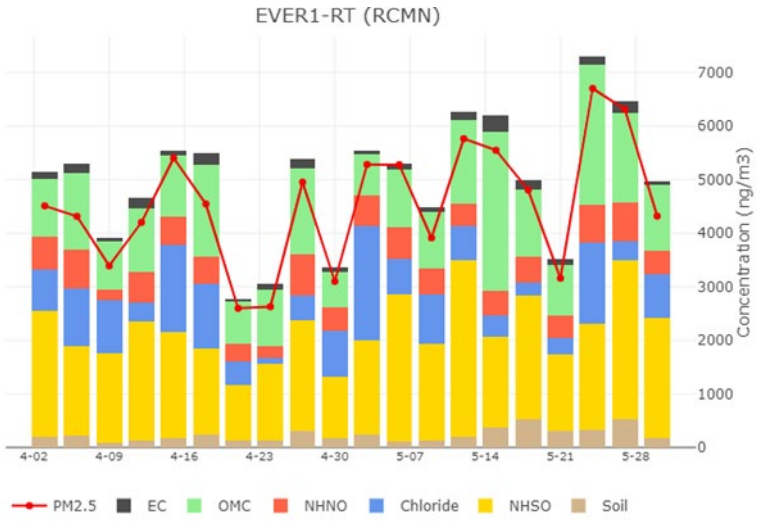


RCMN is also plotted as a bar plot (Figure 6), along with the $PM_{2.5}$ time series, for comparison of RCMN and $PM_{2.5}$ concentrations and to enable the contributions from the various species to be viewed and evaluated.

If $PM_{2.5}$ data is questionable, follow the steps outlined in section 9.3.1.3 to further investigate and identify the cause, including potentially requesting a reweigh.

Figure 6. Time series for RCMN versus Fine mass at EVER1 site.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



9.3.2 Long-Term Network-Wide Checks

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 7). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of chlorine versus chloride mass loadings for the whole network (Figure 8). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of fAbs versus BC (converted from TOR absorption measurements) for the whole network (Figure 9). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of fAbs versus EC for the whole network (Figure 10). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of OC versus EC for the whole network, (Figure 11). This plot is accessible from the IMPROVE Data site, “Early Review” tab.
- Scatter plot of all EC wavelengths for the whole network. (Figure 12). 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 13). 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 14 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.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 7. Scatter plot of sulfur (×3) versus sulfate for the entire IMPROVE network.

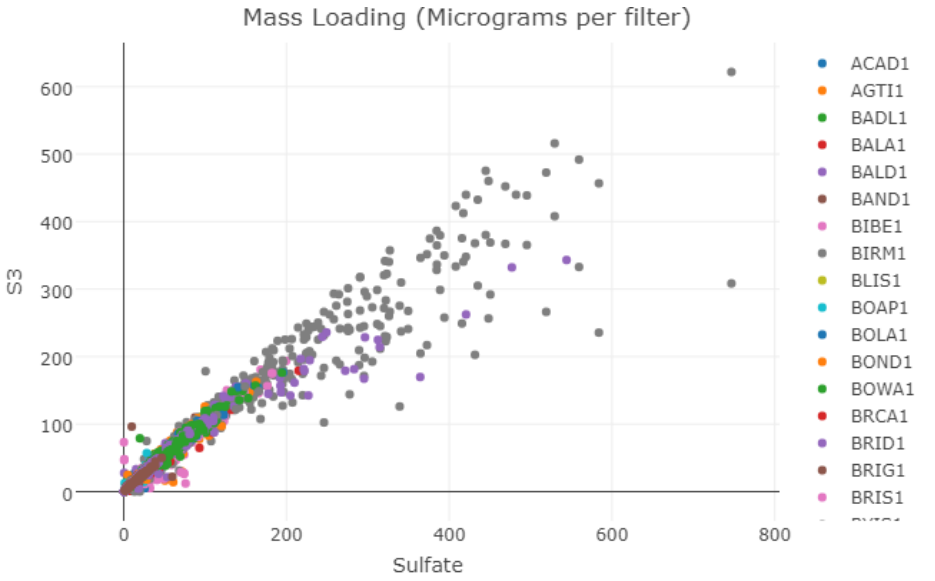
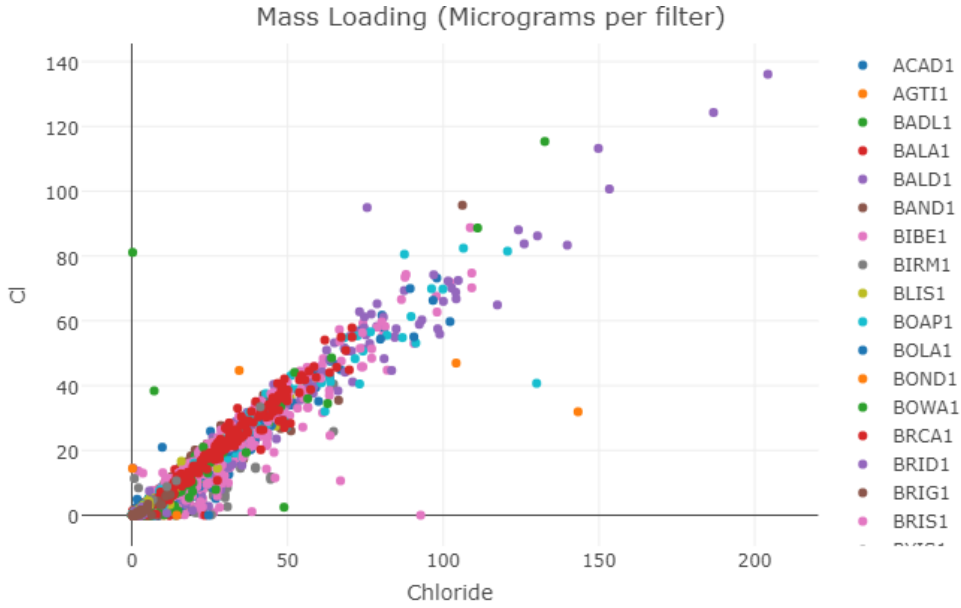


Figure 8. Scatter plot of chlorine versus chloride for the entire IMPROVE network.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 9. Scatter plot of chlorine versus chloride for the entire IMPROVE network.

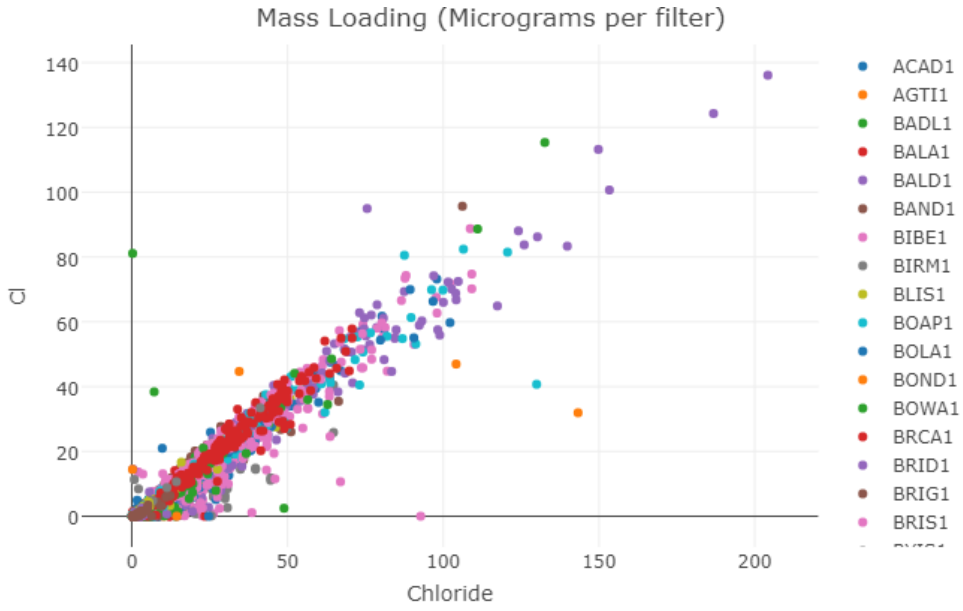
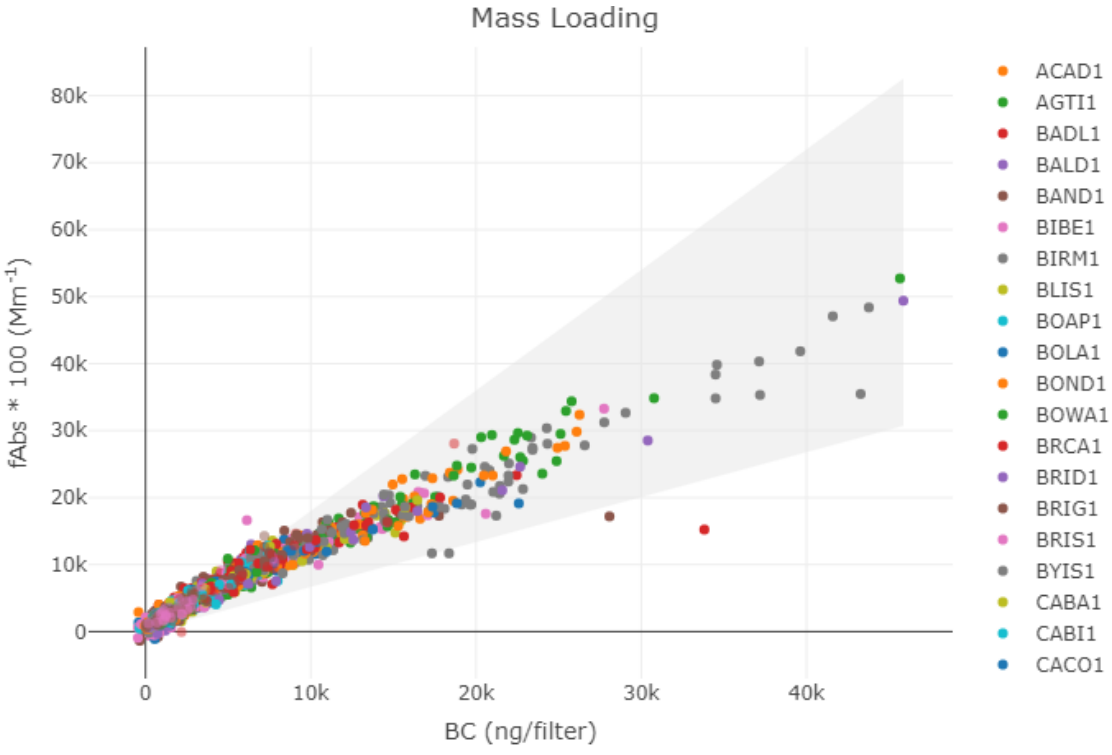


Figure 10. Scatter plot of fAbs versus BC for the whole network.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 11. Scatter plot of fAbs versus EC for the whole network.

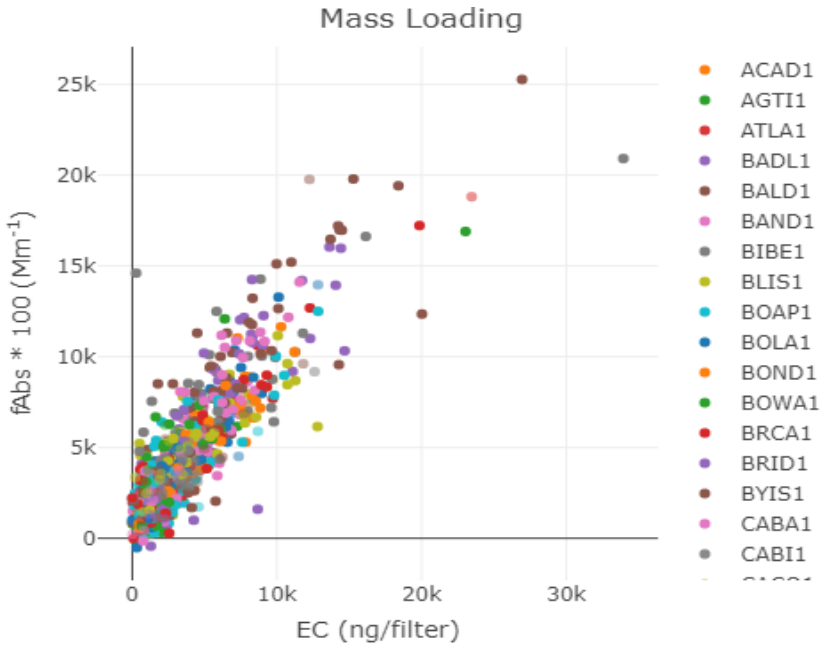
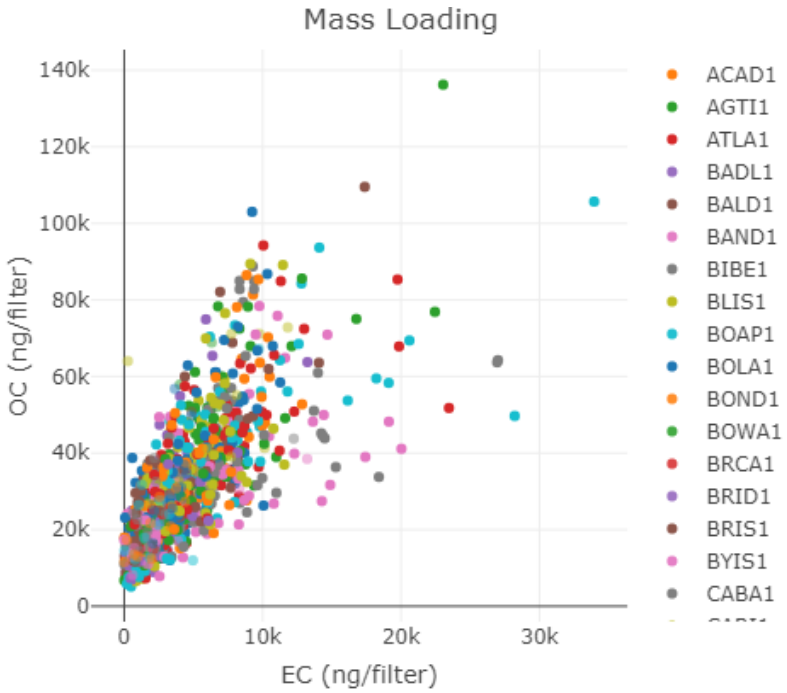


Figure 12. Scatter plot of fAbs versus EC for the whole network.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 13. Scatter plot of ECTR versus other wavelengths for the whole network.

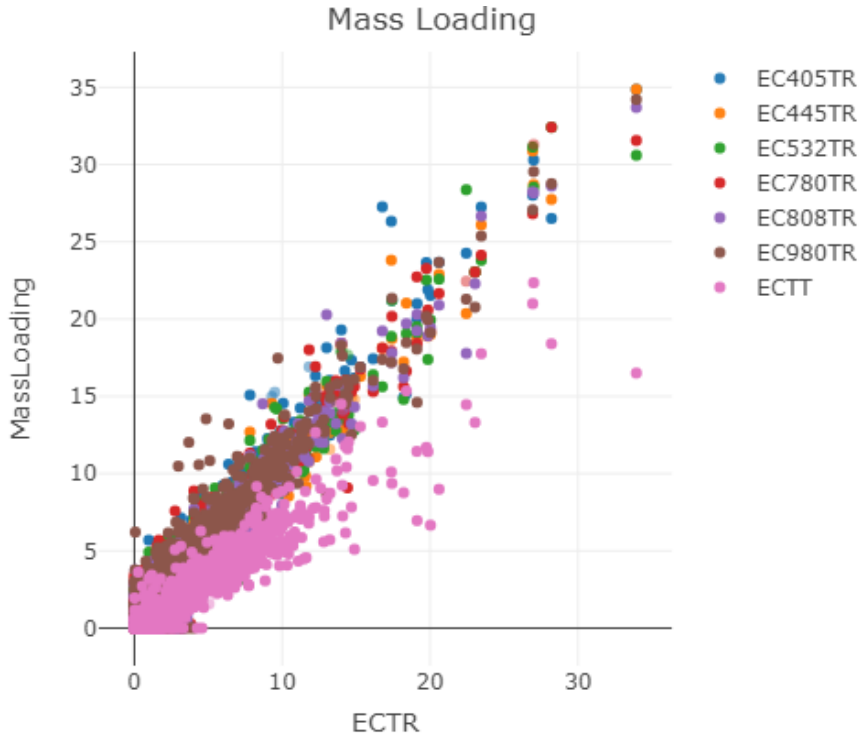
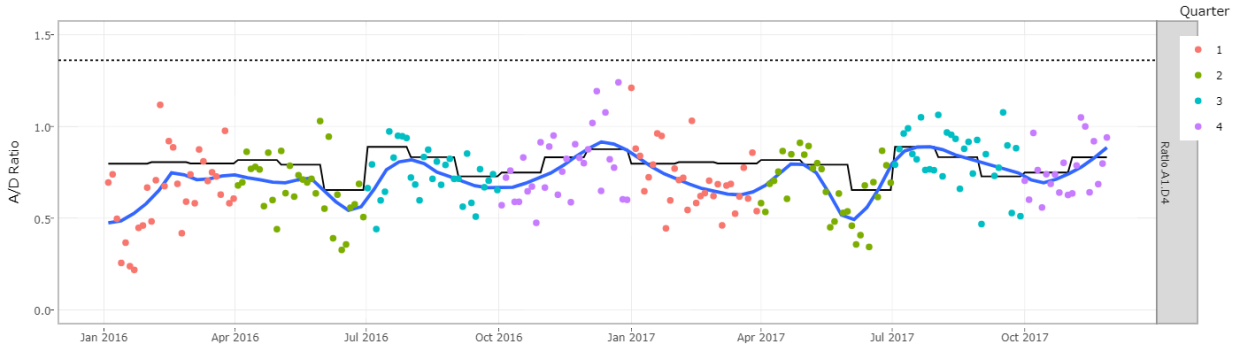
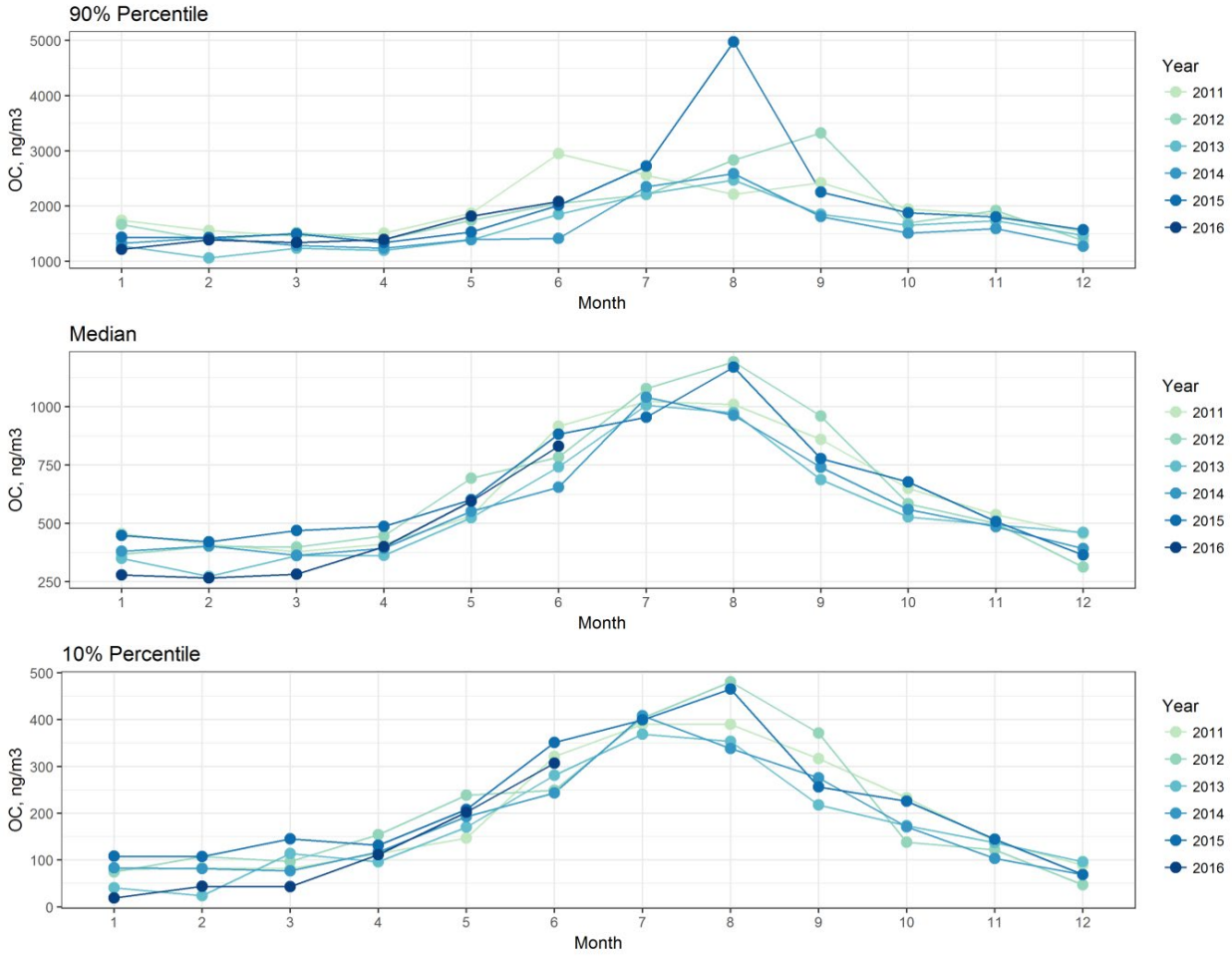


Figure 14. Ratio of $PM_{2.5}$ mass (1A) over PM_{10} mass (4D) at ACAD1 site, represented as raw measurements not adjusted for flow rates. Points are individual sample days (pink = Q1, green = Q2, blue = Q3, purple = Q4). Black line is the multi-year monthly mean. Blue line is the locally weighted average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



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 15. Multi-year monthly 10% percentile (top), median (middle) and 90% percentile (bottom) of organic carbon (OC) concentrations (in ng/m3) for the whole IMPROVE network from 2011 to 2016.



9.3.3 Common Validation Findings

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

9.3.3.1 Filter & Analysis Data Swaps

There are several types of swaps in terms of the filter purposes involved and at what point in the process the swap occurred. Swaps are addressed using the swap tool in the web app (https://improve.aqrc.ucdavis.edu/Swap).

Filter Swaps

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

These types of swaps occurred before sampling (all downstream data are swapped, including flow data and all analysis associated with the filters); also referred to as cartridge position swap. Examples of filter swaps include:

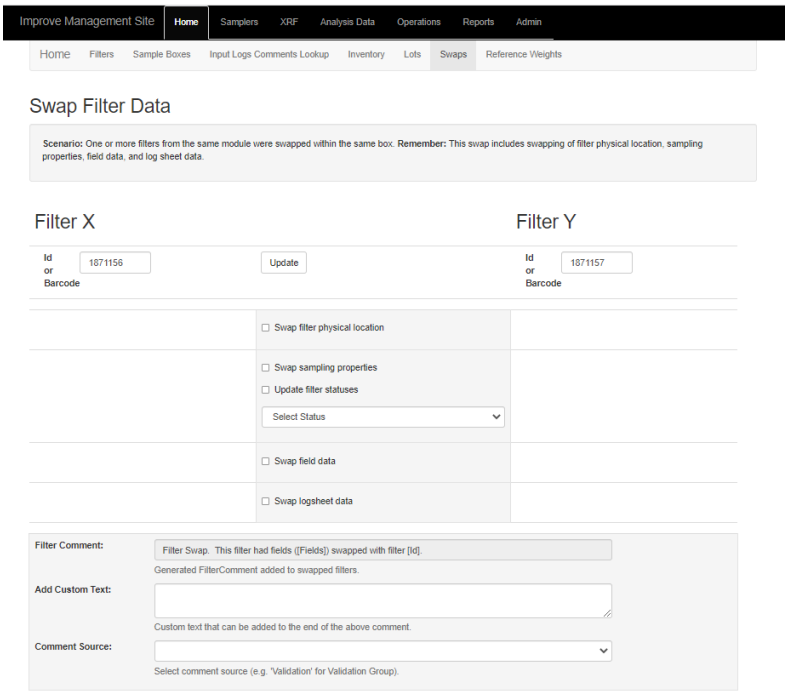
- A routine sample filter was swapped with a field blank filter.
- A routine sample filter was swapped with a collocated sample filter.
- One or more of the same module filters were swapped within the same box (sample date swap).
- A 1A filter was swapped with a 4D filter (uncommon).
- The cartridge was installed incorrectly (rotated clockwise or counterclockwise), and one or more filters sampled on the incorrect day.

For these types of swaps, all data fields are to be swapped relating to the cartridge position between the relevant filters, including filter position properties (Cartridge Position) and log sheet records. Field data also needs to be swapped, specifically flow data. To perform the swap of all of these fields, use the Filter option in the swap tool and follow the steps below:

1. Access the filter swap tool found at https://improve.aqrc.ucdavis.edu/Swap/Filters. The resulting swap page has fields to enter the Filter Id/Barcode of the filters that need to be swapped, where only PTFE filters have barcodes. Enter the filter IDs/Barcode in Filter X and Filter Y fields (Figure 16) and click on the ‘Update’ button. Filter details such as Filter Properties, Physical location, Sampling Properties, Field data (e.g., flow), Log Sheet data, and Analysis data will then show under the relevant filters.
2. Review data shown is as expected.

Figure 16. Filter swap page.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



3. There are four fields available for swapping: filter physical location, sampling properties (including filter purpose), field data, and log sheet data. In the case of a field blank-sample (FB-SA) swap, there is only one filter with flow information; the flow information is assigned to the wrong Filter ID. Also, there is an option to update the filter status. Select ‘SW’ for sample-sample swap and ‘SP’ for sample-field blank swap from the dropdown menu. If one of the filter statuses is invalid, this option should be left unselected. In such cases, update the filter statuses as described in step 5. For all types of filter swaps, select all four fields to be swapped. A comment including the information of filter details swapped is added automatically when a swap is conducted and can be reviewed in the ‘Filter Comment’ section. Use the ‘Add Custom Text’ section to add more details on the nature of the swap. Select ‘Validation’ as the ‘Comment Source’. Click ‘Swap Data’ to do the swap.
4. Check to ensure that the swap was performed by reviewing data in the Early Review tab. The Early Review tab shows data in mass loading from the analysis table; changes are reflected here without data needing to be reprocessed first. If the swap involved a FB filter, also review the Field Blank tab.
5. Using the filter details page in the web app (https://improve.aqrc.ucdavis.edu/Filters/Details), change the filter status to ‘SP – Field Blank/Sample Swap’ for both filters involved in a FB-SA

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

swap and to 'SW – Swapped Sample Dates' for all filters involved in sample-only swaps.

6. After all edits are performed and data is ready to be prepared for delivery, reprocess flows following the steps outlined in *UCD IMPROVE TI #351B: Data Processing*.
7. Reprocess concentrations following the steps outlined in TI #351B.
8. Review the final data in the Validation plots and Field Blank tab.

Analysis Swaps

These swaps occurred after sampling, before all analyses are complete (flow data are confirmed to not be impacted, analysis data are swapped). Swaps can occur between sample-sample (SA-SA) filters or field blank-sample (FB-SA) filters.

To perform the swap, use the Analyses option in the swap tool and follow the steps below:

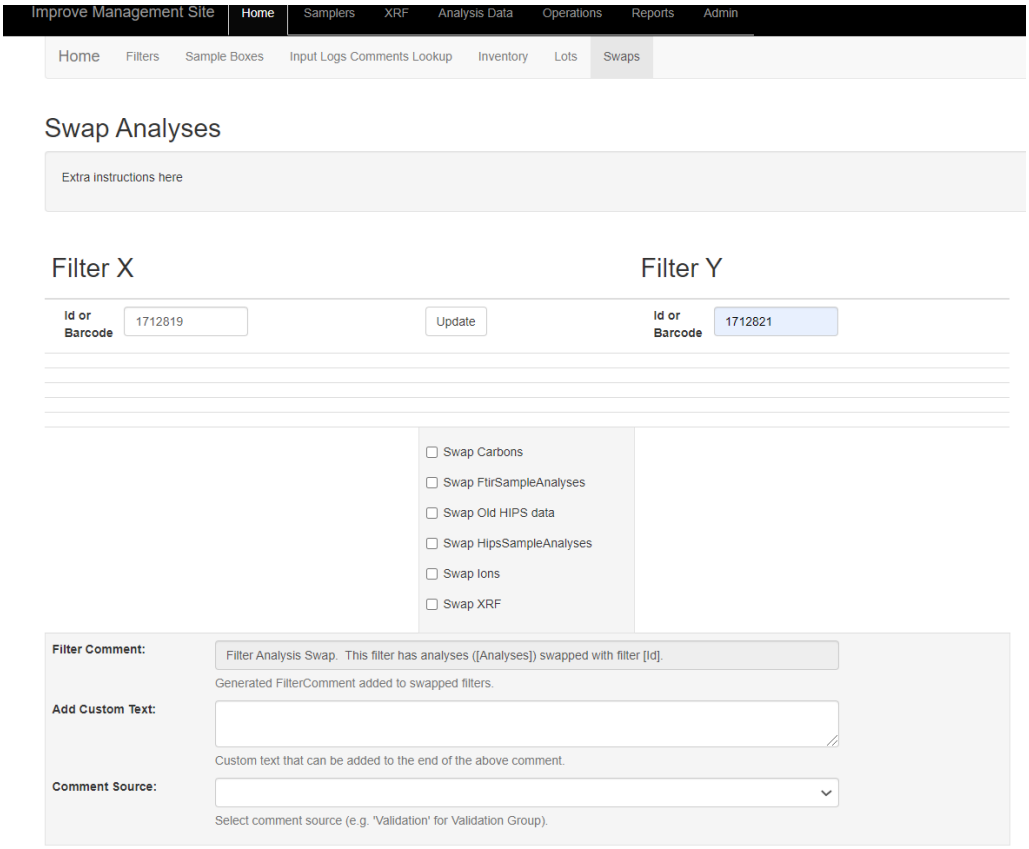
1. Confirm the swap happened for the same module and identify which analysis data are swapped; if multiple analyses are performed on that filter, which is the case for A module filters, identify which sets of analyses have been swapped. Usually, if the filters are swapped at a lab station, all downstream analyses will be swapped.
2. Access the analyses swap tool found at https://improve.aqrc.ucdavis.edu/Swap/Analyses. The resulting swap page has fields to enter the Filter Id/Barcode of the filters that need to be swapped, where only PTFE filters have barcodes. Enter the filter IDs/Barcode in Filter X and Filter Y fields (Figure 17) and click on the 'Update' button. Filter details such as Filter Properties, Physical location, Sampling Properties, Field data (e.g., flow), Log Sheet data, and Analysis data will then show under the relevant filters.
3. Review data shown is as expected.
4. The following fields are available for the swap: Carbons, FtirSampleAnalyses, Old HIPS data (for filters before database change in March 2020), HipsSampleAnalyses, Ions, and XRF. Depending on the filter type and swap point (in the case of A module filters), select the appropriate fields. This is particularly relevant for A module filters where multiple analyses are performed: gravimetric, FTIR, XRF, and HIPS analysis. Be sure to determine after which analysis the swap occurred and only swap the downstream data from that point. For example, if the sample was swapped after gravimetric analysis while placing the filter in a Petri dish, then the FTIR, XRF, and HIPS analysis data will need to be swapped. If the swap occurred after XRF analysis

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

but before HIPS analysis, only HIPS data need to be swapped. The swap tool does not have the option to swap gravimetric mass data; such a swap is unlikely if the filter was weighed in the automated weighing chamber. If the filter was swapped before gravimetric analysis and the filter was weighed on a manual balance, please ask the weigh lab to swap the relevant data. A comment including the information of filter details swapped is added automatically when a swap is conducted and can be reviewed in the 'Filter Comment' section. Use the 'Add Custom Text' section to add more details on the nature of the swap. Select 'Validation' as the 'Comment Source'. Click 'Swap Data' to do the swap.

Figure 17. Analysis swap page.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



5. Inform the relevant analysis labs about the swaps performed.
6. Check to ensure that the swap was performed by reviewing data in the Early Review tab. The Early Review tab shows data in mass loading from the analysis table; changes are reflected here without data needing to be reprocessed first. If the swap involved a FB filter, also review the Field Blank tab.
7. Using the filter details page in the web app (https://improve.aqrc.ucdavis.edu/Filters/Details), change the filter status to ‘SP – Field Blank/Sample Swap’ for both filters involved in a FB-SA swap and to ‘SW – Swapped Sample Dates’ for all filters involved in sample-only swaps.
8. After all edits are performed and data is ready to be prepared for delivery, reprocess flows following the steps outlined in *UCD IMPROVE TI #351B: Data Processing*.
9. Reprocess concentrations following the steps outlined in TI #351B.
10. Review the final data in the Validation plots and Field Blank tab.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

9.3.3.2 Cartridge Swaps

When a cartridge designated for a particular week is set up incorrectly to run in another week or another module, multiple cartridges are likely involved. Examples of cartridge swaps include:

1. A site came back online but does not have a new box:
    a. One or two weeks of an old unused box are used in place of the current box (most common scenario).
2. A site came back online but did not have a new box at the moment. The old Week 3 unused filters were used in place of the current Week 1. A new box was generated and sent out. In the new box:
    a. Week 1 filters were never used
    b. Week 2 & Week 3 sampled correctly
3. Weeks are used in the incorrect order:
    a. Example: Cartridge in the Week 3 bag is used instead of that in the Week 1 bag.
4. Cartridges are input into any wrong module.
    a. This scenario is only possible when an A module cartridge is placed in a D module (as all are PTFE filters) and vice versa.

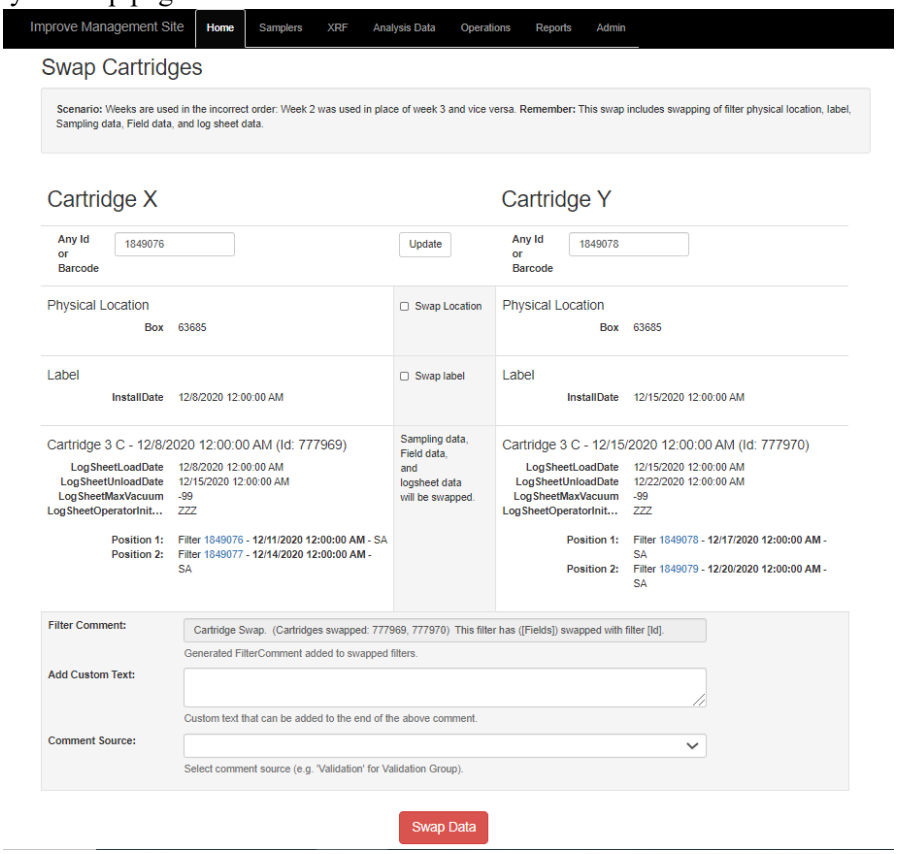
To perform the swap using the swap tool for cartridges, each cartridge pair swap will have to be performed one at a time. A cartridge swap can be performed only if both cartridges have the same number of filters, except for cartridges with field blanks.

- **Cartridge Swaps:** Same module swap or A-D module swap. Resolve by following these steps:
    1. Access the cartridge swap tool found at https://improve.aqrc.ucdavis.edu/Swap/Cartridges. The resulting swap page has fields to add one filter Id/Barcode (only PTFE filters have barcodes) from each cartridge or the cartridge ID that needs to be swapped. Enter the relevant Ids/barcodes in the Cartridge X and Cartridge Y fields and click on the ‘Update’ button. The filter details like Sampling data, Field data, and log sheet data will be shown under the relevant filters/cartridges (Figure 18).
    2. Review data shown is as expected.
    3. The following fields are available for the swap; Label and Location. Select both fields. A comment including information and details of the filter(s) and cartridge(s) swapped is added automatically when a swap is conducted and can be reviewed in the ‘Filter Comment’ section. Use the ‘Add Custom Text’ section to add more details on the nature of the

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

swap. Select ‘Validation’ as the ‘Comment Source’. Click ‘Swap Data’ to do the swap.

Figure 18. Analysis swap page.



4. Check to ensure that the swap was performed by reviewing data in the Early Review tab. The Early Review tab shows data in mass loading from the analysis table; changes are reflected here without data needing to be reprocessed first.
5. Using the filter details page in the web app (https://improve.aqrc.ucdavis.edu/Filters/Details), change the filter status to ‘SW – Swapped Sample Dates’ for all filters involved in sample-only swaps.
6. After all edits are performed and data is ready to be prepared for delivery, reprocess flows following the steps outlined in TI #351B.
7. Reprocess concentrations following the steps outlined in TI #351B.
8. Review the final data in the Validation plots tab.

9.3.3.3 Box Swaps

Swapping filters from entire boxes is sometimes necessary. A box swap becomes necessary when:

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

a) Box X was lost/not delivered, so Box Y of a future cycle was used.
b) Box X was unused from an old cycle; it was used in place of Box Y of a future cycle.
c) Box X was lost/not delivered, so Box Y was assembled for the exact sampling dates but not processed through the Improve database.
d) Box X was assembled and processed through the database but never used/shipped out because the site was offline.
e) Box X was sent to the wrong site and sampled fully in the incorrect site. This usually happens with the same cycle of boxes, but an instance could occur where a 2-3-2 box samples in place of a 3-2-2 box and vice versa. In that case this procedure will not work.

Note that the cartridges have to line up for this to work i.e., if any other swaps occurred within the box, this procedure will not work. In those cases, the procedure is to do a cartridge swap for each pair of cartridges. For example, if week one from the original box was sampled and weeks two and three from the new box were sampled then using this box swap tool is not an option.

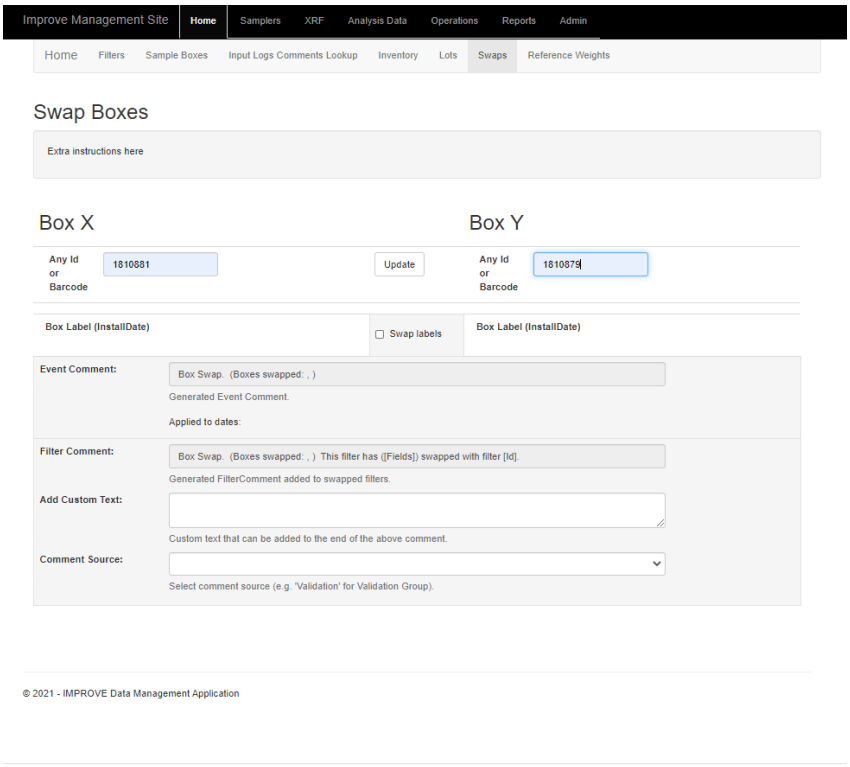
For all of the above examples, the swaps can be performed using the steps outlined below.

1. **Step 1:** Access the box swap tool found at https://improve.aqrc.ucdavis.edu/Swap/Boxes. The resulting swap page has fields to enter one filter Id/Barcode (only PTFE filters have barcodes) from each box or the Box Id that needs to be swapped. Only filter Ids/Barcodes or Box Ids are to be entered here; Cartridge Ids are not to be entered. Enter the relevant Ids/Barcodes in the Box X and Box Y fields (Figure 19) and click on the 'Update' button. All the box properties and cartridge/filter details will be displayed under the box Id fields (Figure 20).
2. **Step 2:** Review data shown is as expected for the boxes, cartridges, and filters. Also, compare and make sure all details match between the boxes (such as 2-3-2 vs. 2-3-2).
3. **Step 3:** Only the Box Label (Install Date) field is available for the swap. Select this field. A comment including the information of filter details swapped is added automatically when a swap is conducted and can be reviewed in the 'Filter Comment' section. Use the 'Add Custom Text' section to add more details on the nature of the swap. Select 'Validation' as the 'Comment Source'. Click 'Swap Data' to do the swap.
4. **Step 4:** Sometimes multiple box swaps need to be performed to address the issue; repeat steps 1-3 for each pair of boxes. In each case, the type of box swap scenario should be assessed to determine which box pairs, if any, are to be swapped.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

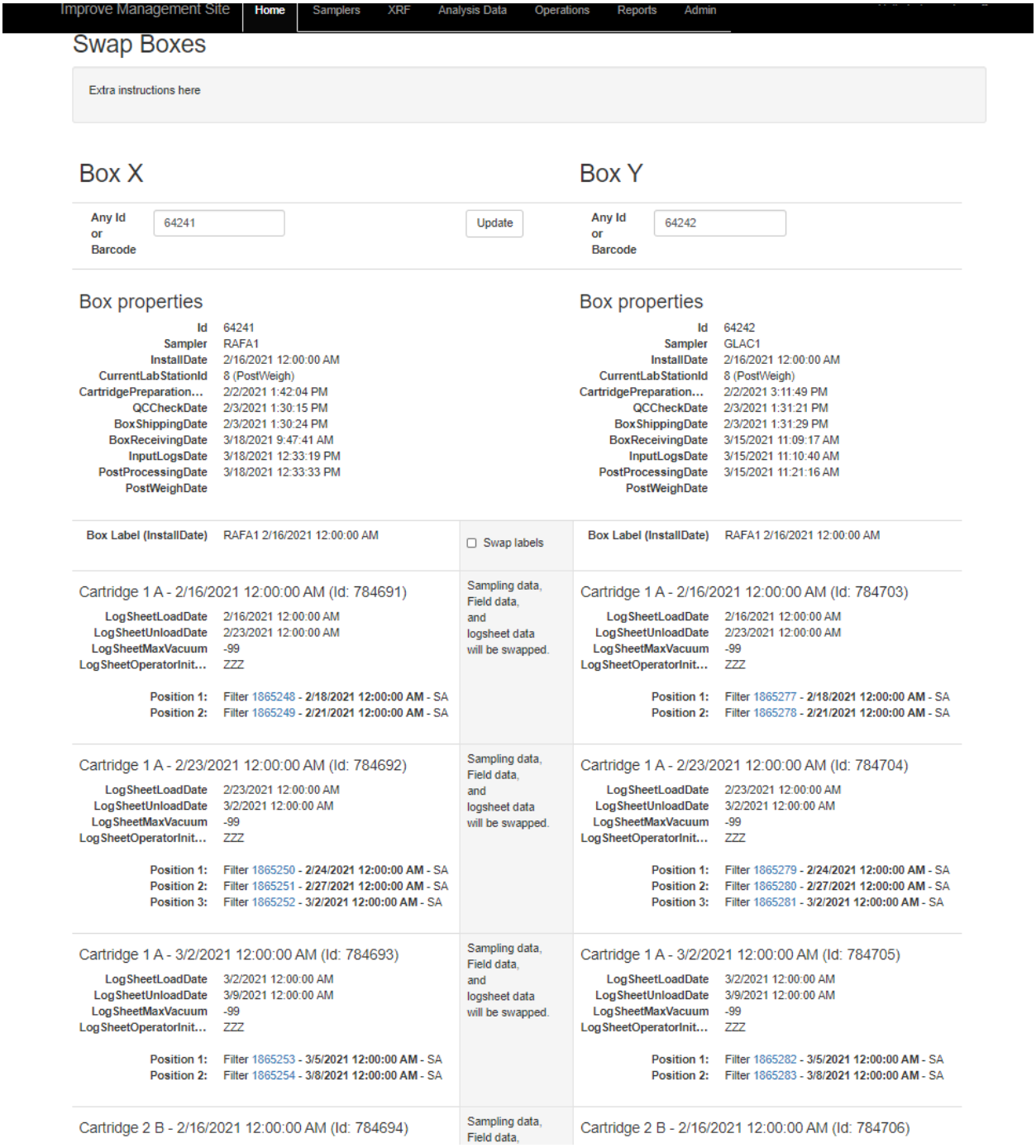
5. **Step 5**: After a box swap is performed, the statuses of all filters in the boxes need to be addressed based on the swap situation. If the box is swapped in place of a lost or undelivered box (example 'a' in the above section), please refer to section 9.1.2 of *UCD IMPROVE TI #351F: Data Preparation and Reporting* to update the filter purpose and current lab station Id of the lost/undelivered box. In cases where a box is swapped with another site (example 'e'), all filter statuses in both boxes need to be updated to 'SW – Swapped Sample Dates' using the filter details page in the web app (https://improve.aqrc.ucdavis.edu/Filters/Details).

Figure 19. Box swap page.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 20. Box swap page after clicking the 'Update' button. Only A module filter details are displayed due to page length.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

9.3.3.4 Sampling Anomalies and Questionable Data

There are several types of sampling anomalies and questionable data commonly observed during validation. Included here are guidelines for addressing and resolving these issues. Note that the NPS treats the SA (sampling anomaly) flag as terminal for Regional Haze Rule purposes; consider the application of the SA flag carefully and apply alternative flags where appropriate. For cases where there is a non-standard sampling but no noticeable data bias a flag other than SA may be used. If a site audit finds any sampling issues, then the SA flag may be appropriate.

- **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. Previous cases have been flagged SA (sampling anomaly) to indicate an operational deviation when the cross-module concentration data agreed.
  - For current cases, review the relevant concentration data and compare with results from other modules. If the cross-module results agree, consider changing the status to NM (normal) or 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 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. However, the SA flag will have been treated as invalid for Regional Haze Rule purposes.
    - 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. The final reported

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

data will have been treated as invalid for Regional Haze Rule purposes, however.

- **Possible manifold open / cartridges not seated correctly**
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. 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 NM, the LF flow status flag would have been reported to end users. If the flow status had been LF and the filter status was SA, the SA flag would have been 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. If the double filter issue is not identified until the filters are in the carbon analysis lab, the analysis lab analyzed the top filter and adds a comment noting the situation.
  - Previous cases may have been flagged SA (sampling anomaly) to indicate an operational deviation when the cross-module concentration data agreed. For current cases, review the relevant concentration data and compare with results from other modules. If the cross-module

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

results agree, consider changing the status to NM (normal) or 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 appropriate comment are applied to the filter to indicate possible impact to the analysis results.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- 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.

- **Problems 1A: Particles (A only)**
  - Typically applies to 1A filter
  - Typically flagged QD by the Sample Handling Laboratory with comment applied. In some cases, the Sample Handling Laboratory is able to remove the debris and reweigh the filter. The QD flag and an appropriate comment are applied to the filter to indicate the possible impact on the analysis results.
  - Review all A module analysis data such as those used in cross-module validation to check if any particular analysis value is elevated or lower than expected or in comparison to the cross-module species. In such cases, reach out to the relevant lab to see if the particle was removed or not before analysis. If the analysis data does not compare with other modules and the particle can not be removed, the filter status will be updated to XX.

- **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.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- **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.
  - Review 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 1 and Table 2 of *UCD IMPROVE TI #351E: Flow Validation*), 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.
- **Egregious sulfur/sulfate discrepancy and corresponding factors**

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- During data validation, the following observtaions may be made for a sample date at a site:
  - Large discrepancy between sulfur and sulfate concentrations, whereby sulfate is higher than sulfur, the 3*sulfur/sulfate ratio is shown to be an outlier, and the respective uncertainties do not overlap;
  - RCMN is higher than $PM_{2.5}$;
  - total sample concentration (RCMN) is high; and
  - the nitrate component is large.
- If such an observation is made a spot check reanalysis of both 'A' and 'B' filters is performed. If there are many sample dates at a single site and/or if there are many samples from many different sites that all meet this criteria, the analyst will identify a subset of the worst cases and request reanalysis of both 'A' and 'B' filters.
- If the reanalysis results do not show any issues with analysis, the data is reviewed again to rule out other potential sampling issues.
- If a collocated CSN site is available, the sulfur and sulfate concentrations should be compared between the two networks. If there are any discrepancies between the sulfur and/or sulfate concentrations from the IMPROVE samples with the CSN samples, the relevant IMPROVE filter should be invalidated using 'XX' (Sample Destroyed, Damaged, or Contaminated) status. If a collocated CSN site is not available and if there are no other issues than the above four criteria, the filter status can be changed to 'NM' (Normal).

For all cases identified, appropriate comments should be added to acknowledge the issue and detail any actions taken.

9.3.4 Analysis Level Flagging Validation Guidelines

If a particular analysis or analysis parameter is questionable, it can be invalidated without invalidating the whole filter. Some scenarios where this action is needed, but not limited to, include:

- Filter is damaged or destroyed between analyses, such as in the case of the A module filter. Remaining analyses can be invalidated.
- A/D module pre mass was incorrect, resulting in incorrect mass data. Other analyses performed on the same filter (e.g., XRF, HIPS of the A module filter) are good. Mass results can be invalidated, leaving other analyses as valid.
- Filter has a hole within the analysis area, impacting particular analyses e.g., for the A module filter, the hole does not impact mass analysis or XRF analysis but HIPS analysis is affected by the hole. The HIPS results can be invalidated, leaving other analyses as valid.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- Chloride contamination during IC analysis is suspected. All other parameters compare well with other modules/nearby sites. Just the chloride parameter can be invalidated, leaving all other ions as valid.

9.3.4.1 Applying a Flag

- Open the web application and search for the filter that needs flagging at https://improve.aqrc.ucdavis.edu/Filters.
- On the filter details page (Figure 21), click on the relevant analysis on the 'Analysis Data' box. For example, to flag gravimetric analysis, select 'GRAV.'
- Selecting the analysis type will lead to the analysis data page. Examples of GRAV, XRF, HIPS, Ions, and Carbon analysis data pages are in figures 22, 23, 24, 25, and 26, respectively.
- The GRAV (mass) analysis data page is the page where flags can be applied, where there is the option to apply flags to either the pre-weight or post-weight analysis. Only one of these weigh types needs to have the relevant flag applied.
- To navigate to the page for applying flags to:
  a. XRF data, click the sample details button to the left of the Filter Id of the relevant filter record.
  b. For HIPS data, click the 'Details' button on the right-hand side of the relevant filter record.
  c. For Ions data, click the ‘Details’ button on the left side of the FilterId of the relevant filter record.
  d. For Carbon data, select the relevant Id under 'Carbon Runs'. If a filter has replicate or reanalysis results, there will be multiple Ids; each Id must be selected separately to add flags.

- For GRAV analysis, the ‘Edit Flags’ button can be selected under the relevant weigh type (‘PREWEIGH’ or ‘POSTWEIGH’) on the analysis details page. For all other analyses, click on the ‘Edit Flags’ button on the filter analysis details page. Figure 27 shows an example of the HIPS details page. This step is the same for XRF, Ions, and Carbon analysis.
  - The next window is the analysis flags page. Figure 28 shows an example of the XRF Analysis Flags page.
  - All analysis types have a section title 'Add New Code', which contains the fields: Status, Comment Source, Comment, Shortcuts, and Parameters (Figure 28).
    a. The Status field has a drop-down list of available analysis statuses. Select the status that is appropriate for the analysis.
    b. A custom comment can be added to the filter using the Comment field. When terminal statuses are applied, a comment is required, while for informational statuses adding a comment is optional. However, it is preferred that a comment is added summarizing the decision for flagging the analysis, whether the status is informational or terminal.
    c. If a custom comment is added, select the relevant group from the Comment Source drop-down menu e.g., select Validation.
    d. To select the parameters needing to be flagged, either use the Shortcuts buttons to flag all parameters or other predefined groups of parameters, where available

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

(e.g., 'CrContam is an option available for XRF analysis) or select individual species by checking the boxes next to relevant parameters in the Parameter list. Once all relevant parameters are selected, click on the 'Add Selected' button to finish the flagging.

- An automated comment with details of the analysis flag, affected parameters, and analysis Id, along with a custom comment, if added, will be added to the filter details page. The comments will be under the respective analysis tab.
  - The processing code does not currently support multiple analysis flags. If the filter already has an analysis flag, contact the user who added the flag and determine which flag is more appropriate. To delete the unwanted flag, go back to the analysis flags page and click on the 'Delete' button (red) under the relevant flag (Figure 29).

Figure 21. Filter details page.

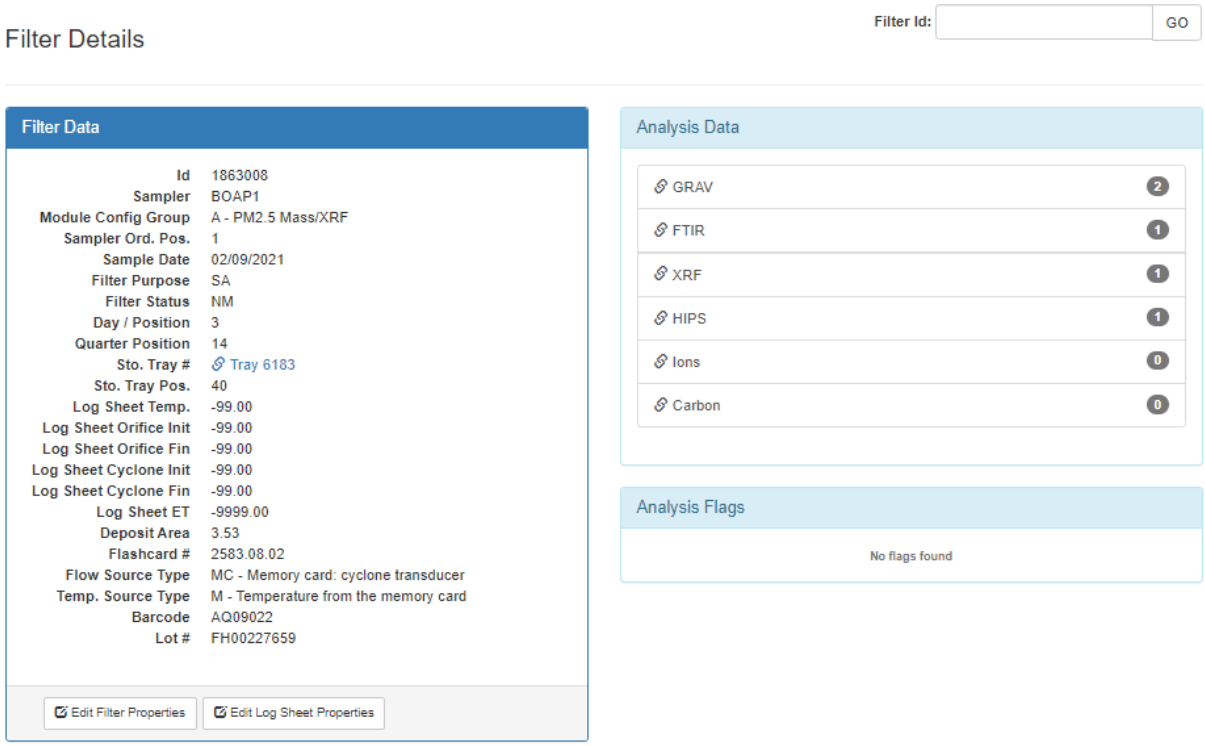
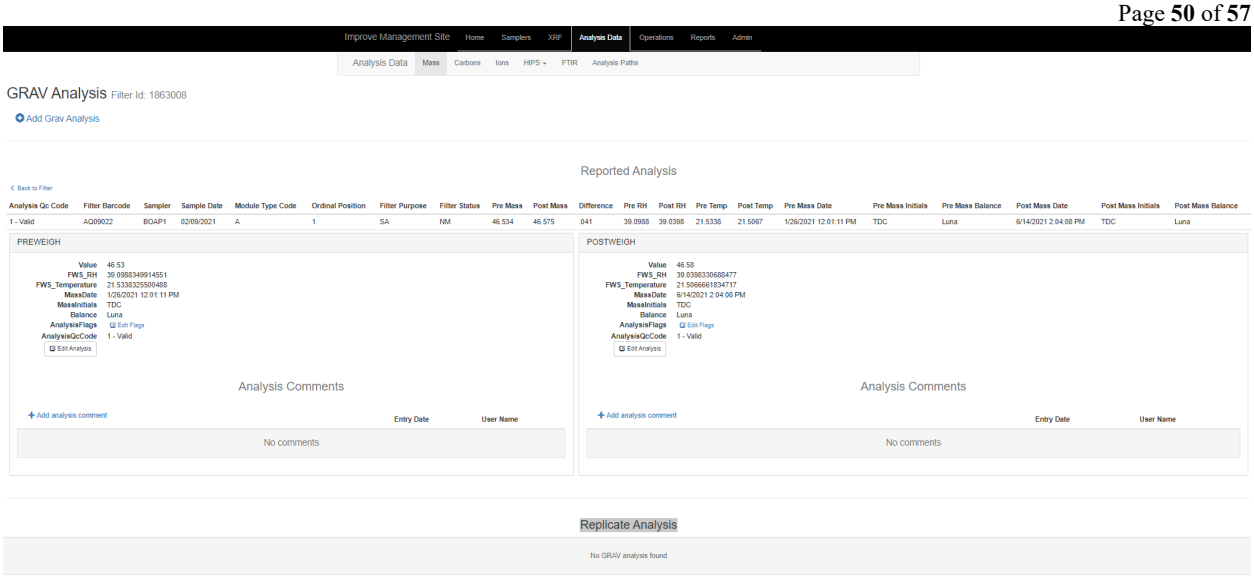


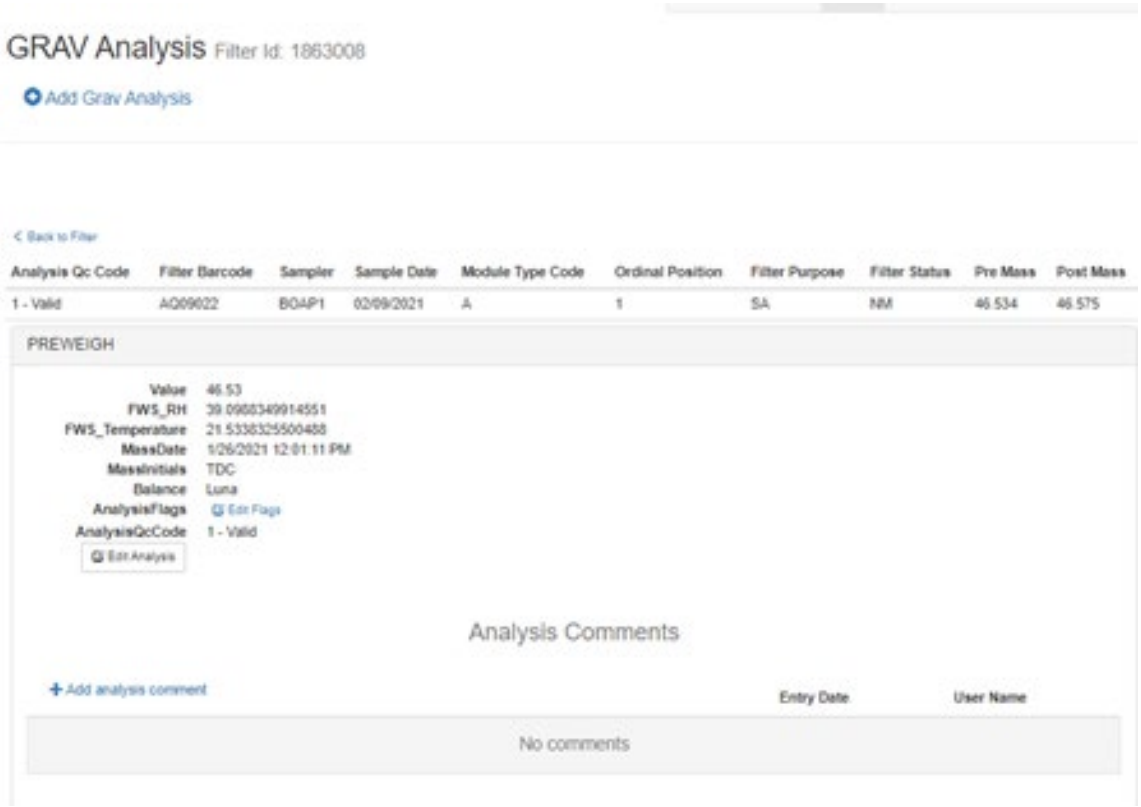
Figure 22. GRAV (Filter Mass; a) analysis page and detail (b, c).

a)

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE



b)



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

c)

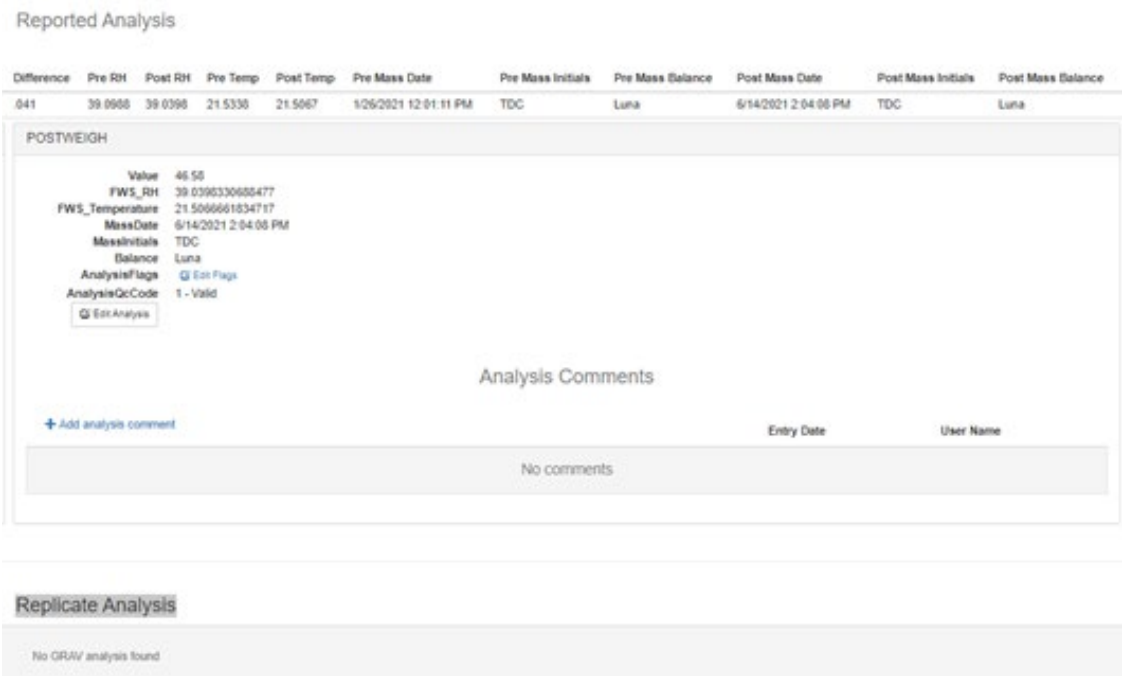


Figure 23. XRF sample analysis page.

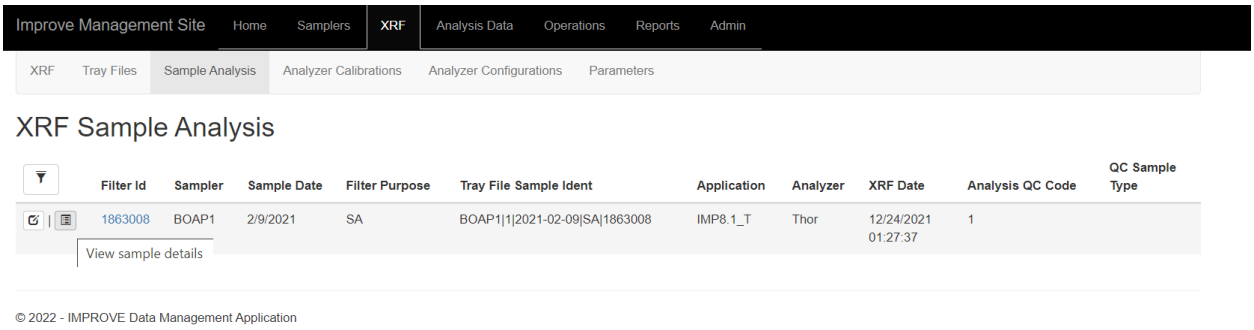


Figure 24. HIPS sample analysis page.

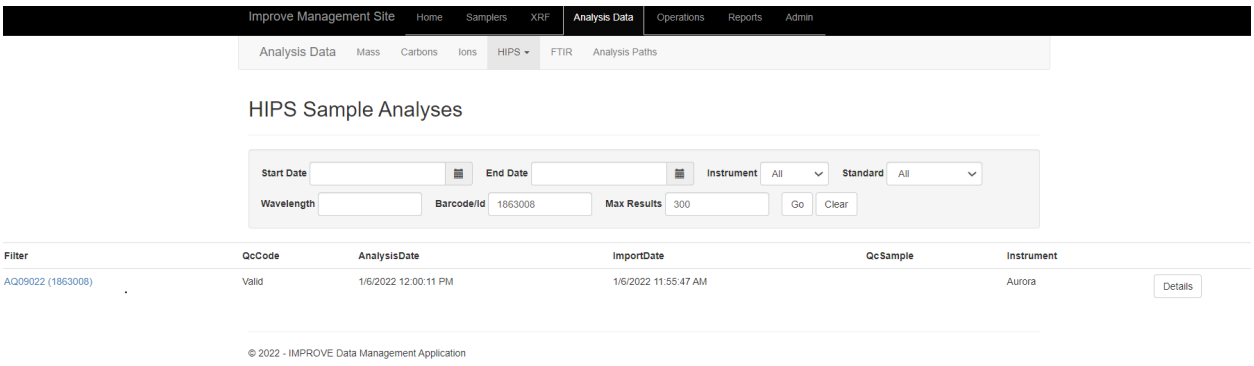


Figure 25. Ions sample analysis page.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

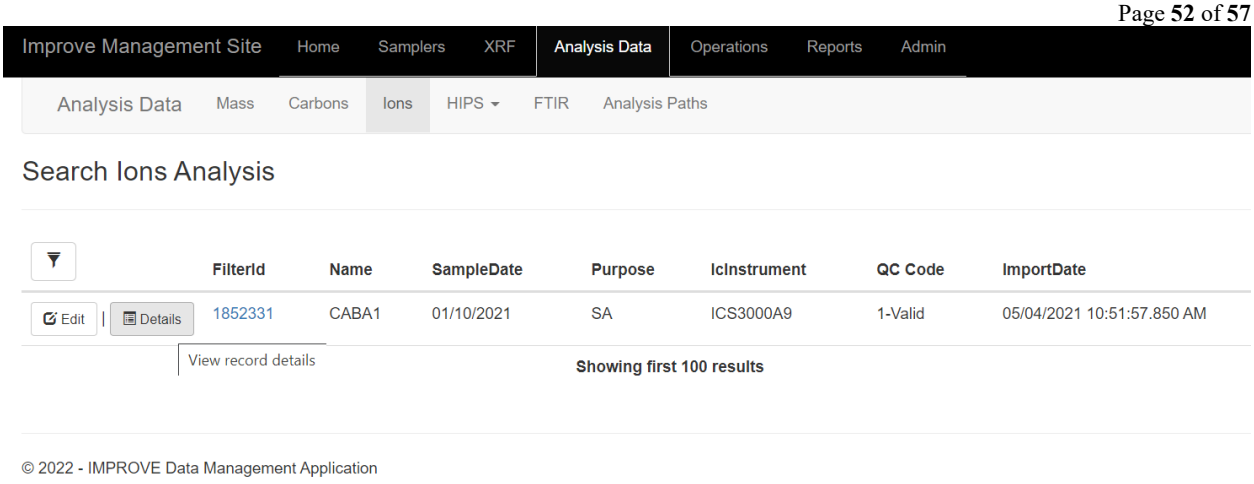


Figure 26. Carbon sample analysis page.

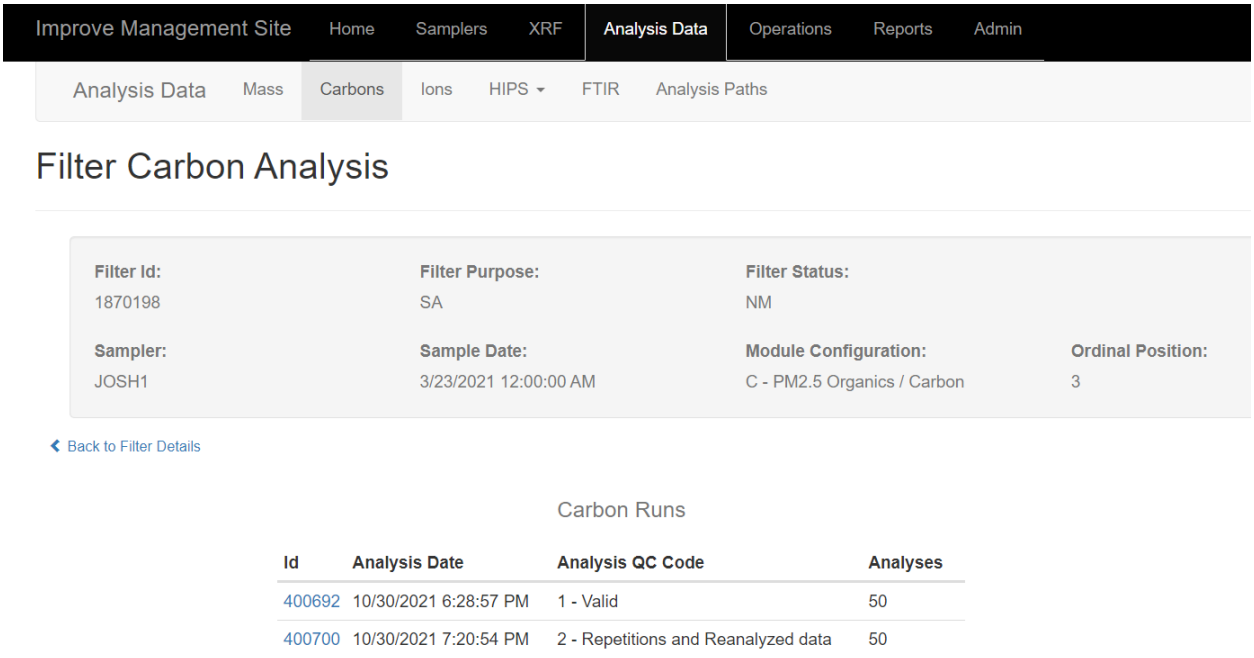


Figure 27. Analysis details page.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

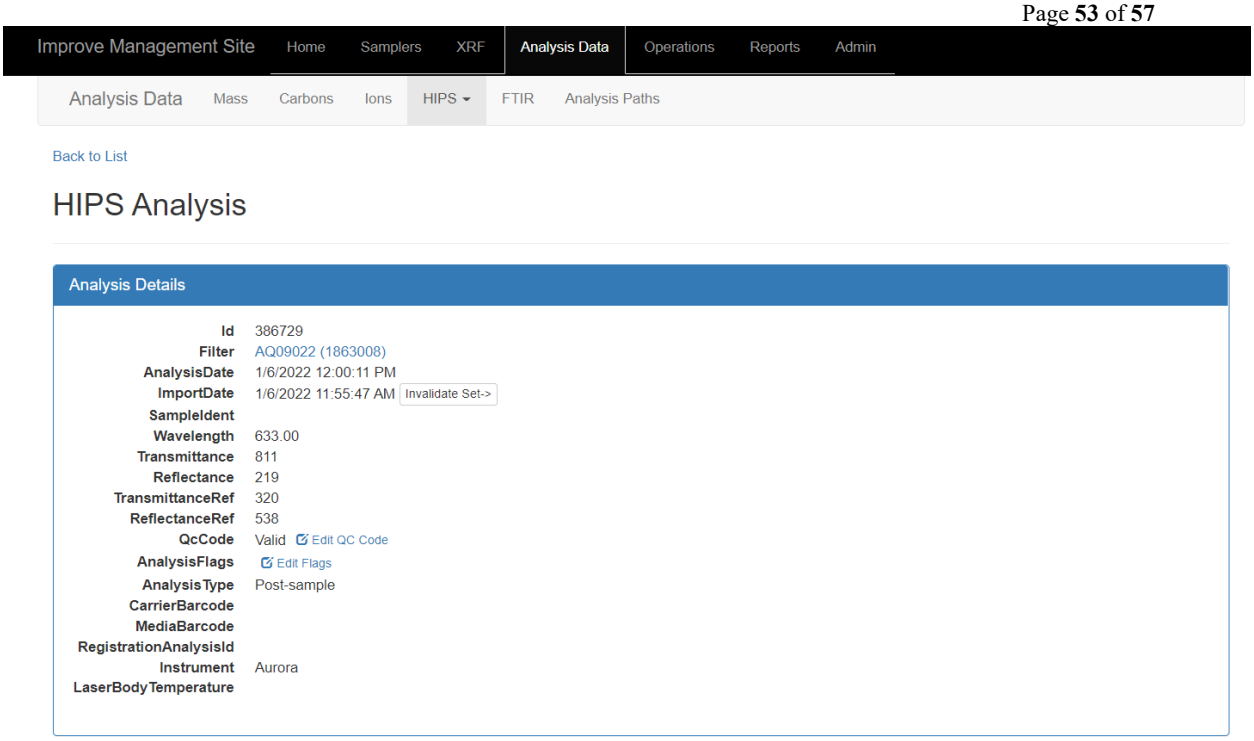
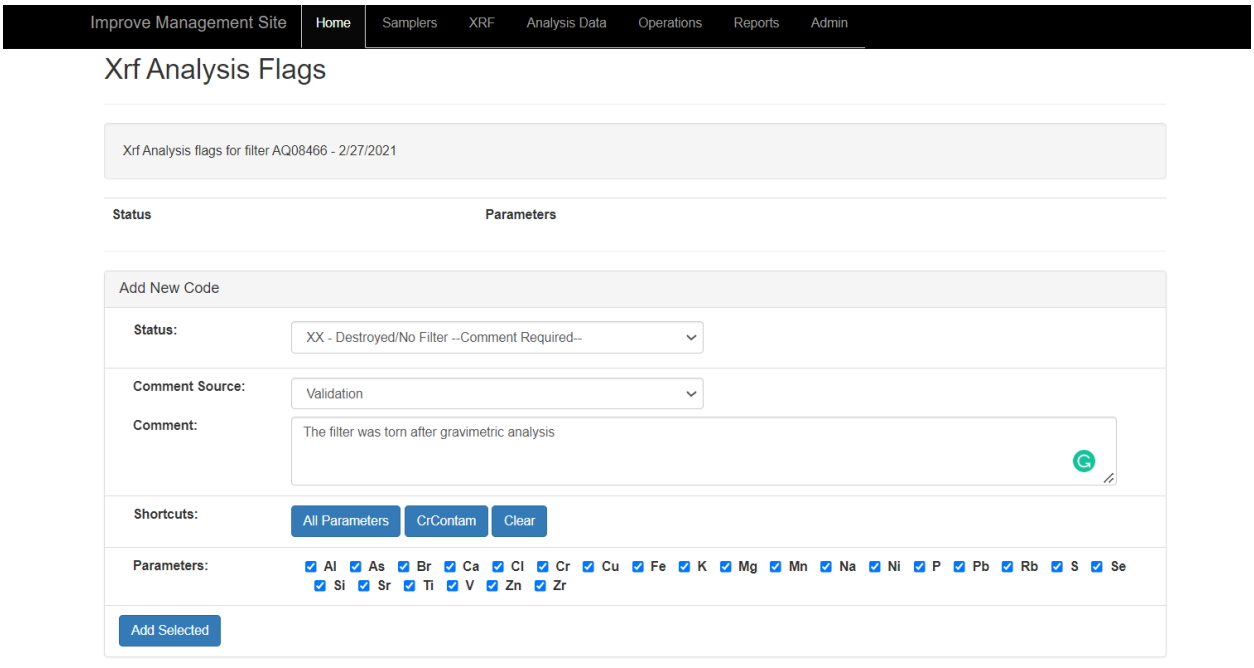
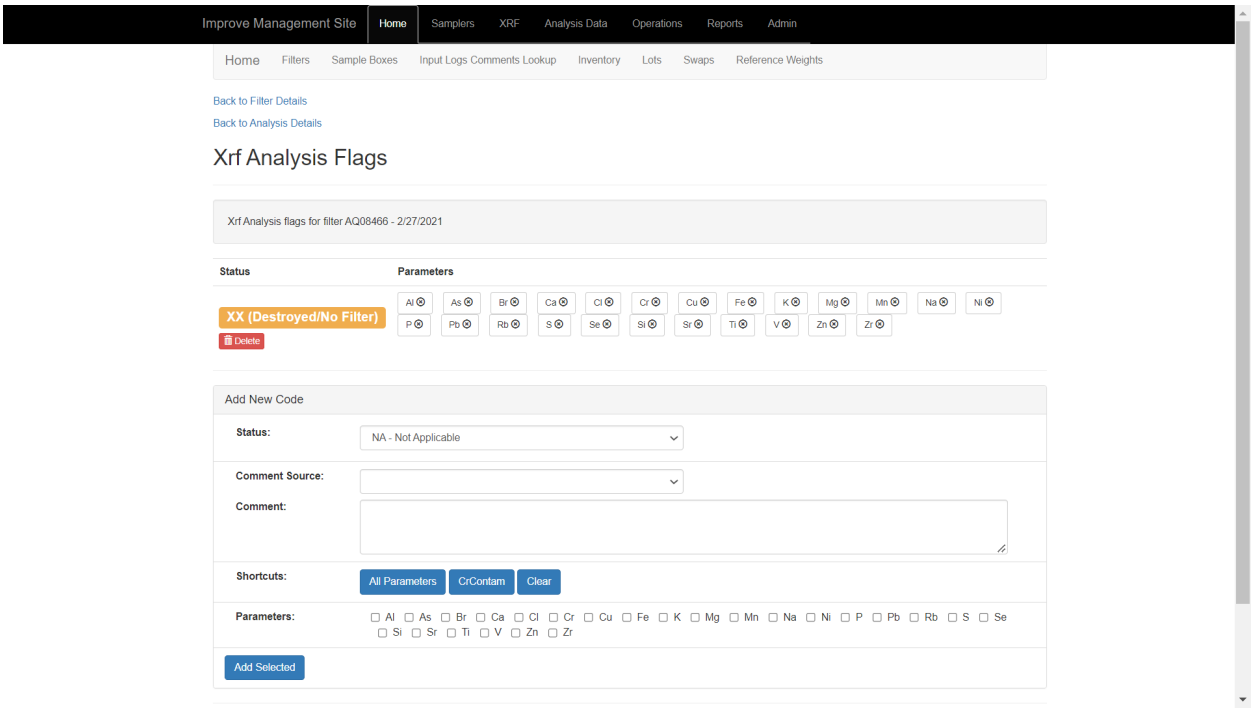


Figure 28. Analysis flags page.



DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

Figure 29. Deleting flags.



9.3.5 Recommended Validation Guidelines

The following section provides guidelines on the approach to validating data to determine if a sample is to be invalidated.

1) Unusual data observation made during validation, typically through reviewing plots on the ImproveData Validation page or from checks performed in R using the validation package e.g.:
    a. Sulfate concentration much higher than sulfur concentration;
    b. Sulfate concentration near zero but sulfur concentration is not;
    c. Negative EC concentration but BC and fAbs are positive and not near zero;
    d. $PM_{2.5}$ much higher than PM_{10}.
2) Review other data for the sample date and check composite variables calculated using the problem species, where available.
    a. E.g. if sulfate >> sulfur, review RCM vs. $PM_{2.5}$ as NHSO (= 4.125 *S) is used in calculating RCM. These relationships can be used to determine if the problem is with sulfur or sulfate, thus the ‘A’ or ‘B’ filter, respectively.
3) Review other anlaysis data from the problematic filter.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

a. E.g. if the problem filter is suspected to be 'B' as sulfate is near zero, check other ions species for similar observations. Is sulfate the only species with near zero concentration?

4) Review adjacent sample days for patterns and compare longer term with historical data.
    a. Use the plots on the Validation page as well as the Explorer page. If this pattern has been seen at the site at similar times in previous years, review the filters for comments and statuses to determine how the sample was handled previously. If the pattern is frequently observed, the current observation may be atmospherically real. If a similar pattern has not previously been observed, the data may still represent the air conditions but further investigation needs to be performed.
5) Review nearby sites for similar patterns.
    a. Local events may impact a subset of sites. Run the back trajectories, if available, in the Explorer page to determine which of the nearby sites may be expected to show similar trends and/or whether the air mass travelled over the ocean.
6) If there is no evidence for a particular issue to explain the observation, request reanlaysis of the the questionable filter(s) to rule out any anlaysis issues. Contact the sample handling lab to determine if there were any sampling or sample handling issues.
7) If no issues are found with the analysis, sampling, or sample handling, thus no changes are made to the data, the analyst should determine how egregious the issue is.
    a. For example, if the sulfate concentration is much higher than the sulfur concentration, the 3*S/SO4 ratio is an outlier, no similar cases have been observed previously, reanalysis results confirm the original anlaysis is valid, flow data does not indicate sampling issues, and surrounding sampling dates also do not show any issues, the analyst should consider invalidating the filter.

If the sulfate concentration is only slightly higher than the sulfur concentration, the 3*S/SO4 ratio is not an outlier and/or the resepctive uncertainties overlap, then perhaps the analyst will consider leaving the filters as valid.

9.3.6 Final Review

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

- The *status_check* function for statuses QD and QV in *datvalIMPROVE* (described in section 9.2.1 and 9.3) is run again after validation is complete to confirm that there are no remaining records with QD or QV status. No records with these statuses in the Status field should exist in the delivery files.

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

- The *ObjCode.check* function in *datvalIMPROVE* (described in section 9.2.1) 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_BadData* function in *datvalIMPROVE* (described in section 9.2.1) is run again after validation is complete to confirm that there are no remaining records with a valid status with values outside of defined normal ranges.
- 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 delivery file to CIRA (FED). 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 CIRA, where the file itself is a .csv file format.

- The *deliverycheck* function in *datvalIMPROVE* checks to determine if there are cases in the delivery file to CIRA where the data are valid but marked with a terminal flag or the data are invalid but marked with a valid flag. Perform this check using the following command in the R environment:

  [month_delivery] <- datvalIMPROVE::deliverycheck(filepath = ['filepath.csv'])

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

As noted in section 9.2.1, many of the functions described above can be performed simultaneously using the *datvalIMPROVE::improve_validate* function. Prior to delivery, some checks performed for initial validation are executed again and some additional final

DocuSign Envelope ID: 5347F36C-A0EC-4901-ABEB-97642D764BAE

checks are performed. Using the following command in the R environment, evaluate the output from the checks described below for delivery:

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

- **output$objective_code –** ObjCode.check
- **output$QD -** status_check
- **output$validsta_null -** ValidSta_NullData
- **output$validsta_bad -** ValidSta_BadData
- **output$mdl_unc -** MDL_UNC

10. QUALITY ASSURANCE AND QUALITY CONTROL

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

11. REFERENCES

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 edition; American Conference of Governmental Industrial Hygienists: Cincinnati, OH, 187-194.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

UCD IMPROVE Technical Information #351D

Data Delivery

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

October 4, 2022
Version 1.0

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 10/7/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 10/7/2022

Approved By: DocuSigned by: Marcus Langston 0A10CECE79B0452... Date: 10/7/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
12/04/21	SRS	All	Document created.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

TABLE OF CONTENTS

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

LIST OF FIGURES

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) concerns the delivery of the resultant data from the Interagency Monitoring of Protected Rural Environments (IMPROVE). This document describes the procedure for preparing and delivering data to the Cooperative Institute for Research in the Atmosphere (CIRA)/Federal Land Manager Environmental Database (FED), the Air Quality System (AQS) database, and the CSN and IMPROVE Archive Database (CIA).

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst prepares a delivery file of the validated IMPROVE data sets using custom tools in the *crocker* R package. Data are formatted into two FED formats and AQS format for delivery. The final data files are checked for correctness and then submitted to CIRA/FED via the National Park Service (NPS) and submitted to the Environmental Protection Agency's (EPA) AQS database. The Sample data are subsequently ingested into the CIA Database.

In addition, an unvalidated preliminary data set is to be processed and delivered to CIRA as single format data file on a routine basis.

3. DEFINITIONS

- **AQMT:** Air Quality Monitoring Team.
- **AQRC:** Air Quality Research Center.
- **AQS:** EPA's Air Quality System database.
- **CSN and IMPROVE Archive (CIA) Database:** A database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.
- **Chemical Speciation Network (CSN)**: EPA's $PM_{2.5}$ sampling network, with sites located principally in urban areas.
- **CIRA:** Cooperative Institute for Research in the Atmosphere.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA)
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (μm) and smaller. PM_{10} is particulate matter with diameters 10 μm or smaller.
- **SQL:** database management system used by AQRC.
- **UCD:** University of CA—Davis.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data ingest are described in the associated *UCD IMPROVE SOP #351: Data Processing & Validation.*

9. PROCEDURAL STEPS

After Level 2 data validation is complete, the data files are submitted to CIRA, AQS, and UCD CIA databases. The date of delivery is added to the AQMT Status IMPROVE Timeline page at https://shiny.aqrc.ucdavis.edu/aqmtstatus/.

9.1 Submission to CIRA

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Export files for CIRA (FED) 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. To generate the "skinny" format export file, run the following command in the R environment:

[FED] <- crocker::improve_export_fed(year = [YYYY], month = [MM], server = 'production', write_file = TRUE)

To generate the "wide" format export file, run the following command in the R environment:

[FED_wide] <- crocker::improve_export_wide(year = [YYYY], month = [MM], server = 'production', write_file = TRUE)

By default, the *write_file* argument is TRUE, so if a normal delivery file is created, no extra step is needed.

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.

To export a subset of the data, in the file export code, set *write_file* = FALSE. Instead of writing the output file directly, the functions return the data frame to the user's R environment. Using the *dplyr* package functions, the user should filter the results to only specific parameters and sites before writing the output. After creating the data frame with the selected subset of data, use the *write.csv* function to export the custom file as follows:

write.csv(df, file = ['filename'], na = '-999', row.names = [FALSE], quote = [FALSE])

The file name argument should specify the path to which you want to save the data as well as other pertinent information regarding the contents of the file. The user can specify the file name as a separate variable in their R environment before running the *write.csv* function. An example of the filename variable is as follows:

filename <- paste0('U:/IMPROVE/FED Export/', 'IMPROVE_Data_', year, '_', month, '_', server, '.csv')

The following checks are performed on the skinny format files:

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

- Open the CSV file and make sure the following columns exist: Id, FilterId, Sampler, ObjectiveCode, SampleDate, Status, Parameter, Value, Uncertainty, MDL, Unit, POC, ModuleTypeCode.
- In the CSV file, filter the Status column to 'UN' and review the parameters listed in the Parameter column: only PM_{10} or $PM_{2.5}$ should be listed. If non-mass parameters are listed, inform the software group to fix the issue.
- For the BYIS1 site, a value for the fAbs parameter is not reported for the routine module. To check this, filter the data in the Sampler column to 'BYIS1' and 'RT' in the ObjectiveCode column, and select 'fAbs in the Parameter field. Review the data in the Value field and confirm values are reported as -999.
- For filters with pending analysis results, use the filter option in the CSV to select the particular filter Id from the FilterId column and confirm the values for all parameters for that filter are reported as -999 and only the 'NS' status is listed in the Status column.
- In the Status column, click on the filter button and inspect the available statuses. Make sure that the listed statuses are currently active.

The following checks are performed on the wide format files:

- Make sure the 'UN' status is only applied to the mass. This can be done by opening the CSV file and using the find and search option for 'UN'. Review the results and confirm 'UN' is only listed under the columns named PM10_flag or PM2.5_flag.

In the CSV file, spot check the number of rows per site equals the number of sample dates in that month. For example, for the month of February, nine rows are expected. Other months are expected to have 10 or 11 rows.

9.2 Submission to AQS

9.2.1 File Generation and Checks

AQS data export files are created using the *improve_export_aqs* function in the *crocker* package. To generate the AQS delivery file, run the following command in the R environment:

[aqs] <- crocker::improve_export_aqs([YYYY], [MM], server = "production", filename = [NULL], action = ["keep"], site = ['XXXXX'], param = [NULL], del_type = ["I"])

This command will generate a formatted text file suitable for AQS delivery containing all data for the year (*[YYYY]*) and month (*[MM]*) and save to the location specified in *filename*, with the default file name and path in the format of '[aqs_path]/AQSResultsOutput_[today]_[year]_[month]_[server].txt'. A typical command for routine monthly data delivery can be run as follows:

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

aqs <- crocker::improve_export_aqs(2020, 04, 'production')

The function has the capability to generate an AQS delivery file down to the parameter level (*param*) and/or particular site(s) (*site*). If *action* is specified as "keep", only the specified records are retained in the delivery file whereas "drop" will remove the specified records. In addition to generating delivery files to add data to the AQS database, the function can also be used to create delivery files to update or delete data within AQS by specifying the *del_type*. To display the helper documentation for the function, the user can run *?improve_export_aqs* in the R Studio console.

Once the file has been generated, various checks should be performed by running the following commands in R Studio:

- Check for existence of duplicate records:
  duplicates <- find_duplicates(aqs, c('StateCode', 'CountyCode', 'SiteID', 'POC', 'SampleStartDate', 'AQSParameterCode', 'AQSMethodCode'))
- Check the number of unique parameters:
  if(count(unique(aqs[c("AQSParameterCode")])) == 45) {print('ok')}
- Check for existence of records with null values and no null code:
  noVal.null <- aqs %>% dplyr::filter(Value == "" & NullCode == "")
- Check for existence of records with non-null values and a null code:
  null.noData <- aqs %>% dplyr::filter(NullCode != "" & Value != "")
- Check for existence of records with a qualifier validity flag and a null code:
  null.noData <- aqs %>% dplyr::filter(NullCode != "" & Qualifier1 != "")

In addition to checking the data file for issues that would result in a failed AQS delivery, the validator also reviews the data further to obtain information on the data set as a whole. The validator can compare this with similar information from previous months of data to determine if the current month of data is reasonable or if there is an unexpectedly large increase in the number of invalid records, for example.

- Count the number of records with a null value and null code:
  noData <- aqs %>% dplyr::filter(NullCode != "" & Value == "")
- Count the number of invalid and valid records:
  no.Nulls <- aqs %>% dplyr::group_by(NullCode)%>%
  dplyr::summarise(n_nulls = n())%>% dplyr::arrange(n_nulls)

For each dataset, the data validator keeps a record of the null codes reported and the number of records with each null code to put each month into context with previous months of data. The log file is located at U:\IMPROVE\Data_Validation\Status change list\AQS_null_code_count_per_month.xlsx.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

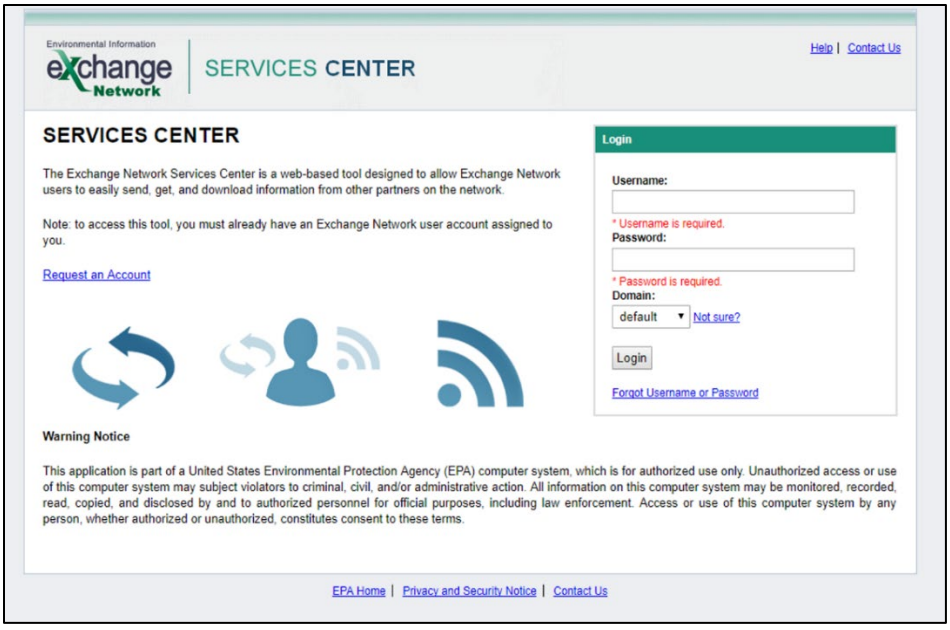
- After running the command above, the validator can confirm the total number of invalid and valid records in the file matches the total number of records in the dataset by running the following command: *print(sum(no.Nulls$n_nulls))*
- Further, the number of invalid records (those with a null code) can be counted and compared with the results from the check for existence of records with non-null values and a null code:

  no.Nulls.total <- no.Nulls %>% dplyr::filter(NullCode != "")
  print(sum(no.Nulls.total$n_nulls))

9.2.2 Submit Data to AQS

Once the checks have been completed, the data can be delivered to AQS. To submit batch data files to AQS, open a web browser and navigate to the EPA Exchange Network Services website, https://enservices.epa.gov/login.aspx (Figure 1). Use credentials to login.

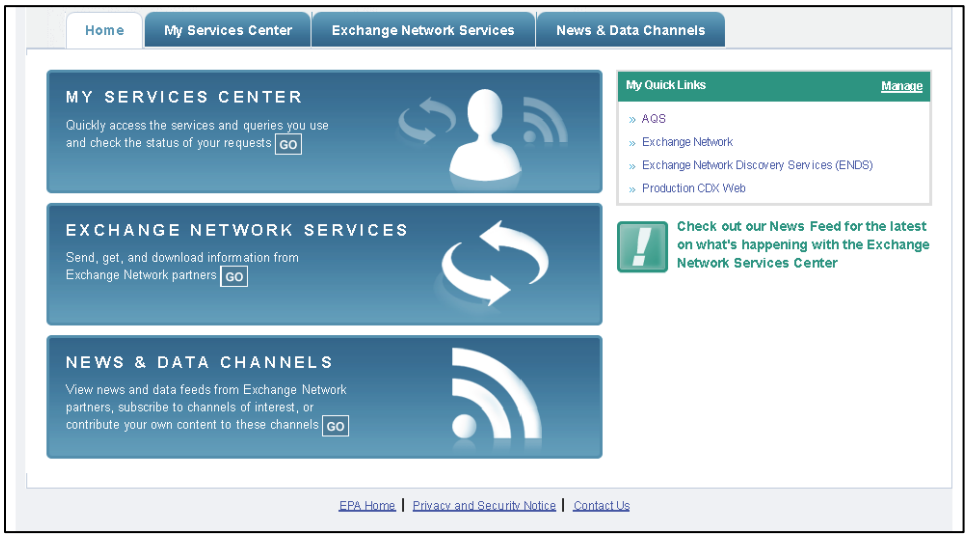
Figure 1. Login screen for the EPA's Exchange Network Services Center website.



Following login, the home screen is accessed (Figure 2). 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.

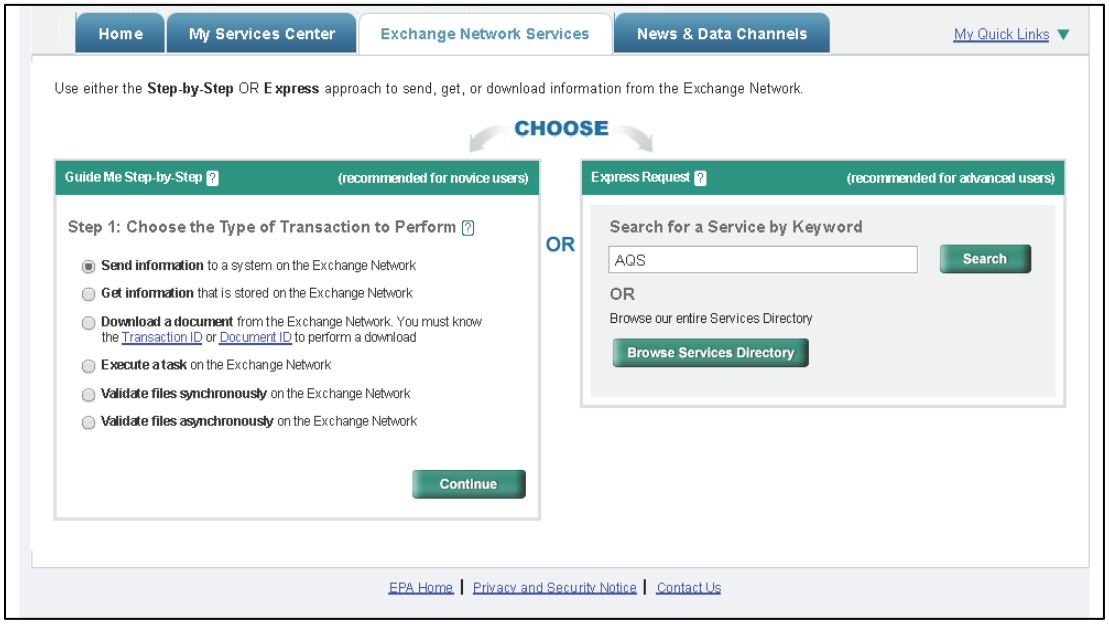
DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 2. Home screen of the Exchange Network Services Center website.



Next, the option for a step-by-step guide and a search bar presented (Figure 3); type **AQS** into the search bar.

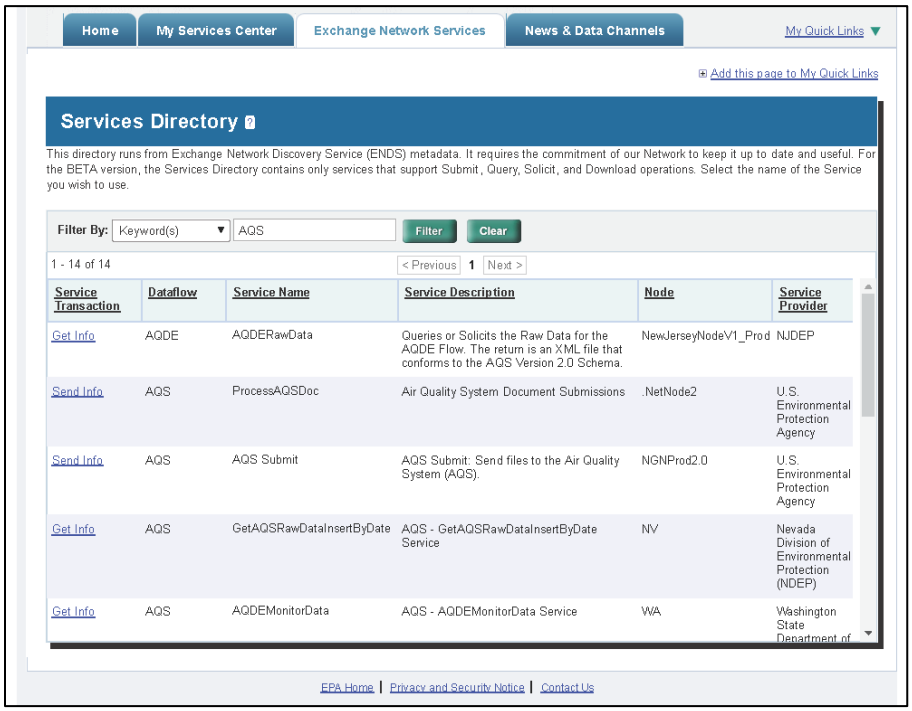
Figure 3. Type AQS into the search bar.



The search results will show all available processes associated with the AQS system (Figure 4). To access the AQS submission form, choose the service that has **AQS Submit** specified in the **Service Name** field (usually the third option listed).

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

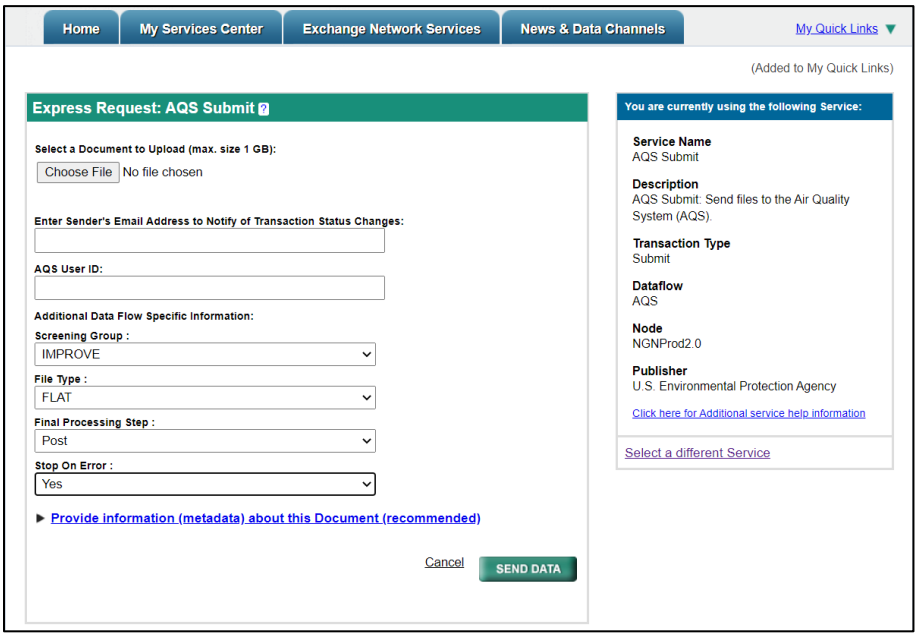
Figure 4. Select the service named AQS Submit.



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 5 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. If the submission to AQS is fully successful, no further actions are necessary; the data have been delivered.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 5. AQS data submission form.



9.2.1 Address Data Errors in AQS

If there are data records that are considered by AQS as statistical evaluation issues or critical review issues, the submission process will stop, despite there being no data errors. The data will only have been staged in AQS and needs to be posted. The user will receive two emails at the email address entered during the submission (Figure 5): one from AQSDEV and one from cdx@epa.gov, which include details of the data submission. The AQSDEV email also includes other details such as links to the load report and statistical and critical review report. Once the user has reviewed the emails to confirm the data are ready to be posted as no row failed to load, the user should log into AQS directly to post the relevant data.

To do this the user should follow these steps:

- Launch the AQS Application at https://www.epa.gov/aqs and enter their AQS credentials (Figure 6);
- Select the Screening Group Access option (Figure 7);
- Select IMPROVE (Figure 8);
- Go to Batch;
- Highlight the relevant row for the submitted file on the Process by File tab;
- Click on the ‘Post File’ button.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 6. EPA AQS Application login screen.

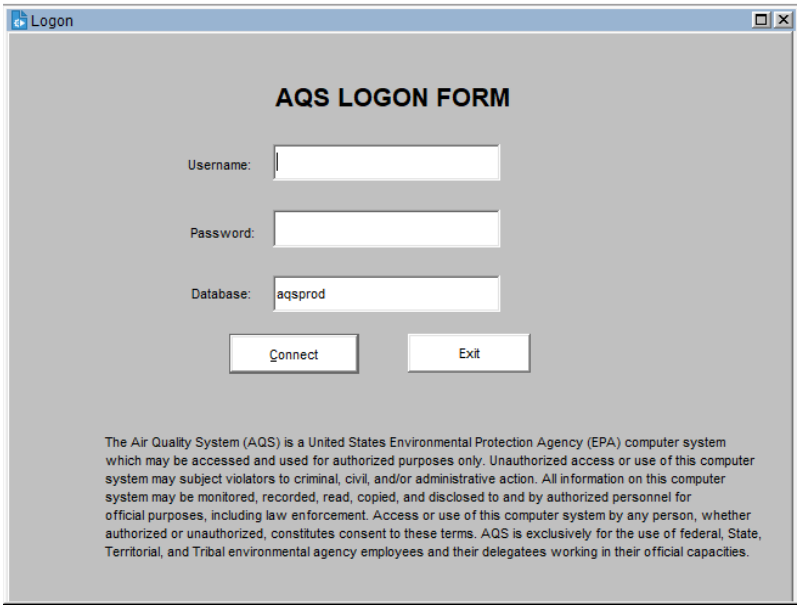
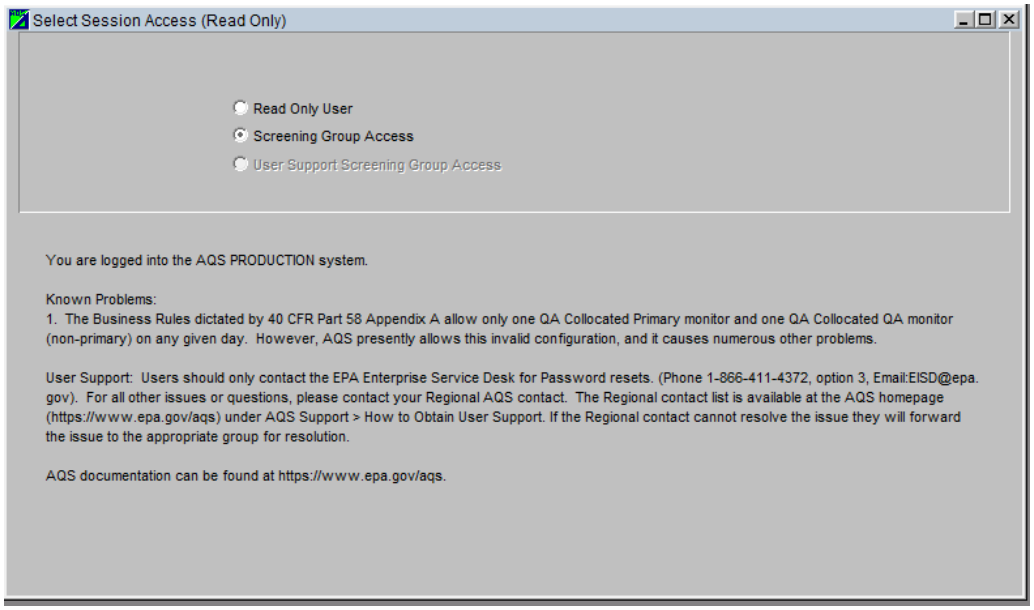


Figure 7. AQS screen after logging into the application; select the Screening Group Access option.



DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 8. AQS pop-up screen after selecting Screening Group Access; select IMPROVE screening group.

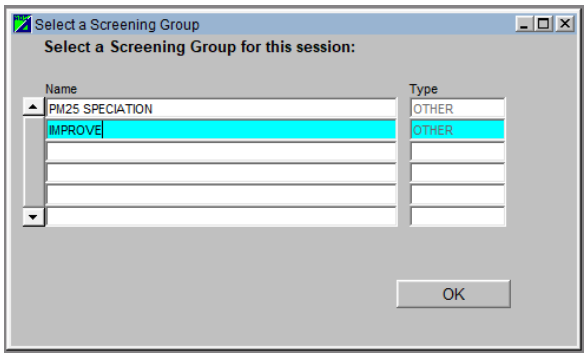
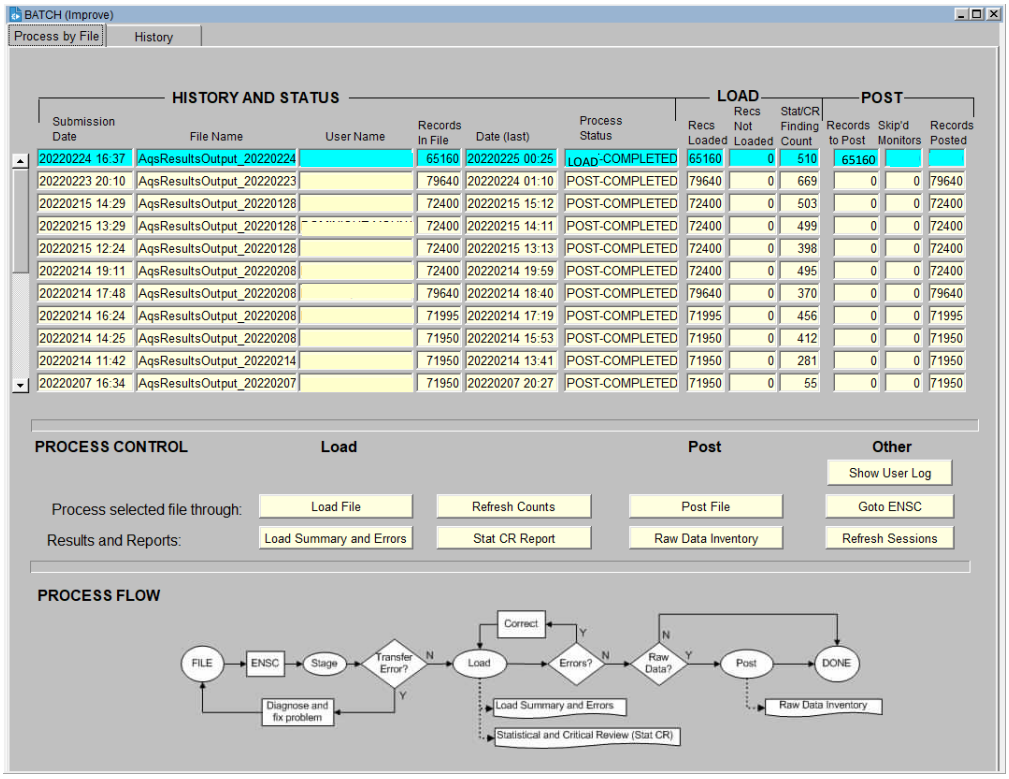


Figure 9. AQS screen after selecting **Batch** from the options. Click on any field of the row for the file which is to be posted. Files to be posted will show a Process Status of 'LOAD-COMPLETED' and the Skip'd Monitors and Records Posted fields will be empty.



If there are data records with errors, the submission process will stop. Again, the user will receive two emails. In the case of data errors, the AQSDEV email will also include a link to a text file containing the records that failed to load. The user should review the data in the text file along with the load summary and detail report to determine the cause of the error and determine how to resolve the issue. Regarding the staged data in AQS, the records that did not load can either be corrected directly in AQS or corrected externally and a new data file be submitted through the process outlined in section 9.2.2. The standard procedure at UCD is to correct the data outside of AQS and resubmit the file for tracking and documentation purposes.

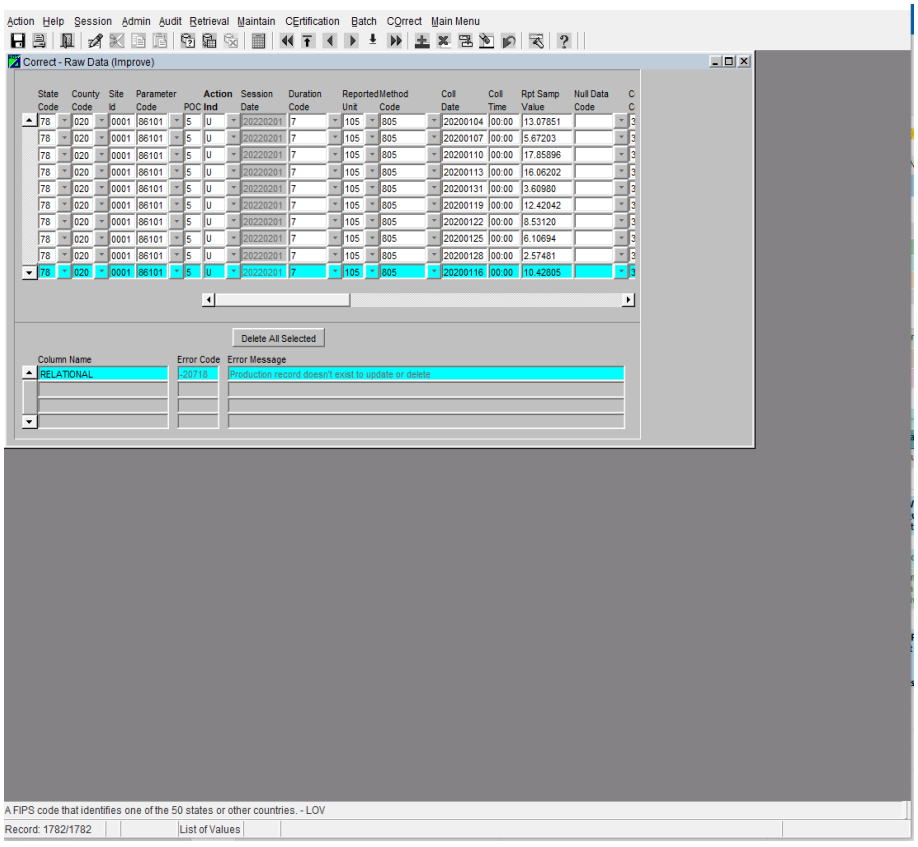
DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Information on how to correct data in AQS can be found in the AQS User Guide (https://www.epa.gov/sites/default/files/2018-07/documents/aqs_user_guide_2018_2.pdf, from page 47) which can be found on the EPA's AQS website at https://www.epa.gov/aqs. Select 'AQS Users Guide' under Documentation and select the AQS User Guide PDF towards the bottom of the webpage.

In brief, to correct data externally and submit another file, the user first needs to delete the data from AQS staging by following these steps:

- Navigate to the Correct option after logging into the IMPROVE screening group;
- Select 'Raw';
- Enter details of known records with issues
    - For example several records for a particular Parameter Code may have failed to load. In this case, enter the relevant Parameter Code in the appropriate field.
- Click on the 'Execute Query' button.

Figure 10. AQS screen after selecting **Correct** -> **Raw** from the options, entering information on failed records (e.g., Parameter Code), and clicking on the 'Execute Query' button.



The data returned are from the AQS staging table and are those that have not yet been loaded into AQS (e.g., Figure 10 shows records that failed to load for the Coarse Mass parameter (code 86101)). The user should click on the down arrow and review the

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

information at the very bottom of the window to confirm the number of pulled records matches what was expected. The user should verify these data are to be corrected and click on 'Delete All Selected'. This button will delete all records that were returned from the query; the records do not all have to be selected.

If there are multiple types of issues, the user should repeat the steps outlined above until all data that failed to load are deleted.

To correct the data, the user may need to fix the data directly using tools and techniques described in UCD 351#C Data Validation, reprocess the data as described in UCD 351#B Data Processing, generate another AQS file as described in section 9.2, and submit to AQS.

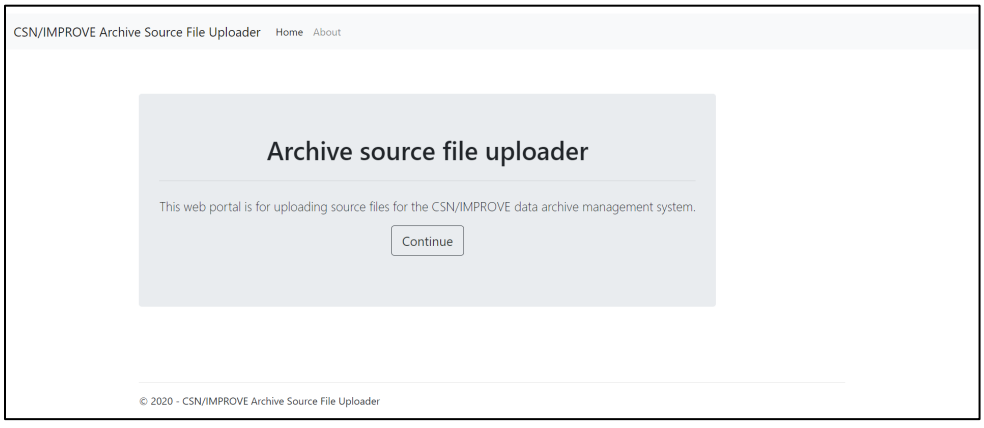
Alternatively, if the data can be easily corrected in bulk, such as if the correction is to change the Action Indicator to "I" from "U", the user can utilize the records in the failed records link in the AQSDEV email. The user should click on the link to the text file, which opens a web browser tab. The user should copy the data – except the last line of the page that notes the number of records that were posted to the file – and paste in a text file such as Notepad. The user can correct the data using 'find and replace' and save as a new AQS file using the same naming convention as is used for regular files, with additional information to note the contents of this particular file. The file can then be submitted to AQS as a typical data file.

9.3 Submission to UCD CIA

The CSN/IMPROVE Archive (CIA) is a database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.

1. Open a web browser and navigate to the UCD CIA submission website, https://cia-uploadportal.azurewebsites.net/ (Figure 11).

Figure 11. UCD CIA submission website home page.

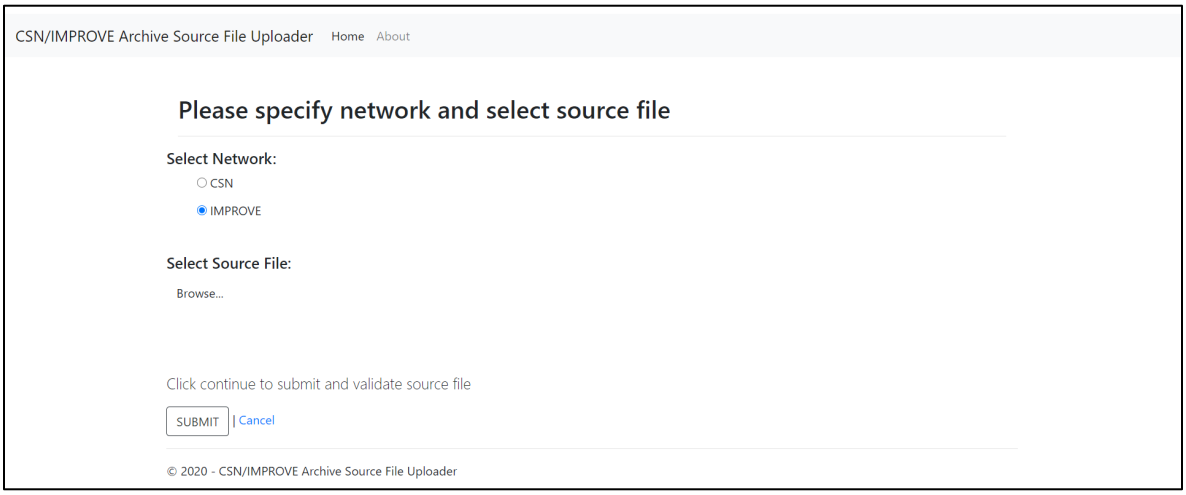


2. Click the **Continue** button in the center of the page.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

3. Specify the network of choice that you will be delivering the data for, which in this case is **IMPROVE**. See Figure 12 for an example.

Figure 12. 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.

Note that the CIA upload portal will only accept files with a single action indicator. Most initial delivery files have only one action indicator (I - ingest). But the redelivery files can have more than one action indicator. If the upload fails, the file needs to be separated based on the action indicator and uploaded separately.

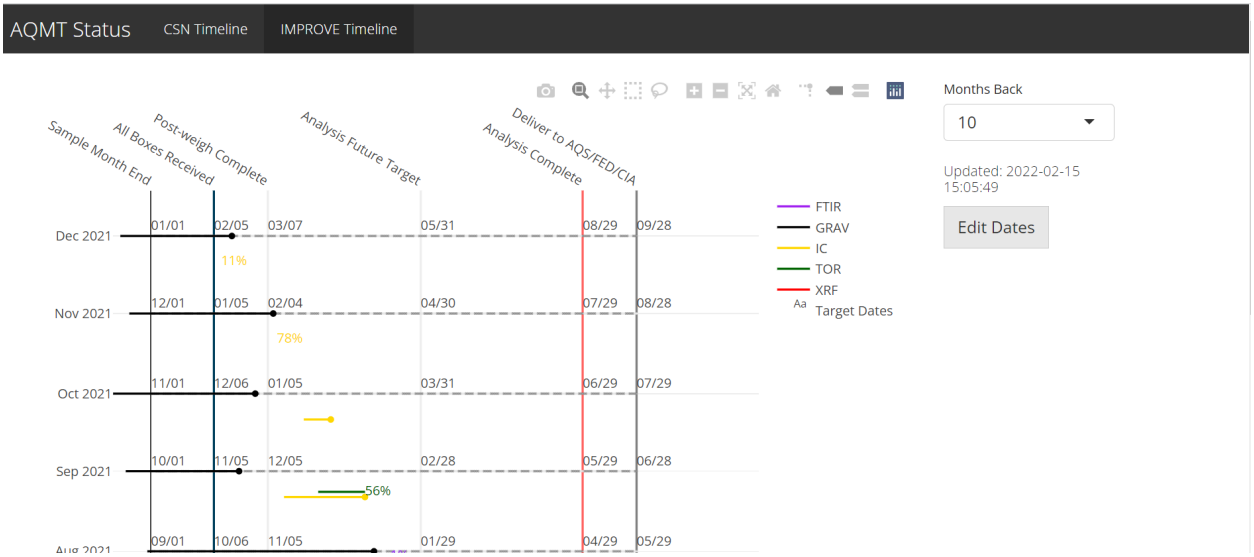
9.4 Add Data Delivery Dates Delivery Dates to AQMT Status Page

After delivering data to CIRA, the user should perform the following steps:

1. Add the IMPROVE delivery dates to the IMPROVE Timeline on the AQMT Status page (https://shiny.aqrc.ucdavis.edu/aqmtstatus/).
2. After accessing the web page, go to the 'IMPROVE Timeline' tab (Figure 13) and click on the Edit Dates button.

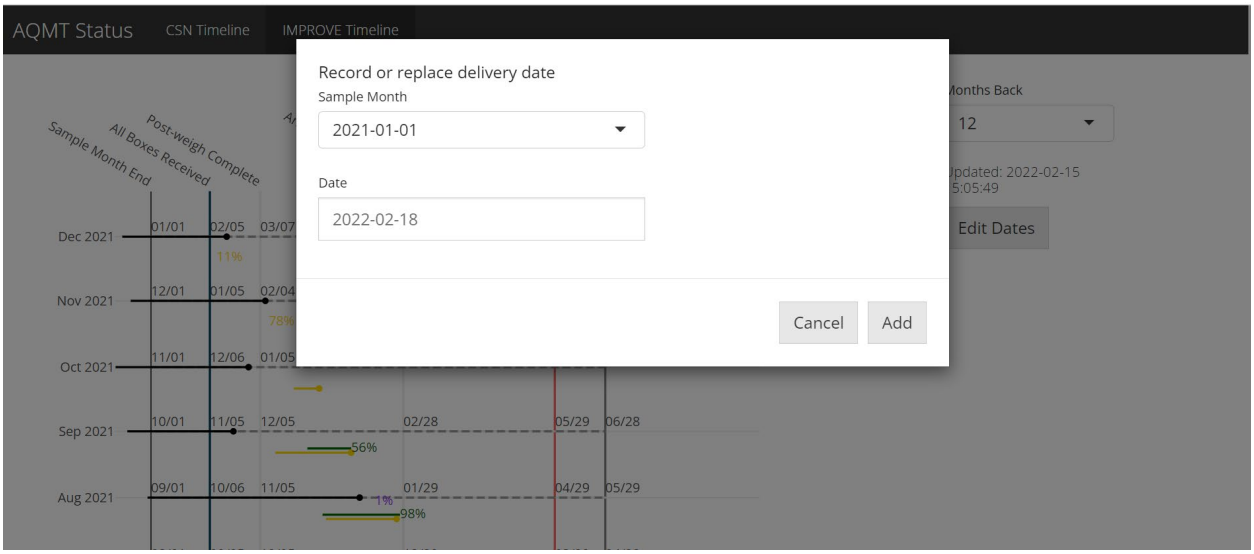
DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Figure 13. 'IMPROVE' Timeline tab on AQMT status page.



3. Select the sample month and add the delivery date and click on the Add button (Figure 14).
4. After adding the file delivery date, confirm that a tick mark is added to the timeline for the month.

Figure 14. File submission date update window.



9.5 Submission of Preliminary Data to CIRA

An unvalidated preliminary data set will be delivered to CIRA every month. Preliminary data differ from validated data in several ways:

- Validation has not been performed or is incomplete.
- Blank corrections are estimated from past data.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

- MDL and uncertainty values are not calculated or reported.
- Statuses may include those used for internal validation purposes only and are not delivered in final data.

9.5.1 File Generation

The improve_export_preliminary function in the *crocker* package is used to create the preliminary data export file. To generate the preliminary delivery file, run the following command in the R environment:

[preliminary_data] <- crocker::improve_export_preliminary(start_year=[YYYY], start_month= [MM], end_year=[YYYY], end_month=[MM], server = 'production', write_file = FALSE)

The argument *start_year* is the numeric Year of the first sample month to process and *start_month* is the numeric month of the sample month to process. To determine the starting Year and Month to process, visit the IMPROVE timeline tab on the AQMT status page as mentioned in section 9.4 (https://shiny.aqrc.ucdavis.edu/aqmtstatus/) and look for the last month of data that was delivered to CIRA, AQS, and CIA. The start year and month of preliminary data to be processed and delivered will then be the month after that. The argument *end_year* is the numeric Year of the last sample month to process. The *end_month* is the numeric Month of the last sample month to process. To determine these end dates, the IMPROVE timeline tab can be reviewed again. The most recent month with completed 'GRAV' analysis can be used as the end date for processing and delivering the preliminary data, however gravimetric analysis does not have to be completed to run preliminary processing.

By default, the *write_file* argument is FALSE. If the write file argument is not specified or set to FALSE, but the function is assigned to a variable then the command generates a data frame containing all data for the year and month range. The user can review the data in their R environment. If the *write_file* argument is set to TRUE, a delivery file in CSV format will be saved to *U:/IMPROVE/FED Export/Preliminary/*, named 'IMPROVE_Preliminary_Unvalidated_YYYY-MM-DD' (e.g., IMPROVE_Preliminary_Unvalidated_2022-05-05'), where YYYY-MM-DD is the date the file was created.

At the time of processing, many analyses will be incomplete. When running the code, warning messages for missing analyses along with other usual validation messages will be printed to the R console. These warnings will not interrupt the processing. The user is not to act on these messages. However, if the processing stops because of an error, the user is to inform the Software Group about the issue.

Once the data frame is generated, review the 'SampleDate' column to confirm all expected months are included. If data for any months are missing, especially the more recent months being queried, it could be because processed flows do not exist for those

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

samples. The status of processed flows can be checked with the Data & Reporting Group Manager or the Sample Handling Laboratory to determine if the flow can be processed for the missing months and subsequently proceed with re-processing and delivering the data. If the flow cannot be processed for more recent months, only report data through to the last month of data that was processed for flow. If the flow data can be processed, use the SQL query or the improve_process_flow function in R as described in Section 9 of IMPROVE TI 351B_Data Processing to derive the daily average flow rate and elapsed time (ET).

Once the data have been finalized, the data frame can be saved into a CSV file by running the following command in the R environment, where *YYYY-MM-DD* in the file name is the date the file was created:

write.csv(preliminary_data,' U:/IMPROVE/FED Export/Preliminary/IMPROVE_Preliminary_Unvalidated_YYYY-MM-DD.csv')

A typical command for standard monthly data delivery can be run as follows:

preliminary_data <- crocker::improve_export_preliminary(2021, 05, 2021, 12, server = "production", write_file = TRUE)

Once a file has been generated, the user is to compress the CSV file into a zip folder and email to the CIRA correspondent(s) as an attachment.

10. DATA AND RECORDS MANAGEMENT

The IMPROVE data are stored in Microsoft SQL Server Databases at UC Davis. The production database is run on a dedicated Windows Server with a RAID array for storage and with offsite backups. Our development and test database environments are virtual machines. To test back up recovery, our development and testing environments are regularly restored from the production backups.

Data management is handled through custom software that interfaces with the UCD IMPROVE database. The primary applications for data ingest and management were developed on the .NET platform. Data processing and calculations were developed as R software packages. In addition, to support data validation and operational monitoring, several interactive visualizations have been developed using the R Shiny platform.

10.1 Disaster Recovery Plan

The scope of recovery activities will depend on the nature of the disaster. Please refer to *UCD IMPROVE SOP #351: Data Processing and Validation* for specific information.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

10.1.1 Facility Recovery

Private security services patrol the laboratory building on a regular basis (including nights, weekends, and holidays). In addition, campus facilities and maintenance staff are on call at all times.

Databases, file servers, and web server virtual and dedicated machines operate primarily out of the Metro IT data center in Hoagland Hall on the UCD campus. Metro IT has a highly-available, disaster recoverable virtualization environment. Weekly backups of the virtual hard drives are taken offsite and stored in the Campus Data Center. In the event of a disaster in Hoagland, critical machines will be mounted at the Campus Data Center. The Drew Avenue laboratory is directly connected to the main campus internet. In the event that connection is disrupted (such as through a construction accident), connections will be switched to a local backup server until service can be restored.

10.1.2 Hardware Recovery Plan

The campus network of IT Administrator staff allows for rapid response to server failure and recovery issues.

10.1.3 Software and Data Recovery Plan

10.1.3.1 UCD Laboratories

Raw and processed analysis data produced with the UCD laboratories are saved and available for use at any time on the computers associated with each instrument, including the PANalytical Epsilon 5 EDXRF, MTL Automated Weighing System (gravimetric mass), Hybrid Integrating Plate and Sphere (HIPS).

Operational flow rate information from samplers in the field is automatically transferred nightly to a file processing server. As a backup, the flow data are stored on SD cards and delivered to the sample handling lab along with the exposed filters.

Data from all analyses, along with the flows, are scheduled to automatically transfer to a central Microsoft SQL Server database located at a data center on the UCD campus. Differential backups are performed daily, and full backups are performed weekly.

10.1.4 Data Security

UCD access policies: Access to databases and computers associated with this project is limited to authorized project personnel by use of access control lists for files, programs, and database access. Access to laboratory and office space is controlled by keycards.
Password policies: Unique passwords are issued to each employee by the UCD campus system administrator. Password integrity is monitored by the UCD campus system administrator.
Termination policies: System access is revoked for terminated personnel. The IT Administrator disables domain accounts and passwords upon termination of employment.
Virus protection: Microsoft Endpoint Protection is used for virus scanning and protection. All staff are required to complete annual cyber security awareness training.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

11. QUALITY ASSURANCE AND QUALITY CONTROL

11.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

11.2 Bug Reporting

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

11.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this SOP.

12. REFERENCES

Not applicable.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

UCD IMPROVE Technical Information #351E

Flow Validation

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

October 4, 2022
Version 1.0

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 10/7/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 10/7/2022

Approved By: DocuSigned by: Marcus Langston 0A10CECE79B0452... Date: 10/7/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
03/14/22	SRS	All	Previously anthologized version separated into individual TIs; content updated.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

TABLE OF CONTENTS

LIST OF TABLES

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide information regarding the steps to process and validate the flow data from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. Flow data from the network are reviewed and validated using various tools.

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst uses the UCD Flow Plotter website along with custom software in the R language to perform flow data processing and validation.

3. DEFINITIONS

- **AQRC:** Air Quality Research Center.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **datvalIMPROVE:** A custom software package in the R language that contains the data validation code used to collect, compare, and flag the final results.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.
- **Hybrid Integrating Plate/Sphere (HIPS):** An analytical technique for optical absorption.
- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (μm) and smaller. PM_{10} is particulate matter with diameters 10 μm or smaller.
- **SOP:** Standard Operating Procedure.
- **SQL:** database management system used by AQRC.
- **TI:** Technical Information; subset document paired to an SOP.
- **UCD:** University of CA—Davis.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data and Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data validation are described in the associated *UCD IMPROVE SOP #351: Data Processing and Validation*.

9. PROCEDURAL STEPS

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. The SHL receives flow data from the V2 controllers 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 Type Code for the affected sample is changed from the default (MC/MO) to log sheet (LC/LO) or nominal value (NF) to ensure an accurate calculation of the average flow rate. Detailed procedures on flow data ingestion and Level 1A validation can be found in *UC IMPROVE TI #251E: Entering Log Sheets and Simple Problem Diagnosis*.

9.1 Processed Flow Data

Prior to checking flow data, the quality assurance officer processes flow data using the SQL query or the *improve_process_flow* function in R as described in section 9 of *UCD IMPROVE TI #351B: Data Processing* 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 with flow rates that deviate from the nominal values. Table 1 and 2 list the

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

types of flow flags and the associated criteria for applying them to $PM_{2.5}$ and PM_{10} samples, respectively.

Table 1. 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
TO	Timing Outside normal bounds	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:

$$D_{50} = 52.5 * Q^{-0.99} \quad \text{(351E-1)}$$

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. 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 $PM_{2.5}$.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

A similar set of flags is applied to the PM_{10} data (Table 2), 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 2. 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
TO	Timing Outside normal bounds	Informational	Elapsed time between 1080 minutes (18 hours) - 1380 minutes (23 hours) or 1500 minutes (25 hours) – 1800 minutes (30 hours)

Once the flow data have been processed, the data are to be validated. The following sections describe the procedure for generating a report containing flow related items that have met check criteria and require further investigation as well as some commonly observed scenarios.

Several Level 1B checks (see *UCD IMPROVE TI #351C: Data Validation* for details on Level 1B) on the 15-minute raw flow data are performed by running the *flow.check* function (for both the V2 and V4 controller data) from 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’)

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

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 three asterisks (***) are generated automatically in the output from the *flow.check* function to indicate data issues.

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.

9.2 Generating the Flow Validation Report

The flow validation report is generated as an Excel spreadsheet. It is populated using the data returned from running several checks on the flow data. As the first step of validation, check for valid filters with missing flow data. The *flow.completeness* check will return a list of filters with missing flow data.

- No Flow data: To generate the list, run the following command in the R environment:

```
[No flowdata] <- datvalIMPROVE::flow.completeness(startdate =
['YYYY-MM-DD'], enddate = ['YYYY-MM-DD'], server = 'production')

write.csv(No flowdata, "U:/IMPROVE/Data_Validation/Flow/Nofllow.csv",
row.names = TRUE)
```

  Once the list is generated, coordinate with the Sample Handling Laboratory to investigate the reason(s) behind the missing flow data and resolve as appropriate. Once all the filters have the correct flow data attached, reprocess the flow using the SQL query or the *improve_process_flow* function in R as described in Section 9 of IMPROVE TI 351B.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

The next tab of the spreadsheet is populated using the data returned from running the *flow.check* function as described in section 9.1 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)

  Save the data frame as a CSV file using the following R command:

  write.csv(month_flow$OldController, "U:/IMPROVE/Data_Validation/Flow/Monthflow_OldController.csv", row.names = TRUE)

Once all the flow validation-related data frames are exported (the steps are below) in CSV format, they can be combined to Excel format to make the flow validation report. Once in Excel format, color code the modules (A = red, B = Yellow, C = Green, and D = Blue). The three asterisks (***) generated automatically in the output from the *flow.check* function (see Section 9.1) 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)

  As described in the previous step, export the data frame and color code the modules.

- 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)

- Flow flags (CG, CL, LF, PO, EP, TO): These sheets contain lists of samples where the flow status is flagged as CG, CL, LF, PO, EP, or TO and require confirmation of appropriate flagging (see Tables 1 and 2). 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', 'EP', 'TO')], server = 'production')

  To generate a list with only one of the flow flags, set the *flowflag* argument to equal one of the six flags. Export the results and add it to the appropriately labelled sheet in the flow validation report.

9.3 Flow Validation

To further investigate the data returned from the flow checks and to validate flow data, flow plots are carefully reviewed (IMPROVE Flow Graphs; https://shiny.aqrc.ucdavis.edu/FlowRates/). The Flow Source Code is assigned if the primary source (MC for A, B, C modules and MO for D module; automatically

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

transmitted flow data or flash card) is not reliable. Guidelines for validating flow data include:

- Review the flow graph 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 [Improve_2.1].[ops].[ControllerFilterReadings] or the hand-written records on the paper Field Log Sheets.
- The Flow Source Code or Filter Status Code can be updated as needed from the Filters page of the 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.
- Utilize the Controller Filter readings page (https://improve.aqrc.ucdavis.edu/Operations/ControllerFilterReadings) of the IMPROVE Management Site.

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 flash card files or faulty transducer readings), the flow source code is changed and average flow rate is reprocessed to correct the sample status. The same approach is taken for other flow flags like CG and LF.

9.3.1 Common Flow Review Scenarios

In this section, common scenarios investigated during flow validation are described including guidelines for resolving issues.

Clogged / Clogging (CL or CG) status

- The flow data are flagged with CL or CG status when there is heavy loading on the filter or due to pump malfunction. In the case of heavy filter loading, no further action is needed. If the flow status is CL due to pump malfunction, change the filter status code to EP (Equipment Problem. Refer Table 1 and Table 2) from the Filters page of the IMPROVE Management Site and reprocess the flow data using the SQL query described in section 9 of *UCD IMPROVE TI #351B: Data Processing*.

Power Outage (PO) status

- The PO status gets applied when Elapsed time < 1080 minutes. Check the filter readings table, flow plotter, or log sheet data to ensure a late sample change was

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

not the cause for the elapsed time to fall below the limit. If the sample change was late, update the filter status to NS (No Sample. Refer Table 1 and Table 2) and reprocess the flow data using the SQL query described in section 9 of TI #351B.

Temperature probe malfunction

- If the temperature data is showing extreme values (e.g., 200 degrees Celsius), the Temperature probe could be malfunctioning. If the flow data looks normal and analysis values look good, this can be confirmed as a malfunction. Check the temperature data from nearby sites and or local weather records available online to rule out extreme events. In such cases, we can use the nominal temperature for flow calculations. The temperature source code can be updated to Nominal from the Filters page of the IMPROVE Management Site and reprocess the flow data using the SQL query described in section 9 of TI #351B.

Low Flow (LF) flow status

- Common reasons for the LF status are:
    - Heavy loading which results in a flow value between NM range and CG range as described in Table 1 and Table 2. No action is required in such cases.
    - Swaps between filter types. If consecutive sample dates have an LF flag and the filters are all in the same cartridge, check if any other module is affected by flow fluctuation. If there is, it is suspected that the cartridges have been swapped between the modules. Examine the following dates to see if the flow pattern returned to normal after a sample change. In such cases, the filter statuses can be updated to QD (Questionable Data) from the Filters page of the IMPROVE Management Site for further review after all analyses come back. If the pattern is continuing, request the field group or sample handling lab to contact the site operator to ensure proper installation of cartridges.

Double filter

- Double filters is most commonly observed for 3C Module filters. If low flow (lower range as in Table 1 and Table 2) is isolated to a single day, the cause could be because a double filter was loaded. If the SHL has observed double filters or extra screens on the cassette during download, the QD (Questionable Data) status is applied to the filter. If there is no explanation for low flow, flag the filter with the QD status at this point. In both cases, the data require further review once all analysis results have been received and the data can be processed and validated.

10. DATA AND RECORDS MANAGEMENT

The IMPROVE data are stored in Microsoft SQL Server Databases at UC Davis. The production database is run on a dedicated Windows Server with a RAID array for storage

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

and with offsite backups. Our development and test database environments are virtual machines. To test back up recovery, our development and testing environments are regularly restored from the production backups.

Data management is handled through custom software that interfaces with the UCD IMPROVE database. The primary applications for data ingest and management were developed on the .NET platform. Data processing and calculations were developed as R software packages. In addition, to support data validation and operational monitoring, several interactive visualizations have been developed using the R Shiny platform.

10.1 Disaster Recovery Plan

The scope of recovery activities will depend on the nature of the disaster. Response to an actual disaster may require implementing multiple sections of this SOP.

10.1.1 Facility Recovery

Private security services patrol the laboratory building on a regular basis (including nights, weekends, and holidays). In addition, campus facilities and maintenance staff are on call at all times.

Databases, file servers, and web server virtual and dedicated machines operate primarily out of the Metro IT data center in Hoagland Hall on the UCD campus. Metro IT has a highly-available, disaster recoverable virtualization environment. Weekly backups of the virtual hard drives are taken offsite and stored in the Campus Data Center. In the event of a disaster in Hoagland, critical machines will be mounted at the Campus Data Center. The Drew Avenue laboratory is directly connected to the main campus internet. In the event that connection is disrupted (such as through a construction accident), connections will be switched to a local backup server until service can be restored.

10.1.2 Hardware Recovery Plan

The campus network of IT Administrator staff allows for rapid response to server failure and recovery issues.

10.1.3 Software and Data Recovery Plan

10.1.3.1 UCD Laboratories

Raw and processed analysis data produced with the UCD laboratories are saved and available for use at any time on the computers associated with each instrument, including the PANalytical Epsilon 5 EDXRF, MTL Automated Weighing System (gravimetric mass), Hybrid Integrating Plate and Sphere (HIPS).

Operational flow rate information from samplers in the field is automatically transferred nightly to a file processing server. As a backup, the flow data are stored on SD cards and delivered to the sample handling lab along with the exposed filters.

DocuSign Envelope ID: D2A7C4FB-A926-4DD1-ACCD-8F4006E289EF

Data from all analyses, along with the flows, are scheduled to automatically transfer to a central Microsoft SQL Server database located at a data center on the UCD campus. Differential backups are performed daily, and full backups are performed weekly.

10.1.4 Data Security

UCD access policies: Access to databases and computers associated with this project is limited to authorized project personnel by use of access control lists for files, programs, and database access. Access to laboratory and office space is controlled by keycards.
Password policies: Unique passwords are issued to each employee by the UCD campus system administrator. Password integrity is monitored by the UCD campus system administrator.
Termination policies: System access is revoked for terminated personnel. The IT Administrator disables domain accounts and passwords upon termination of employment.
Virus protection: Microsoft Endpoint Protection is used for virus scanning and protection. All staff are required to complete annual cyber security awareness training.

11. QUALITY ASSURANCE AND QUALITY CONTROL

11.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

11.2 Bug Reporting

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

11.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this TI.

12. REFERENCES

Walter John & Georg Reischl (1980): A Cyclone for Size-Selective Sampling of Ambient Air, Journal of the Air Pollution Control Association, 30:8, 872-876, DOI: 10.1080/00022470.1980.10465122

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

UCD IMPROVE Technical Information #351F

Data Preparation and Reporting

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

November 10, 2022
Version 1.1

Prepared By: DocuSigned by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 11/10/2022

Reviewed By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 11/10/2022

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 11/10/2022

UCDAVIS
AIR QUALITY RESEARCH CENTER

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

DOCUMENT HISTORY

Revision	Release Date	Initials	Section/s Modified	Brief Description of Modifications
1.0	03/14/22	SRS	All	New TI created.
1.1	11/10/22	DEY, ITS	9.2	Updated procedure for compiling report.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

TABLE OF CONTENTS

LIST OF FIGURES

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

1. PURPOSE AND APPLICABILITY

The purpose of this technical information (TI) is to provide guidance on metadata creation, data creation and modification, and compilation of reports. This document is intended to guide users on creating box information in the University of California, Davis (UCD), Interagency Monitoring of Protected Visual Environments (IMPROVE) database and changing box metadata. In addition, the user is guided in the process of adding new site information to the UCD IMPROVE database and the EPA's Air Quality System (AQS) database. The procedure for compiling reports on the current status of the IMPROVE data is also included in this document.

2. SUMMARY OF THE METHOD

Requests are received from the UCD sample handling laboratory to assist with data preparation including creating boxes in the UCD IMPROVE database and modifying existing box information. Boxes are created and modified in the UCD IMPROVE database using the UCD IMPROVE Data Management website.

On a quarterly basis, the current status of the IMPROVE samples is reported to IMPROVE related personnel. The report is compiled using the IMPROVE Status Views website and is in an Excel spreadsheet format.

New IMPROVE sampling sites may start within the network; the site metadata is to be added to the UCD IMPROVE database directly using SQL queries, to AQS directly in the AQS application, if the data from the site is to be delivered to AQS, and the National Park Service (NPS) is to be informed of the new site metadata.

3. DEFINITIONS

- **AQRC:** Air Quality Research Center.
- **AQS:** EPA's Air Quality System database.
- **CSN and IMPROVE Archive (CIA) Database:** A database of the complete record of CSN and IMPROVE data coupled with a web-based visualization and analysis tool.
- **Chemical Speciation Network (CSN)**: EPA's $PM_{2.5}$ sampling network, with sites located principally in urban areas.
- **CIRA:** Cooperative Institute for Research in the Atmosphere.
- **CSV:** a comma-separated value file that is the common format for delivery files.
- **Federal Land Manager Environmental Database (FED):** a database of environmental data managed by Cooperative Institute for Research in the Atmosphere (CIRA)
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- **Interagency Monitoring of Protected Visual Environments (IMPROVE):** Federal $PM_{2.5}$ and PM_{10} sampling network directed by the National Park Service, with sites located principally in remote rural areas.
- **IMPROVE database:** A SQL Server database that is the central warehouse of IMPROVE preliminary and final data at UCD.
- **NPS:** National Park Service.
- **PM:** Particulate Matter. $PM_{2.5}$ is particulate matter with diameters 2.5 micrometers (µm) and smaller. PM_{10} is particulate matter with diameters 10 µm or smaller.
- **SQL:** database management system used by AQRC.
- **UCD:** University of CA—Davis.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. EQUIPMENT AND SUPPLIES

The hardware and software used for IMPROVE data ingest are described in the associated *UCD IMPROVE SOP #351: Data Processing & Validation.*

9. PROCEDURAL STEPS

9.1 Data Preparation Tasks

9.1.1 Box Creation

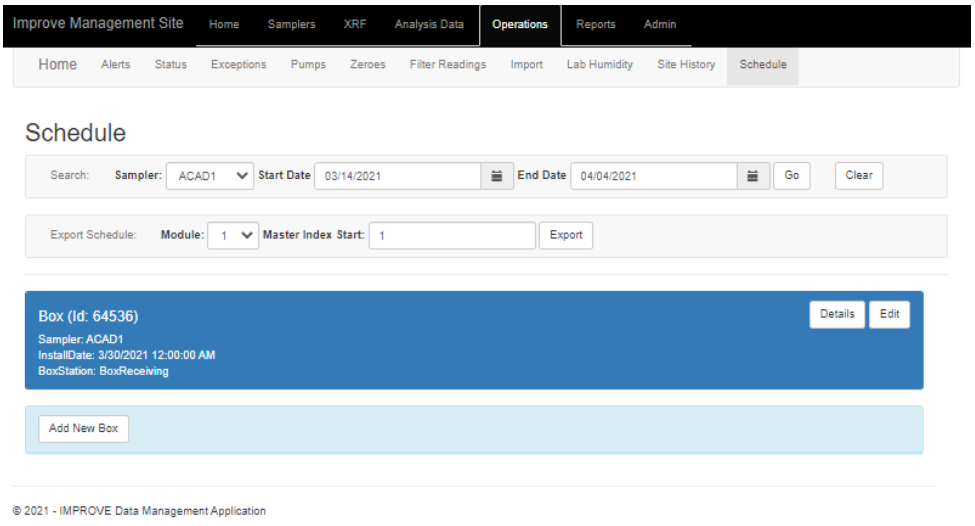
Occasionally, the box sent to the site by the sample handling lab is lost either before it reaches the site or after sampling and before being received back at the sample handling lab at UCD. For boxes that are lost prior to sampling, a replacement box is created and sent to the site as soon as the sample handling lab is alerted to the lost box. If the Data &

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Reporting group are requested to assist in the creation of a new/replacement box, the following tool can be used: https://improve.aqrc.ucdavis.edu/Operations/BoxSchedules, which can be accessed by going to the IMPROVE Management Site, selecting the 'Operations' tab and the sub-tab of 'Schedule'. To create a box the following steps should be taken:

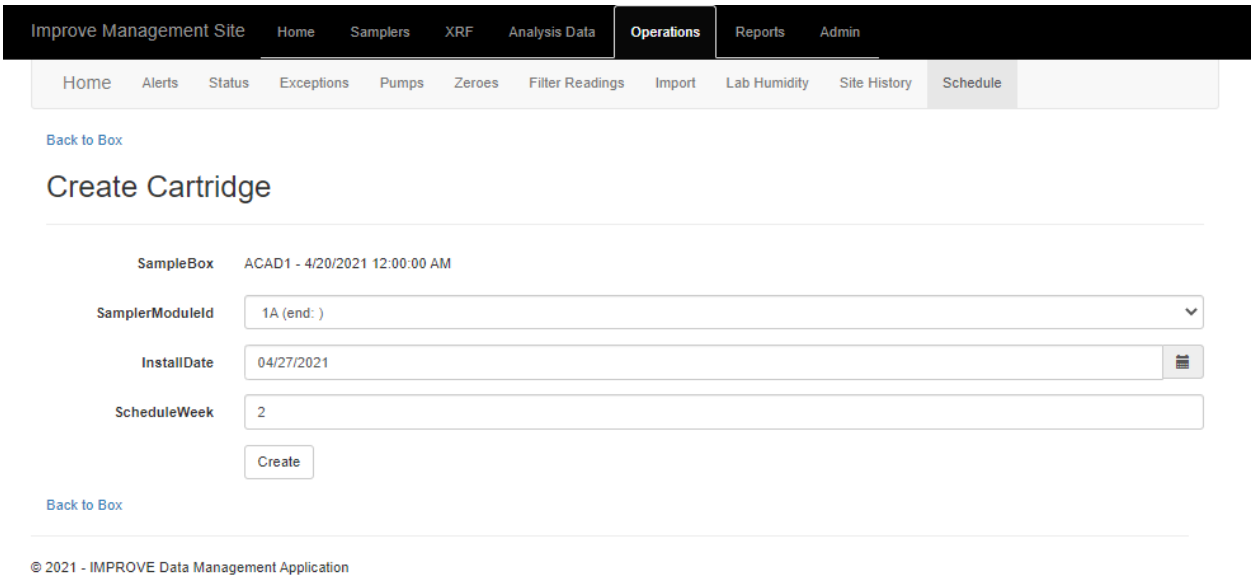
1. Go to the Box Schedules page and select the site via the drop-down menu next to 'Sampler' for which the new box is needed. Click on the 'Go' button to the right (Figure 1).
2. Scroll to the bottom of the page and click on the 'Add New Box' (Figure 1) button on the left.
3. This will lead to a 'Create Box' page as shown in figure XX. Enter the relevant date in the 'InstallDate' option and check 'this is a replacement box' option and click 'Create'. By selecting the replacement box option, the lab station Id will be set to pre weigh.
4. Select 'Add New Cartridge' and on the 'Create Cartridge' page that is subsequently opened, various cartridge information can be added including Sampler Module ID (e.g., 1A, 2B, 3C, 4D), Install Date (this is the cartridge install date and can be found in the IMPROVE calendar), and Schedule Week (i.e., Week 1, Week2, or Week 3). An example of install dates and schedule weeks is as follows: next upcoming box install date is on 04/20. The week 1 installment is 04/20, week 2 is 04/27, and week 3 is 05/04. The cartridge installment is always on a Tuesday regardless of it being a 2-3-2 or 3-2-2 box. Figure 2 shows an example of a Cartridge that is ready to be created. Once all relevant and required information is added, click on 'Create'.

Figure 1. Box creation page.



DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Figure 2. Cartridge creation page.

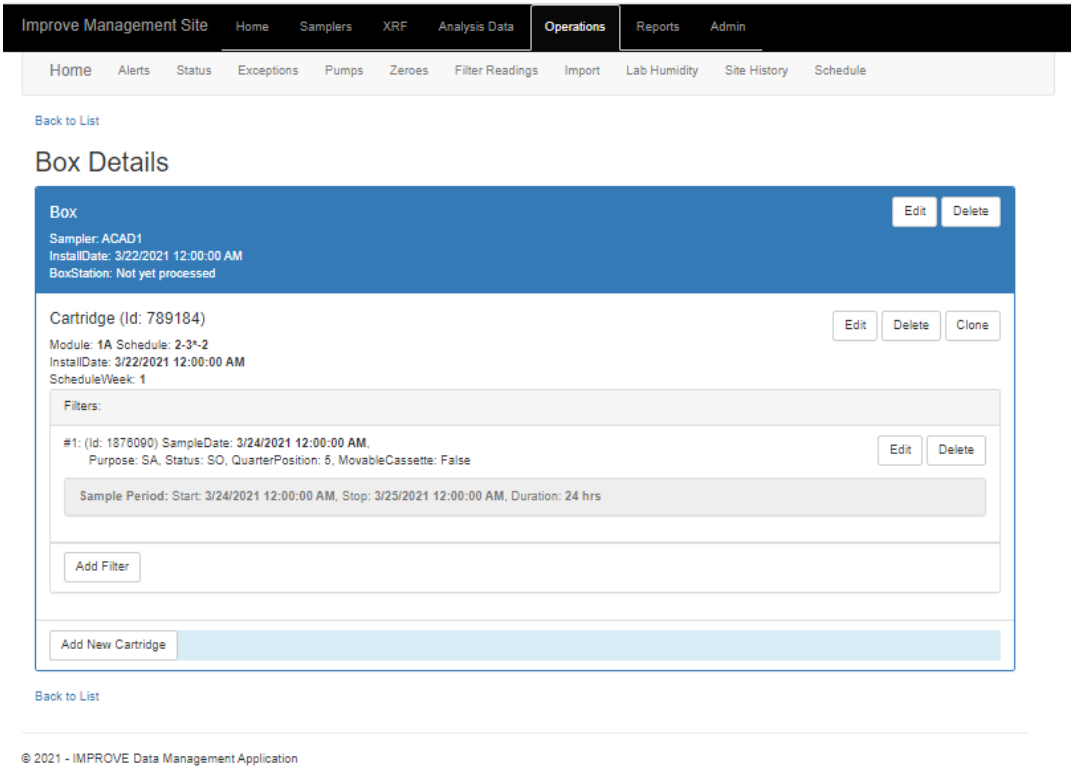


5. Once the cartridge is added, individual filters can then be added, one-by-one, using the Add Filter feature on the page loaded after creating the cartridge. After clicking on 'Add Filter', the 'Create Filter' page is loaded. Every filter added requires the following information to be added: Cartridge Position (1, 2, or 3), Sample Date, Quarter Position (which can be found from the details of the lost box), Lot ID, and an indicator for whether it is a Moveable Cassette ('o-ring'). For the 2-3-2 boxes the Moveable Cassette will always be the third position of the second week. For the 3-2-2 boxes the Moveable Cassette will always be the third position of the first week. All four modules will have the Moveable Cassette on the same date (position); as the cartridges are cloned, this information will automatically be transferred. Once all information for a single filter is filled, click on the 'Create' button at the bottom of the page. A summary of the added filter is then displayed (Figure 3).
6. To add more filters, click 'Add Filter' at the bottom of the summary information. Repeat the instruction from step 5 to add the filter information. The number of filters to be added depends on whether it is a three- or two-position week. The box schedule can be found under the Cartridge Id details (near Module details; Figure 3). For a three-position week, three filters will need to be added and for a two-position week, two filters will need to be added.
7. Once a cartridge is created and all relevant filters have been added, the cartridge can be cloned to create cartridges for other modules by clicking the 'Clone' button to the right of the cartridge information on the Box Details page (Figure 3). The user should select the relevant Destination Module and Destination Lot from the drop-down menus and click 'Create'.
8. Repeat steps 4-7 to create the cartridges for the second and third week.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

9. Once the entire box has been created, the box details are to be send to the sample handling lab to add filter pre-weights and filter barcodes for the PTFE filters. Make sure the current lab station Id (as described in section 9.7.2) is set to 2 so the sample handling lab can assign the pre-weights accordingly. If the box was created in place of a lost box, please proceed to section 9.7.2 for further actions that need to be taken.
10. If an item needs editing or deleting at any point of the box creation, the edit/delete options on the right-hand side can be used accordingly (Figure 3). To delete a box, all cartridges must first be deleted. To delete a cartridge, all filters within the cartridge must first be deleted.

Figure 3. Box creation page; after addition of filter.



9.1.2 Changing Current Lab Station ID and Assigning UF

When a replacement box is created in the case when boxes are lost prior to sampling, there are several additional steps to be performed to correctly assign data and other information to both the new box and lost box.

- For the new box, the sample handling lab assigns filter pre-weights. The current lab station of the box needs to be PreWeigh (Station ID = 2) to enable the sample handling lab to assign the weights. The box creation tool has the option to set the current lab station Id to pre weigh. The field could be updated using the following

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

query if it wasn't selected during box creationit can be changed by running the following SQL update query in the UCD database, where *NewBoxID* is the ID of the newly created box:

UPDATE [Improve_2.1].[filter].[SampleBoxes]

SET CurrentLabStationId = 2

WHERE Id = NewBoxID

Check the update was successful by performing a SQL select query e.g.:

*SELECT **

FROM [Improve_2.1].[filter].[SampleBoxes]

WHERE Id = NewBoxID

- For the lost box, the current lab station needs to be updated to Finished (Station ID = 9). To do this, run the following SQL update query in the UCD database, where *LostBoxId* is the ID of the lost box:

  UPDATE [Improve_2.1].[filter].[SampleBoxes]

  SET CurrentLabStationId = 9

  WHERE Id = LostBoxId

  To check the update was successful, perform a SQL select query e.g.:

  *SELECT **

  FROM [Improve_2.1].[filter].[SampleBoxes]

  WHERE Id = LostBoxID

- For the filters in the lost box, the filter purposes are to be updated to UF (Unused/Lost Filter (Filter Purpose ID = 16) and can be updated using the following SQL update query, where *LostBoxId* is the ID of the lost box:

  UPDATE f

  SET f.FilterPurposeId = 16

  FROM [Improve_2.1].[filter].[Filters] f

  LEFT JOIN [Improve_2.1].[filter].[SampleCartridges] sc ON sc.Id = f.SampleCartridgeId

  WHERE sc.SampleBoxId = LostBoxID

  After updating the filter purpose, review and confirm the filter purpose Id for the whole box is correct by running the following query.

  *SELECT **

  FROM [Improve_2.1].[filter].[Filters] f

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

LEFT JOIN [Improve_2.1].[filter].[SampleCartridges] sc ON sc.Id = f.SampleCartridgeId

WHERE sc.SampleBoxId =BoxID

9.2 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. First, process flow data using the SQL execution code or the Crocker flow processing function as described in section 9 of UCD TI# 351 B and process flows for the relevant date range to be covered in the quarterly field status report by changing the Start Date and End Date fields. Processing flows at this point ensures the most up-to-date flow data and subsequent statuses are reported.
2. 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. In the Site status report tab, update the year in the Terminal samples by quarter and percent complete by quarter columns to the current year. For example, for 2022, it should be A22, B22, C22, and D22.
    - 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]* and *[Improve_2.1].[module].[Modules]* tables in the production database. The *sampler.Samplers* table gives the site installation date, while the *module.*Modules table lists the first sampling date.
3. Populate the report:

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- From the IMPROVE Status page (https://shiny.aqrc.ucdavis.edu/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. The status page has options to choose "filter status only", "filter and flow statuses", and "final statuses (filter, flow, and analysis statuses)". For this report, the user is to only consider filter and flow statuses.
    - Change the Year and Ending Quarter fields to align with the reporting period.
    - Select the "filter and flow statuses" option from the Status Types drop-down menu.
    - Download the full table by clicking on the 'Download' button. A .csv file is downloaded.
    - Compare the columns in the Site Status Report tab to the content of the downloaded spreadsheet; add new columns to the Site Status Report tab as needed.
    - Compare the sampler details in the Site Status Report tab to the content of the downloaded spreadsheet; add sampler details to or remove sampler details from 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 downloaded 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
          Note that the Annual Completeness column should only be colored for the fourth quarter (Q4) report. Report uncolored values for the first, second, and third quarter reports.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- Check formatting for consistency, including font style and type, coloring, and shading.
- Consecutive Terminal Samples may require merged cells 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, check the box to "Include flow statuses".; Change the Year to be relevant.
- Do a search for NF statuses. If any NF statuses are found, process the flows again using the SQL execution code or the Crocker flow processing function. If the NF statuses are for the most recent quarterly period, run the code in SQL or R, changing the Start Date and End Date fields; accordingly, if successful, a date/time of completion will show in the window:

  If the NF statuses are for a small set of filters/sites/dates, confirm why this is the case and edit the SQL execution code or input the appropriate arguments in the Crocker flow processing code to run on the specified filter, date range, and/or site.

- Do a search for no statuses. For sites with no statuses, determine if it is a new site or a site that's closed or if there is a reason such as paused shipments or the site temporarily offline.
  If the site is new or closed, there may be blank records prior to the start date/ end date; if so, leave as-is but make sure the site is not falsely reported as failing the Regional Haze Rule criteria. Also check and ensure that the annual completeness column in Site status report tab is reported correctly.
  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 are found, the data can be downloaded to be included in the report. There is a checkbox option defaulted to include the collocated module data on the network timeline page. As the Network Status page does not include statuses of filters from collocated modules, the data downloaded from the Network Timeline page should also only use filter statuses from routine modules. To do this, unselect the 'Include collocated modules' option and download the data by clicking on the 'Download' button. The default name of the downloaded .csv filter is 'IMPROVE_network_timeline.csv'. As the PHOE5 site is a

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

collocated site with all four modules, it is considered an independent site in the Network Status table. To include the data from this site in the quarterly report, re-select the 'Include collocated modules' option and download the file by clicking on the 'Download' button. The file name is 'IMPROVE_network_timeline_cl.csv'. Copy the line containing the PHOE5 information from this file and paste/insert it into the IMPROVE_network_timeline.csv file in the row under PHOE1.

- Compare the site list from the IMPROVE_network _timeline.csv with the sites listed in the Sampler Locations and Status Flag Table tabs in the Site Status Report tab. Update site and date information as needed after confirming with the Lead QA Officer; some sites are for special studies and are not included in this report.
- Copy/paste all content from the IMPROVE_network_timeline.csv to the Status Flag Table tab.
- Color the relevant fields:
  - QD, TO, CG, and LF flags, yellow
  - Null/terminal flags, red

4. 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.
5. Send to the Data & Reporting Group Manager for review.
6. Once reviewed and approved, the Data & Reporting Group Manager will deliver to various personnel including IMPROVE site operators, NPS, and EPA staff via email. A summary of site losses is to be included in the body of the email.

9.3 Adding AQS Site Information

Whenever a new site starts within the network and the data will be delivered routinely to AQS, both the UCD database is to be updated with AQS related information and the AQS database is to be updated to add the new site and its associated monitors (where monitors are the AQS term for parameters). The site details should also be sent to NPS prior to data delivery when a new site starts sampling.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

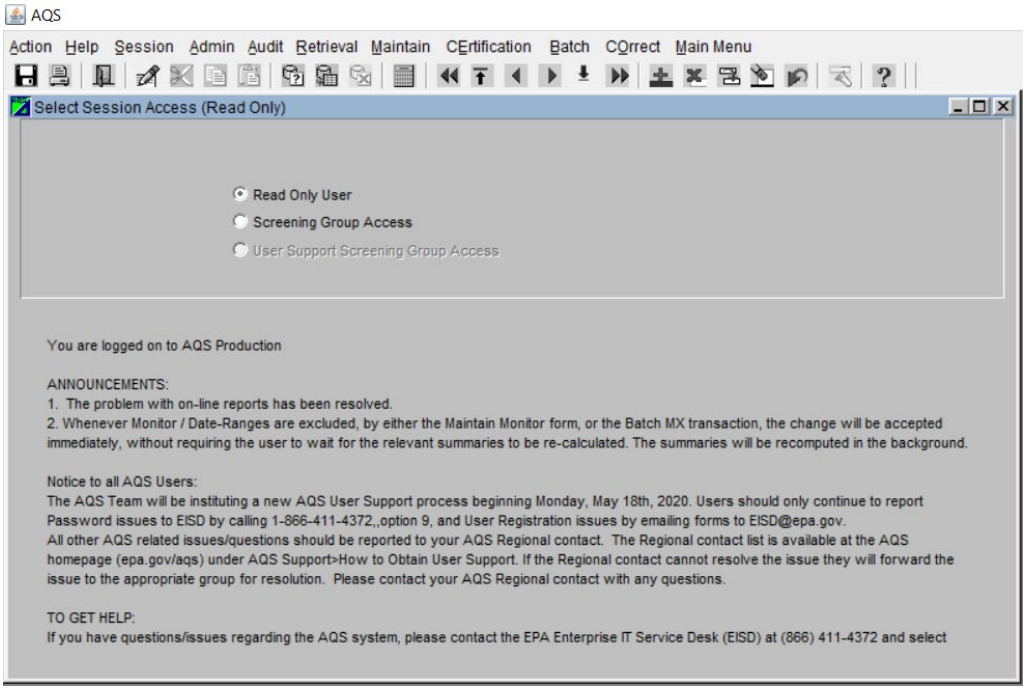
9.3.1 Updating UCD Database

9.3.1.1 AQS Site ID

An AqsSiteId needs to be assigned to the new site in the sampler.Samplers table in the UCD database (Improve_2.1), which consists of State Code, County Code, and Site Id. Further information can be found in sections 3.2.3-3.2.5 of the AQS Data Coding Manual (https://www.epa.gov/sites/production/files/2015-09/documents/aqs_data_coding_manual_0.pdf). If the state and county where the site is located are known, then the associated codes can be found by searching FIPS online (Federal Information Processing Standards/Series, e.g. https://geonames.usgs.gov/apex/f?p=138:1:::NO:1:P1_SHOW_FIPS55,P1_SHOW_ADV,P1_SHOW_ANTAR:Y,,). The user should query the AQS database to determine if a site already exists at the same location. To do this, the user should follow these steps:

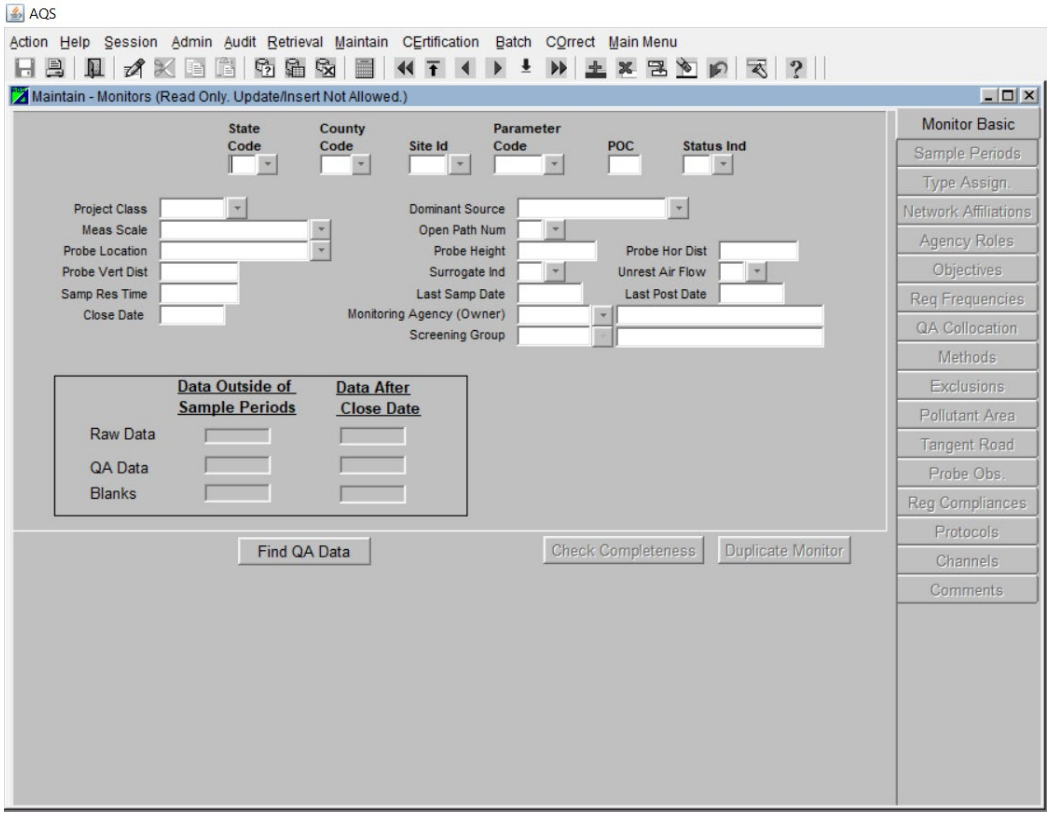
- Log into the AQS application
- Select the Read Only User option (Figure 4)
- Go to Maintain
- Select Monitor (Figure 5)
- Type in the State Code and County Code
- Click on the 'Execute Query' button.

Figure 4. AQS screen after logging into the application; select the Read Only User option.



DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Figure 5. AQS screen after selecting Monitor from the Maintain option.



The site does not already exist in AQS if no results are returned. If any results are returned, the user should review the details to confirm if the parameter code and/or the POC is one that is reported as part of the IMPROVE network. By convention, if the site does not already exist in AQS, the Site Id assigned is '9000'. The user should repeat the steps above to include the Site Id of '9000' and executing the query again to confirm the site and/or monitors does not already exist.

Another way to check what site IDs are already in use and thus what are 'available' is to review the 'Site Listing' spreadsheet via https://aqs.epa.gov/aqsweb/airdata/download_files.html#Meta. Note that this is updated only every 6 months so it may not be the most up-to-date. By convention

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

CDPHE uses the lower site numbers and NPS has used the 9000 series. Reach out to NPS, CIRA and air-resources to confirm what site Id's are already in use.

If 9000 is already in use, get the next available count, make sure it's not already in use, and confirm with various agencies (eg. CDPHE) that we are not in conflict with anyone.

Once the State Code, County Code, and Site Id are known, the UCD database can be updated as follows:

- Query the database using the following SQL query to find the relevant site record, where 'XXXX#' represents the four-character site name plus the number, typically 1.

  *SELECT **

  FROM [Improve_2.1].[sampler].[Samplers]

  WHERE Name = 'XXXX#'

- Update this site record with the newly generated AQS site ID, a nine-digit ID comprising the State Code, County Code, and Site ID, with no separation.

  UPDATE [Improve_2.1].[sampler].[Samplers]

  SET AqsSiteId = '#########'

  WHERE Name = 'XXXX#'

9.3.1.2 AQS Parameters and POCs

In addition to updating the Samplers table in the database with the AQS site ID, POCs (Parameter Occurrence Code) need to be added to the analysis.AqsPOCs table. In AQS, POCs are assigned per parameter. If there was no existing site in AQS, POC = 1 for all parameters, except for the coarse mass parameter (PM_{10}-$PM_{2.5}$), which is assigned POC = 5 by convention. If there are existing collocated sites in AQS, the next smallest different number is to be used, e.g. POC = 2.

To add the parameters and POCs to the database, specifically the *analysis.AqsPOCs* table, a SQL insert query can be written using the starting format below where each set of values is for a different parameter:

INSERT INTO [Improve_2.1].[analysis].[AqsPOCs] (SamplerName, ObjectiveCode, Parameter, POC)

Values (SamplerNameX, ObjectiveCodeX, ParameterX, POCX), (SamplerNameX, ObjectiveCodeX, ParameterY, POCX), …

Alternatively, to add to the database in bulk, an R script can be written and used, ensuring that the outputs from each step of the script is reviewing along the way.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

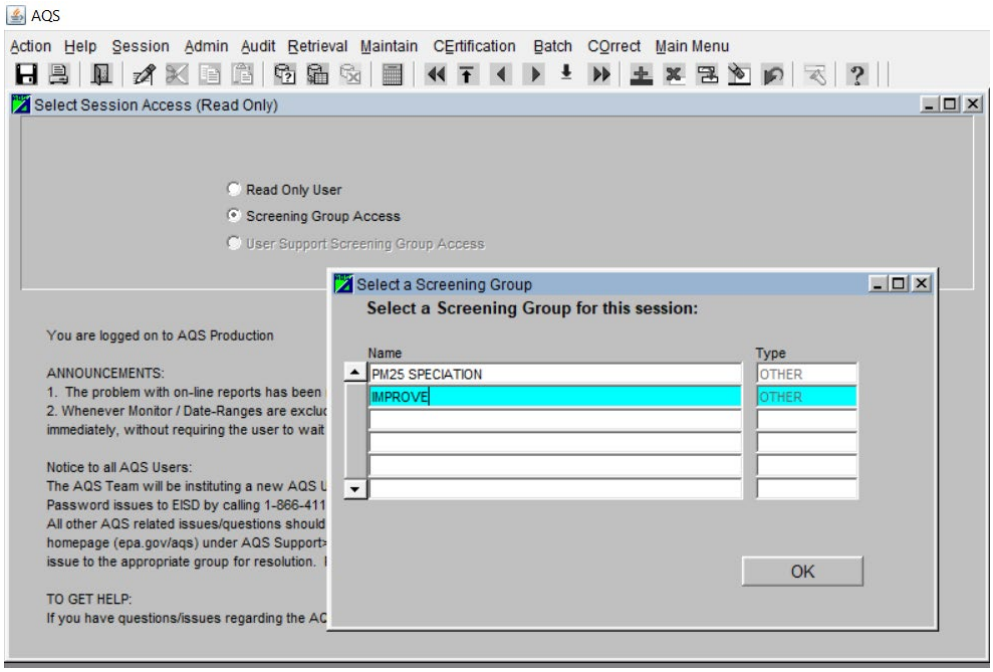
9.3.2 Adding Site and Monitors to AQS

9.3.2.1 Adding a New Site to AQS

To add a new site to AQS the user should follow the steps below:

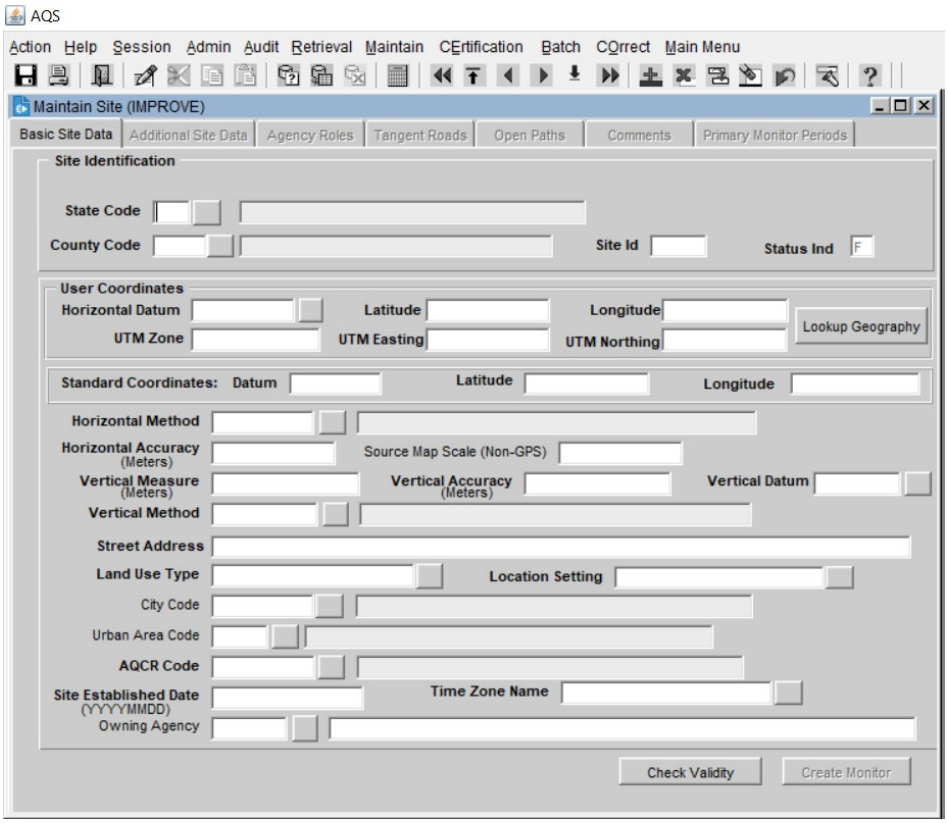
- Log into the AQS application
- Select the IMPROVE Screening Group Access option (Figure 6)
- Go to Maintain
- Select Site
- Click 'Cancel Query' (Figure 7).
    - This allows the user to click on the 'Check Validity' button at the bottom of the window once various details have been entered.

Figure 6. AQS screen after logging into the application; select the Screening Group Access option and then IMPROVE.



DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

Figure 7. AQS screen after selecting Site from the Maintain option and clicking Cancel Query.



The fields to be completed are detailed below:

- Site Identification
  - The State Code, County Code and Site Id should all be known from the previous section. Enter the codes; the associated names will fill automatically.
- User Coordinates
  - Horizontal Datum: WGS84 (by convention).

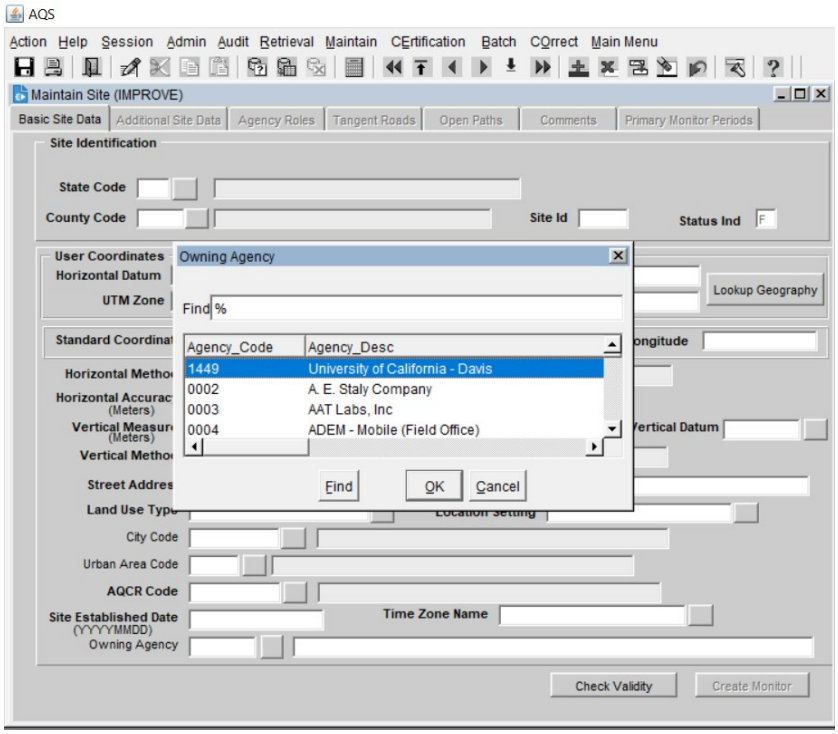
DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- Latitude and Longitude: find in the IMPROVE Data page, under Sites, select the relevant site. Note that W longitude should be entered as negative. i.e., 105 W is entered as -105.
- Horizonal Method: 103(GPS) (by convention).
- Horizontal Accuracy: 5 meters (by convention).
- Source Map Scale (Non-GPS): 10000 (by convention)
- Vertical Measures: Site elevation; find in the IMPROVE Data page, under Sites, select the relevant site.
- Vertical Accuracy: 5 (by convention).
- Vertical Datum: NAVD88 (by convention).
- Vertical Method: 001 (by convention).
- Street Address: If not already known, the site operator may have to be contacted to obtain this information.
- Land Use Type: If not already known, the site operator may have to be contacted to obtain this information. Options for this field are: Residential, Commercial, Industrial, Agricultural, Forest, Desert, Mobile, Blighted Areas, Military Reservation.
- Location Setting: If not already known, the site operator may have to be contacted to obtain this information. Options for this field are: Urban and center city, Suburban, Rural.
- AQCR Code: Use the drop-down menu, select the code listed; there should only be one.
- Site Established Date: Find in the sampler.Samplers table in the UCD database.
- Time Zone Name: Use the drop-down menu, select the option listed.

Owning Agency: If not already known, the site operator may have to be contacted to obtain this information. To search for the agency and obtain the relevant agency code, use the drop-down menu and in the 'Find' box, type in the details of the agency (Figure 8) and select the appropriate affiliation from the returned results.

Figure 8. AQS screen when searching for Owning Agency details.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF



Once all the fields have been entered, the user should click 'Check Validity' at the bottom of the screen. A 'Row_Errors' window may appear with Error Descriptions left blank. If not blank, the user should return to the Site's window and correct the errors. Upon completion, the details are saved by clicking on the save button at the top left of the AQS window (under 'Action'). Follow any additional prompts, e.g. click 'Lookup Geography', and save again.

For confirmation the site is saved, the user should navigate to a fresh Maintain Site window, enter the State Code, County Code, and Site Id and execute query. The full site details should be displayed and the Status Ind should have changed from 'F' to 'P', meaning the site is now in production.

9.3.2.2 Adding Monitors to AQS

In addition to opening the parent site, the monitors (parameters) need to be opened. There are two methods for adding monitors: Batch processing (preferred because of high efficiency) and manually adding monitors, one at a time.

For batch processing, the user should follow these steps:

- Navigate to the template text file, monitor_template.txt, at U:\IMPROVE\AQS\AQS_Documentation.
- Save this as a new file, with a file name that indicates which site it is for.
- Update the State Code, County Code, Site Id, and POC, if necessary by performing a 'Find and Replace' in Notepad (or other application).

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- If this is a new site, then the only date in the file should be updated to be the site start date. If the site is being closed, then an end date needs to be added as well (this field is currently blank in the template).
- Navigate to the Exchange Network website and submit the file in the same way that raw data is delivered as described in section 9.4.2.

To add monitors manually, one at a time, the user should follow these steps:

- After adding the site to AQS as detailed in section 9.6.2.1, the user should ensure they are in the IMPROVE Screening Group Access session and navigate to the Maintain Monitors window (Figure 5).

The following details the information that must be added to each specified tab to open a parameter (monitor):

- Monitor Basic: enter State Code, County Code, Site Id, Parameter Code, and POC for the parameter that is being opened.
- Sample Period: enter the date used for 'Site Established Date' when creating the site in the 'Begin Date' field.
- Type Assign: enter 'EPA' as the Monitor Type and Begin Date is the same date as the Site Established Date.
- Network Affiliations: enter 'IMPROVE' for the Monitor Network Code and the Begin Date.
- Agency Roles: enter a row each for Agency roles of ANALYZING, COLLECTING, REPORTING and PQAO, list the Agency Code as '0745', and the Begin Date. If the Site Established Date is before 2007-01-01, the Begin Date should be entered as '20070101'.
- Objectives: select 'GENERAL/BACKGROUND and enter '0000' for the UA Represented field.

The user should save the entry and confirm that the 'Status Ind' in the Monitor Basic tab is 'P'. To add more monitors for the same site, the user should click on the 'Duplicate Monitor' option at the bottom of the Monitor Basic tab and enter the appropriate parameter code details. The user should follow these steps until all relevant monitors are opened.

9.3.3 Updating NPS

The NPS needs to have the site details for any new site that starts sampling in advance of data delivery. The NPS requires the following information be sent:

- Site Name
- State
- County
- AQS Code
- Latitude

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

- Longitude
- Elevation
- Start Date
- End Date (assumed to be blank at present)
- Sponsor/Agency
- Location Description
- Rural or Urban (or other demographic code)
- Land Use code, if any
- Photos of the site.

9.3.4 Updating UCD-CIA Database

Whenever a new sampler is added, the sites table in the UCD-CIA database needs to be updated. The request should be directed to the software group. The software group uses the EPA's master AQS site list to add information to the database. If a new sampler is not added to the UCD-CIA database prior to data submission, the data will stage but will not migrate. Once the relevant sampler information is added to the database, the staged data will successfully be posted to the UCD-CIA database when the SQL query to migrate data is next run (typically every night).

10. QUALITY ASSURANCE AND QUALITY CONTROL

10.1 Code Development

Software for data management, processing, and validation is developed in-house by professional software engineers. Source code is managed through a code repository. Development of code changes and new applications is conducted on a development environment that parallels the production environment. Prior to deployment in production, all code changes undergo testing within a separate test environment. The testing, which is conducted by developers, managers, and users, is targeted both at the identification of software bugs and the confirmation of valid data equivalent to the production system.

10.2 Bug Reporting

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

10.3 Data Validation

Data integrity is enforced within the UCD IMPROVE database via unique primary keys and non-nullable records. Data completeness and data quality are thoroughly checked through the data validation process, as described elsewhere in this SOP.

DocuSign Envelope ID: 5C00D2FC-D495-4089-987F-71027B7FA5EF

11. REFERENCES

UCD IMPROVE SOP #351: Data Processing and Validation