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

1. Purpose and Applicability 5
2. Summary of the Method 5
3. Definitions 5
4. Health and Safety Warnings 6
5. Cautions 6
6. Interferences 6
7. Personnel Qualifications 6
8. Equipment and Supplies 6
9. Procedural Steps 6
   9.1 Definition of Status Flags 8
   9.2 Level 1 Validation Procedures 10
      9.2.1 Level 1B Checks 10
   9.3 Level 2 Validation Procedures 15
      9.3.1 Cross-Module Comparison 15
      9.3.2 Long-Term Network-Wide Checks 26
      9.3.3 Common Validation Findings 31
      9.3.4 Analysis Level Flagging Validation Guidelines 47
      9.3.5 Recommended Validation Guidelines 54
      9.3.6 Final Review 55
10. Quality Assurance and Quality Control 57
11. References 57

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

LIST OF FIGURES

Figure 1. Screen shot of the IMPROVE Data website Field Blanks tab. 12
Figure 2. S3/SO4 comparison plot for the GRGU1 site 16
Figure 3. Comparison plot of light absorption coefficient measurements 19
Figure 4. Time series plot of PM_{10} and $PM_{2.5}$ masses and their ratio at BOWA1 site. 23
Figure 5. Time series plot of RCMC, RCMN, and $PM_{2.5}$ concentrations 25
Figure 6. Time series for RCMN versus Fine mass at EVER1 site. 25
Figure 7. Scatter plot of sulfur (×3) versus sulfate for the entire IMPROVE network. 27
Figure 8. Scatter plot of chlorine versus chloride for the entire IMPROVE network. 27
Figure 9. Scatter plot of chlorine versus chloride for the entire IMPROVE network. 28
Figure 10. Scatter plot of fAbs versus BC for the whole network. 28
Figure 11. Scatter plot of fAbs versus EC for the whole network. 29
Figure 12. Scatter plot of fAbs versus EC for the whole network. 29
Figure 13. Scatter plot of ECTR versus other wavelengths for the whole network. 30
Figure 14. Ratio of $PM_{2.5}$ mass (1A) over PM_{10} mass (4D) at ACAD1 site 30
Figure 15. Multi-year monthly 10% percentile, median and 90% percentile 31
Figure 16. Filter swap page. 32
Figure 17. Analysis swap page. 35
Figure 18. Analysis swap page. 38
Figure 19. Box swap page. 40
Figure 20. Box swap page after clicking the ‘Update’ button 41
Figure 21. Filter details page. 49
Figure 22. GRAV (Filter Mass; a) analysis page and detail (b, c) 49
Figure 23. XRF sample analysis page 51
Figure 24. HIPS sample analysis page. 51
Figure 25. Ions sample analysis page. 51
Figure 26. Carbon sample analysis page. 52
Figure 27. Analysis details page. 52
Figure 28. Analysis flags page. 53
Figure 29. Deleting flags 54

LIST OF TABLES

Table 1. Status flags and their definitions. 9
Table 2. Details of the IMPROVE database table names 13
Table 3. Replicate criteria for carbon analysis results. 19
Table 4. Qualitative checks and criteria for carbon (OC, EC, and TC) validation. 19

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.

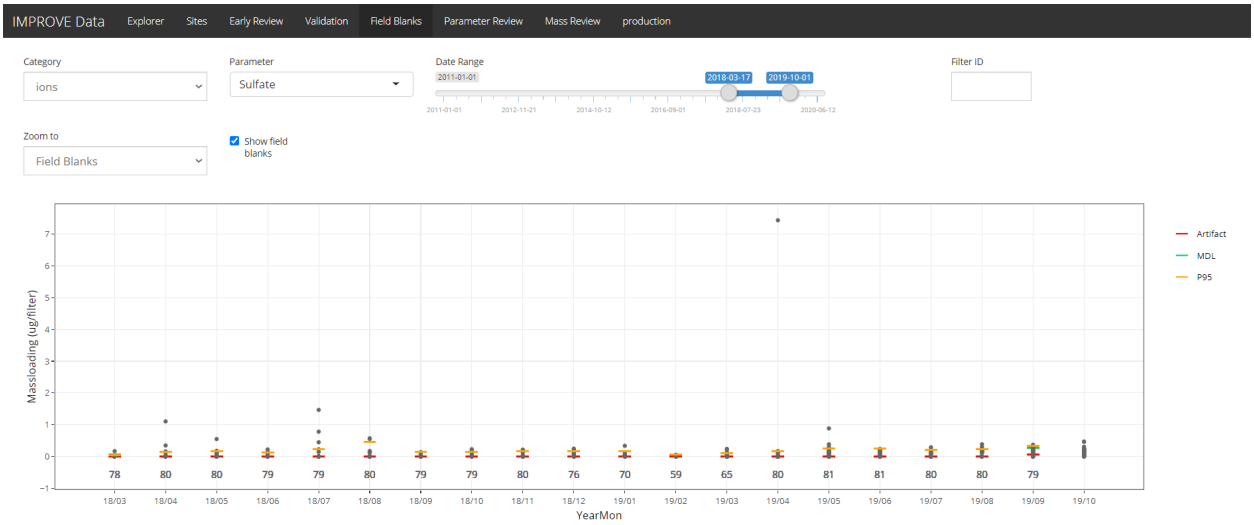
  For all analysis types (ions, carbon, elements, and mass), field blank data across the network can be compared using the Field Blanks tab in the IMPROVE Data website (https://shiny.aqrc.ucdavis.edu/ImproveData/; Figure 1). The mass loading of a specified parameter should be compared to field blank data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

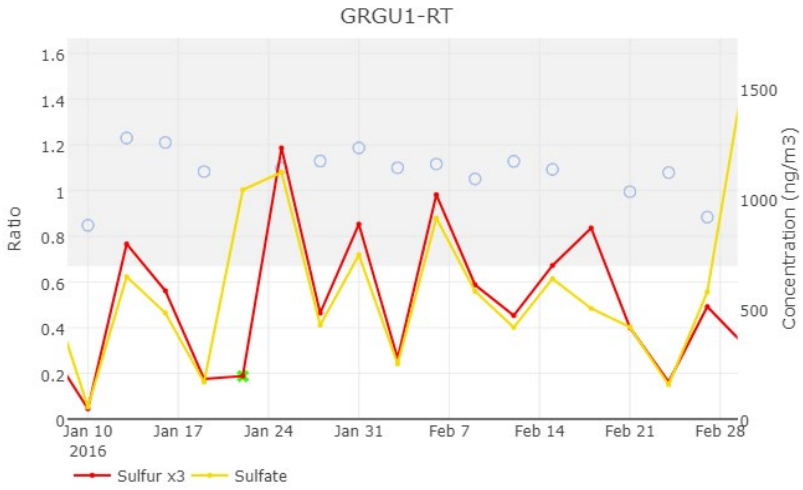
$$Index1 = \left(\frac{S3_1}{SO4_1} - 1\right) \times \left(\frac{S3_2}{SO4_2} - 1\right) \qquad Index2 = \left(\frac{S3_1}{SO4_2} - 1\right) \times \left(\frac{S3_2}{SO4_1} - 1\right) \qquad \text{(351C-1)}$$

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.

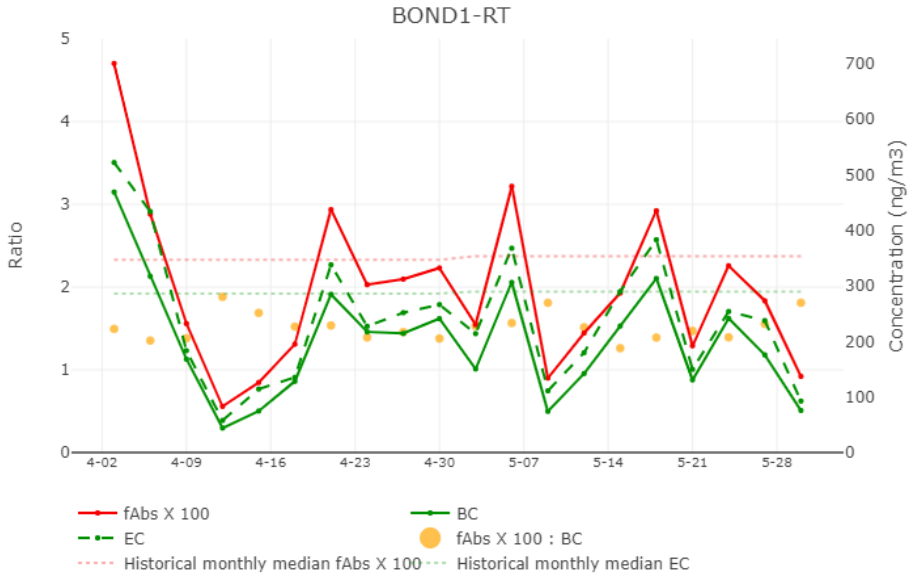


Table 3. Replicate criteria for carbon analysis results.

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'])

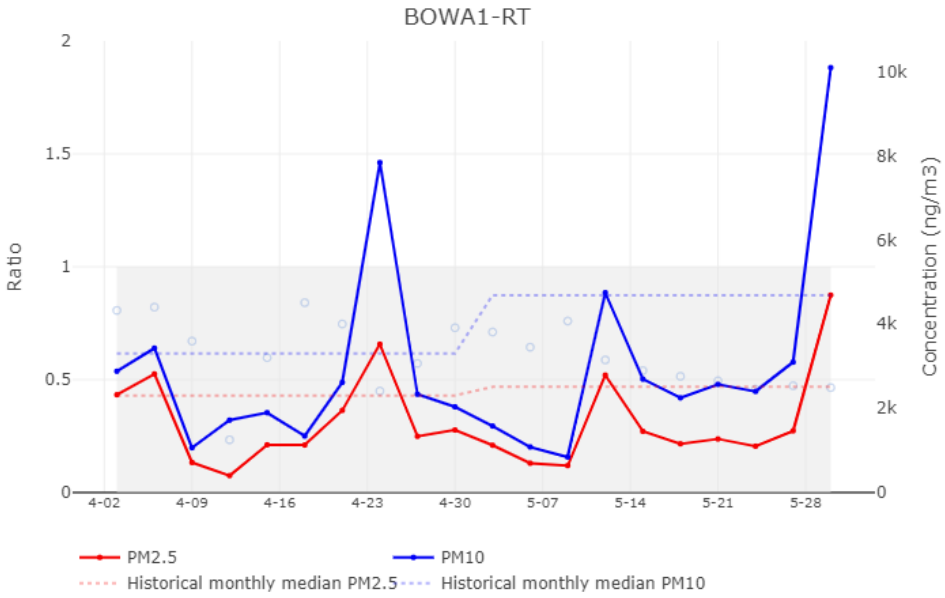
where *inputpath* is the file path and file name of the reweigh list and *input_sheet* denotes the relevant sheet within the reweigh list spreadsheet. The user can specify the name and location of the output file (*output_path*) as well as the sheet name (*output_sheet*), where the default sheet name 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".

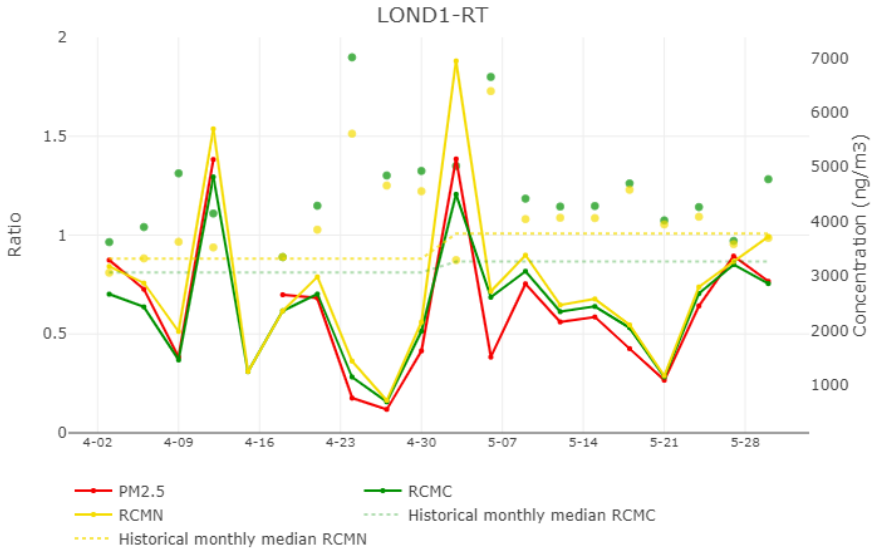
Z scores are calculated as follows:

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

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.

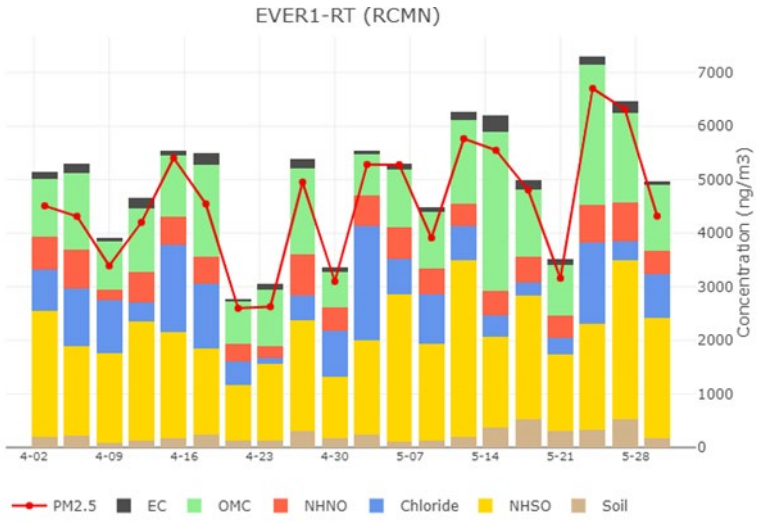


RCMN is also plotted as a bar plot (Figure 6), along with the $PM_{2.5}$ time series, for comparison of RCMN and $PM_{2.5}$ concentrations and to enable the contributions from the various species to be viewed and evaluated.

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.

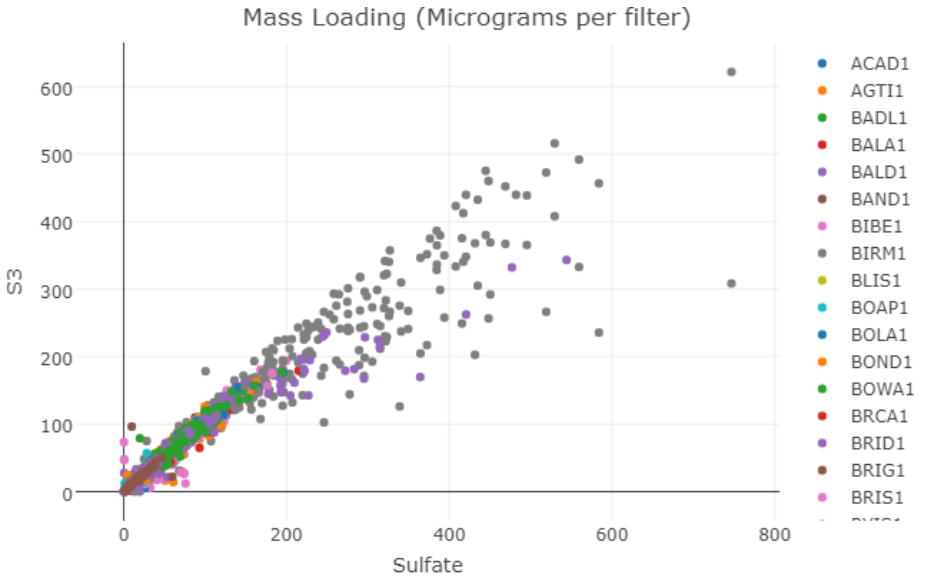
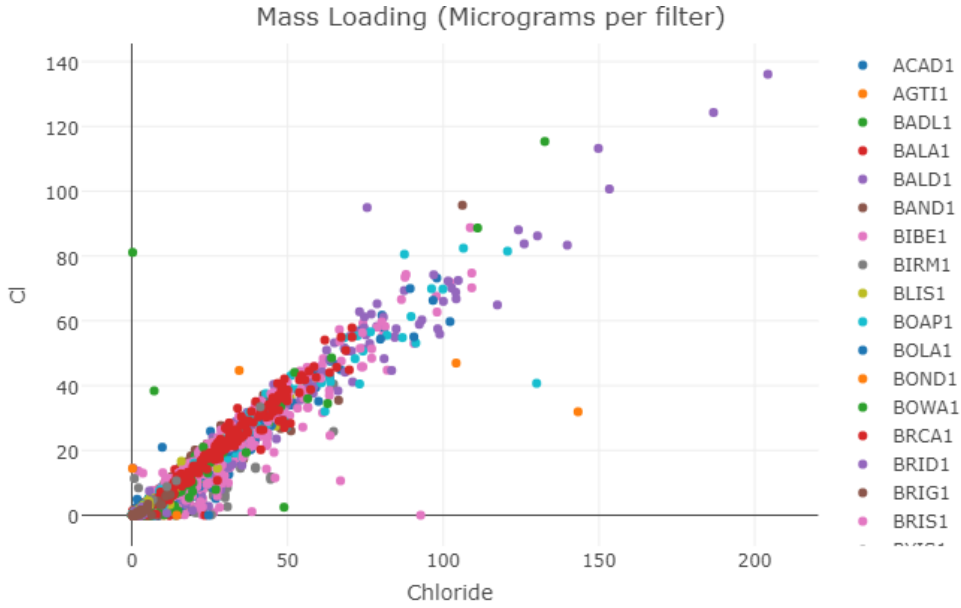


Figure 8. Scatter plot of chlorine versus chloride for the entire IMPROVE network.



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

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

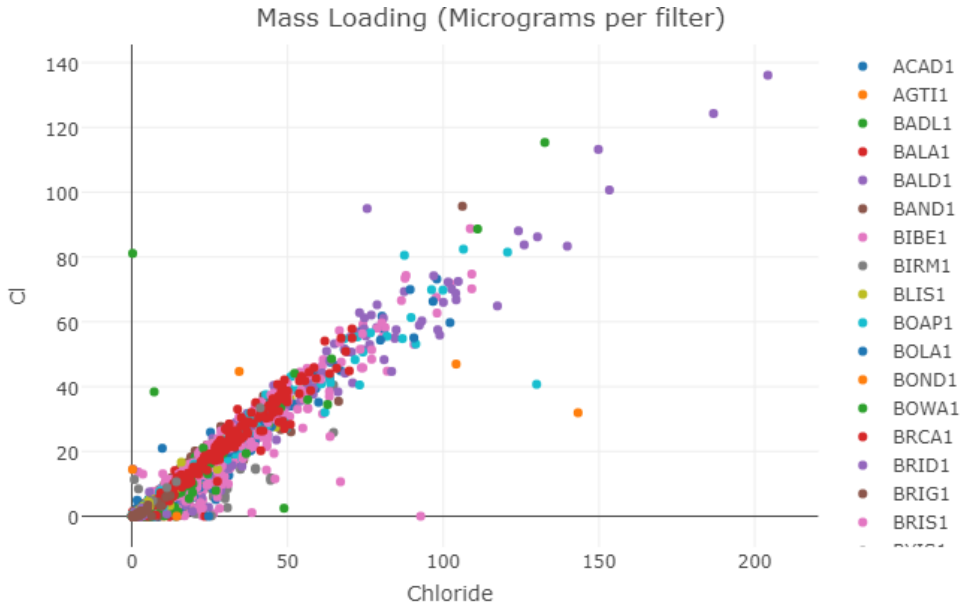
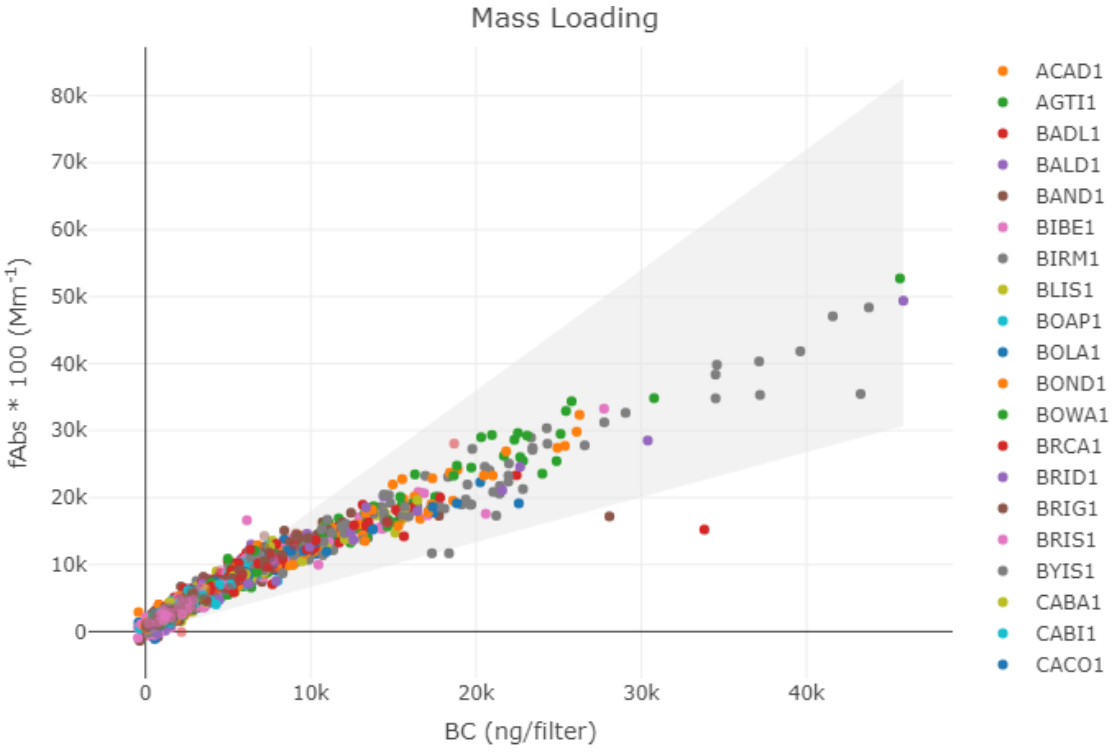


Figure 10. Scatter plot of fAbs versus BC for the whole network.



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

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

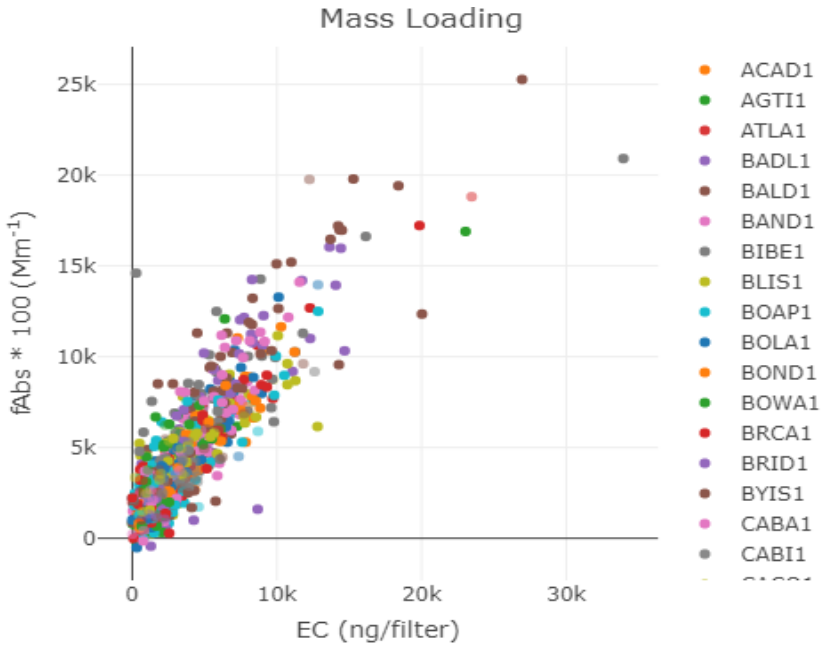
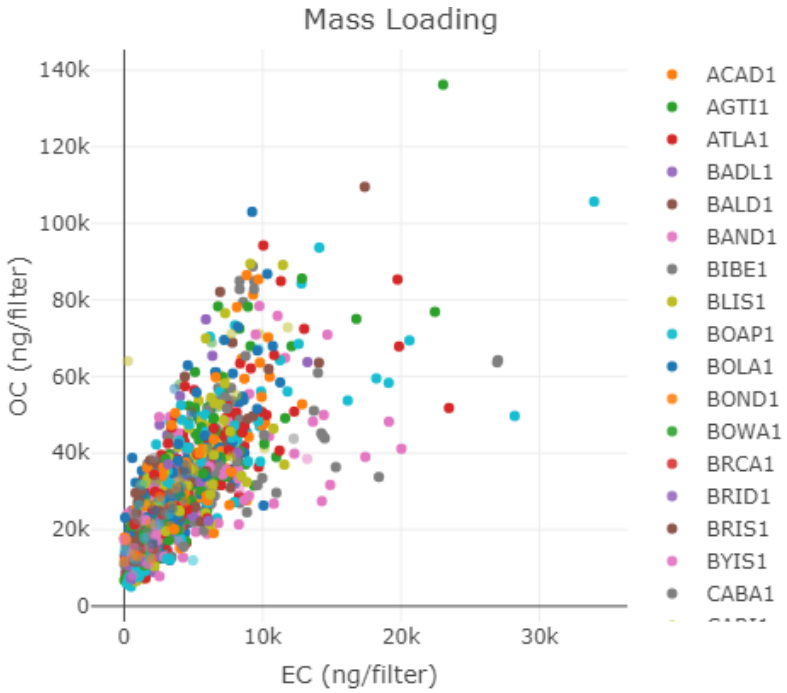


Figure 12. Scatter plot of fAbs versus EC for the whole network.



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

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

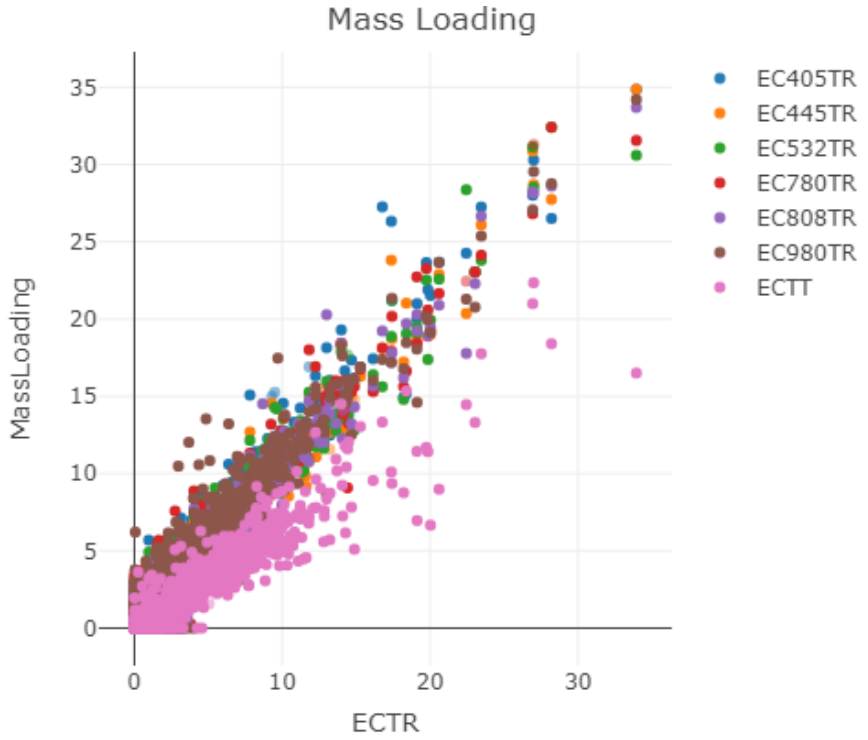
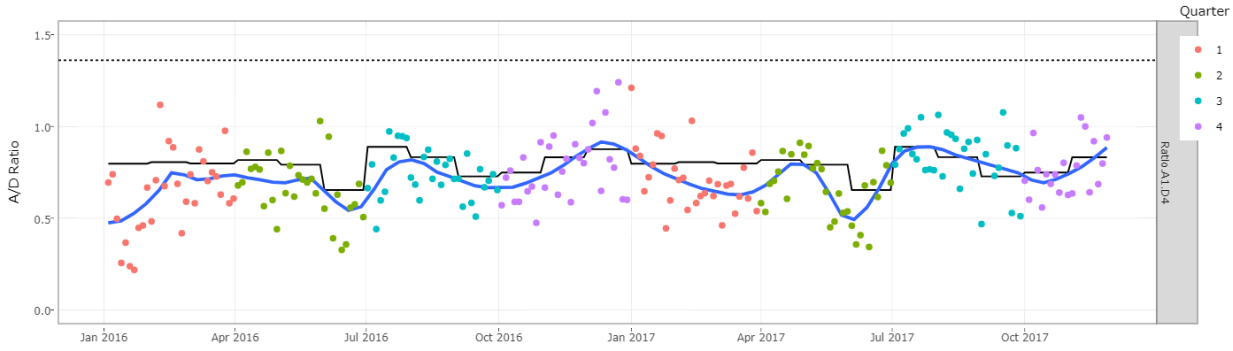
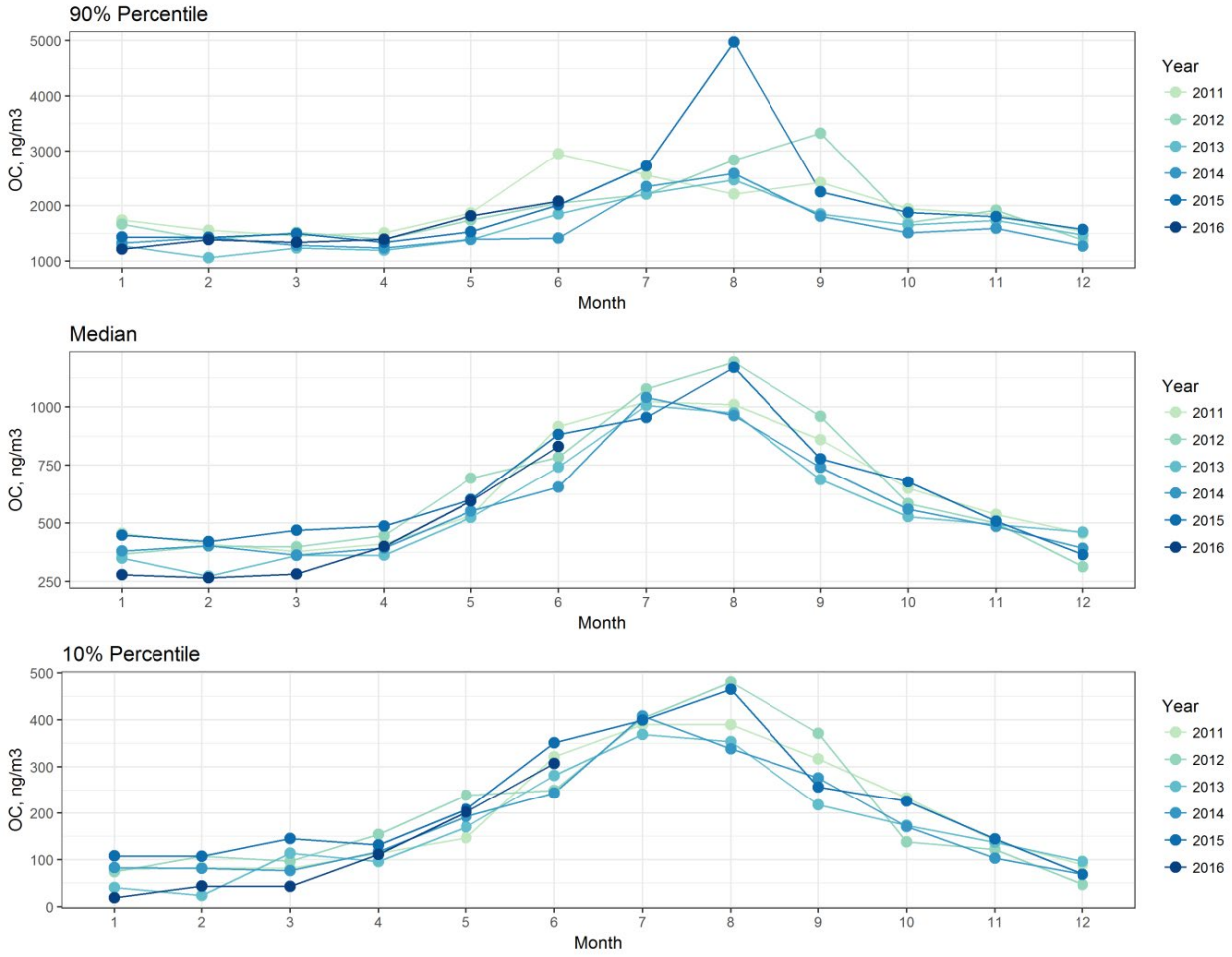


Figure 14. Ratio of $PM_{2.5}$ mass (1A) over PM_{10} mass (4D) at ACAD1 site, represented as raw measurements not adjusted for flow rates. Points are individual sample days (pink = Q1, green = Q2, blue = Q3, purple = Q4). Black line is the multi-year monthly mean. Blue line is the locally weighted average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

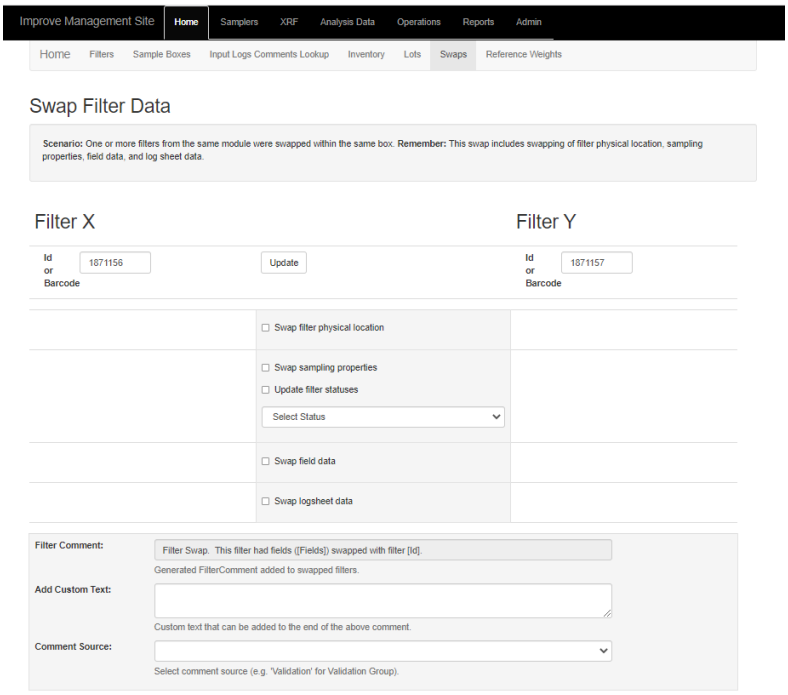
- A routine sample filter was swapped with a field blank filter.
- A routine sample filter was swapped with a collocated sample filter.
- One or more of the same module filters were swapped within the same box (sample date swap).
- A 1A filter was swapped with a 4D filter (uncommon).
- The cartridge 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:

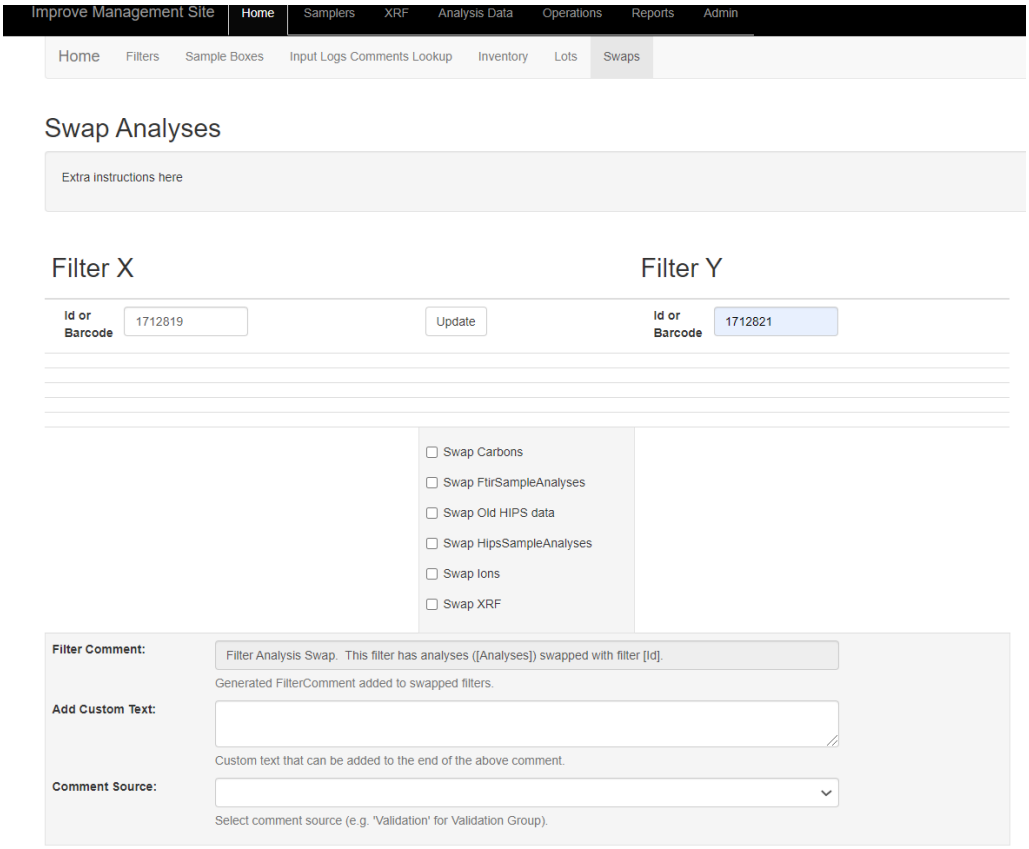
1. Confirm the swap happened for the same module and identify which analysis data are swapped; if multiple analyses are performed on that filter, which is the case for A module filters, identify which sets of analyses have been swapped. Usually, if the filters are swapped at a lab station, all downstream 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.

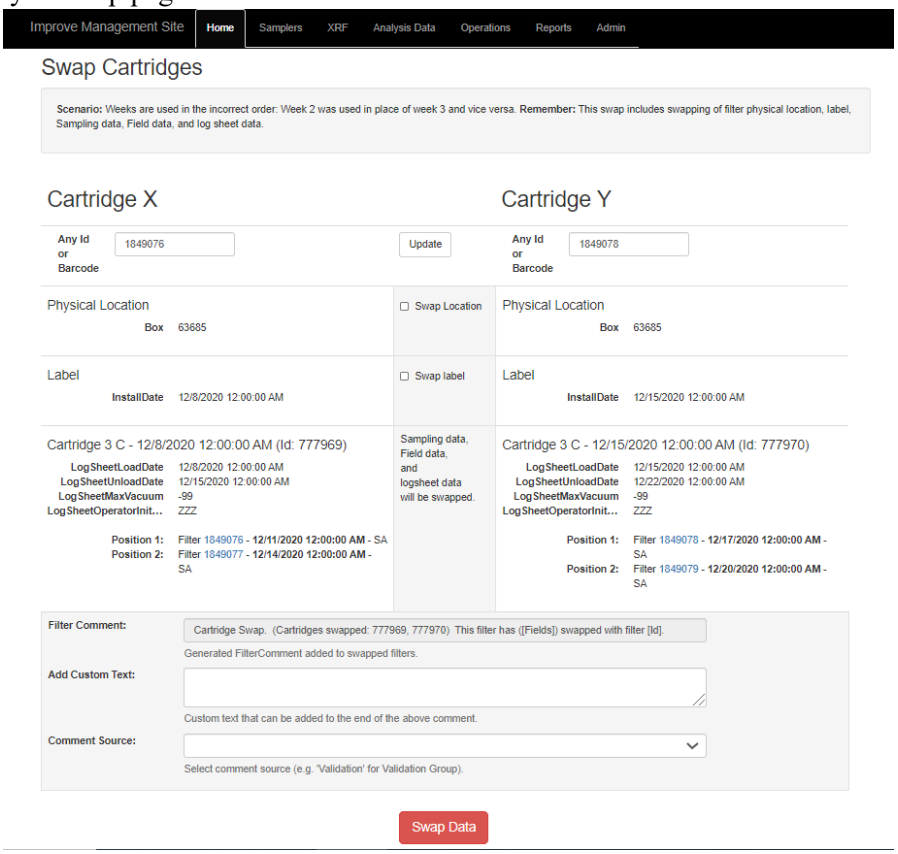
To perform the swap using the swap tool for cartridges, each cartridge pair swap will have to be performed one at a time. A cartridge swap can be performed only if both cartridges have the same number of filters, except for cartridges with field blanks.

- **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.

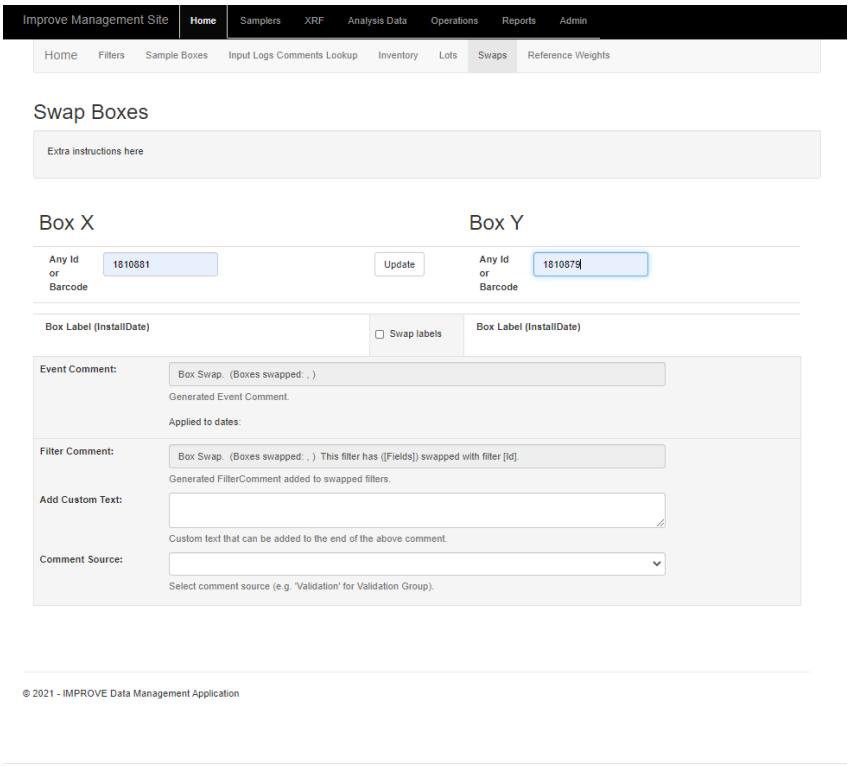
For all of the above examples, the swaps can be performed using the steps outlined below.

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

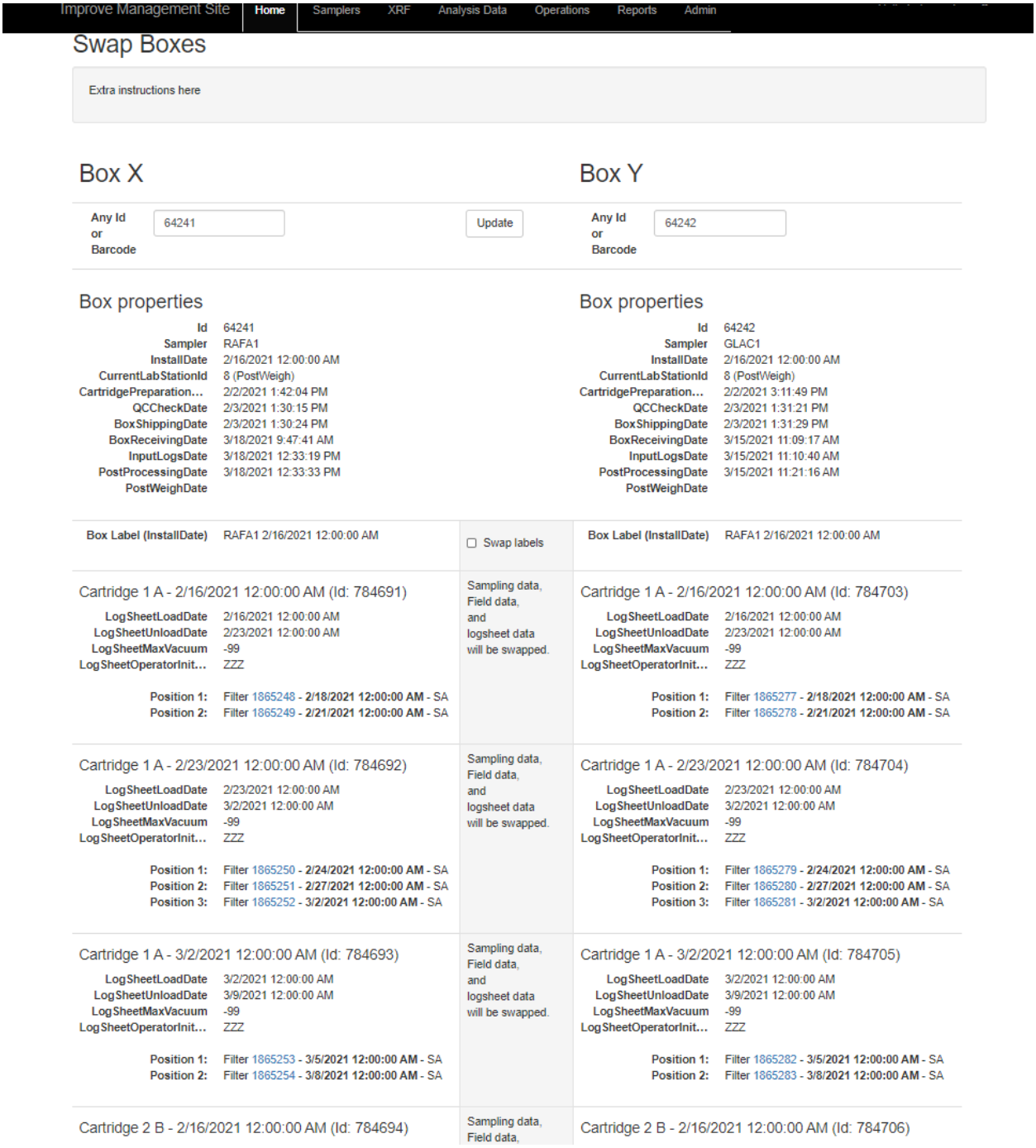
5. **Step 5**: After a box swap is performed, the statuses of all filters in the boxes need to be addressed based on the swap situation. If the box is swapped in place of a lost or undelivered box (example 'a' in the above section), please refer to section 9.1.2 of *UCD IMPROVE TI #351F: Data Preparation 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.

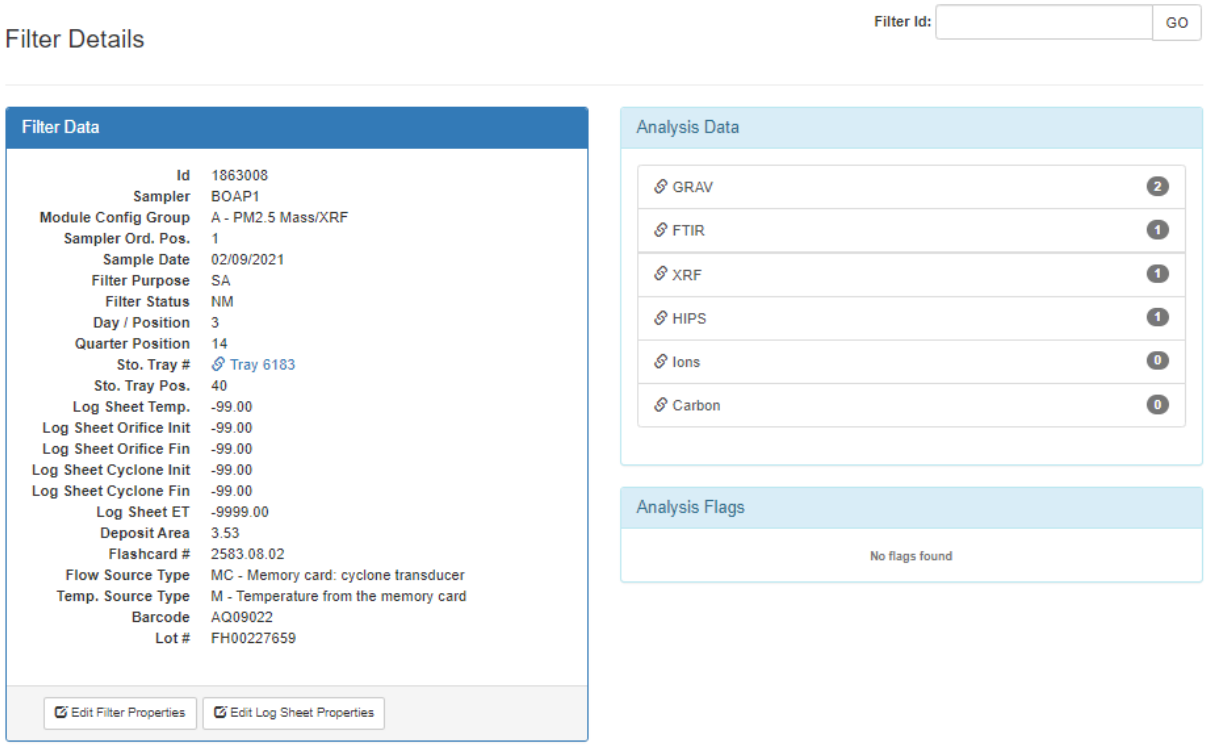
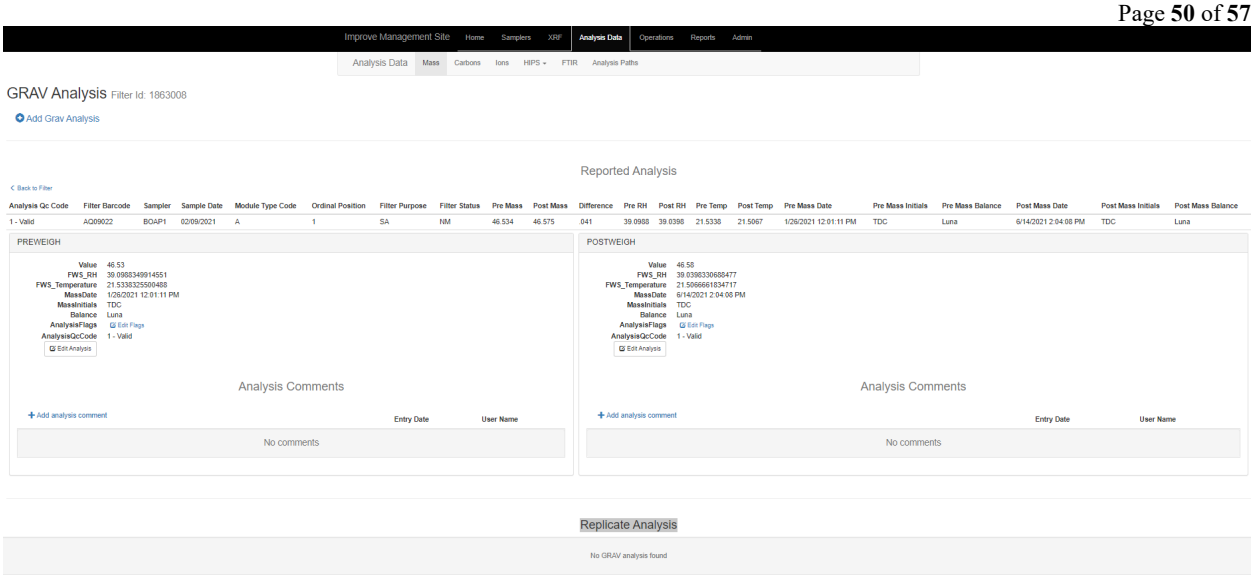


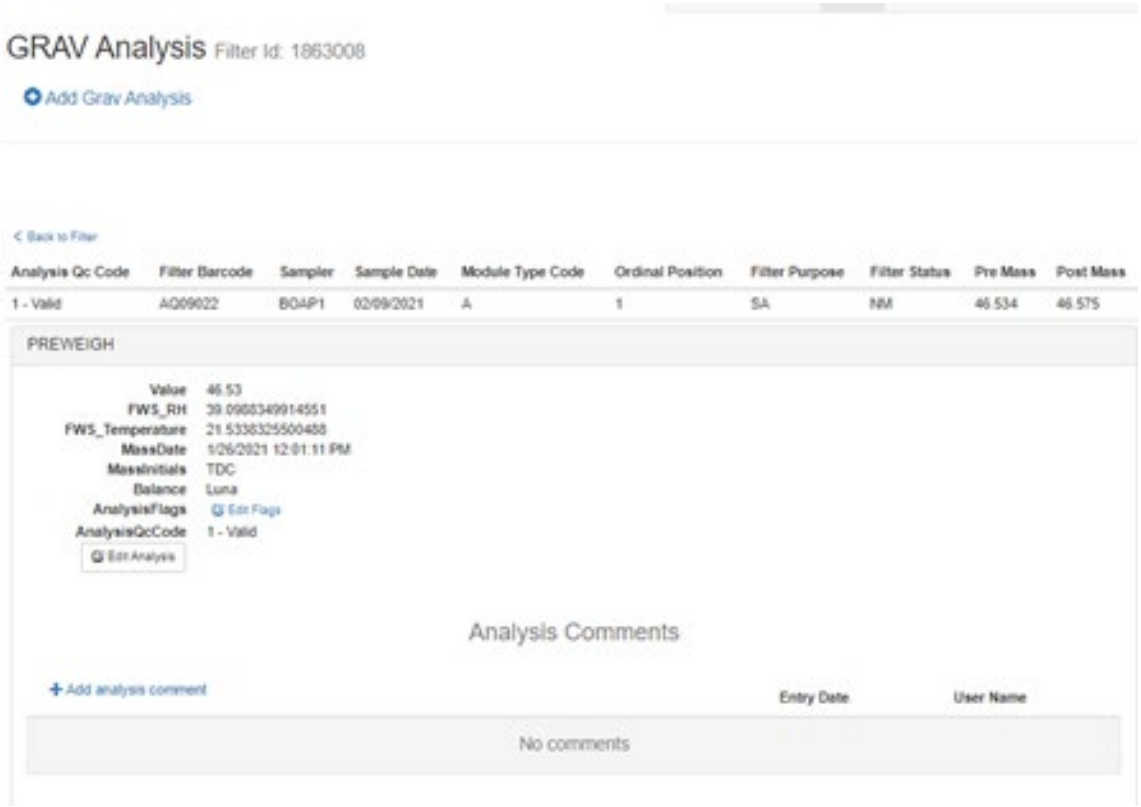
Figure 22. GRAV (Filter Mass; a) analysis page and detail (b, c).

a)

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



b)



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

c)

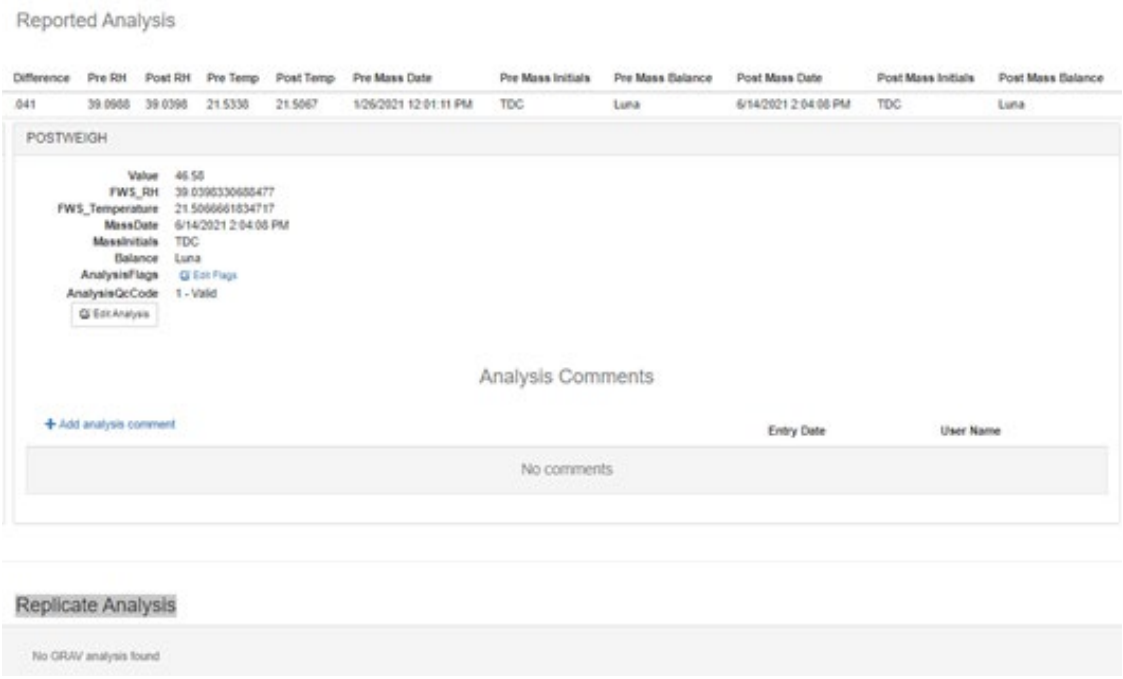


Figure 23. XRF sample analysis page.

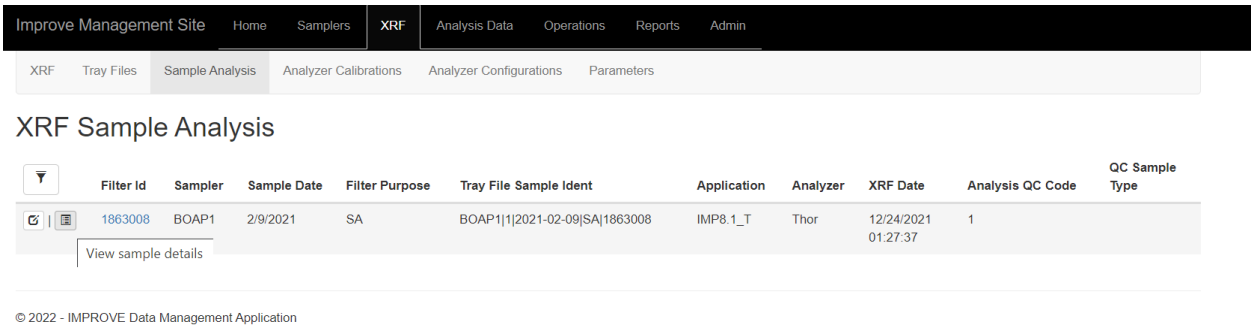


Figure 24. HIPS sample analysis page.

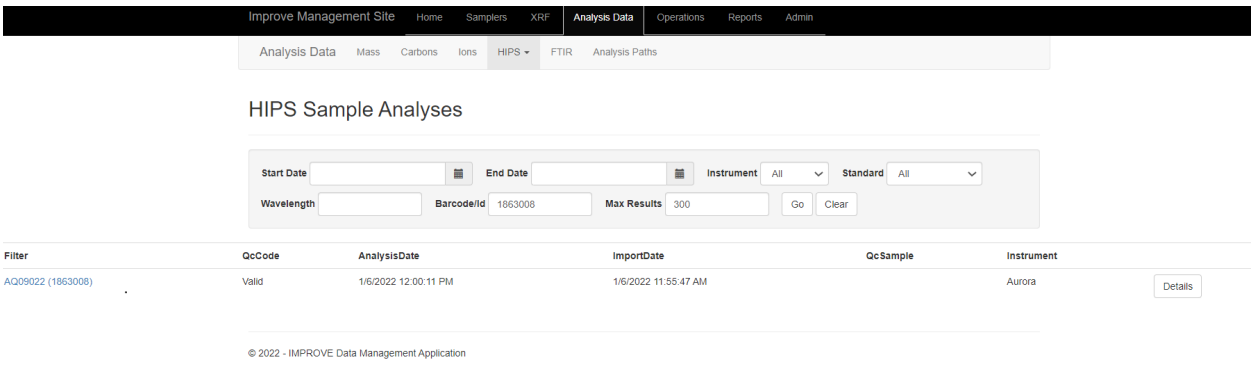


Figure 25. Ions sample analysis page.

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

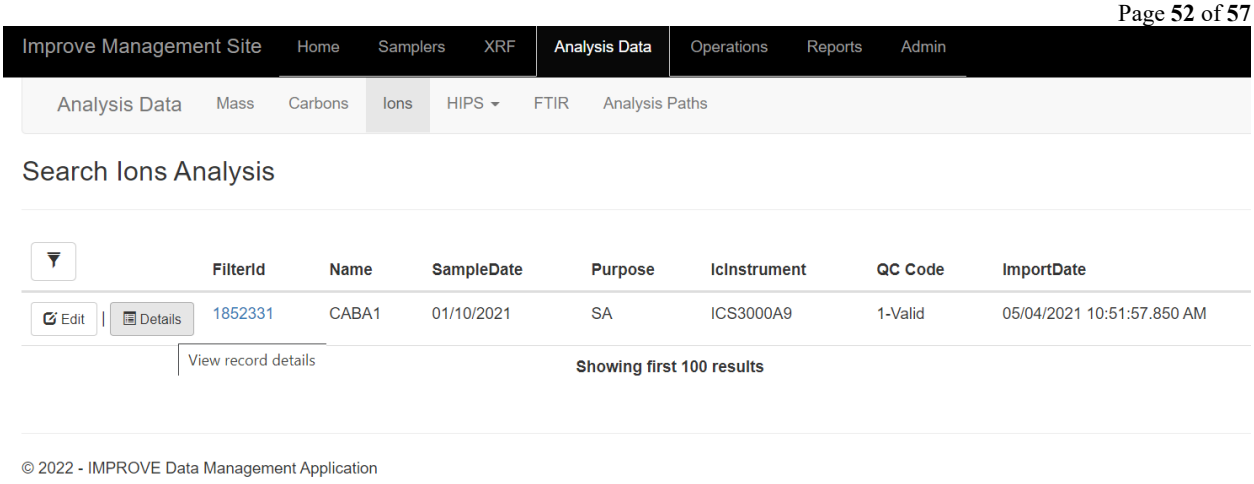


Figure 26. Carbon sample analysis page.

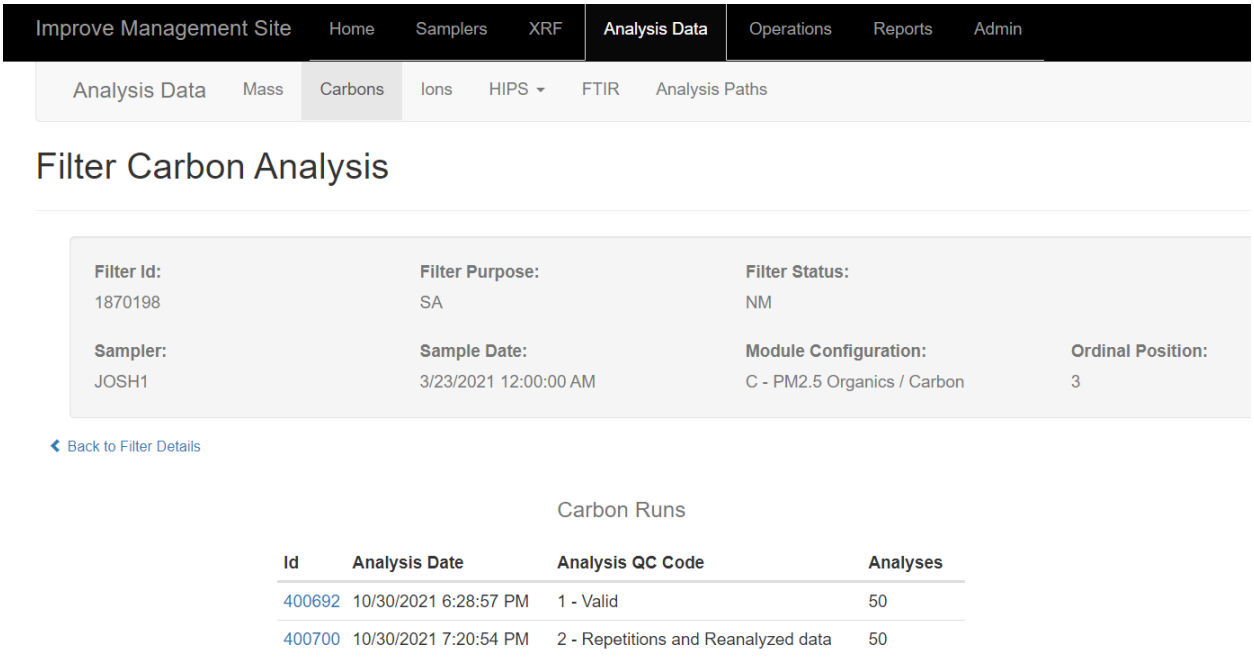


Figure 27. Analysis details page.

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

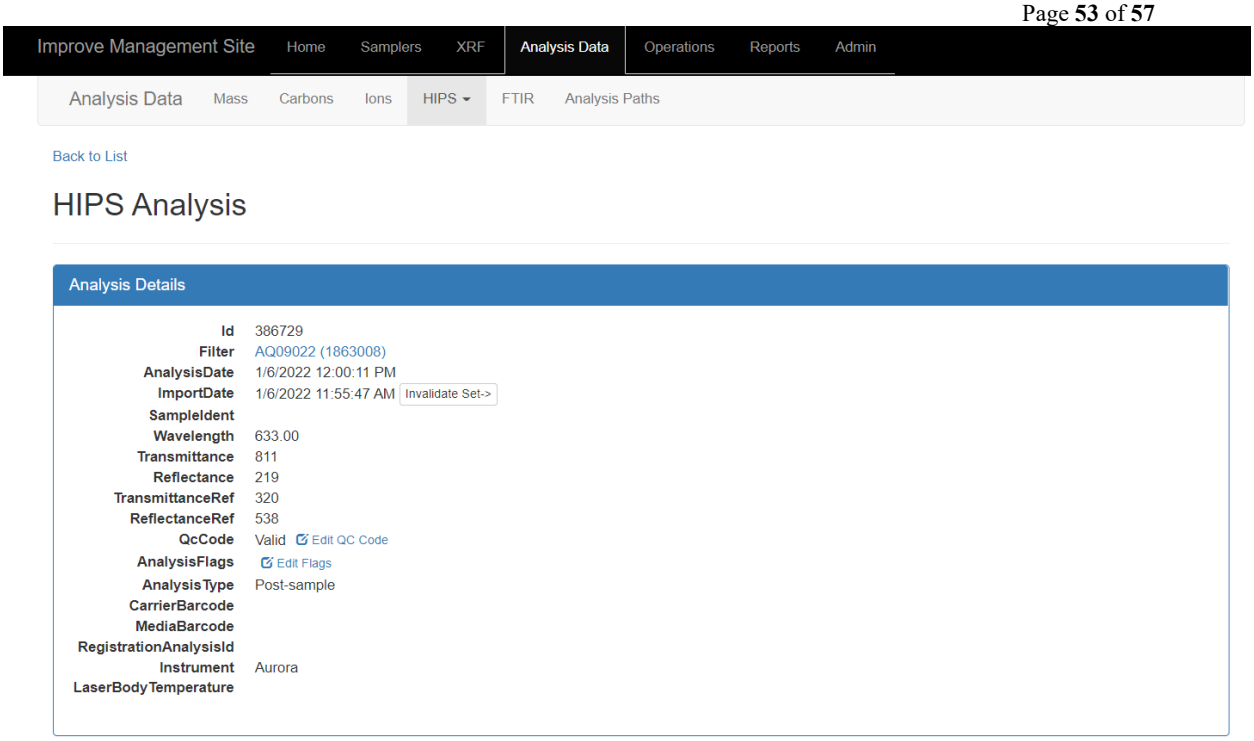
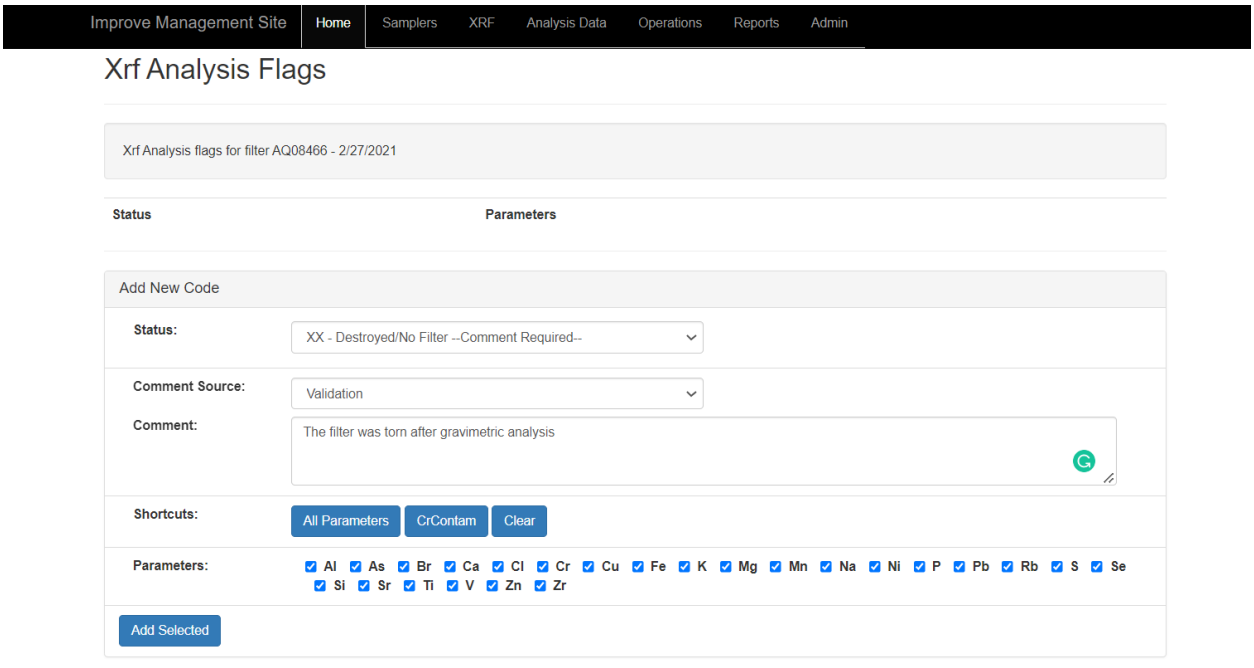
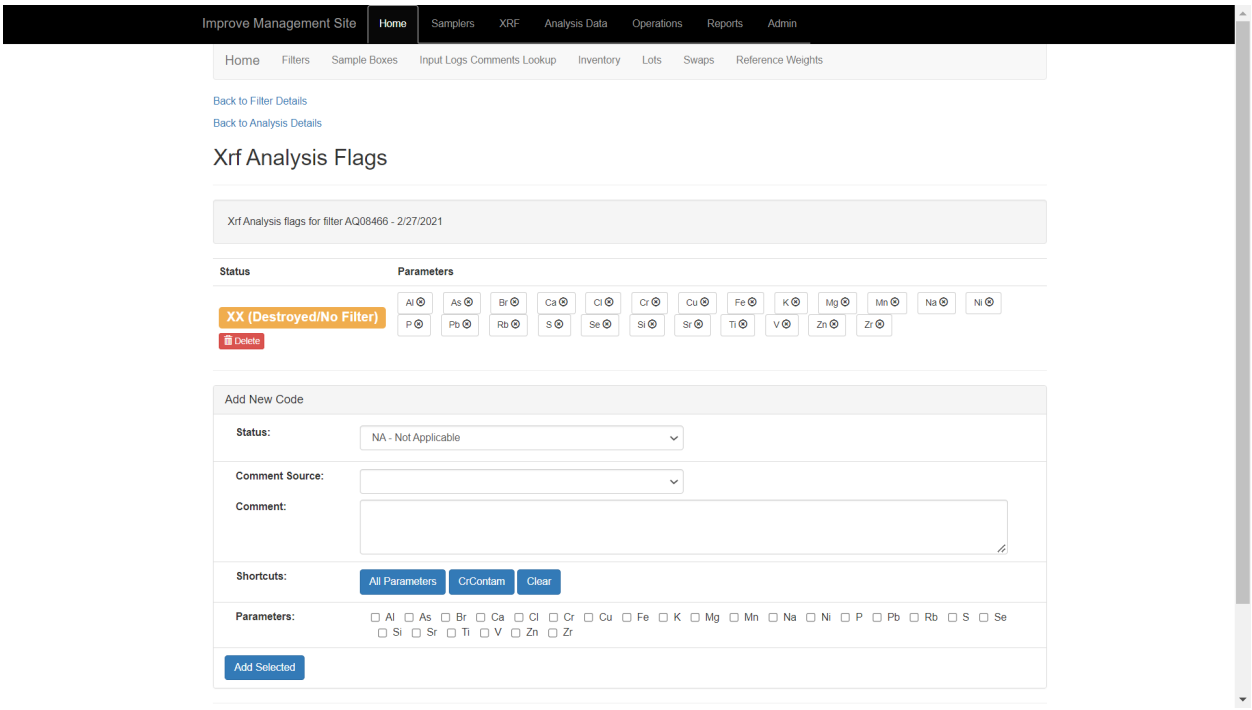


Figure 28. Analysis flags page.



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.