Section 508 Compliant ☐ Yes ☑ No

UCD IMPROVE Technical Information #351E

Flow Validation

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

January 7, 2025
Version 1.2

Prepared By: Signed by: Indu Thekkemeppilly Sivakumar 19F6B63B1B17443... Date: 1/10/2025

Reviewed By: DocuSigned by: Alex Murrain 2B7AC07297404D6... Date: 1/10/2025

Approved By: DocuSigned by: Marcus Langston 0A10CFCF79B0452... Date: 1/10/2025

UC DAVIS
AIR QUALITY RESEARCH CENTER

DOCUMENT HISTORY

Revision	Release Date	Initials	Section/s Modified	Brief Description of Modifications
1.0	03/14/2022	SRS	All	Previously anthologized version separated into individual TIs; content updated.
1.1	03/03/3023	DEY, ITS	9	Updated information on volume calculations for old controllers. Updated codes for Level 1 B checks.
1.2	01/07/2025	ITS, AMM	3, 9	Updated definitions, added flow status and validation information, including examples.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. PURPOSE AND APPLICABILITY

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

2. SUMMARY OF THE METHOD

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

3. DEFINITIONS

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

- **UCD IMPROVE Data Management Website:** A web-based application developed with .NET framework to interact with the IMPROVE database. Also referred to as IMPROVE web app.

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

Flow data from the V4 controllers is automatically transmitted daily to the UCD IMPROVE database for near real-time review by the Sample Handling Laboratory (SHL) and Field Group. Field log sheets and flashcards (with raw pressure transducer readings) are also available as backup flow data and are shipped with the physical sampled filters from the field sites to the UCD SHL. The details of flow data ingest are given in section 9.3 of the data processing TI (IMPROVE_TI_351B Data Processing). The information pertaining to V2 controllers has been excluded from this document as the last sampler utilizing the V2 controller ceased operations in December 2022. The relevant procedure can be referenced in version 1.1 of this document (IMPROVE_TI_351E_v1.1_Flow_Validation). As part of the Level 1A validation process, flow data are reviewed for inconsistency resulting from sampling anomaly and/or sampler malfunction. In these cases, the sample status is changed from NM to a terminal or temporary flag, and filter/sample event comments are provided. When automatically transmitted flow data are not available, the flashcard, log sheet, or nominal value can be used instead. The Flow Source Type Code for the affected sample is changed from the default (MC/MO) to the log sheet (LC/LO) or nominal value (NF) to ensure an accurate calculation of the average flow rate. An accurate average temperature is necessary to calculate the average flow rate accurately. The TemperatureSourceTypeCode for the affected sample is changed from the default (M) to the log sheet (L) or nominal value (N). Detailed procedures on flow data ingestion and Level 1A validation can be found in *UC IMPROVE TI #251E: Entering Log Sheets and Simple Problem Diagnosis*.

9.1 Processed Flow Data

Prior to checking flow data, the Data Validator processes flow data using the SQL query or the *improve_process_flow* function in R as described in section 9 of *UCD IMPROVE TI #351B: Data Processing* to derive the daily average flow rate and elapsed time (ET). The flow processing code automatically assigns non-normal flow status flags to the samples with flow rates that deviate from the nominal values. Tables 1 and 2 list the types of flow flags and the associated criteria for applying them to $PM_{2.5}$ and PM_{10} samples, respectively.

Table 1. Definitions and application criteria of automatic flow flags for $PM_{2.5}$.

Automatic Flow Flag	**Definition**	**Type**	**Criteria for Application for $PM_{2.5}$ Samples**
CL	Clogged Filter	Terminal	Flow rate < 15 L/min for more than 6 hours if controller or flashcard data are used. Average flow rate < 15 L/min if log sheet values are used.
CG	Clogging Filter	Informational	Flow rate < 18 L/min for more than 6 hours if controller or flashcard data used.

Automatic Flow Flag	**Definition**	**Type**	**Criteria for Application for $PM_{2.5}$ Samples**
			Average flow rate < 18 L/min if log sheet values are used.
LF	Low/high flow rate	Informational	Average flow rate < 19.7 L/min or > 24.1 L/min
PO	Power Outage	Terminal	Elapsed time < 1080 minutes (18 hours)
EP	Equipment Problem	Terminal	Elapsed time > 1800 minutes (30 hours) or is missing
TO	Timing Outside normal bounds	Informational	Elapsed time between 1080 minutes (18 hours) - 1380 minutes (23 hours) or 1500 minutes (25 hours) – 1800 minutes (30 hours)
SD	Short Duration Sample	Informational	1 second < Elapsed time < 1080 minutes (18 hours); controller stopped sampling when flow rate < 15L/min for at least 15 minutes while the vacuum is low. The status is terminal for the Regional Haze Rule but remains valid for other purposes.

The 2016 IMPROVE $PM_{2.5}$ cyclone characterization test yielded results consistent with the characterization performed by John and Reischl (1980). The particle size cut of the cyclone at any operating flow rate can be determined from the following equation:

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

Where,

D_{50} = 50% cutoff diameter (in µm)

Q = flow rate (in L/min)

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

The criteria for the CL, CG, and LF flags are determined based on calculation limitations, performance testing, and particle size cut. If >24 15-minute (6 hours in total) flow rate readings are below 15 L/min, or if the average flow rate is below 15 L/min when log sheet data are used, the sample is flagged as CL, and no concentration data are reported. The $PM_{2.5}$ cyclone cut point is 3.6 µm at 15 L/min.

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

A similar set of flags is applied to the PM_{10} data (Table 2), but with several differences in the criteria, due principally to the lower flow rate at which the PM_{10} sampler operates.

The relationship between the PM_{10} Sierra cyclone and particle size cut is not well characterized, so the criteria are determined somewhat arbitrarily. It is important to note that under the circumstance of a failing pump that produces less vacuum, equation (351E-1) is no longer true, and the calculated flow rates for the PM_{10} module are not valid.

Table 2. Definitions and application criteria of automatic flow flags for PM_{10}.

Validation Flag	**Definition**	**Type**	**Criteria for Application for PM_{10} Samples**
CL	Clogged Filter	Terminal	Flow rate < 10 L/min for more than 6 hours if controller or flashcard data are used. Average flow rate < 10 L/min if log sheet values are used.
CG	Clogging Filter	Informational	Flow rate < 14 L/min for more than 6 hours if controller or flashcard data are used. Average flow rate < 14 L/min if log sheet values are used.
LF	Low/high flow rate	Informational	Average flow rate < 15 L/min or > 18 L/min
PO	Power Outage	Terminal	Elapsed time < 1080 minutes (18 hours)
EP	Equipment Problem	Terminal	Elapsed time > 1800 minutes (30 hours) or is missing
TO	Timing Outside normal bounds	Informational	Elapsed time between 1080 minutes (18 hours) - 1380 minutes (23 hours) or 1500 minutes (25 hours) – 1800 minutes (30 hours)
SD	Short duration flow	Informational	1 second < Elapsed time < 1080 minutes (18 hours); controller stopped sampling when flow rate < 10L/min for at least 15 minutes. The status is terminal for the Regional Haze Rule but remains valid for other purposes.

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

At several IMPROVE sites, active flow control is implemented to ensure consistent flow rates. In this system, if the flow rates fall below nominal values, the variable speed pump will automatically increase its speed to maintain the flow rate at the desired nominal levels.

A clogging shutoff mechanism has been implemented at various IMPROVE sites to ensure that valid flow rates can still be obtained from samples, even when the filter becomes clogged due to heavy loading before a valid sample can be collected (18 hours ET). This feature proves particularly advantageous during the wildfire season when the PM 2.5 Teflon filter becomes clogged prematurely due to heavy loading. In such instances, the flow receives an 'SD' status, denoting a short-duration flow, rendering it invalid for the Regional Haze Rule. However, shutting off the pump contributes to accurately calculating the concentrations from the sample.

The specific conditions that lead to the pump shutting off and result in an 'SD' status for both the PM 2.5 and PM 10 modules are clearly outlined in Tables 1 and 2. If the clogging shutoff is triggered after eighteen hours of sampling, it results in the shutdown of all four modules, leading to a 'TO' flow status for all modules. If the elapsed time is less than eighteen hours, only the affected clogged module will be deactivated. The occurrence of the clogging shutoff can be identified through a metadata line in the flow plotter.

Several Level 1B checks (see *UCD IMPROVE TI #351C: Data Validation* for details on Level 1B) on the 15-minute raw flow data are performed by running the *flow.check* function from the *datvalIMPROVE* R package. To perform these checks, open an R environment (such as RStudio) and run the following command:

[month_flow] <- datvalIMPROVE::flow.check(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’], site = [‘%’], list_all = [‘FALSE’], exclude_objective_code = [“(‘TS’, ‘RS’)”], server = ‘production’)

The *exclude_objective code*, by default, will exclude the 'TS' and 'RS' objective codes. A single objective code can be provided in the command line if only one needs to be omitted.

When *list_all* is set to FALSE, the function returns a report that lists all the samples during the date period specified with abnormal flow variability, abnormal sampling temperature, and number of records for further investigation. If the *list_all* argument is set to TRUE, only the sample events with relative standard deviation out of range will be returned. The three asterisks (***) are generated automatically in the output from the *flow.check* function to indicate data issues.

The Data Validator can perform the checks for all active sites in the network by setting *site = ‘%’* or just for a particular site by specifying the site name. Several criteria are checked:

- Abnormal flow variability: > 8% during a 24-hour sampling period; can be caused by equipment installation problems or steady pressure drop from a heavily loaded filter.

- Abnormal sampling temperature: relative standard deviation of temperature < 0.01% or > 10%; average temperature < -20 °C or > 40 °C.
- Abnormal number of records: number of 15-minute flow readings is < 72 rows (equivalent to 18 hours of run time) or > 104 rows (equivalent to 26 hours of run time).

Additional criteria implemented for the V4 controller include:

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

9.2 Generating the Flow Validation Report

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

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

  [No_flowdata] <- datvalIMPROVE::flow.completeness(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’], exclude_objective_code = "('TS', 'RS')", server = ‘production’)

  The *exclude_objective code,* by default, will exclude the 'TS' and 'RS' objective codes. A single objective code can be provided in the command line if only one needs to be omitted.

  write.csv(No_flowdata, "U:/IMPROVE/Data_Validation/Flow/Nofllow.csv", row.names = TRUE)

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

The next tab of the spreadsheet is populated using the data returned from running the *flow.check* function as described in section 9.1 above. The spreadsheet has several tabs as described below:

Once all the flow validation-related data frames are exported (the steps are below) in CSV format, they can be combined to Excel format to make the flow validation report.

Once in Excel format, color code the modules (A = red, B = Yellow, C = Green, and D = Blue). The three asterisks (***) generated automatically in the output from the *flow.check* function (see Section 9.1) indicate data issues.

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

View([month_flow]$NewController$MainCheck)

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

write.csv (month_flow]$NewController$MainCheck, "U:/IMPROVE/Data_Validation/Flow/ Monthflow_NewController.csv", row.names = TRUE)

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

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

View([month_flow]$NewController$SolenoidCheck)

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

write.csv (month_flow]$NewController$SolenoidCheck, "U:/IMPROVE/Data_Validation/Flow/ Monthflow_Solenoidcheckr.csv", row.names = TRUE)

- Flow flags (CG, CL, LF, PO, EP, TO, SD): These sheets contain lists of samples where the flow status is flagged as CG, CL, LF, PO, EP, or TO and require confirmation of appropriate flagging (see Tables 1 and 2). Generate this data by running the following command in R:

  [month_flowflag] <- datvalIMPROVE::flow.status(startdate = [‘YYYY-MM-DD’], enddate = [‘YYYY-MM-DD’], flowflag = [(‘CG’, ‘CL’, ‘LF’, ‘PO’, ‘EP’, ‘TO’)], exclude_objective_code = "('TS', 'RS')", server = ‘production’)

  The *exclude_objective code,* by default, will exclude the 'TS' and 'RS' objective codes. A single objective code can be provided in the command line if only one needs to be omitted.

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

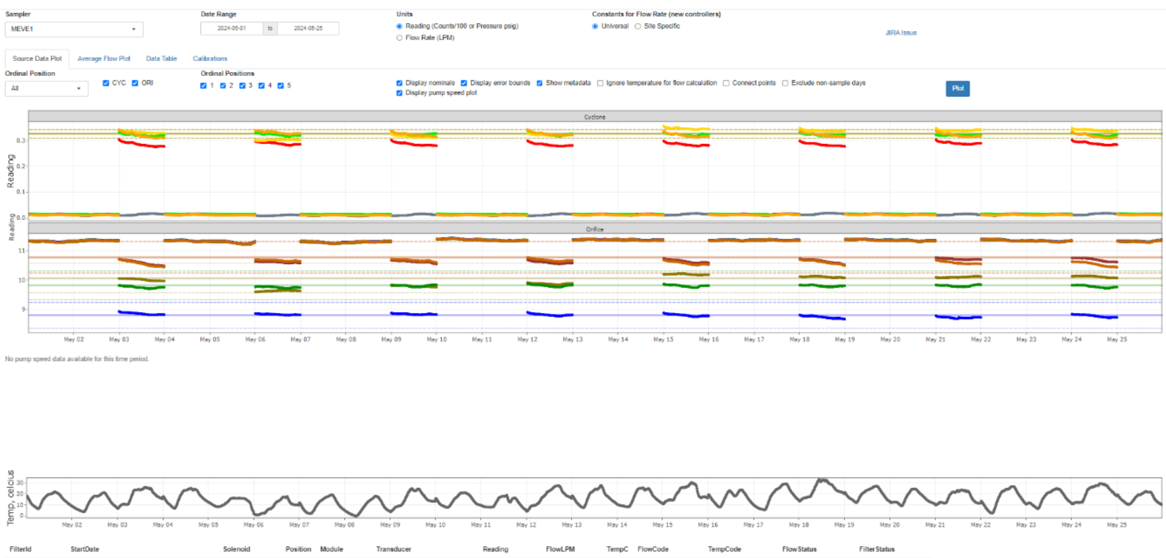
write.csv (month_flowflag,, "U:/IMPROVE/Data_Validation/Flow/ Month flow flowflag.csv", row.names = TRUE)

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

9.3 Flow Validation

To further investigate the data returned from the flow checks and to validate flow data, flow plots are carefully reviewed (IMPROVE Flow Graphs; https://shiny.aqrc.ucdavis.edu/FlowRates/). The Flow Source Code is assigned if the primary source (MC for A, B, and C modules and MO for D module.; automatically transmitted flow data or flash card) is not reliable. IMPROVE Flow Graphs comprise multiple tabs. The Source Data Plot tab (Figure 1) serves as the primary tool for flow validation and contains three main sections: Cyclone reading (CYC), orifice reading (ORI), and temperature readings. In cases where flow control is active at the site, the pump speed is also presented in the flow plotter (Figure 2). The flow plotter has another section that shows important details about a selected filter, including FilterId, StartDate, Solenoid, Position, Module, Transducer, Reading, FlowLPM, TempC, FlowCode, TempCode, FlowStatus, and FilterStatus. This section is only visible when a filter Id is selected by clicking the CYC or ORI readings.

Figure 1. Source Data Plot without active flow control.



The CYC reading section utilizes a color scheme for visual representation as follows: 1A (Position 1) in red, 2B (Position 2) in yellow, and 3C (Position 3) in green. The PM 10 module (Position 4) does not have a CYC reading and consistently displays a base value in gray. Additionally, any fifth module, irrespective of filter type and position, is consistently depicted in orange.

The ORI reading section utilizes a color scheme for visual representation as follows: 1A (Position 1) in reddish brown, 2B (Position 2) in yellow-green, and 3C (Position 3) in dark green, 4D in blue. Any fifth module, irrespective of filter type and position, is consistently depicted in orange. To determine the corresponding module for each line, hover over the flow lines. Position 1 is designated as Channel 1, position two as Channel 2, position three as Channel 3, position four as Channel 4, and position five as Channel 5.

Figure 2. Source Data Plot with active flow control.

Guidelines for validating flow data include:

- Review the flow graph to identify unstable flow readings. Evaluate the flow readings and temperature to determine if there are any abrupt changes or if the flow is changing gradually throughout the sampling day. Abrupt changes may indicate equipment issues or power interruptions, while a gradual change could be caused by heavy loading. If there is no identifiable pattern, it may signal a potential issue that requires further investigation.If automatically transmitted flow and flashcard data are not available or reliable, use log sheet data which can be retrieved from [Improve_2.1]. [ops].[ControllerFilterReadings] or the hand-written records on the paper Field Log Sheets.
  - The Flow Source Code or Filter Status Code can be updated as needed from the Filters page (https://improve.aqrc.ucdavis.edu/Filters) of the IMPROVE web app.
  - Utilize the Average Flow Plot in the Flow Graphing App to further evaluate flow data.
  - Utilize the Early Review page in the IMPROVE Data App (https://shiny.aqrc.ucdavis.edu/ImproveData/) to view site-by-site analysis data, which can be used to help evaluate flow issues.

- Utilize the Controller Filter readings page (https://improve.aqrc.ucdavis.edu/Operations/ControllerFilterReadings) of the IMPROVE web app.

Finally, all samples flagged as terminal (i.e., CL, EP and PO) by the flow processing code are manually reviewed for errors. In cases where valid samples are flagged as invalid (e.g., corrupt flash card files or faulty transducer readings), the flow source code is changed, and the average flow rate is reprocessed to correct the sample status. The same approach is taken for other flow flags like CG and LF.

9.3.1 Common Flow Review Scenarios

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

Equipment Problem (EP) Status

The EP status is assigned to flow data when no flow data is linked to the filter ID. This situation may arise due to the absence of existing flow data in the database or incorrect association of flow data with the designated filter.

Check the flow plotter (Figure 1 and Figure 2) to determine if there is any available flow data for the affected sample dates. If no flow data is visible, please contact SHL to verify if there are any errors in the automated data upload or to confirm whether the internet connection is down. If necessary, request to upload flow data from the backup flashcard or use the values from the log sheet.

If flow data is present in the plotter, the reason for the EP status could be that the flow data is not correctly linked to the appropriate filter in the module.flowsourcedatav2 table. This often occurs when a replacement box is sent to replace a lost one, causing the flow data to be incorrectly associated with the wrong box. In such situations, the box swap tool in section 9.3.3.3 Box Swap of the IMPROVE_TI_351C__Data_Validation TI should be used to swap the sampling data. Please note that the SHL may have already updated the filter purpose of the lost box to 'UF' and updated the filter statuses to a terminal value. Performing a box swap will undo these alterations. Please use the provided query to update the statuses accordingly in such cases.

For the filters in the lost box, the filter purposes are to be updated to UF (Unused/Lost Filter (Filter Purpose ID = 16) and the filter status id to a terminal value. XX (Id:23) is the most commonly used terminal status.

SHL may have already assigned filter statuses, so it's recommended to save them. To obtain the current statuses, use the query below and save the results for review.

*SELECT **

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

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

WHERE sc.SampleBoxId = LostBoxID

The following SQL update query is used, where LostBoxId is the ID of the lost box:

UPDATE f

SET f.FilterPurposeId = 16, f. FilterStatusId = 23

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

WHERE sc.SampleBoxId = LostBoxID

After updating the filter purpose and filter status, review and confirm the filter purpose Id for the whole box is correct by running the SELECT query mentioned above.

If the filter purpose is designated as 'UF' for the filters in the replacement box, a similar update query will be executed to rectify the undesired modification. Per best practices, the sample handling lab should not include field blanks in the replacement box. In addition, all filter purposes (28 for routine and 35 for sites with collocated modules) will be updated to one. If filter statuses have already been assigned by the sample handling lab and differ from SO or NM, please utilize the list generated by the select query to identify the correct status of each filter. For a bulk update, the following query will be employed. If the filter status of all filters in the box is not the same, omit the filter status ID from the update query. The filter status can be modified using the Filters page (https://improve.aqrc.ucdavis.edu/Filters) in the IMPROVE web app.

UPDATE f

SET f.FilterPurposeId = 1, f. FilterStatusId = 8

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

WHERE sc.SampleBoxId = ReplacementBoxID

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

*SELECT **

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

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

WHERE sc.SampleBoxId = ReplacementBoxID

When the flow data exists in the flow plotter but is not linked to any filter ID in the module's FlowSourceData table, you can use the following select query to link the filter ID with the flow data. The list of affected filter IDs is obtained from the results of the *No flowdata* check.

UPDATE fs

SET fs.FilterId = f.FilterId

FROM [Improve_2.1].module.FlowSourceData as fs

INNER JOIN

(select s.Name,f.SampleDate, m.SamplerOrdinalPos ,f.id as FilterId, m.Id as ModuleId

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

LEFT JOIN [Improve_2.1].module.Modules m ON m.Id = f.SamplerModuleId

LEFT JOIN [Improve_2.1].sampler.Samplers s ON m.SamplerId = s.Id

where f.id in (Filter ID))

f ON fs.SamplerModuleId = f.ModuleId AND fs.SampleDate = f.SampleDate

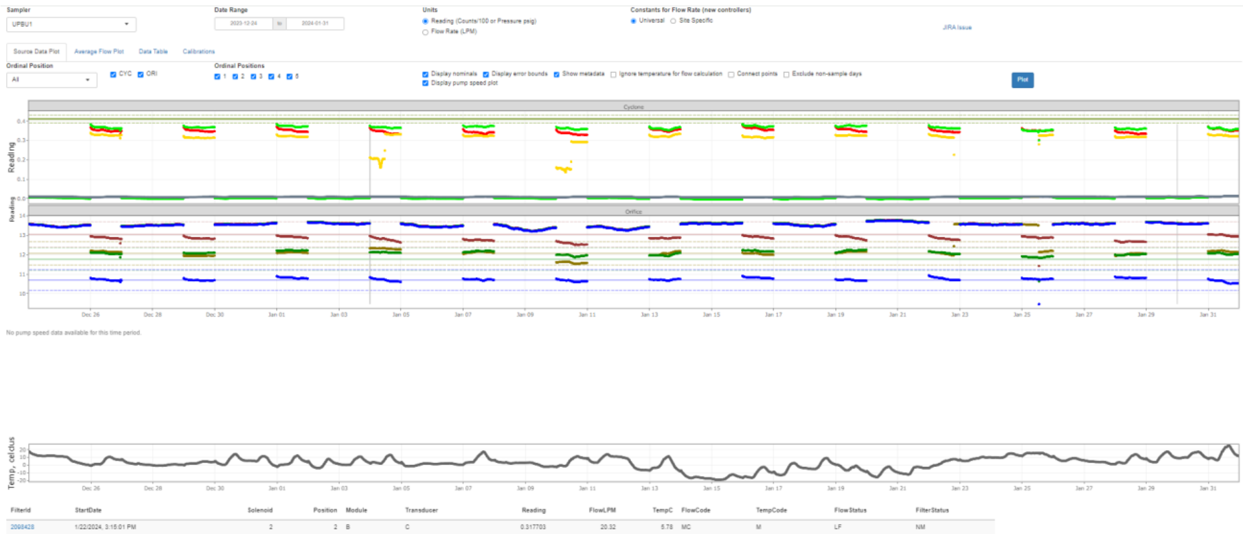
Clogged / Clogging (CL or CG) Status

- The flow data are flagged with CL or CG status when there is heavy loading on the filter or due to pump malfunction. In the case of heavy filter loading, no further action is needed. An example of CG flow status is shown in Figure 3.

Figure 3. CG flow status.

- The CL, CG, or even LF flow status can also be incorrectly assigned due to an incorrect elapsed time. This could be due to a sampling period of less than 24 hours because the pump started late, stopped during the designated sampling period, or for other reasons. An example is depicted in Figure 4, where the UPBU1 sample dates 1/22/24 and 1/25/24 have less than 24 hours of elapsed time for the 2B module due to pump malfunction. In such cases, the flow source values should be updated in the log sheet. To obtain the correct elapsed time, please check the flow plotter or obtain the values from the log sheet. If the flow status is CL due to pump malfunction, change the filter status code to EP (Equipment Problem. Refer Table 1 and Table 2) from the Filters page (https://improve.aqrc.ucdavis.edu/Filters) of the IMPROVE web app and reprocess the flow data using the SQL query described in section 9 of *UCD IMPROVE TI #351B: Data Processing*.

Figure 4. CG, CL, and LF due to incorrect elapsed time and manifold issues.



- Another reason is the manifold was not closed properly, which led to a leakage, or the manifold was open at some point during the sample collection. An example is depicted in Figure 4, where the UPBU1 sample dates 1/4/24 and 1/10/24 have CG flow status for the 2B module due to an open manifold for part of the sample date. In these situations, the flow status determined by the flow processing code is accurate. The filter status is changed to QD, and the data needs to undergo additional review once all analysis results are available, allowing the data to be processed and confirmed.
- If there is a temperature probe malfunction resulting in an extreme average temperature, the flow status may get a CG status. To address this, it is recommended that the temperature source code be updated to Nominal or log sheet values through the Filters page (https://improve.aqrc.ucdavis.edu/Filters)of the IMPROVE web app. The temperature log sheet should be utilized only if it exhibits reasonable data and if the adjacent sample dates show discernible trends. The flow data can then be reprocessed using the SQL query in section 9 of TI #351B. Blockage of the inlet may occur due to the presence of small objects, which can result in reduced flow rates. It is important to assess the severity of the blockage and continually monitor for any consistent deterioration. The issue typically begins with flow rates in the LF range and can worsen into CG status. Usually, the field group is responsible for addressing this issue. Therefore, it is recommended that they check if corrective actions were already taken and documented in a Jira issue. If there are no such reported issues, it is advisable to contact the field group to confirm any measures taken.

Power Outage (PO) Status

- The PO status gets applied when Elapsed time is < 1080 minutes or greater than 1560 minutes. Check the filter readings table, flow plotter, or log sheet data to ensure a late sample change was not the cause for the elapsed time to fall below the limit. If the sample change was late, update the filter status to NS (No Sample. Refer to Table 1 and Table 2) and reprocess the flow data using the SQL query described in section 9 of TI #351B. If the elapsed time is low due to a power outage, update the filter status to PO using the Filters page (https://improve.aqrc.ucdavis.edu/Filters) of the IMPROVE web app. Another reason might be that the flow ingest has generated duplicate lines, resulting in an elapsed time ranging from 26 to 30 hours. In such instances, the flow source should be updated with the values from the log sheet.

Temperature Probe Malfunction

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

Low Flow (LF) Flow Status

- Some reasons for LF status overlap with CL and CG flow status and are already covered. Other reasons include: Heavy loading which results in a flow value between NM range and CG range as described in Table 1 and Table 2. No action is required in such cases.
- Swaps between filter types. If consecutive sample dates have an LF flag and the filters are all in the same cartridge, check if any other module is affected by flow fluctuation. If there is, it is suspected that the cartridges have been swapped between the modules. Examine the following dates to see if the flow pattern returned to normal after a sample change. In such cases, the filter statuses can be updated to QD (Questionable Data) from the Filters page of the IMPROVE web app for further review after all analyses come back. If the pattern continues, request the field group or sample handling lab to contact the site operator to ensure proper installation of cartridges.
- One prevalent reason for the Low Flow status is filter variability. The filter's thickness may vary, even within the same lot, thereby impacting the CYC reading. In the event of consistent low flow observed across multiple sites, it is advisable to notify the sample handling laboratory to conduct a thorough

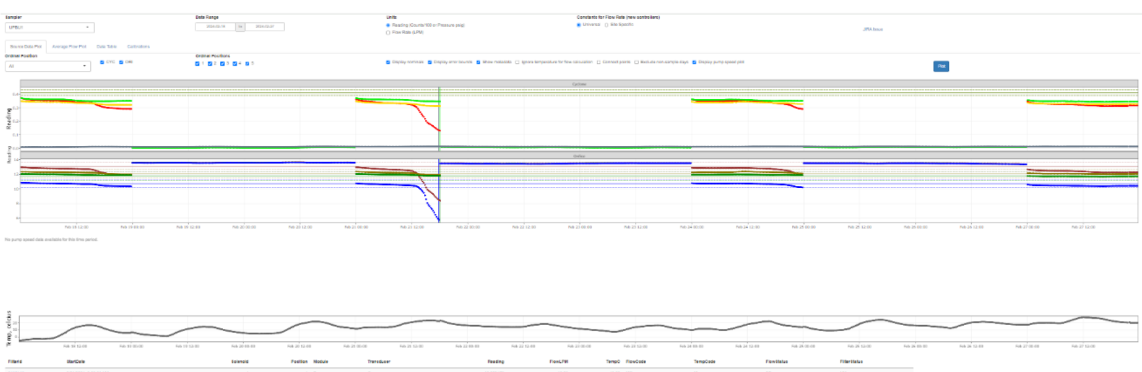
investigation to ascertain if this issue is linked to a specific lot and to evaluate its impact on the data quality.

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

Short Duration Sample (SD) Status

- The SD status gets applied when the Elapsed time is < 1080 minutes, but the controller automatically stops sampling when a filter becomes clogged from heavy loading. The upgrade allows for calculating accurate concentrations. The data is not valid for the regional haze rule. A meta data line is added to the flow plotter when the flow shutoff is triggered (Figure 5).

Figure 5. SD flow status.



10. DATA AND RECORDS MANAGEMENT

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

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

10.1 Disaster Recovery Plan

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

10.1.1 Facility Recovery

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

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

10.1.2 Hardware Recovery Plan

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

10.1.3 Software and Data Recovery Plan

10.1.3.1 UCD Laboratories

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

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

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

10.1.4 Data Security

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

11. QUALITY ASSURANCE AND QUALITY CONTROL

11.1 Code Development

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

11.2 Bug Reporting

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

11.3 Data Validation

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

12. REFERENCES

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