UCD CSN Technical Information #801C

CSN Data Validation

Chemical Speciation Network
Air Quality Research Center
University of California, Davis

July 31, 2020
Version 1.3

Prepared By: DocuSigned by: Dominique Young BB55DBA34BAB407... Date: 8/13/2020

Reviewed By: DocuSigned by: katrina Gorham 1B2408CC2DFE4FF... Date: 8/13/2020

Approved By: DocuSigned by: Nicole Hyslop BCDBAE03A95C40A... Date: 8/13/2020

UCDAVIS
AIR QUALITY RESEARCH CENTER

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
11/30/18	NJS	1-3, 7-11	Rewording for clarity and updating name changes.
7/31/19	DEY, KAG	1-3, 7-8	Additional step and refinement to the validation process. Wording changes for clarity.
3/6/20	KAG	4-17	Updated content to reflect CSN Data Explorer changes. Wording changes for clarity.

TABLE OF CONTENTS

LIST OF FIGURES

1. PURPOSE AND APPLICABILITY

The subject of this technical information (TI) document is validating the analytical data from the Chemical Speciation Network (CSN). Data from the network are reviewed and validated using a variety of tools. Qualifier codes and/or null codes are applied as appropriate.

2. SUMMARY OF THE METHOD

The University of California, Davis (UCD) analyst uses the UCD CSN 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

- **Chemical Speciation Network (CSN)**: EPA's $PM_{2.5}$ sampling network, with sites located principally in urban areas.
- **AQS:** EPA's Air Quality System database.
- **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
- **CSN database:** An SQL Server database that is the central warehouse of CSN preliminary and final data at UCD.
- **Data Analysis and Reporting Tool (DART):** A web application for environmental data visualization and validation procedures.
- **datvalCSN:** A custom software package in the R language that contains the data validation code used to collect, compare, and flag the final results.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING

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. PROCEDURAL STEPS

Data validation is performed using the *crocker* and *datvalCSN* R packages, which are developed and maintained by UCD specifically for data processing, monitoring, and validation of the CSN data. Data validation is performed by the UCD Data and Reporting Group on monthly batches of data (a calendar month of sample start dates). Validation occurs in four steps:

- Collect necessary measurement data (Section 8.1)
- Compare results (Section 8.2)
- Review summary plots (Section 8.3)
- Review summary data via web application (Section 8.4)

UCD is responsible for the first two levels (Level 0 and Level 1) of the data validation process. CSN data delivered to DART is Level 1 data while data delivered to the AQS database is Level 2 data. The state, local, and tribal (SLT) agency validators perform Level 2 data validation while the data is available in DART. The levels are defined as follows:

Level 0

Data at this level are raw data obtained directly from the data-acquiring instruments. These data can be reduced or reformatted, but are unedited and not reviewed, 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 (e.g., ambient temperature of sampler every fifteen minutes). Average times represent the minimum intervals recorded, and these data may need to be aggregated to obtain averages for the sampling periods (e.g., 24-hour average). Level 0 data have not been edited for instrument downtime, nor have procedural adjustments for baseline shifts, span changes, or known problems been applied. Examples of CSN Level 0 data include:

- 24-hour averaged pressure, temperature, and flow data recorded during sample change procedures;
- EDXRF raw spectra;
- Sample date and sampling time before consistency checks.

Level 1A

Data at this level have passed several automatic and manual qualitative reviews for accuracy and completeness. The focus of Level 1A validation is to obtain as complete a dataset as possible. CSN Level 1A data validation consists of:

- Adding records for expected filter samples that were never generated, for completeness purposes;
- Setting status codes to reflect sampler malfunctions, site or laboratory analyst errors, or power outages;
- Reviewing analyst, sample handling laboratory, and measurement laboratory comments to verify consistency between records and correct any typographical errors;
- Verifying analyst, sample handling laboratory, and measurement laboratory comments for questionable records;
- Identifying, investigating, and/or flagging data that are beyond reasonable bounds or that are unrepresentative of the variable being measured, including:
    - Sulfur/sulfate ion ratios outside the range 0.784 to 1.731 and non-soil potassium/potassium ion ratios outside the range 0.656 to 6.76 (where non-soil potassium is defined as K minus (2/27)×Si),
    - Collocated mass/reconstructed mass ratios outside the range of 0.5 to 2,
    - Anion/cation ratios outside the range 0.86 to 2.82,
    - Z-scores of the OC/EC (organic carbon/elemental carbon) ratios outside the range -1 to 1;
- Examining daily flow rates based on the results of the *flow.check* function that identifies abnormal flow rates and significant variations over 24-hours;
- Setting qualifier codes when deviations from nominal operational settings have occurred (e.g., temperature or pressure outside instrumental tolerances);
- Examining the field blank analyses for evidence of swaps with sample filters;
- Examining the sample analyses for evidence of swaps with other sample filters on different days.

Level 1B

Data at this level have passed additional quantitative and qualitative reviews for accuracy and internal consistency. Comparison with collocated measurements and internal consistency tests are applied by the analyst. Discrepancies that cannot be resolved are reported to the measurement laboratories and sample handling laboratory for investigation. Data that deviate from consistency objectives are individually examined for errors. Extreme outliers (e.g., -85 °C temperature) are noted with qualifier codes or null codes depending on the parameter and value. Changes to the data (e.g., swapping dates on consecutive samples) are recorded and documented by providing comments to the SLT agency validators. Level 1B time-series data review is conducted on a site-by-site basis using a combination of *datvalCSN* package functions and Shiny web applications. The historical archive of CSN data is used to place new measurement values in context with previous measurements. CSN Level 1B data validation includes:

- Comparing sulfur and sulfate concentrations as well as non-soil potassium and potassium ion concentrations;
- Comparing organic carbon to elemental carbon for both blank corrected and uncorrected concentrations;
- Comparing anions to cations in a molar mass balance;

- Comparing reconstructed mass and collocated $PM_{2.5}$ mass;
- Examining individual data points identified by the various checks as potential sample swaps;
- Reviewing network level field blank mass loadings for a given month to identify outliers, and further comparing with associated sample filter data to identify intended purpose swaps or issues with sample set-up or individual samplers;
- Reviewing values of each parameter at the network level for a given month to look for anomalies;
- Comparing parameter values and concentrations with those at other nearby sites, within a specified radius, and across the network for a given month;
- Comparing the analytical data to expectations based on prior years.

Level 2

Level 2 validation is performed by the SLT validators after submission of data files to DART. At this level, the data are reconciled with local events, knowledge, and expertise. Upon completion of the DART review period, data are sent back to UCD for a final review before data are submitted to AQS. The final review at UCD ensures consistency between user comments and data transfers. UCD does not make changes to DART-reviewed data unless requested by the SLT validators.

8.1 Collect Measurement Data, Perform Checks and Comparisons

In the previous step, laboratory results for elements, ions and carbon results are processed into mass per volume of air and posted into the UCD CSN database (see UCD TI 801B). Additionally, collocated mass data should be collected from the AirNowTech API web service via file transfer protocol (FTP). All data can be collected using functions in the *datvalCSN* package. In addition to collecting measurement data, records for expected sample filters that were never generated are added for completeness purposes.

The analyst begins validation by opening an R environment (such as RStudio) and running the following command to review a dataframe of the missing sample filter records:

[missingRecords] <- datvalCSN::find_missing_filters(Month = ['MM'], Year = ['YYYY'])

This function predicts missing filters for the specified month and year based on the sampling frequency for a given site. Missing sample filter records are reconstructed for posting into the database with the null code "AF - Scheduled but not collected". The analyst will review the list of missing records and determine which records are to be added to the database based on the reason for the records originally being missed. For example, a sampler may be offline for repairs and shipment of filters to the site is paused during this time. The following command will add the records to the database:

[postRecords] <- datvalCSN::post_missing_filters(missing.filters = [missingRecords], username = ['userCredentials'])

Once all expected records are in the database, the analyst will run the following command to collect records for all measured and derived parameters from the laboratories, and to begin validation by running automated checks[1]:

[monthData] <- datvalCSN::csn_validate(Month = ['MM'], Year = ['YYYY'])

The command form shown here is appropriate for typical validation procedures; more options can be included by changing defaults. To find out more information on this command, run *?csn_validate* in the R environment.

Within the *csn_validate* command, records are collected from the *analysis.Results* table that are pertinent to the month and year specified. Mass data from the AirNowTech API web service are collected when the function parameter *with.mass* is set to TRUE, whereby a function within the *csn_validate* command extracts the desired data from the posted values on the AirNowTech server. All of the results will be returned in memory to the variable *monthData*. The qualifier codes and null codes assigned from various checks performed within the *csn_validate* function are associated accordingly with the relevant records for delivery.

8.2 Data Validation Checks: Automated and Analyst Performed Checks

The validation process begins with automated checks. Each check compares resulting values against pre-defined limits as well as comparable parameters. All automated checks are performed as part of the *csn_validate* command. The resulting data tables from these checks are formatted for further graphical interpretation. The individual checks are described below. As each check is a function within the *datvalCSN* R package, it is possible to execute them outside of the *csn_validate* command, if required. In the following descriptions of the checks, the function name is documented; further details can be obtained for each function from the help documents in the R environment by entering a question mark before the function name in the R console.

Operational parameters such as sample flow rate and operating temperature should have been checked both at the sampling site as well as at the Sample Handling Laboratory. However, there are two checks included within the *csn_validate* function for flow and ambient data (function names: *Flow.check* and *Amb.check*, respectively) to compare reported values with instrumental limits to determine if additional flagging is needed. Both functions use the *filter.Filters* table from the CSN database.

The *Flow.check* function generates a dataframe within the R environment containing the records that have aberrant flow values but have not been invalidated. The null code "AH" will be applied to all of these filters, which represents "Sample Flow Rate or CV out of Limits" in the AQS database. The null code "SV" will be applied when the sample volume is out of limits. If either the "AH" or "SV" null codes are applied, all parameters except the ambient temperature and ambient pressure will be invalidated.

[1] Text in [brackets] indicates values that can be changed by the user. Other values should be typed as written.

The ambient check will not invalidate samples based on ambient temperature, ambient pressure, or transport temperature, but instead will append an appropriate code to appropriate parameters notifying end users that these operational parameters were outside the specifications only if the appropriate code has not already been applied. If the ambient pressure value is outside of the 600-810 mmHg range then the "QP – Pressure Sensor Questionable" qualifier code will be applied to the ambient pressure parameter only. If the ambient pressure value is outside of the 450 to 1000 mmHg range for the URG or 450 to 850 mmHg for the SASS/SuperSASS – allowable limits as defined by AQS – then the "AN – Machine Malfunction" null code will be applied so the record can be delivered to AQS. If the ambient temperature is outside of the -20 to 45 °C range for the URG or -30 to 50 °C range for the SASS/SuperSASS then the "QT – Temperature Sensor Questionable" qualifier code will be applied to the ambient temperature parameter only. Further, if the ambient temperature is outside of the -40 to 55 °C range for both samplers – allowable limits as defined by AQS – the "AN – Machine Malfunction" null code will be applied so the record can be delivered to AQS. If the transport temperature is greater than 4 °C – as measured upon receipt at the Sample Handling Laboratory – then the "TT – Transport Temperature is Out of Specs" qualifier code will be applied to all analytical species parameters for the given filter. The "X – Filter Temperature Difference or Average out of Spec" qualifier code is assigned by the site operator and is specific to temperature of the filter; it is only applied to analytical species parameters for a given filter.

Analytical data are checked using various functions that are automatically run when the *csn_validate* command is executed. The three main checks performed, which exclusively use the data collected from the database during the first step of the *csn_validate* command, are for ions (function name: *ions.check*), carbon (function name: *carbon.check*), and a cross-module comparison, hereafter referred to as the XRF-IC check (function name: *xrfic.check*).

The ions check compares the anions and cations data to aid identification of outlier ions measurements. The carbon check compares organic carbon and elemental carbon to aid identification of outlier carbon measurements. The XRF-IC check compares sulfur and potassium elemental concentrations from the PTFE filter to sulfate and potassium ion concentrations from the nylon filter. The potassium elemental concentrations are used to calculate non-soil potassium (where non-soil K is K minus (2/27) × Si) for comparison with the potassium ion because the soil-bound potassium is not represented in potassium ion. The ratio of each corresponding element/ion pair is compared to aid identification of questionable samples. For each of the checks, a summary of the results is produced with outliers indicated. Sample records are not invalidated during this process; however, a qualifier code may be automatically applied if comparison results lie outside the predefined limits. If flagged as an outlier, a comment is added to the record to indicate which check produced the qualifier code. An example of automatic flagging is where the "5 – outlier" qualifier code is applied to all analytical species from the PTFE and nylon filters (i.e. elements and ions, respectively) when the sulfur/sulfate ratio is outside of the predefined limits. This is the only qualifier code that is applied from the automated analytical checks for ions, carbon, and XRF-IC.

Another automated check attempts to identify swaps of sample data. The swap check calculates two indices, one assuming two samples were not swapped and the other assuming the samples were swapped. These indices use the sulfur/sulfate ratio to identify potentially swapped samples for PTFE and nylon filter samples. Similar to the *XRF-IC check* output, the swapped variable will contain concentrations for sulfur and sulfate, outliers by the sulfur/sulfate ratio, and potentially swapped samples based on the calculated indices.

The other functions performed by the *csn_validate* command include sample-field blank swap check (function name: *get_blank_swaps*), a mass check (function name: *mass.check*), a completeness check (function name: *csn_complete*), qualifier code checks (function name: *qualifiers*), and a function to process invalid filters (function name: *csn_X*).

The sample-field blank swap check compares field blanks with sample filters to determine if a filter swap may have occurred. The *mass.check* uses the collocated mass concentrations retrieved from AirNowTech to compare with reconstructed mass. When available, it also compares gravimetric mass with reconstructed mass (though gravimetric mass is not typically included in the CSN measurement suite). No codes are applied during the mass check; however, the output is useful for the analyst to identify potential issues. The completeness check ensures that all reportable parameters for every sample event have complete records. If any expected parameter records are missing a value and null code, the function fills in with a "AM - Miscellaneous Void" null code for clarification during review in DART by the SLT validators. The qualifier code check and invalid filters check ensure that codes are associated with appropriate parameters and output lists of all records with qualifier codes and null codes, respectively. Each of the available qualifier codes and null codes are applied for either a single parameter, a group of parameters (such as all analytical or all operational parameters), or all parameters relating to a given filter type. Some parameters have both program defined and AQS defined (outside of which AQS will not accept the data record) acceptable value ranges, requiring qualifier code or null code application for cases where values fall outside of the acceptable ranges. As part of the qualifier code check, the Intended Use Date and Sample Start Date of filters are compared. If the dates do not match, the "2 – Operational Deviation" qualifier code and an appropriate comment are added to the filter parameters. As part of the invalid filters check, the "AI – Insufficient Data (cannot be calculated)" null code is applied to composite parameters when it is not possible to calculate the concentrations.

Although these checks are performed automatically within *csn_validate*, the analyst should review the outputs from all of the checks to further investigate issues and/or confirm that codes have been applied appropriately. To further investigate issues, the analyst may need to work with the Sample Handling Laboratory or analytical laboratories and/or use other available tools, such as regional concentration comparisons, to determine if additional codes or comments should be applied manually.

8.3 Review Summary Plots

To generate a summary plot, the analyst can use a variety of R plotting tools or custom plotting tools in the d*atvalCSN* package. The analyst should review the help pages for each of the available plotting functions. For example, to view the results of the XRF-IC check function at a particular CSN site AA-BBB-CCCC, the analyst can run:

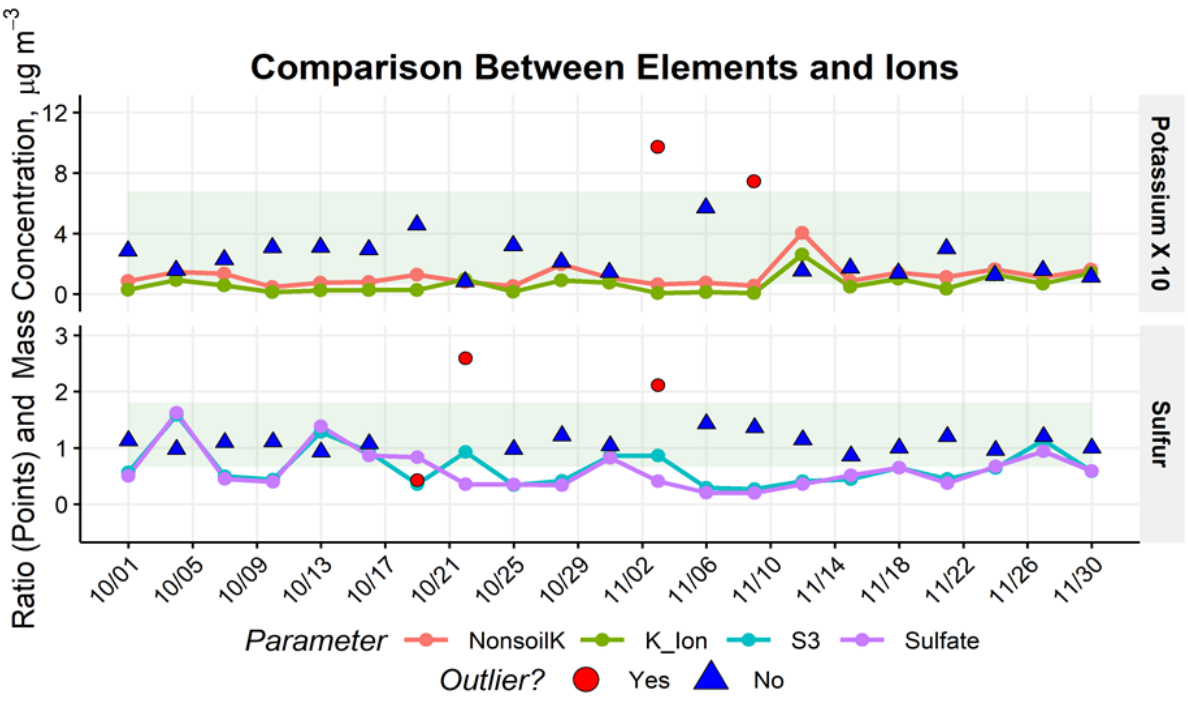
[results] <- datvalCSN::csn_get(start.date = ['YYYY-MM-DD'], end.date = ['YYYY-MM-DD'], Site = ['AA-BBB-CCCC'], FBs = [FALSE])

[xrfic] <- datvalCSN::xrfic.check([results])

datvalCSN::plot_xrfic([xrfic], AqsSiteId =['AA-BBB-CCCC'], POC = ['D'], y.limits = c([y1, y2]))

where "*AA"* is the state code, "*BBB*" is the county code, "*CCCC*" is the site code, and "D" is the Parameter Occurrence Code (POC). To reduce database querying time, typically only the sample data is plotted; the option to retrieve field blank data, *FBs*, is set to FALSE. If start and end dates are not specified then the number of days back in time from the last sampling date of the processed data available in the database is used. This can be specified by the *csn_get* function parameter *days*. To view the current month of data in context, it is recommended that at least 90 days of data is reviewed; for example, see Figure 1.

Figure 1. Sample output of *plot_xrfic* for comparing EDXRF and IC results. The blue and pink lines are sulfur and elemental non-soil potassium measurements by EDXRF; the purple and green lines are 3× sulfate and potassium ion measurements by IC. The points are the EDXRF/IC ratio. Outliers are indicated per ratio for cases that fall outside of the defined limits (shaded areas). The '5 - outlier' qualifier code is applied for sulfur/sulfate outliers where concentrations are above the MDL.



It may be necessary to change the axis ranges on a plot to view the data more closely in which case the *y.limits* variable can be utilized. However, as data is reviewed on a monthly basis, for convenience, the analyst can enter the following command:

create_monthly_plots(['Month'], [MonthData])

to produce plots for each site and each check. The default directory for saved plots is *U:/CSN/QA/Temporary Plots/*, with a subfolder based on the month of interest (e.g., *Month Plots*). The *['MONTHYY']* parameter is a character string that will be used to name the plot folder while the *[MonthData]* parameter is the "validation" object created by the *csn_validate* function. The analyst should review plots for each site to identify potentially swapped samples as well as aberrant data that has not yet been flagged.

8.4 Review all data via web application

In order to facilitate efficient review of all data for a given month, custom web applications (webapps) were developed. Three important webapp tools for the CSN analyst include:

- CSN Data Management – csn.aqrc.ucdavis.edu
- CSN Data Explorer – analysis.crocker.ucdavis.edu:3838/csnData/
- CSN Status Explorer – analysis.crocker.ucdavis.edu:3838/csnStatus/

The CSN Data Management page is an online access point to interact with the UCD CSN database. With this tool, the user can look up filter, batch, and site information. Qualifier codes and null codes can be assigned here with an explanatory comment. Each time a code is changed, a timestamped record is included. This is also the portal for importing data from external laboratories or sources. Files from the respective folders in the networked U drive are uploaded following the instructions in TI 801A.

The CSN Data Explorer page has multiple tabs for viewing the resultant data in various forms. The analyst should familiarize themselves with the intuitive controls, especially within the "Early Review" (Figure 2), "Explorer" (Figure 3), "Field Blanks" (Figure 4), and "Parameter Review" (Figure 5) tabs. The data displayed in the "Early Review" tab enables the user to identify anomalies in the PTFE and nylon filters by reviewing the sulfur/sulfate and non-soil potassium/potassium ion ratios. It is important to note the "Early Review" tab may be used prior to receiving a complete, monthly dataset. This tool utilizes reported sulfur and sulfate mass loadings before the post processing steps described in TI 801B. Similarly, the "Carbon" tab displays raw OC and EC concentrations, providing a tool for the user to identify anomalies in the quartz filters and/or TOA analysis. The "Explorer" tab is a comprehensive tool enabling full chemical composition to be assessed using spatial and temporal comparisons. The "Field Blanks" tab compares field blanks with associated sample filters. This tool enables rapid identification of unusually high field blank mass loadings that may indicate a swap between field blank and sample filters. The "Parameter Review" tab must be reviewed by the analyst for each monthly dataset. With this tool, every reported data point can be inspected quickly and efficiently within the context of the monthly and historical network data.

Figure 2. The Early Review tab of the CSN Data Explorer webapp showing the sulfur/sulfate and non-soil potassium/potassium ion time series for the specified time range at a specified site.

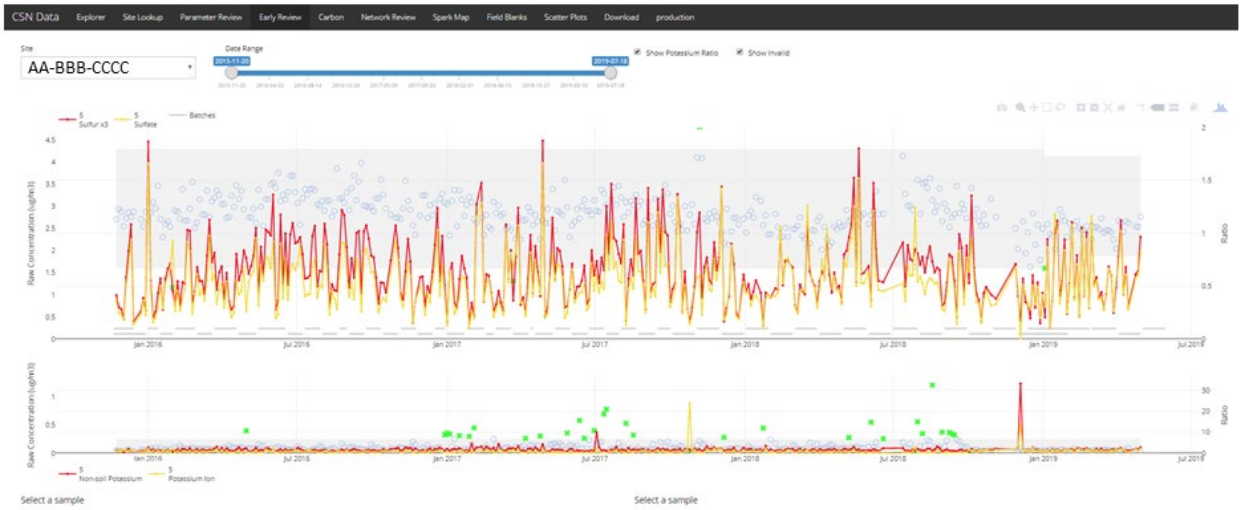


Figure 3. The Explorer tab of the CSN Data Explorer webapp showing various tools available to investigate and compare selected data.

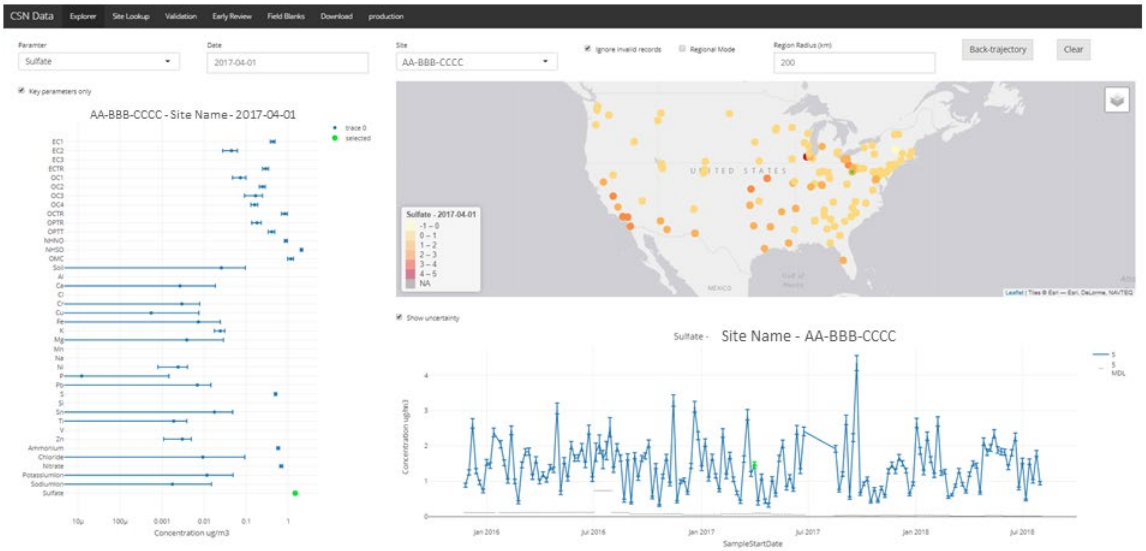
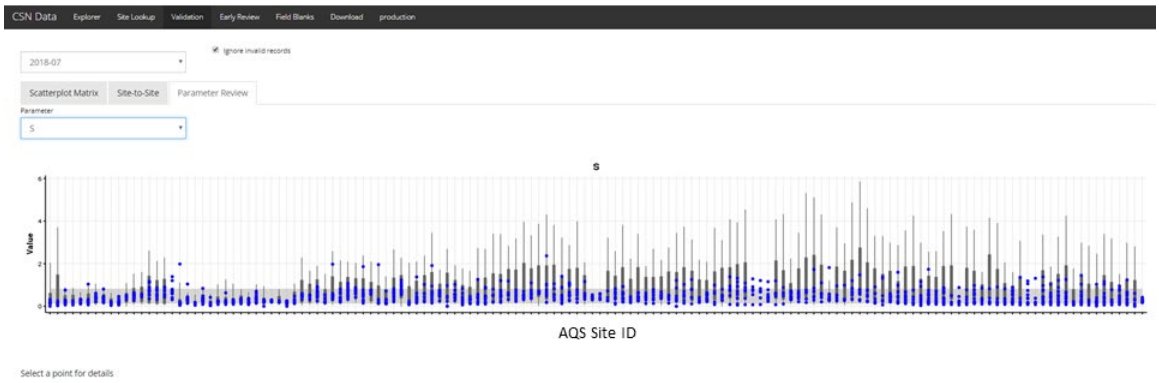


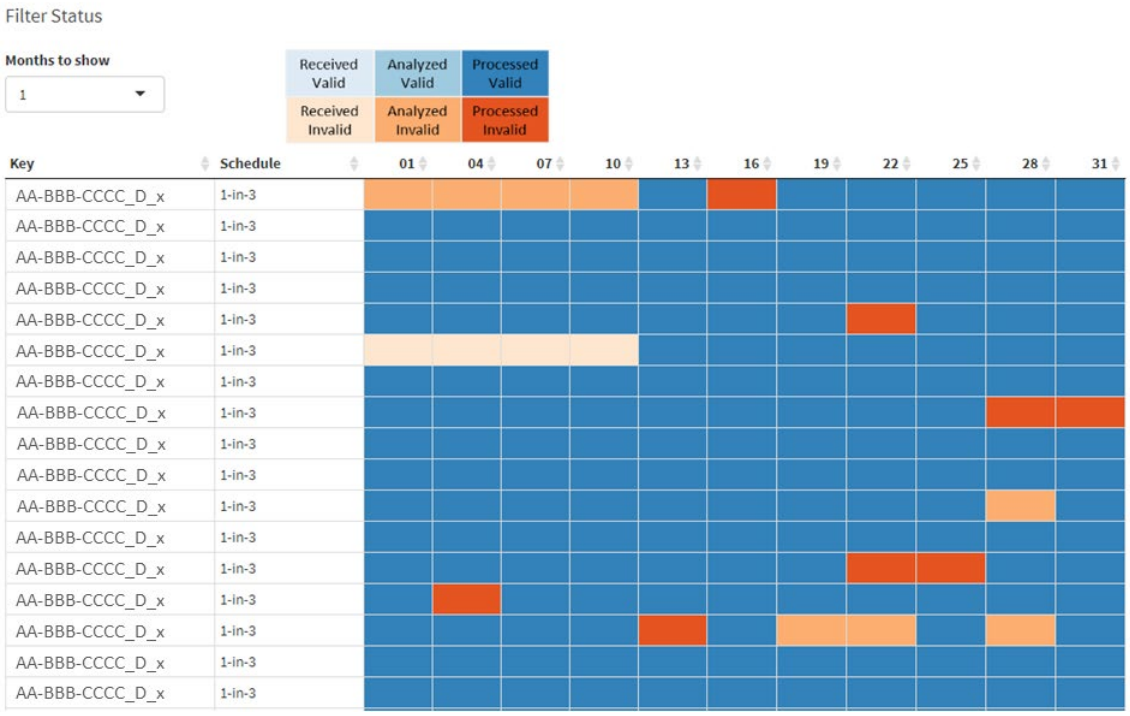
Figure 4. The Field Blanks tab of the CSN Data Explorer webapp showing the mass loading of a selected species from one of the three filter types for all field blank filters across the network.

Figure 5. The Parameter Review tab of the CSN Data Explorer webapp showing the sulfur concentrations for the entire CSN. Sites ordered from west to east on the x-axis and designated by the AQS Site ID. The blue points represent the data for the month selected while the tops of the gray box and whisker indicate historical 90th and 99th percentiles, respectively.



The CSN Status Explorer page is useful for monitoring the progress of analyses and timelines for delivery. There are myriad data views and the user should explore the various figures and tables. An important portion to review for validation is the “Analysis Completeness” tab, where the analyst should inspect filter records not analyzed; all unanalyzed filters should either be flagged as invalid, or be queued for analysis. Similarly, the “Status Grid” tab provides a useful visualization to identify consecutive invalid samples, missing analysis data for a given filter record, and duplicate issues. This grid can also be used to check overall network issues such as bulk errors in the electronic data. Figure 6 is a screenshot of part of the “Status Grid” tab from the CSN Status Explorer page.

Figure 6. Snapshot of the Status Grid tab of the CSN Status Explorer webapp. The grid is generated for each of the three filter types and shows information for each site from the selected month. Each cell denotes an intended sampling date for a given site and is color coded according to filter status.



Other important views from the CSN Status Explorer page include the "Level 0 Checks" tab, a visual tool of various checks performed dynamically, which can be used to identify issues with the data. Filter records require further investigation by the analyst for these scenarios:

- The start date of a sample filter is missing and there are no comments associated with the filter;
- The end date of a sample filter is missing and there are no comments associated with the filter;
- The start date of a sample filter is not NULL, is the same as the end date, and there are no comments associated with the filter;
- The filter sampled for more than 24 hours and there are no comments associated with the filter;
- The calculated sample volume is more than 10% different from the reported sample volume and there are no comments associated with the filter;
- The sample is marked as invalid but there are no comments associated with the filter;
- The Filter Analysis ID is missing and there are no comments associated with the filter;
- The start date is outside of the expected range for the month of interest;
- The Filter Type does not match the expected Analysis Type;

- The "TT" qualifier code is inconsistent with the transport temperature reported;
- The sample volume is missing and there are no null codes.

9. DATA VALIDATION EQUATIONS

The following section presents the equations used to calculate swap check indices. These calculations are performed by the *datvalCSN* R package.

The ratio of sulfur to sulfate in the ambient atmosphere is generally well known. Since the majority of sulfur-bearing aerosols by mass are in the form of sulfate, the stoichiometric ratio of 3×sulfur / sulfate is typically close to one. The calculated indices of the swap check function utilize this tendency. The first index assumes that two subsequent samples are not swapped. The ratio of sulfur by EDXRF to sulfate by IC is subtracted by one for each sample, and the two results are multiplied. Since both ratios are expected to be close to unity, a small number is expected. Inversely, the second index assumes two samples have been swapped. This irregularity would result in a significantly larger number than index 1. Mathematically,

$$Index1 = \left(\frac{S3_1}{SO4_1} - 1\right) * \left(\frac{S3_2}{SO4_2} - 1\right) \quad (1)$$

$$Index2 = \left(\frac{S3_1}{SO4_2} - 1\right) * \left(\frac{S3_2}{SO4_1} - 1\right) \quad (2)$$

Where,

$S3_x$ = sulfur concentration ($\mu g/m^3$) multiplied by 3

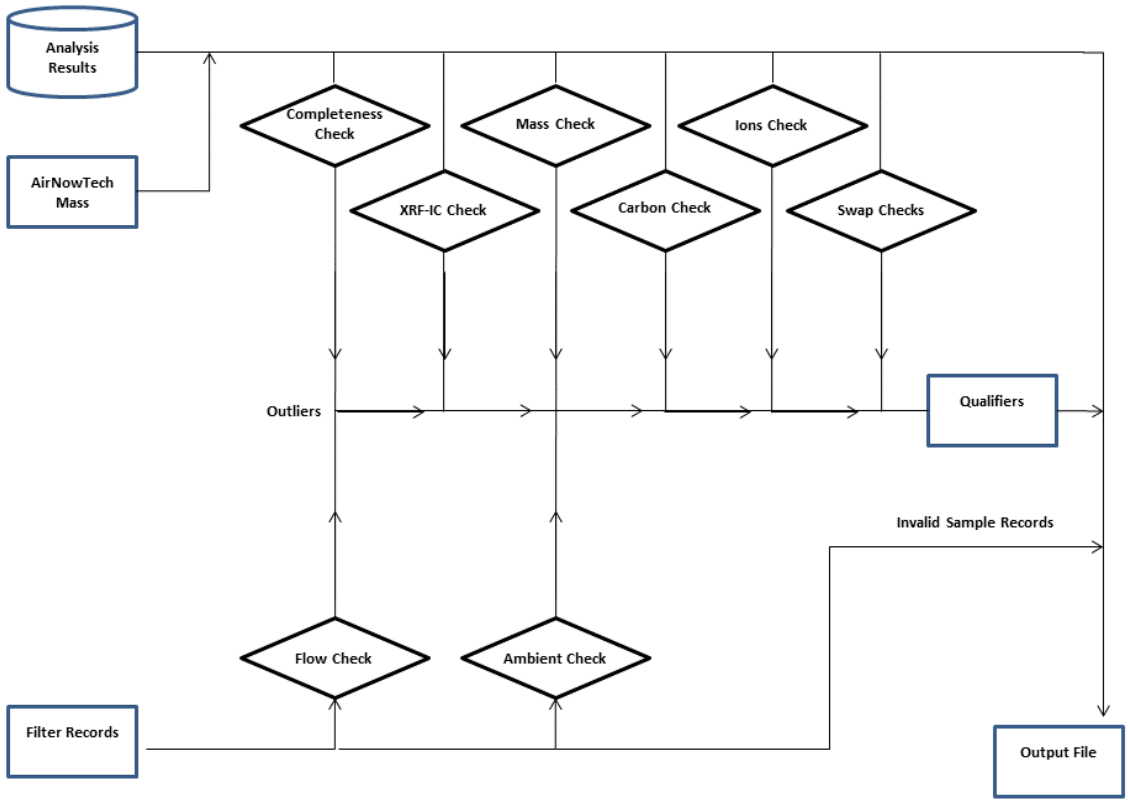
$SO4_x$ = sulfate concentration ($\mu g/m^3$)

Subscripts denote the sampling event of interest (1 for target event, 2 for subsequent event). Empirically, potential swaps are indicated by an Index1 < -0.03 and an absolute value of Index2 < 0.05.

10. DATA PROCESSING CODE

This section describes the data flow through the data validation code used to execute CSN validation checks. Figure 7 outlines the flow of data from the filter and analysis results database tables to final results. The wrapper function *csn_validate* is the only function executed directly by the analyst (see Section 8.1); *csn_validate* in turn calls several functions sequentially to generate data frames with outliers identified. Source code for the functions shown in Figure 7 is stored in the AQRC source repository.

Figure 7. Flow diagram of the validation code in datvalCSN::csn_validate. Rectangles represent data files, diamonds represent R functions, cylinders represent databases, and lines represent inputs and outputs.



11. EQUIPMENT AND SUPPLIES

The hardware and software used for CSN data validation are described in the associated *UCD CSN SOP #801: Processing & Validating Raw Data*.

12. 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.

13. REFERENCES

Not Applicable.