

Data Validation for the Chemical Speciation Network

A Guide for State, Local, and Tribal Agency Validators

Version 2.0

July 25, 2019

Prepared for:

*U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711*

Prepared by:

*Air Quality Research Center
University of California, Davis
Davis, CA 95616*



Table of Contents

1. Purpose	1
2. Validation Pathway.....	2
3. Validation Philosophy	3
4. Validation Process	3
4.1. Level 1B Data Import Validation.....	3
4.2. Level 1B Automated Flagging	4
4.3. Level 1B Manual Review.....	6
4.4. Submission to DART	7
5. Level 2 Validation by SLT Agencies	8
5.1. Notifying End Users of Local Impacts.....	9
5.2. Making Use of Local Knowledge	11
6. The Do's and Don'ts of DART	11
6.1. Suggested Analyses.....	14
6.1.1. Regression Analysis.....	14
6.1.2. Multi-Parameter Trends	14
6.1.3. Geospatial analysis.....	15
7. Explanation of Terms	17
7.1. Field Blanks	18
7.2. Value Limits.....	18
8. Further Information	19
9. Appendix I – AQS Codes	21
10. Appendix II – Parameter List	24

Table of Figures

Figure 1. Graphical representation of CSN data flow.....	2
Figure 2. Comparison of network-wide sulfur and sulfate measurements.	5
Figure 3. Inter-method comparison check.	7
Figure 4. Elevated concentrations for several elements observed on January 1, 2016.....	11

Figure 5. A year-long time series of organic carbon, potassium, and reconstructed mass concentrations from a single site.....	15
Figure 6. Time series of sulfate concentrations from seven sites within a 200km radius.	16

Table of Tables

Table 1. Qualifier flags applied to data during level 1B data validation.	4
Table 2. Calculated parameters delivered to DART.	17
Table 3. Limits assigned for operational parameters.	19
Table 4. Limits for inter-parameter comparisons.	19

1. Purpose

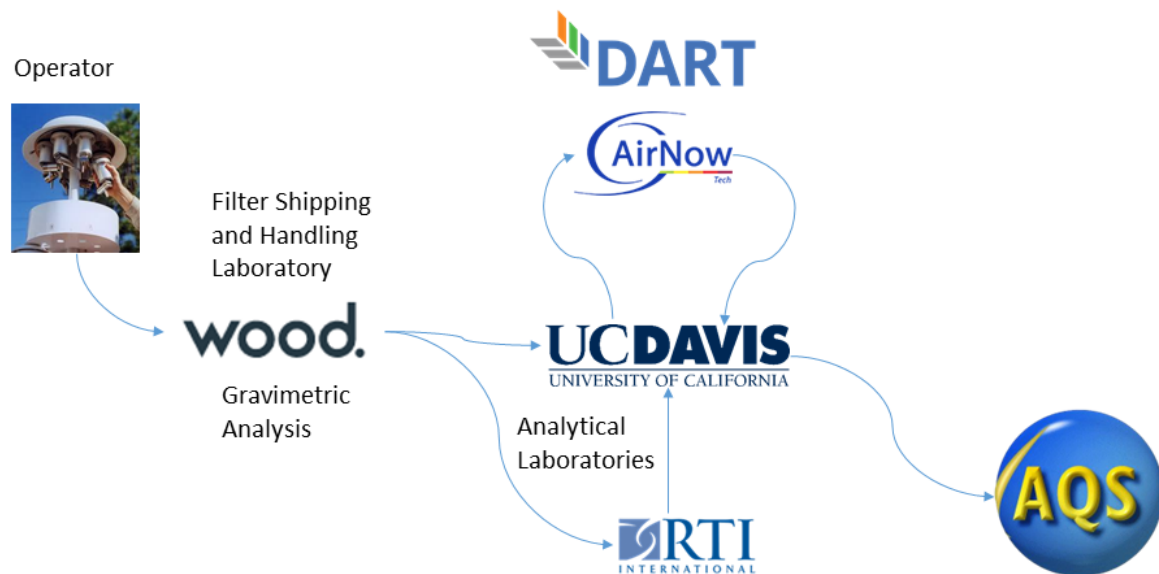
The Chemical Speciation Network (CSN) is a complementary network to the National PM_{2.5} Mass Monitoring Network. The goal of CSN is to measure ambient PM_{2.5} concentrations to be compared against the PM_{2.5} National Ambient Air Quality Standard (NAAQS), although CSN data are not used for attainment or nonattainment decisions related to PM_{2.5} mass. The programmatic objectives of the CSN network are to provide annual and seasonal spatial characterization of aerosols, air quality trends information for analysis, tracking the progress of State Implementation Plan (SIP) control programs, and development of emission control strategies.

Since CSN data are intended for public distribution, there are many other potential uses of the dataset. Primary stakeholders in the CSN are the decision-makers of State, Local, and Tribal (SLT) agencies, who use the data for model inputs as well as for development of emission control strategies and determination of their long-term effectiveness. Secondary stakeholders include EPA analysts who may use CSN data to determine trends of PM_{2.5} chemical species over time, relate the data to health effects, and to develop and evaluate air quality models. Other users may include public health officials, epidemiological researchers, and air quality researchers.

The primary objective of this document is to inform SLT data validators of the validation practices performed during sample handling and data processing for the CSN, as well as provide some guidance on potential validation strategies. The initial stages of the validation process are performed by Wood, PLC (Wood) and University of California, Davis (UCD) (Sections 2-4). SLT validators are then able to use the Data Analysis and Reporting Tool (DART) interface (Section 5) to perform higher level validation and data analysis (Sections 6 and 7). Meant as a tool to clarify and define the validation process, this document is one of many resources available (Section 8).

2. Validation Pathway

Figure 1. Graphical representation of CSN data flow.



Data generation begins at sample preparation and collection. The filter shipping and handling laboratory at Wood generates data associated with filter sampling events. The site operator provides feedback from the sampling instruments as well as notes on sampling conditions. This information is passed back to Wood via the Field Sampling Chain of Custody (FSCOC) forms where it is entered into a database and compiled for export to UCD for data processing. Wood also analyze the PTFE filters for gravimetric mass before shipping to the analytical laboratories. The analytical laboratories receive the filters from Wood, analyze the filters to generate concentration data along with additional notes on filter and analysis conditions, and deliver the data to UCD for further processing. UCD ingests all data into a database and calculates the results for each measurement result, hereafter referred to as “parameter.”

Early stage validation (Level 0 and 1A, see Section 4) is performed at each point that a person handles data or physical samples. Once the preliminary data are prepared at UCD, further validation (Level 1B, see Section 4) is performed prior to delivering the dataset to DART. All critical parameters are passed into DART where the data owners (SLTs) can review and provide additional information and/or corrections. Data in DART has some differences relative to data delivered to the EPA Air Quality System (AQS) database. As discussed in Section 4.4, some of the parameters reported to DART are not delivered to AQS.

After data have been validated by the SLT agencies in DART, it is returned to UCD for reformatting and delivery to AQS. At this stage, data are reviewed to ensure consistency between the final data and SLT validator comments. If discrepancies are found, the validators are contacted and every effort is made to resolve the issue before final delivery. The AQS format does not include

comments; therefore, it is important that observations are communicated to the end users through AQS qualifier and null flags, see Section 5, *Level 2 Validation by SLT Agencies* for a list of flags.

3. Validation Philosophy

The core principles guiding the data validation performed at UCD are:

- All data must be validated;
- Definitive evidence is required to invalidate records;
- Data are not censored; and
- Validation procedures are revisited and improved over time.

When reviewing data, it can be tempting to remove points that do not “fit” an observed profile. However, it is critical to retain all viable data points for an unbiased dataset. Clear evidence, such as sampler malfunction or sample damage, is required for a sample or data to be invalidated. There are numerous qualifier flag options to inform the end user. Similarly, negative values are included in the final dataset as they arise from the subtraction of background signals and can be understood as “indistinguishable from zero.” Censoring negative values artificially biases the dataset and can be problematic for statistics.

Data validation is an evolving process. Over time, as the data is better understood, methods for checking data integrity are updated and improved.

4. Validation Process

Validation of CSN data is divided into four steps:

- Level 0 – Performed by SLT site operators, Wood technicians, and UCD data analysts
- Level 1A – Performed by laboratory technicians and analysts
- Level 1B – Performed by UCD data analysts (Sections 4.1, 4.2 and 4.3)
- Level 2 – Performed by SLT agency analysts (Section 5)

This document describes Level 1B and Level 2 validation procedures. The validation protocols for Level 0, Level 1A, and Level 1B can be found in UCD Processing & Validating Raw Data standard operating procedure (SOP) documents, available at <https://aqrc.ucdavis.edu/documentation>.

The Level 1B validation performed at UCD focuses on data integrity and completeness using a combination of automated and manual checks. The automated checks find simple errors such as typographical errors or low flow rates. The manual checks involve control charts and graphical visualizations used to focus on identifying potential filter swaps or inconsistencies in reporting.

4.1. Level 1B Data Import Validation

When electronic data from Wood are ingested, the import program performs several checks to verify the form of the electronic file (i.e. field names), linking of data to existing sites, format of

dates, and unique sample event IDs. The ingest process is performed by a UCD data analyst who identifies and investigates potential issues with filter-specific information. Since the data from the FSCOC forms are entered manually, some typographical mistakes are inevitable. The automated check compares the intended use date, the sample start date, and the sample end date for each sample event record. Operational parameters are reviewed manually to identify inconsistencies such as issues with ambient temperature and pressure, flow rate coefficient of variation (CV), and sample volume. UCD notifies Wood of inconsistencies identified during Level 1B validation, and these are typically resolved prior to submission to DART. In cases where the inconsistencies are not resolved by UCD and Wood, comments are left for the DART reviewer for final determination.

4.2. Level 1B Automated Flagging

Parameter limits are defined for the CSN which enable assignment of some automated null (invalidated data) or qualifier flags. All automated flags assigned by UCD data analysts are accompanied by comments. The only null flags assigned automatically are “AH” for flow rate issues and “AM” for miscellaneous voids when no data exists for an expected result. The “AH” flag is applied to samples that were 1) collected with an air flow outside of $\pm 10\%$ of the nominal flow rate and 2) have a flow CV greater than 5% for Met One SASS and 2% for URG 3000N instruments.

Automated qualifier flag assignments include “TT”, “MD”, “QP”, “QT”, and “5”, as detailed in Table 1.

The “TT” code identifies samples received at Wood above 4 °C. The “MD” flag indicates a species parameter was measured below the method detection limit. The “QP” and “QT” qualifier codes indicate issues with either the pressure or temperature sensor, respectively, and are applied to samples that were collected outside the bounds of the manufacturer’s specifications. The CSN samplers use mass flow controllers that calculate and adjust flow based on sampling temperature and pressure. If the sensors measuring these parameters are faulty or are operating outside the bounds of the manufacturer’s specifications, then the sample flow is suspect; erroneous sample flow is propagated to the reported concentrations; therefore, it is important to notify the end user that the data may be affected.

Table 1. Qualifier flags applied to data during level 1B data validation.

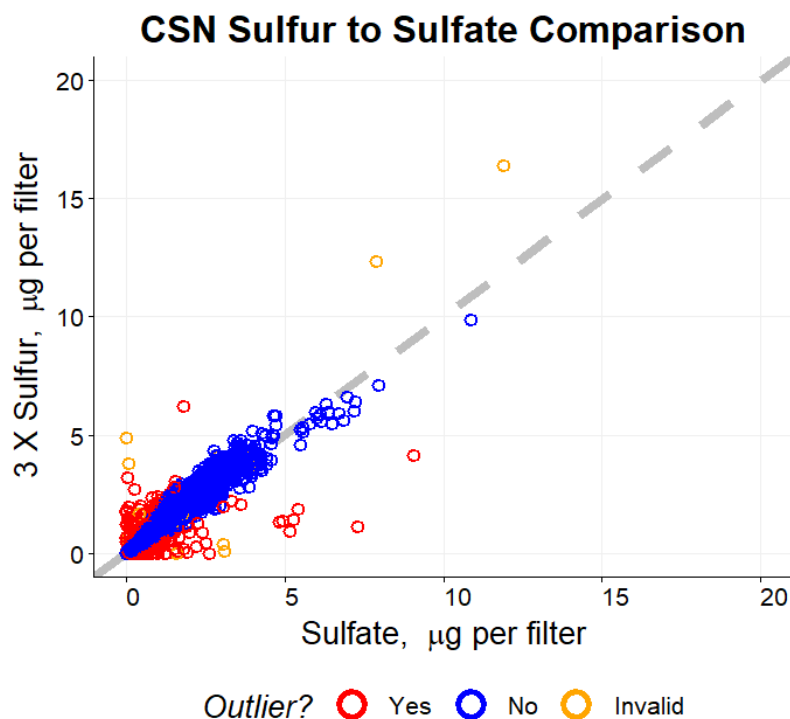
Flag	AQS Description	Usage
TT	Transport Temperature is Out of Specs.	Samples received at Wood > 4 °C
MD	Value less than MDL	Measurement value < parameter MDL
QP	Pressure Sensor Questionable	Pressure readings outside instrument limits*
QT	Temperature Sensor Questionable	Temperature readings outside instrument limits*
5	Outlier	3×S / SO ₄ ratio is outside expected limits†

* See Table 3 for further details.

† Note that this applies to January 2019 data onwards. Prior to January 2019, the “5” qualifier was applied when the 3×S / SO₄ and potassium/potassium ion ratios were outside expected limits. See Table 4 for further details.

The sulfur and non-soil potassium concentrations from the polytetrafluoroethylene (PTFE) filter are compared to the sulfate and potassium ion concentrations from the nylon filter (see Figure 2 for an example of the comparison of sulfur and sulfate). Non-soil potassium, defined as $K - (2/27) * Si$, is used to compare with the potassium ion because the soil-bound potassium is not represented in potassium ion concentrations. However, the “5” qualifier flag is only applied when the sulfur to sulfate comparison is outside of pre-defined limits (detailed in Table 4) and at least one of the two species is greater than its respective MDL. This check compares the PTFE and nylon filters but there is not currently an inter-comparison for the quartz filter. In the future, optical measurements of the PTFE filter may be used to compare results with the elemental carbon concentrations from the quartz filter. The “5” qualifier flag is also applied when the gravimetric mass differs greatly from the reconstructed mass (defined in Section 7). Gravimetric mass is not routinely measured at CSN sites. However, gravimetric mass from the PM_{2.5} NAAQS monitoring network and the gravimetric mass reported to AirNowTech is used for comparison with reconstructed mass determined from the CSN data.

Figure 2. Comparison of network-wide sulfur (S) and sulfate (SO₄) measurements. The molecular weight of SO₄ (96 g/mol) is three times the atomic weight of S (32 g/mol), so the concentration ratio (3×S)/SO₄ should be one if all particulate sulfur is present as water-soluble sulfate. In practice, real measurements routinely yield a ratio greater than one. The grey dashed line represents the 1:1 ratio between 3×S:SO₄. Data points are marked as outliers if the ratio is outside of pre-defined limits and at least one of the two species is above its respective MDL. Outliers require further inspection.

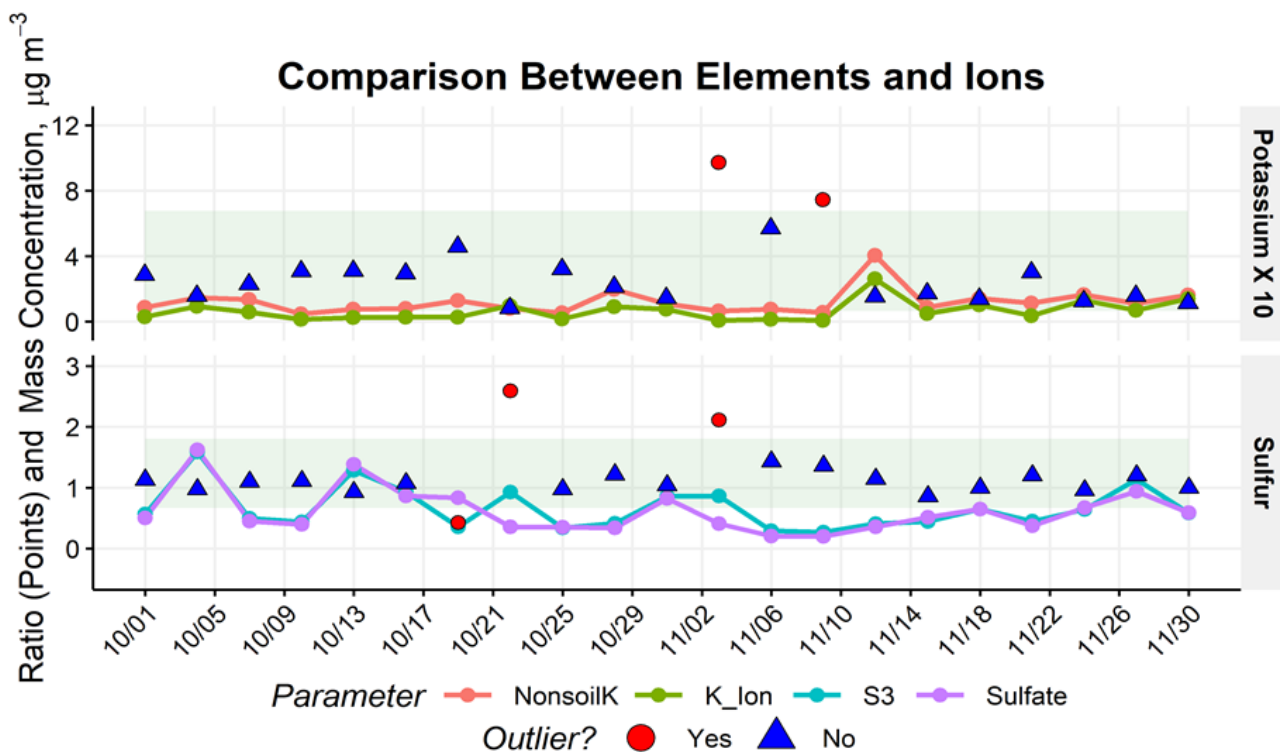


4.3.Level 1B Manual Review

The UCD data analysts perform a series of checks to ensure data completeness and integrity prior to submission to DART, including:

- Field blank swap check – comparison of the field blank measurement with both the associated sample measurement as well as statistical benchmarks for each month of sampling (95th percentile of field blanks and 5th percentile for sample measurements) to determine if two filters have been accidentally switched, or “swapped.” A list of suspect filters is reviewed manually by the data analyst.
- Sample swap check – potential swaps are inspected based on proximity in analysis time as well as sampling time and site. Time-series of each parameter are reviewed and compared with collocated instruments and/or near-neighbor measurements to provide evidence for a physical sample swap.
- Outlier check – measurements are compared with historical results as well as nearest neighbors to place them in context. **It is important to note that data are not invalidated due to high or low values.** The “5” qualifier flag may be used during Level 2 validation to alert the data end user that a value is abnormally high or low.
- Invalid sample check – review samples flagged as invalid to ensure correct application of null flags and investigate any potentially valid samples.
- Flow check – although flow rates are checked automatically, these are checked again manually to ensure proper calculation of ambient concentrations.
- Inter-method comparison check – these checks use collocated comparisons to detect sampling or analysis problems (see Figure 3 as an example).
- Contamination check – contamination is detected by reviewing network-wide results of individual species and confirmed through laboratory re-analysis, as requested by the UCD data analyst.

Figure 3. Inter-method comparison check used by UCD data analysts to identify potential sample swaps or sampling issues. The colored lines show the mass concentration results while the points indicate the inter-method ratios ($3\times S / SO_4$ and non-soil potassium / potassium ion). The potassium concentrations (lines) are multiplied by 10 in order for the concentrations to be plotted alongside the ratio values (points). The green shaded area denotes the pre-defined limits for validity. When the $3\times S / SO_4$ comparison is less than 0.784 or greater than 1.731, all analytical species from the PTFE and nylon filters (elements and ions species, respectively) receive the “5” qualifier flag. Evidence of a potential swap is seen for the samples on 10/19 and 10/22, particularly evident in the $3\times S / SO_4$ time series.



4.4.Submission to DART

After UCD data analysts complete Level 1B validation, the data is submitted to DART for SLT validator review. Additional parameters (e.g. reconstructed mass) are provided in DART to assist in data validation, even though they are not delivered to AQS. Only the blank subtracted elemental and organic carbon concentrations (parameters 88321 and 88320, respectively) measured using thermal/optical reflectance (TOR) are reported to DART. Other carbon parameters, including carbon fractions and non-subtracted concentrations, are excluded from DART to reduce redundancy in data validation. The differences between data available in DART and AQS are summarized below and in Appendix II.

Reported to DART but not AQS, defined in Section 7:

- Summed Elements – Parameter 00001 (calculated in DART)
- Summed Ions – Parameter 00002 (calculated in DART)
- Transport Temperature – Parameters 00007, 00008, and 00009 (reported by Wood)
- PM_{2.5} Mass Difference – Parameter 00010 (calculated in DART)
- Ammonium Sulfate – Parameter 88339
- Ammonium Nitrate – Parameter 88344
- Soil – Parameter 88348
- Organic Mass by Carbon – Parameter 88350
- Reconstructed Mass – Parameter 88401
- PM_{2.5} Raw Data – Parameter 88501

Not reported to DART but reported to AQS:

- Blank-Subtracted Carbon Fractions – Parameters 88324 to 88327 and 88329 to 88331
- Unsubtracted (Raw) Carbon Fractions – Parameters 88374 to 88378 and 88383 to 88385
- Blank-Subtracted Thermal Optical Reflectance (TOR) Pyrolyzed Carbon – Parameter 88328
- Raw TOR Pyrolyzed Carbon – Parameter 88378
- Blank-Subtracted Thermal Optical Transmission (TOT) Pyrolyzed Carbon – Parameter 88379
- Raw TOT Pyrolyzed Carbon – Parameter 88388
- Raw TOR Elemental Carbon – Parameter 88380
- Blank-Subtracted and Raw TOT Elemental Carbon – Parameters 88381 and 88357, respectively
- Raw TOR Organic Carbon – Parameter 88370
- Blank-Subtracted and Raw TOT Organic Carbon – Parameters 88382 and 88355, respectively

5. Level 2 Validation by SLT Agencies

SLT validators review site specific data in DART using their knowledge of sampling conditions, significant sources, emission events, and historical context.

The usage of DART is covered in depth during DART training sessions, available online at AirNow Tech (login required):

<https://dart.airnowtech.org/welcome>

Additional resources exist on YouTube:

<https://youtu.be/bNSjMgVSdj0> (July 2019 webinar)

and, as archived content from the 2018 National Ambient Monitoring Conference:

https://projects.erg.com/conferences/ambientair/conf18/Young_Chemical%20Speciation%20Network.pdf

SLT validators are encouraged to use DART as well as other techniques to validate data. Section 5.1. through 6.1.3. provide an overview of suggested techniques for data validation.

5.1. Notifying End Users of Local Impacts

Local impacts can cause data abnormalities from specific events or sampling conditions. **Local impacts do not result in invalid data.** It is the responsibility of the data end user to determine what data to include in their analysis. It is the responsibility of the data validators to provide detailed information concerning the sampling conditions to enable the end user to make informed decisions. Informational flags (AQS data qualifier codes) exist to communicate local impacts, including:

- IA – African dust
- IB – Asian dust
- IC – Chemical spills and industrial accidents
- ID – Cleanup after a major disaster
- IE – Demolition
- IF – Fire - Canadian
- IG – Fire - Mexico/Central America
- IH – Fireworks
- II – High pollen count
- IJ – High winds
- IK – Infrequent large gatherings
- IL – Other
- IM – Prescribed Fire
- IN – Seismic activity
- IO – Stratospheric ozone intrusion
- IP – Structural fire
- IQ – Terrorist act
- IR – Unique traffic disruption
- IS – Volcanic eruptions
- IT – Wildfire - US
- J – Construction

Although the “IL” flag is sufficient to indicate a local impact on the data, additional flags may be added as needed. Recommendations for new flags can be submitted to the EPA AQS team online: <https://www.epa.gov/aqs/forms/aqs-contact-us>

AQS accepts up to ten qualifier flags per data record.

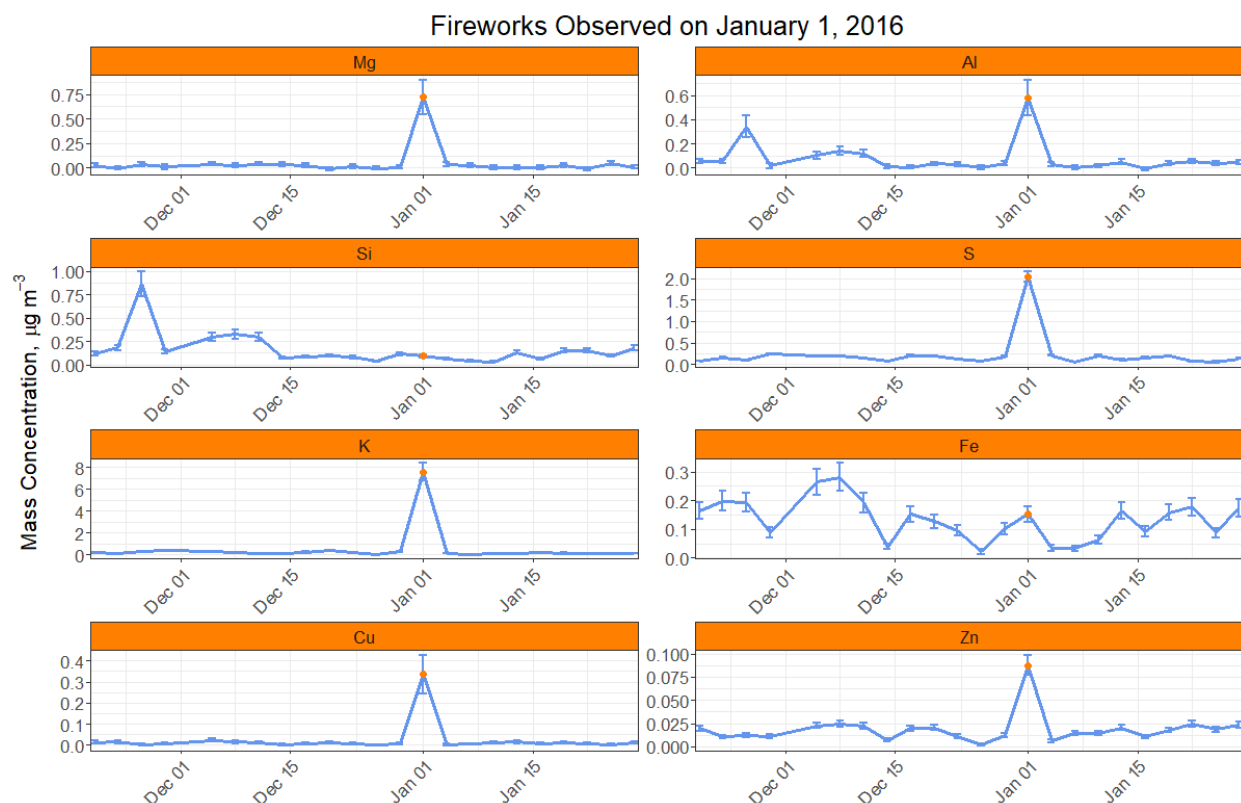
Some AQS qualifier flags have very similar descriptions to null flags. These include:

- Sample time – the “AG” null flag (Sample Time out of Limits) should be used if the elapsed sample time is <23 hours or >25 hours. The “Y” qualifier flag (Elapsed Sample Time out of Spec.) should be used if the elapsed sample time is 24±1 hours.
- Construction – the “AC” null flag (Construction/Repairs in Area) should be used if samples cannot be collected due to construction. The “J” qualifier flag (Construction) should be used if construction activities were present during the collection of the sample.
- Flow Rate – the “AH” null flag (Sample Flow Rate out of Limits) indicates that either the sample flow rate was insufficient to collect an appropriate sample or the sample flow variability (Flow CV) was out of limits (currently 2% for the URG 3000N quartz filter sampler and 5% for the MetOne SASS/SuperSASS PTFE and nylon sampler). The “W” qualifier flag (Flow Rate Average out of Spec.) indicates that the sample flow rate may have been high or low, but was within 10% of the nominal flow rate.
- Filter Integrity – the “AJ” null flag (Filter Damage) is used when a filter cannot be analyzed due to damage. The “FX” qualifier flag (Filter Integrity Issue) indicates a physical issue with filter (e.g. small hole, minor contamination, uneven sample loading) but analysis was performed.

Example

On January 1st, 2016, abnormally high potassium concentrations were observed at several sites, as well as elevated concentrations for several other elements (magnesium, aluminum, silicon, sulfur, iron, copper, and zinc). Upon further investigation, the spike in elemental concentrations was determined to be from local fireworks for New Year’s celebrations. The “IH - fireworks” qualifier flag was applied to these data.

Figure 4. Elevated concentrations for several elements observed on January 1, 2016.



5.2. Making Use of Local Knowledge

The SLT validators have the benefit of communication with site operators and access to field records for investigation of site specific issues and events. Additionally, local knowledge of site history can provide important context. For example, a site may have elevated levels of specific industrial emissions under certain meteorological conditions. To Wood technicians or UCD data analysts, this may appear to be an outlier and receive the “5” qualifier flag. However, the SLT validator may have local knowledge and apply a more appropriate qualifier flag such as “NS – Influenced by Nearby Source”.

Ambient aerosol concentrations are intimately connected to meteorology. Meteorological events such as storms, heavy fog, windy or stagnant conditions may be observed and recorded as qualifier flags by the site operator. These flags are propagated and delivered as part of the data to DART. It is beneficial for the SLT validator to incorporate their knowledge of local weather events into assessment of the data as well as include this information as flags for the benefit of the end user.

6. The Do’s and Don’ts of DART

The following list of common practices are recommendations to aid the SLT validation process. **The data produced by the CSN are owned by the SLT air quality agencies.** The SLT data validators have ultimate authority regarding any flags applied to the data.

PLEASE DO

- Check all invalid filters – It is possible for a sample to be erroneously invalidated at any stage in the process. Please review the null flags and cross-check with field records. The measured values for invalid samples have been included in DART for them to be reviewed alongside valid samples.
- Change the “AM” null flag to a more appropriate flag – The “AM” (miscellaneous void) null flag is automatically applied to data records that were not received from either the analytical labs or Wood. SLT validators may have further insight based on field records and site operator input, and may be able to apply a more appropriate flag.
- Add qualifiers flags – Up to ten qualifier flags are allowed per data record in AQS (see Appendix I for flag options). There is no need to overwrite an existing qualifier flag unless it is erroneous or there are already ten flags used. A searchable list of flags is available at <https://aqs.epa.gov/aqsweb/documents/codetables/qualifiers.html>.
- Invalidate samples with a serious sampling problem – Serious sampling problems include poor flow rate ($> \pm 10\%$ from specification) and inadequate sampling time ($> \pm 1$ hour). These issues should have been identified during one or more of the previous validation steps but should be confirmed in DART. Do not invalidate a sample if it lost less than one hour of sampling time; for these cases, the “Y” qualifier flag (Elapsed Sample Time out of Spec.) should be used.
- Check field blank concentrations – Field blanks may have higher than expected concentrations (where concentrations are calculated using a nominal sample volume), which could indicate sampler issues, filter set-up and run issues, or contamination. Compare field blank concentrations with historical field blank concentrations, field blank concentrations from other sites, where available, and associated sample concentrations.
- Check flow rates, dates, and operational parameters carefully – There are cases where flow values were incorrectly reported to UCD. When this happens, all associated concentrations must be recalculated and re-examined. The SLT validator should correct erroneous flow rates and operational parameter values using DART.
- Respond to questions in the comments section – Sometimes discrepancies in the data cannot be answered prior to DART. For these cases, the UCD data analysts ask questions to the SLT validator through the DART interface. In order for these issues to be resolved, the SLT validator must leave a clear and detailed comment responding to the question.
- Write clear and detailed comments – Abbreviations and cryptic references make it difficult to assess what is being communicated. The SLT validator must clearly indicate when action is required by the UCD data analysts, and specify the parameter(s), date(s), and describe the changes that need to be made (for example, “Please change the sampling date for all parameters relating to the quartz filter at site 12-345-6789 from 2017-01-01 to 2017-01-02 and remove the AG null flag.”)

- Be careful when applying flags to multiple parameters – DART provides a convenient option to propagate a selected flag to multiple parameters based on sampling event or filter type. The SLT validator must confirm that the appropriate propagation type has been selected to avoid accidental over-application of flags. For example, if flags only apply to the sample filter data, be sure to check that the field blank data are not also marked, and vice versa. Determine whether all parameters relating to a single filter are flagged or not flagged accordingly. For example, fireworks may result in higher than normal concentrations of some species so the analytical parameters may need to be flagged “IH” but perhaps the operational parameters such as sample flow rate CV may not need to be flagged. Finally, check that the propagated changes made are consistent with expectations.
- Review data flagged with the “A1”, “B1”, and/or “C1” flags – These flags, and associated comments, are only delivered to DART to aid validation and are removed prior to delivery to AQS. The flags denote changes made by Wood (“A1”) and UCD (“B1”) that will result in differences between the data reported in DART and the data recorded on the FSCOC. The SLT validator should review these flags and comments and confirm the data. The “C1” flag is manually applied by UCD data analysts and is used to highlight data that needs further review by the SLT validator to determine if and what actions are needed. The UCD data analyst also provides a comment detailing what aspect of the data is questionable.

PLEASE DO NOT

- Invalidate samples with the “FX” or “MX” qualifier flags (see Appendix 1) unless additional information support invalidation – These flags are selected during visual inspection of the filters at Wood and the analytical laboratories. If a filter defect results in poor analysis results, it is invalidated. When a filter is damaged or contaminated (e.g. grass, hair, insect) outside of the analysis area, a qualifier flag is applied. Contaminated areas can often be avoided during quartz filter analysis, but to a lesser extent for PTFE filters, and not for nylon filters.
- Remove the “TT” and “MD” flags – These flags are automatically generated from measurement results. The “TT” flag is applied when samples are received at Wood with a temperature greater than 4 °C. Removing the “TT” flag reduces the amount of information about the sample that may be pertinent to an end user. The “MD” flag is applied when the reported value is below the corresponding detection limit. The “MD” flag should only be applied to the analytical parameters for a given filter.
- Use the “Rx” Request Exclusion flags – These flags are available for use for NAAQS related data but they are not relevant to CSN data. All “Rx” flags have an “Ix” Informational Only equivalent flag (as described in Section 5.1); these “Ix” flags should be applied to CSN data where applicable.

6.1. Suggested Analyses

The DART interface provides many useful tools for visualizing the datasets. In addition to the online DART interface, DART allows users to export data for further analysis using a spreadsheet or other software. There are innumerable methods for data analysis, which are outside the scope of this document. Offered here are some examples of analyses that may be useful for SLT validators both using the DART interface and offline. Other useful analyses not covered in this document may incorporate wind rose plots, trajectory analysis, and AirNowTech data.

6.1.1. Regression Analysis

Regression analysis is a common tool used for evaluating the relationship between two or more interrelated variables. For data validation, linear regression can be used to determine the comparability of two types of results, such as reconstructed mass versus gravimetric mass or titanium versus silicon (typical soil elements). For measurements we expect to be very similar (as in $3\times$ sulfur versus sulfate; Figure 2), a regression slope near one would indicate subjectively good agreement. A slope that deviates from one may indicate a methodological bias by one instrument or analysis. Likewise, a non-zero intercept indicates a bias between the two variables. The analyst should take uncertainties, methodological differences, and local factors into account when interpreting results.

6.1.2. Multi-Parameter Trends

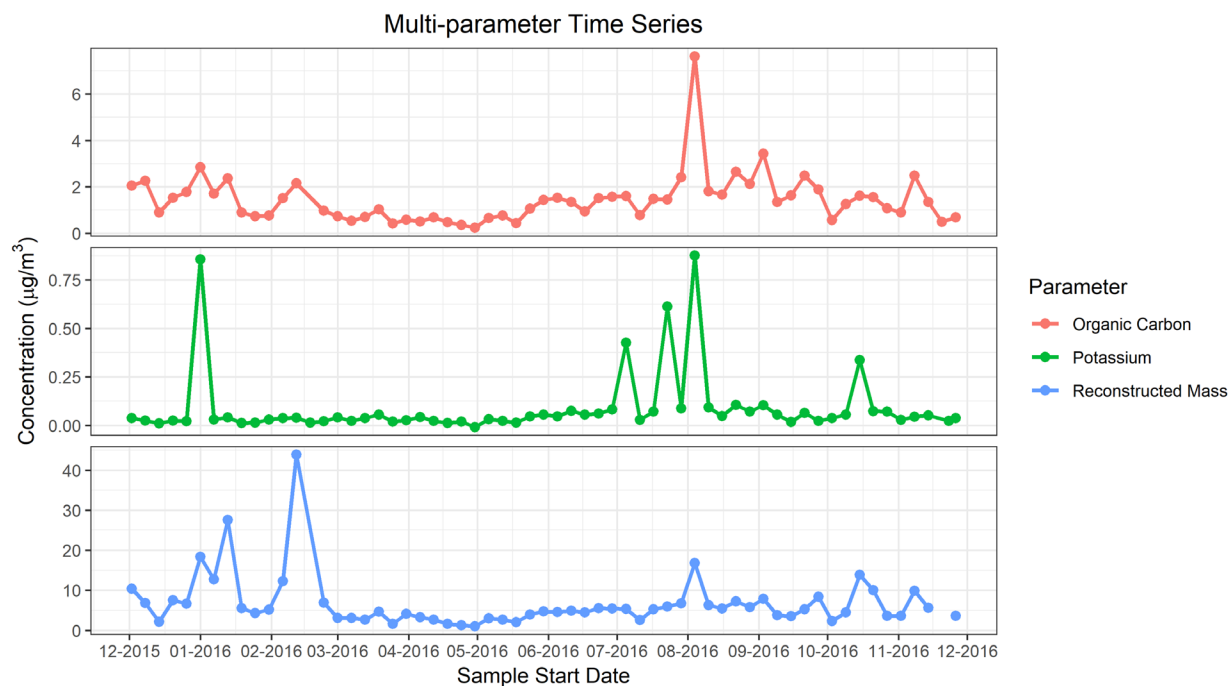
In DART, time-series plots can be viewed for one or more parameters simultaneously. Additionally, several calculated parameters are provided to aid in interpreting the data. These include reconstructed mass, summed elements, summed ions, ammonium sulfate, ammonium nitrate, soil, and organic carbon by mass (see Section 7).

Viewing multiple parameters across different filter types can be a useful tool for determining whether a data point is atmospherically real or a sampling anomaly. For species with common sources, comparing their concentrations can provide supporting evidence if they exhibit similar trends. For example, if sodium ion and chloride both have high concentrations it can be indicative of sea salt influence. However, if the concentrations of all ions species are high, this could indicate an issue with the sample itself, as not all of the measured species from the nylon filter have a common source. Comparing the sulfate from the nylon filter with the sulfur from the corresponding PTFE filter could be used to corroborate the results.

As another example, shown in Figure 5 are time series for organic carbon, potassium, and reconstructed mass. There are several instances of high potassium concentrations, with distinct peaks in January 2016 and August 2016. However, during this time period the only clear organic carbon peak is in August 2016. Using local knowledge, the January 2016 potassium peak can likely be attributed to celebratory fireworks, and is further supported by comparing trends with species such as iron and copper (common components of firework emissions). The high potassium and organic carbon concentrations in August 2016 are likely related to other sources, possibly fires.

To support this interpretation, MODIS satellite data (<https://modis.gsfc.nasa.gov/>) can be used to identify smoke from fires, and reconstructed mass can be used to evaluate overall particulate matter concentration relative to other days.

Figure 5. A year-long time series of organic carbon, potassium, and reconstructed mass concentrations from a single site.

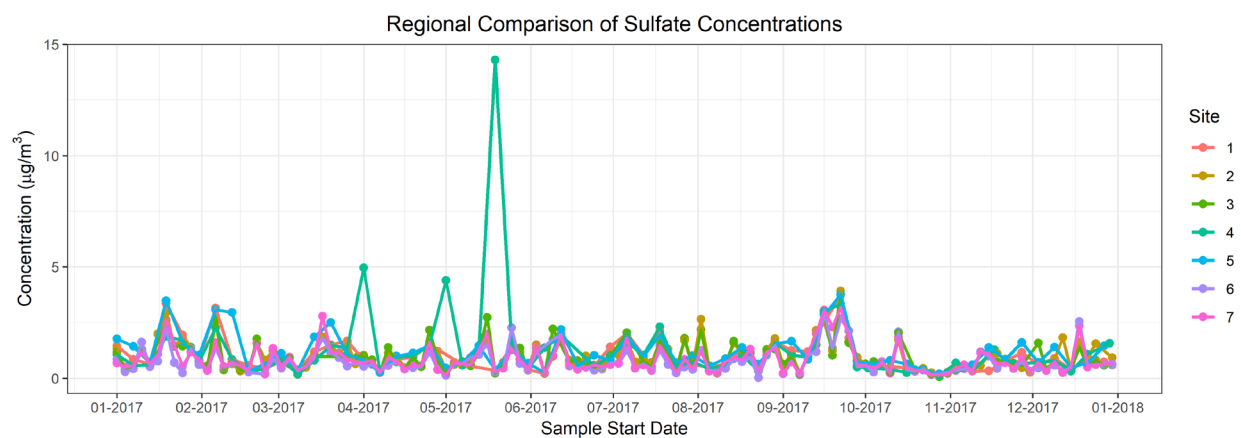


6.1.3. Geospatial analysis

Comparing species concentrations between sites can provide insight into local and regional events as well as highlight potential issues with the data. When performing a geospatial analysis it is important to consider the site location relative to other CSN sites and potential local influences. It is also important to understand the sources and atmospheric lifetimes of the species being compared; for example, sulfur is generally regional whereas organic carbon can have local and regional sources and evolves in the atmosphere with time.

Figure 6 shows time series of sulfate concentrations at seven sites within a 200km radius. Site #4 has three outlier samples; the highest on May 19, 2017 with sulfate concentration approximately 50 times higher relative to the other sites. This is a case where the SLT validator could further investigate the sample from Site #4 to determine if issues in the field, sampler malfunction, laboratory errors, or possible contamination impacted the sample. Alternatively, a localized event could cause the elevated sulfate at Site #4; obtaining meteorological data could help to support the interpretation as well as having knowledge of local events, activities, and/or industries.

Figure 6. Time series of sulfate concentrations from seven sites within a 200km radius.



7. Explanation of Terms

In DART, some parameters are provided that **are not delivered to AQS** (Table 2). These parameters do not need to be validated. UCD and STI calculate and provides these parameters to aid in visualization of the dataset and validation.

Table 2. Calculated parameters delivered to DART.

Parameter code	Parameter name	Calculation	Notes
00001	Elements	Na + Mg + Al + Si + P + S + Cl + K + Ca + Ti + V + Cr + Mn + Fe + Co + Ni + Cu + Zn + As + Se + Br + Rb + Sr + Zr + Ag + Cd + In + Sn + Sb + Cs + Ba + Ce + Pb	Sum of elements measured by XRF
00002	Ions	chloride + ammonium + sodium ion + potassium ion + nitrate + sulfate	Sum of ions measured by IC
00010	PM _{2.5} Mass Difference	PM _{2.5} Raw Data - Reconstructed Mass	Difference between PM _{2.5} Raw Data (Parameter 88501) and Reconstructed Mass (Parameter 88401)
88339	Ammonium sulfate	4.125× sulfur	Estimated ammonium sulfate using XRF sulfur rather than measured sulfate
88344	Ammonium nitrate	1.29× nitrate	-
88348	Soil	2.2× aluminum + 2.49× silicon + 1.63× calcium + 2.42× iron + 1.94× titanium	Estimated from stoichiometric relationships of crustal soil oxides
88350	Organic Carbon Mass (OMC)	1.4× organic carbon	Organic carbon by TOA
88401	Reconstructed Mass	ammonium sulfate + ammonium nitrate + soil + 1.8× chloride + elemental carbon + OMC	Elemental carbon by TOA

PM_{2.5} Raw Data (parameter 88501) does not go to AQS and should not be validated in DART. This mass value is retrieved from the AirNowTech web service and provided in DART for comparison with reconstructed mass (parameter 88401), a composite variable estimated from speciated chemical measurements. The data in the AirNowTech service is dynamic, so the value available in DART may not correspond with the most recent data available from AirNowTech. The SLT validator should download recent data from AirNowTech (<https://www.airnowtech.org/>) when reviewing the data.

7.1. Field Blanks

Field blanks are collected for quality assurance purposes and to calculate blank correction and method detection limits (MDLs); one field blank per filter type per month is scheduled at each CSN site. The use of trip blanks was discontinued as of February 2018.

With this frequency of field blank collection, statistically meaningful data is obtained for blank correction and MDL calculations. All field blanks from a given month are used in the calculations. A percentile method that is less sensitive to the impact of outliers, compared to a standard deviation method, is currently used to estimate MDLs. Outliers can occur if air is passed through the field blank in error during the collection process; they may have been accidentally sampled or there may have been low flow through the sampler channel. Enough evidence is not always available to confirm these cases; however field blanks should be carefully evaluated and invalidated if there is evidence of flow. The SLT validator should compare field blank data with associated sample data as well as field blanks from other sites and historical values to determine if the filter truly ran as a field blank. Since field blank data is used to calculate network wide blank corrections and MDLs, erroneous results can have network wide impact.

7.2. Value Limits

Parameter range value limits have been assigned, and samples falling outside these ranges receive either null or qualifier flags (Table 3). For the narrower ranges, the acceptable values were determined from the operational manuals, such as for temperature, historical CSN data, such as for pressure and flow rate CV, or being $\pm 10\%$ from the nominal value. For the wider ranges, the acceptable values were defined in AQS; outside of these ranges, the parameter record will not be successfully ingested into the AQS database unless the record is marked with a null flag. It is important to note that the application of the flags, whether qualifier or null, in terms of which parameters the flags are applied to for a given filter, is not the same for all variables in the automated flagging process at UCD. For example, the QT qualifier flag is applied only to the temperature parameter whereas the AG null flag is applied to all parameters relating to a given filter.

Table 3. Limits assigned for operational parameters, listed by instrument, and corresponding flag applied for cases outside of limits. For parameters where two ranges are specified, the broader range is the acceptable range in AQS for a given parameter and outside of that range a null flag is applied (where applicable) to enable delivery of data to AQS.

Variable	Instrument		Flag Applied
	URG 3000N	Met One SASS	
Temperature, °C	-20 to 45	-30 to 50	QT qualifier flag
	-40 to 55	-40 to 55	AN null flag
Pressure, mmHg	600 to 810	600 to 810	QP qualifier flag
	450 to 1000	450 to 850	AN null flag
Nominal Flow Rate, L min ⁻¹	22	6.7	-
Lower Flow Rate Limit, L min ⁻¹	19.8	6.03	AH null flag
Upper Flow Rate Limit, L min ⁻¹	24.2	7.37	
Flow rate CV, %	0 to 2	0 to 5	AH null flag
	0 to 20	0 to 20	AN null flag*
Nominal Sample Volume, m ³	31.7	9.6	-
Lower Sample Volume Limit, m ³	28.5	8.6	SV null flag
	0	0	AN null flag*
Upper Sample Volume Limit, m ³	34.9	10.6	SV null flag
	35	25	AN null flag*
Nominal Sample Time, minutes	1440	1440	-
Lower Sample Time, minutes	1380	1380	AG null flag
Upper Sample Time, minutes	1500	1500	

* Null flag applied if not already invalid.

Parameter ratio limits have been assigned for specific inter-parameter comparisons that are expected to have well defined relationships. The outlier flag, “5”, is only assigned for the comparison of reconstructed mass to measured gravimetric mass (where available) and when the $3 \times S / SO_4$ ratio is out of bounds, where either $3 \times S$ or SO_4 is above its respective MDL.

Table 4. Limits for inter-parameter comparisons.

Ratio	Upper Limit	Lower Limit	“5” Flag Assigned
$3 \times S / SO_4$	1.731	0.784	Yes
Non-soil potassium* / potassium ion	6.76	0.656	No
gravimetric mass / reconstructed mass	2	0.5	Yes
sum of anions / sum of cations	2.82	0.86	No

* Non-soil potassium = K - (2/27) * Si

8. Further Information

Additional resources are available for SLT validators:

- The AirNowTech website contains training and help documents for using the DART interface: <https://airnowtech.org>.
- The most recent DART training module can be found at: <https://youtu.be/bNSjMgVSdj0>

- The Ambient Monitoring Technology Information Center (<https://www.epa.gov/amtic>) contains many useful documents concerning CSN methods, parameters, and network changes.
- A searchable list of AQS codes is available at: <https://www.epa.gov/aqs/aqs-code-list>.
- Various documentation, including the CSN Annual Reports, UCD Standard Operating Procedures (SOPs), UC Quality Assurance Project Plan, and Data Advisories can be found at: <https://aqrc.ucdavis.edu/documentation>.

For assistance:

- Send questions regarding the DART interface to DART@sonomatech.com.
- Contact CSNsupport@sonomatech.com to reach members of EPA, UCD, and STI.

9. Appendix I – AQS Codes

Qualifier Flag	Qualifier Description	Qualifier Type
1	Requirement	Quality Assurance Qualifier
2	Operational Deviation	Quality Assurance Qualifier
3	Field Issue	Quality Assurance Qualifier
4	Lab Issue	Quality Assurance Qualifier
5	Outlier	Quality Assurance Qualifier
6	QAPP Issue	Quality Assurance Qualifier
7	Below Lowest Calibration Level	Quality Assurance Qualifier
9	Negative value detected - zero reported	Quality Assurance Qualifier
CB	Values have been Blank Corrected	Quality Assurance Qualifier
CC	Clean Canister Residue	Quality Assurance Qualifier
CL	Surrogate Recoveries Outside Control Limits	Quality Assurance Qualifier
DI	Sample was diluted for analysis	Quality Assurance Qualifier
EH	Estimated; Exceeds Upper Range	Quality Assurance Qualifier
FB	Field Blank Value Above Acceptable Limit	Quality Assurance Qualifier
FX	Filter Integrity Issue	Quality Assurance Qualifier
HT	Sample pick-up hold time exceeded	Quality Assurance Qualifier
LB	Lab blank value above acceptable limit	Quality Assurance Qualifier
LJ	Identification Of Analyte Is Acceptable; Reported Value Is An Estimate	Quality Assurance Qualifier
LK	Analyte Identified; Reported Value May Be Biased High	Quality Assurance Qualifier
LL	Analyte Identified; Reported Value May Be Biased Low	Quality Assurance Qualifier
MD	Value less than MDL	Quality Assurance Qualifier
MS	Value reported is 1/2 MDL substituted.	Quality Assurance Qualifier
MX	Matrix Effect	Quality Assurance Qualifier
ND	No Value Detected, Zero Reported	Quality Assurance Qualifier
NS	Influenced by nearby source	Quality Assurance Qualifier
QP	Pressure Sensor Questionable	Quality Assurance Qualifier
QT	Temperature Sensor Questionable	Quality Assurance Qualifier
QX	Does not meet QC criteria	Quality Assurance Qualifier
SQ	Values Between SQL and MDL	Quality Assurance Qualifier
SS	Value substituted from secondary monitor	Quality Assurance Qualifier
SX	Does Not Meet Siting Criteria	Quality Assurance Qualifier
TB	Trip Blank Value Above Acceptable Limit	Quality Assurance Qualifier
TT	Transport Temperature is Out of Specs.	Quality Assurance Qualifier
V	Validated Value	Quality Assurance Qualifier
VB	Value below normal; no reason to invalidate	Quality Assurance Qualifier
W	Flow Rate Average out of Spec.	Quality Assurance Qualifier
X	Filter Temperature Difference out of Spec.	Quality Assurance Qualifier
Y	Elapsed Sample Time out of Spec.	Quality Assurance Qualifier

AA	Sample Pressure out of Limits	Null Data Qualifier
AB	Technician Unavailable	Null Data Qualifier
AC	Construction/Repairs in Area	Null Data Qualifier
AD	Shelter Storm Damage	Null Data Qualifier
AE	Shelter Temperature Outside Limits	Null Data Qualifier
AF	Scheduled but not Collected	Null Data Qualifier
AG	Sample Time out of Limits	Null Data Qualifier
AH	Sample Flow Rate or Flow Rate CV Out of Limits	Null Data Qualifier
AI	Insufficient Data (cannot calculate)	Null Data Qualifier
AJ	Filter Damage	Null Data Qualifier
AK	Filter Leak	Null Data Qualifier
AL	Voided by Operator	Null Data Qualifier
AM	Miscellaneous Void	Null Data Qualifier
AN	Machine Malfunction	Null Data Qualifier
AO	Bad Weather	Null Data Qualifier
AP	Vandalism	Null Data Qualifier
AQ	Collection Error	Null Data Qualifier
AR	Lab Error	Null Data Qualifier
AS	Poor Quality Assurance Results	Null Data Qualifier
AT	Calibration	Null Data Qualifier
AU	Monitoring Waived	Null Data Qualifier
AV	Power Failure	Null Data Qualifier
AW	Wildlife Damage	Null Data Qualifier
AX	Precision Check	Null Data Qualifier
AY	Q C Control Points (zero/span)	Null Data Qualifier
AZ	Q C Audit	Null Data Qualifier
BA	Maintenance/Routine Repairs	Null Data Qualifier
BB	Unable to Reach Site	Null Data Qualifier
BC	Multi-point Calibration	Null Data Qualifier
BD	Auto Calibration	Null Data Qualifier
BE	Building/Site Repair	Null Data Qualifier
BF	Precision/Zero/Span	Null Data Qualifier
BG	Missing ozone data not likely to exceed level of standard	Null Data Qualifier
BH	Interference/co-elution/misidentification	Null Data Qualifier
BI	Lost or damaged in transit	Null Data Qualifier
BJ	Operator Error	Null Data Qualifier
BK	Site computer/data logger down	Null Data Qualifier
BL	QA Audit	Null Data Qualifier
BM	Accuracy check	Null Data Qualifier
BN	Sample Value Exceeds Media Limit	Null Data Qualifier
BR	Sample Value Below Acceptable Range	Null Data Qualifier
CS	Laboratory Calibration Standard	Null Data Qualifier

DA	Aberrant Data (Corrupt Files, Aberrant Chromatography, Spikes, Shifts)	Null Data Qualifier
DL	Detection Limit Analyses	Null Data Qualifier
FI	Filter Inspection Flag	Null Data Qualifier
MB	Method Blank (Analytical)	Null Data Qualifier
MC	Module End Cap Missing	Null Data Qualifier
QV	Quality Control Multi-point Verification	Null Data Qualifier
SA	Storm Approaching	Null Data Qualifier
SC	Sampler Contamination	Null Data Qualifier
ST	Calibration Verification Standard	Null Data Qualifier
SV	Sample Volume Out of Limits	Null Data Qualifier
TC	Component Check & Retention Time Standard Holding Time Or Transport Temperature Is Out Of Specs.	Null Data Qualifier
TS		Null Data Qualifier
XX	Experimental Data	Null Data Qualifier
IA	African Dust	Informational Only
IB	Asian Dust	Informational Only
IC	Chem. Spills & Indust Accidents	Informational Only
ID	Cleanup After a Major Disaster	Informational Only
IE	Demolition	Informational Only
IF	Fire - Canadian	Informational Only
IG	Fire - Mexico/Central America	Informational Only
IH	Fireworks	Informational Only
II	High Pollen Count	Informational Only
IJ	High Winds	Informational Only
IK	Infrequent Large Gatherings	Informational Only
IL	Other	Informational Only
IM	Prescribed Fire	Informational Only
IN	Seismic Activity	Informational Only
IO	Stratospheric Ozone Intrusion	Informational Only
IP	Structural Fire	Informational Only
IQ	Terrorist Act	Informational Only
IR	Unique Traffic Disruption	Informational Only
IS	Volcanic Eruptions	Informational Only
IT	Wildfire-U. S.	Informational Only
J	Construction	Informational Only

10. Appendix II – Parameter List

Parameter Name	Parameter Code	Reported To DART	Reported To AQS
Elements	00001	YES	NO
Ions	00002	YES	NO
Transport Temperature	00007	YES	NO
Transport Temperature	00008	YES	NO
Transport Temperature	00009	YES	NO
PM _{2.5} Mass Difference	00010	YES	NO
Avg. Ambient Temperature	68105	YES	YES
Avg. Ambient Pressure	68108	YES	YES
Sample Flow Rate CV	68111	YES	YES
Sample Flow Rate CV	68112	YES	YES
Sample Flow Rate CV	68113	YES	YES
Sample Volume	68114	YES	YES
Sample Volume	68115	YES	YES
Sample Volume	68116	YES	YES
Avg. Ambient Temperature	68117	YES	YES
Avg. Ambient Pressure	68118	YES	YES
ANTIMONY PM2.5 LC	88102	YES	YES
Arsenic PM2.5 LC	88103	YES	YES
Aluminum PM2.5 LC	88104	YES	YES
BARIUM PM2.5 LC	88107	YES	YES
Bromine PM2.5 LC	88109	YES	YES
CADMIUM PM2.5 LC	88110	YES	YES
Calcium PM2.5 LC	88111	YES	YES
Chromium PM2.5 LC	88112	YES	YES
Cobalt PM2.5 LC	88113	YES	YES
Copper PM2.5 LC	88114	YES	YES
Chlorine PM2.5 LC	88115	YES	YES
CERIUM PM2.5 LC	88117	YES	YES
CESIUM PM2.5 LC	88118	YES	YES
Iron PM2.5 LC	88126	YES	YES
Lead PM2.5 LC	88128	YES	YES
INDIUM PM2.5 LC	88131	YES	YES
Manganese PM2.5 LC	88132	YES	YES
Nickel PM2.5 LC	88136	YES	YES
Magnesium PM2.5 LC	88140	YES	YES
Phosphorus PM2.5 LC	88152	YES	YES
Selenium PM2.5 LC	88154	YES	YES
TIN PM2.5 LC	88160	YES	YES
Titanium PM2.5 LC	88161	YES	YES

Vanadium PM2.5 LC	88164	YES	YES
Silicon PM2.5 LC	88165	YES	YES
SILVER PM2.5 LC	88166	YES	YES
Zinc PM2.5 LC	88167	YES	YES
Strontium PM2.5 LC	88168	YES	YES
Sulfur PM2.5 LC	88169	YES	YES
Rubidium PM2.5 LC	88176	YES	YES
Potassium PM2.5 LC	88180	YES	YES
Sodium PM2.5 LC	88184	YES	YES
Zirconium PM2.5 LC	88185	YES	YES
Chloride PM2.5 LC	88203	YES	YES
AMMONIUM ION PM2.5 LC	88301	YES	YES
SODIUM ION PM2.5 LC	88302	YES	YES
POTASSIUM ION PM2.5 LC	88303	YES	YES
Total Nitrate PM2.5 LC	88306	YES	YES
Total Carbon PM2.5 LC TOT	88312	NO	NO
OC PM2.5 LC TOR	88320	YES	YES
EC PM2.5 LC TOR	88321	YES	YES
OC1 PM2.5 LC	88324	NO	YES
OC2 PM2.5 LC	88325	NO	YES
OC3 PM2.5 LC	88326	NO	YES
OC4 PM2.5 LC	88327	NO	YES
OP PM2.5 LC TOR	88328	NO	YES
EC1 PM2.5 LC	88329	NO	YES
EC2 PM2.5 LC	88330	NO	YES
EC3 PM2.5 LC	88331	NO	YES
Ammonium Sulfate PM2.5 LC	88339	YES	NO
Ammonium Nitrate PM2.5 LC	88344	YES	NO
Soil PM2.5 LC	88348	YES	NO
Organic Carbon Mass PM2.5 LC	88350	YES	NO
OC CSN Unadj. PM2.5 LC TOT	88355	NO	YES
EC CSN Unadj. PM2.5 LC TOT	88357	NO	YES
OC CSN Unadj. PM2.5 LC TOR	88370	NO	YES
OC1 CSN Unadj. PM2.5 LC	88374	NO	YES
OC2 CSN Unadj. PM2.5 LC	88375	NO	YES
OC3 CSN Unadj. PM2.5 LC	88376	NO	YES
OC4 CSN Unadj. PM2.5 LC	88377	NO	YES
OP CSN Unadj. PM2.5 LC TOR	88378	NO	YES
OP PM2.5 LC TOT	88379	NO	YES
EC CSN Unadj. PM2.5 LC TOR	88380	NO	YES
EC PM2.5 LC TOT	88381	NO	YES
OC PM2.5 LC TOT	88382	NO	YES
EC1 CSN Unadj. PM2.5 LC	88383	NO	YES

EC2 CSN Unadj. PM2.5 LC	88384	NO	YES
EC3 CSN Unadj. PM2.5 LC	88385	NO	YES
OP CSN Unadj. PM2.5 LC TOT	88388	NO	YES
Reconstructed Mass PM2.5 LC	88401	YES	NO
Sulfate PM2.5 LC	88403	YES	YES
PM2.5 Raw Data	88501	YES	NO
PM2.5 mass	88502	YES	YES