Data Validation for the Chemical Speciation Network
A Guide for State, Local, and Tribal Agency Validators

Version 1.0

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

UCDAVIS
AIR QUALITY RESEARCH CENTER

Table of Contents

# Table of Figures

Table of Tables

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 as input to models and 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 and epidemiological 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 CSN as well as provide some guidance on potential validation strategies. The initial stages of the validation process are performed by Wood, PLC 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. The Validation Pathway

Figure 1. Graphical representation of CSN data flow.
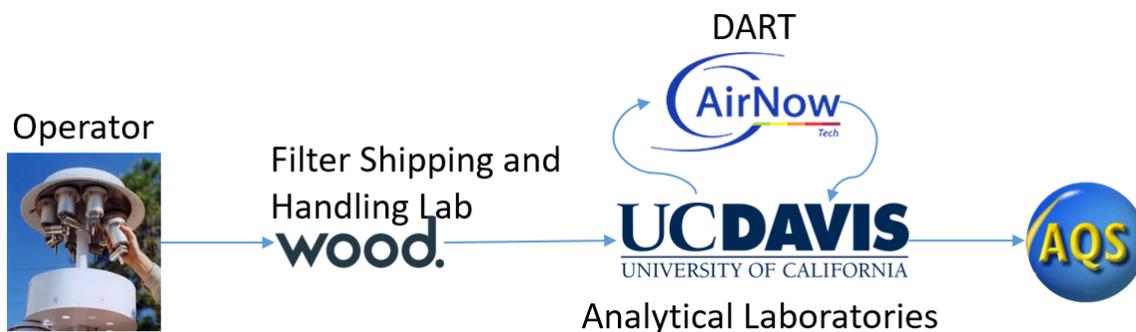
Data generation begins at sample preparation and collection. The filter shipping and handling laboratory at Wood, PLC generates data associated by filter sampling events. The site operator provides feedback from the sampling instruments as well as notes on sampling conditions. This information is passed back to the filter shipping and handling laboratory via Field Sampling Chain of Custody (FSCOC) forms where it is entered into a database and compiled for export to UCD for data processing. The data at this point reflects the physical conditions that the sample represents (i.e. sampling location, sampling timeframe, sample volume, and ambient conditions at time of sampling and during transport). The analytical laboratories receive the filters from Wood, PLC and generate the concentration data along with additional notes on filter and analysis conditions. UCD ingests all data into a central 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. It is important to note that data in DART has some differences relative to data delivered to the EPA Air Quality System (AQS) database. Calculated parameters (e.g. reconstructed mass) are included in DART while others are omitted (e.g. carbon fractions) in order to increase the efficiency of DART validation. The specific differences are described in Section 4.4, *Submission to DART*.

After data have been validated by the SLT agencies in DART, it is returned to UCD for reformatting for AQS delivery. During this process, the data are reviewed a final time to ensure consistency between the data's final form and all comments left by the SLT data validators. If any 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 made during all previous validation steps are communicated to the end users through AQS qualifier and null codes, see Section 5, *Level 2 Validation by SLT Agencies* for a list of codes.

3. Validation Philosophy

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

- 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 flags options to notify the end user of any concerns about the data. Similarly, negative values should be 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 time-series statistics.

Finally, we recognize that data validation is an evolving process. Over time, we will learn more about the data and will be able to devise new methods for checking data integrity. The groundwork laid in the DART interface and in the UCD data validation software packages has been carefully constructed so that experimental analysis methods can be investigated.

4. Validation Process

Validation of CSN data is divided into four steps. Specifically,

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

This document describes the validation procedures for levels 1B and 2. The validation protocols for Levels 0 and 1A can be found in the associated standard operating procedures (SOP) documents, available at the EPA Ambient Monitoring Technology Information Center (AMTIC, https://www3.epa.gov/ttnamti1/specsop.html). The Level 1B procedures are also outlined in a published SOP (UCD CSN Technical Information #801C CSN Data Validation) but are re-iterated here to provide context for the Level 2 discussion.

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 focus on identifying potential filter swaps or inconsistencies in reporting. The description of these checks below proceed in a logical progression starting with receipt of data from the analytical laboratories.

4.1. Level 1B Data Import Validation

When electronic data from Wood, PLC 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 ingestion process is overseen and checked by a data analyst at UCD to identify potential issues with filter-specific information. Since the data from the FSCOC forms are entered manually, some typographical mistakes are inevitable. The automated check at this point compares the intended use date, the sample start date, and the

sample end date for each sample event record. Ideally, the intended use date and the sample start date would be the same and the sample end date would be 24 hours later than the former dates. Typographical errors and insufficient elapsed sample times are also discovered during manual review of operational parameters (i.e. ambient temperature and pressure, flow rate coefficient of variation or CV, sample volume). UCD notifies Wood, PLC of inconsistencies identified during Level 1B validation, and these are typically resolved prior to submission to DART. In cases where the inconsistencies are not resolvable by UCD and Wood, PLC, comments in the form of questions 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 codes. All automated flags assigned by UCD validators are accompanied by comments explaining the reason the flag was applied. The only null codes assigned automatically are "AH" for flow rate issues and "AM" for miscellaneous voids when no data exists for an expected result. The "AH" code is currently being applied to samples that were 1) collected with an air flow outside of ±10% of the nominal flow rate and 2) have a flow rate variability (flow rate CV) greater than 5% for Met One SASS and 2% for URG 3000N instruments.

Automated qualifier code assignments include "TT", "MD", "QP", "QT", and "5". The "TT" code identifies samples that have been received at Wood, PLC above 4 °C while the "MD" flag indicates that a measured parameter was below the method detection limit. The "QP" and "QT" qualifier codes indicates 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 codes applied to data during level 1B data validation.

AQS Code	AQS Description	Usage
TT	Transport Temperature is Out of Specs.	Samples received at Wood, PLC > 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	Sulfur/sulfate and potassium/potassium ion ratios are outside expected limits

Finally, the "5" qualifier code indicates a sample is an "outlier" based on an inter-comparison of filter-specific measurements. The sulfur and potassium concentrations from the Teflon™ filter

are compared to the sulfate and potassium ion concentrations from the Nylon filter. The code is only applied when both comparisons are outside of pre-defined limits. The limits were used by the previous analytical laboratory contractor and are being re-assessed considering the full CSN historical data record. This check compares the Teflon™ and Nylon filters but there is not currently an inter-comparison for the quartz filter. In the future, optical measurements may be used to compare results from the Teflon™ filter with the EC concentrations from the quartz filter. The final case for application of the "5" code occurs when the gravimetric mass differs greatly from the reconstructed mass (defined in Section 7, *Explanation of Terms*). As of April 2016, gravimetric mass is only measured at one site in the CSN. However, gravimetric mass from the PM$_{2.5}$ NAAQS monitoring network and the gravimetric mass reported to AirNowTech can also be used for comparison with reconstructed mass determined from the CSN data.

Figure 2. Comparison of network-wide sulfur and sulfate measurements. In theory, 3 times the sulfur concentrations should always be greater than or equal to sulfate; however, in practice, uncertainty in both measurements allows sulfate concentrations to exceed sulfur. Note the outliers with high sulfate and relatively lower sulfur concentrations. These sampling events require greater inspection.
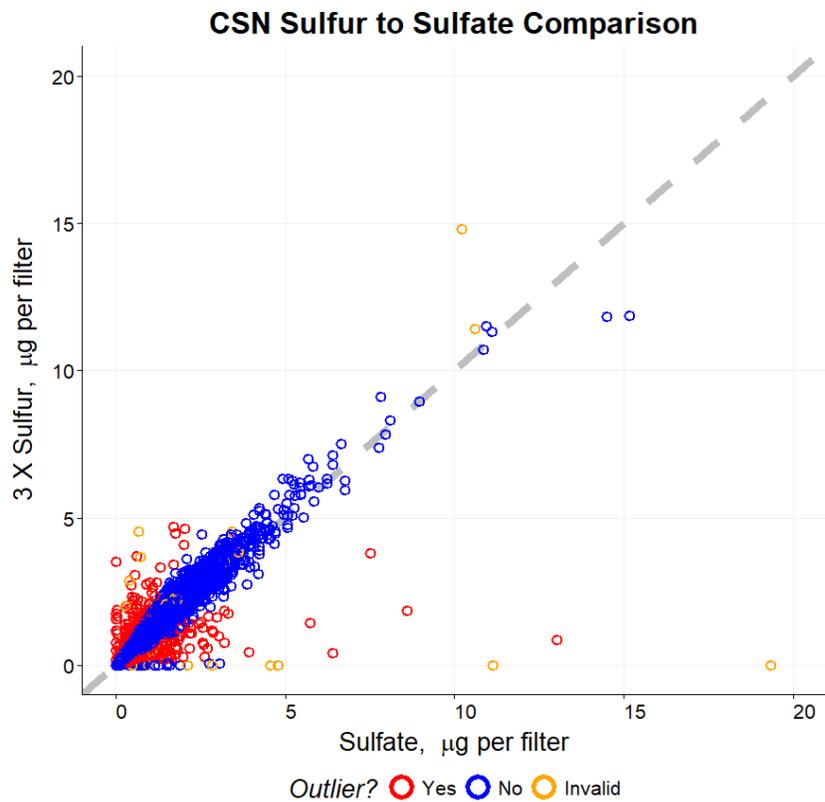
4.3. Level 1B Manual Review

The data analysts at UCD perform a series of checks to ensure data completeness and integrity prior to submission to DART. These checks include:

- 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.
- Outliers 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 samples check – review samples flagged as invalid to ensure correct application of null flags and investigate any potentially valid samples.
- Flow checks – although flow rates are checked automatically, these are checked again manually to ensure proper calculation of ambient concentrations.
- Inter-method comparison checks – these checks use collocated comparisons to detect sampling or analysis problems. An example of a control chart used by the UCD analyst is shown in Figure 3.
- Contamination checks – contamination is detected by reviewing network-wide results of individual species and confirmed through re-analysis. An example of zinc contamination is shown in Figure 4.

Figure 3. Summary figure used by UCD analysts to identify potential sample swaps or sampling issues. The colored lines show the mass concentration results while the points indicate the results of the inter-method ratios (3 x sulfur / sulfate and potassium / potassium ion). The potassium concentrations (lines) are multiplied by 10 in order for the concentrations to be plotted alongside the ratio values (points). It is important to note that the bounds (shown in the green shaded area) of the potassium/potassium ion comparison are temporary. A full, historical analysis of these relationships is underway. Also, the "S3" parameter refers to sulfur by XRF multiplied by 3, which is a shorthand calculation for sulfate from elemental sulfur based on molecular weight.
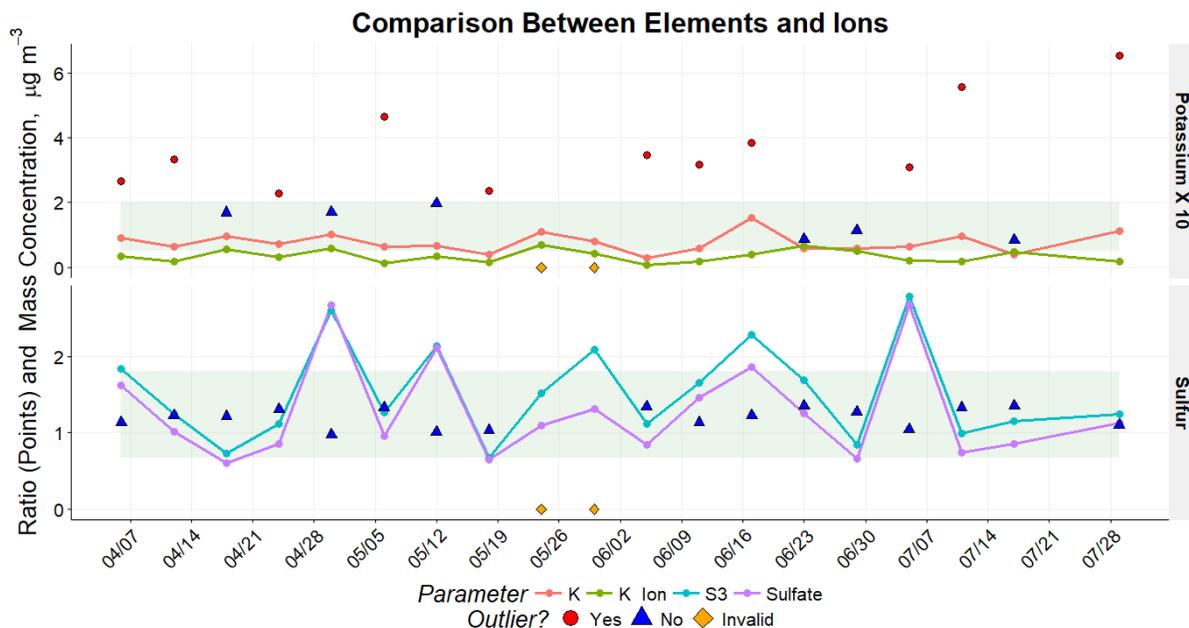
Figure 4. Monthly summary of zinc concentrations, with two high concentrations circled in red. The x-axis is each site in the network (not shown) and the y-axis is the measurement result for each parameter, in this case mass concentration of zinc in micrograms per cubic meter. The circled outlier samples were re-analyzed and identified as affected by zinc contamination.
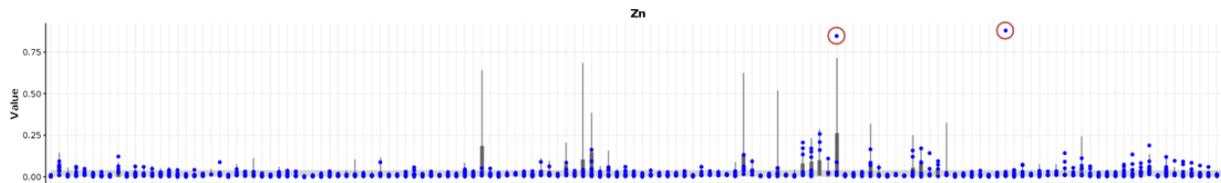
4.4. Submission to DART

After validation is complete through Level 1B, and identifiable issues resolved, the data are prepared for delivery to DART. As noted previously, the data in DART is a modified dataset

from the final file uploaded to the AQS database. Additional parameters (e.g. reconstructed mass) are provided in DART to assist in data validation, even though they are not delivered to AQS. Also, only the blank subtracted elemental and organic carbon concentrations measured using thermal/optical reflectance (TOR) are reported to DART. The rest of the carbon parameters, including carbon fractions and non-subtracted concentrations, are excluded from DART to reduce redundancy in data validation. These differences are summarized below. A table of all parameters and where they are reported is included in **Appendix II – Parameter List**.

Reported to DART but not AQS, defined in Section 7, *Explanation of Terms*:

- Summed Elements – Parameter 00001 (calculated in DART)
- Summed Ions – Parameter 00002 (calculated in DART)
- Transport Temperature – Parameters 00007, 00008, and 00009 (reported by Wood, PLC)
- 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) Pyrolized Carbon – Parameter 88328
- Raw TOR Pyrolized Carbon – Parameter 88378
- Blank-Subtracted Thermal Optical Transmission (TOT) Pyrolized Carbon – Parameter 88379
- Raw TOT Pyrolized 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

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

DART review provides a critical check on the processes of the analytical laboratories and the UCD data management team. The usage of DART is covered in depth during DART training sessions and available online:

https://www.epa.gov/sites/production/files/2016-09/documents/dart_training_for_pams_and_csn.pdf

The data validator may use any number of techniques to validate their data, both inside and outside of the DART interface. The remainder of this document will provide some examples and options for SLT data validators.

5.1. Notifying End Users of Local Impacts

Local impacts can cause data abnormalities from specific events or sampling conditions. An example would be a local festival or fireworks. **It is important to note that local impacts do not invalidate 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. Various informational flags exist that communicate local impacts via qualifier codes, 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" code will be sufficient to notify the end user that there was a local impact on the data, additional codes may be added when needed. Recommendations for new codes can be submitted to the EPA AQS team online: https://www.epa.gov/aqs/forms/aqs-contact-us

Keep in mind that **AQS accepts up to ten qualifier codes**.

Additionally, some AQS qualifier codes have very similar descriptions to null codes. These include:

- Sample time – the "AG" null code (Sample Time out of Limits) should be used if the sample elapsed time is less than 23 hours or more than 25 hours. The "Y" qualifier code (Elapsed Sample Time out of Spec.) should be used if the sample elapsed time is within one hour of a full 24 hour sampling time.
- Construction – the "AC" null code (Construction/Repairs in Area) should be used if samples cannot be collected due to construction. The "J" qualifier code (Construction) should be used to notify the end user that construction activities were present during the collection of the sample.
- Flow Rate – the "AH" null code (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%). The "W" qualifier code (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 code (Filter Damage) is used when a filter cannot be analyzed due to damage. The "FX" qualifier code (Filter Integrity Issue) indicates that some issue (e.g. small hole, minor contamination, uneven sample loading) was identified but analysis was able to be performed.

Example

On January 1st, 2016, abnormally high potassium was observed at several sites. This was accompanied by peaks in several other elements (magnesium, aluminum, silicon, sulfur, iron, copper, and zinc). Upon further investigation, it was discovered that the spike in elemental concentrations was due to local fireworks for New Year's celebrations. These data were then flagged with the appropriate "IH - fireworks" qualifier code.

Figure 5. Elevated concentrations for several elements observed on January 1st, 2016.

Fireworks Observed on January 1, 2016
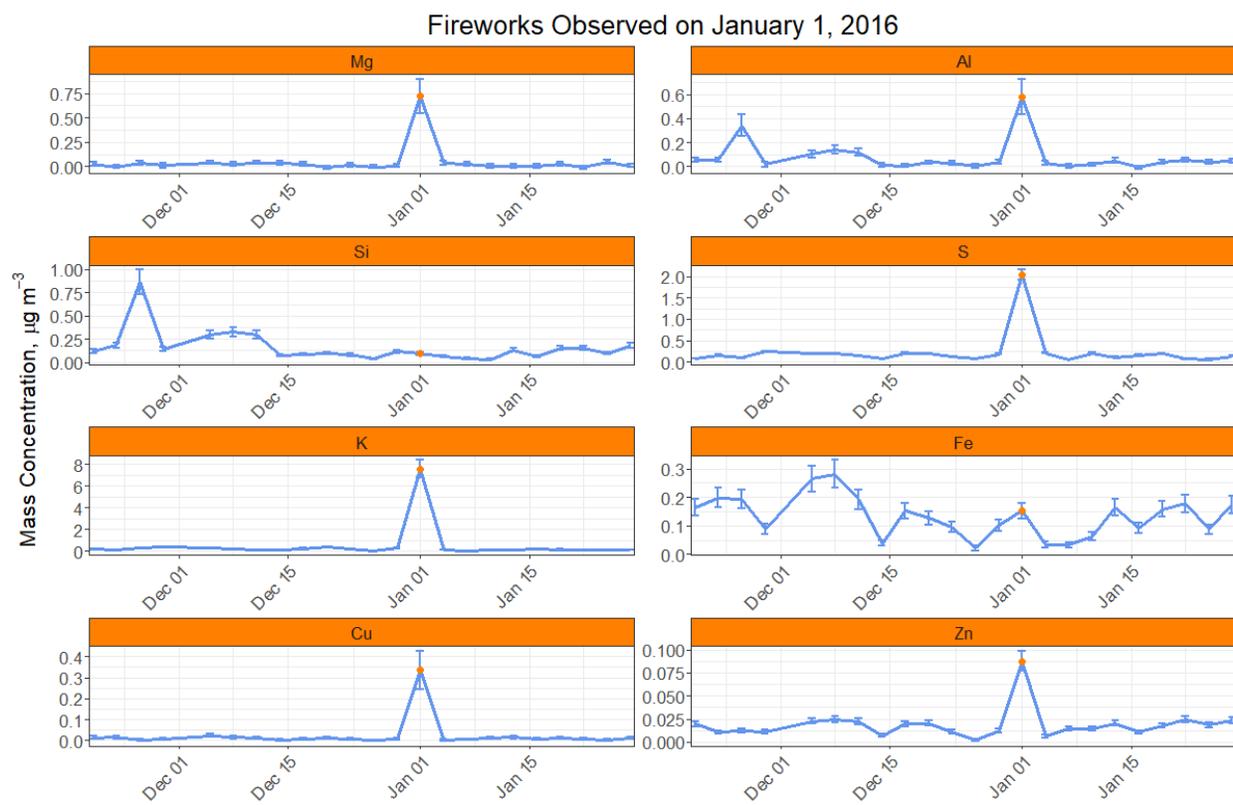
5.2. Making Use of Local Knowledge

The SLT data validators have the benefit of communication with site operators and field logs to investigate site specific issues or events. Additionally, local knowledge of a sampling site's history and situation can provide important context for measurement results. For example, a site may receive elevated levels of specific industrial emissions under certain meteorological conditions. To Level 0 and Level 1 reviewers, this may appear to be an outlier and be flagged with a "5" code. The data validator may understand the situation more clearly and adjust the qualifier flag to "NS – Influence 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 along with the data from the sample shipping and handling lab through delivery to DART. It is beneficial for the data 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

As with any new tool, there is a learning curve associated with effective use. We have compiled a list of common practices that aid in validation. These are merely recommendations. **The data produced by the CSN are owned by the SLT air quality agencies**. The SLT data validators have ultimate authority regarding any flags they wish to be applied to their data. UCD can only offer advice, such as the following:

PLEASE DO

- Check all invalid filters – the UCD analysts have 30 days to validate the data for the entire network prior to submission into DART. This does not afford enough time to investigate the invalid filters as well as all the valid ones. It is possible for a sample to be erroneously invalidated at any stage in the process. Please review the null codes and cross-check with your field logs. The measured values for invalid samples have been included in DART so that they can be reviewed alongside valid samples.
- Change the "AM" null code to a more appropriate code – the "AM" (miscellaneous void) null code is automatically applied for data records that were not received from either the analytical labs or the filter shipping and handling lab. UCD does not have enough information to apply a more informative code but your field logs and site operators might. Please check those resources while reviewing the invalid filters.
- Add qualifiers – in AQS, up to ten qualifier codes are allowed, with many codes currently to choose from (see Section 9, **Appendix I – AQS Codes**). There is no need to overwrite an existing qualifier code unless you believe the code to be erroneous or there are already ten codes attached to a data record, which is unlikely. A searchable list is available through the EPA website:
  https://aqs.epa.gov/aqsweb/documents/codetables/qualifiers.html.
- Invalidate samples with a serious sampling problem – serious sampling problems include poor flow rate (> +/- 10% from specification) and inadequate sampling time (> +/- 1 hour). These issues should have been identified during one or more of the previous validation steps but should be re-checked in DART. Alternatively, do not invalidate a sample if it experienced a power outage less than one hour in duration. For these cases, you can use the "Y" qualifier code (Elapsed Sample Time out of Spec.).
- 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 will need to be recalculated and re-examined. Flow rates and operational parameter values can be corrected within DART by the SLT data validator.
- Respond to questions in the comments section – sometimes discrepancies in the data cannot be answered prior to DART delivery. In these cases, the UCD analysts ask questions to the DART reviewers. In order for these issues to be resolved, there should be some communication between the DART reviewers and UCD. The preferred method of

communication is for DART reviewers to 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. Please clearly indicate whether your comment requires action by the UCD analysts, such as updating a flow rate or pressure reading. Please specify the parameter(s), date(s), and a clear description of the changes that need to be made (i.e. "Please change the flow rate for the quartz filter at site 12-345-6789 on 2017-01-01 from 25.0 to 22.0 and remove the AH null flag.")
- Be careful when applying flags to multiple parameters – the DART tool provides a convenient option to propagate a selected flag to multiple parameters based on sampling event or filter type. Double check that you have selected the appropriate propagation type so that you do not accidentally flag Teflon™ and Nylon parameters for something that only affects the quartz parameters. Additionally, check that the propagated changes made are consistent with your expectations.

PLEASE DO NOT

- Invalidate samples with the "FX" or "MX" qualifier codes (see Section 9 **Appendix I – AQS Codes**) unless additional information support invalidation – these codes are selected by the sample shipping and handling laboratory and analytical laboratory technicians during visual inspection of the filters. If a filter defect results in poor analysis results, it is invalidated. When a filter experiences damage or contamination (e.g. grass, hair, insect) outside of the analysis area, it is instead flagged with a qualifier code. Contaminated areas can often be avoided during analysis of Quartz filters, but to a lesser degree for Teflon™ 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 temperatures greater than 4 °C. The "MD" flag is applied when the reported value is below the corresponding detection limit. Removing the "MD" flag will not change the fact that the value is below detection. Removing the "TT" flag only reduces the amount of information about the sample that may be pertinent to an end user.

6.1. Offline Analyses

The DART interface provides many useful tools for visualizing the datasets. In addition to the online tools, DART enables users to export datasets to their local computer for offline analysis using a spreadsheet or other data analysis software. There are innumerable methods for data analysis, which is outside the scope of this document. Instead, we will offer some examples of offline analysis that may be useful for validation. 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 for understanding the relationship between two or more interrelated variables. In Figure 2, simple linear regression was used to compare sulfur and sulfate concentrations across the entire CSN network. For data validation, linear regression can be used to determine the comparability of two types of results, such as reconstructed mass versus collocated mass or titanium versus silicon (typical soil elements). For measurements we expect to be very similar (as in 3 x sulfur versus sulfate), 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.

Care should be taken when interpreting results of regression. There are no hard rules governing expectations. For the sulfur/sulfate comparison, it is logical to assume that the sulfate concentrations will be similar to or below the sulfur results because all of the sulfur in the sulfate molecules should be detected. However, there are multiple factors that may affect the measurements of both sulfur and sulfate. Furthermore, uncertainties are present in both measurements, so there will be some variation in the end result. Thus, measured sulfate may be higher than measured sulfur (multiplied by three). While sulfate usually dominates the mass concentration of sulfur in ambient air, other forms of sulfur may be present (e.g. sulfuric acid, hydrogen sulfide, and organosulfates). The analyst should take uncertainties, methodological differences, and local factors into account when interpreting results.

Figure 6. Comparison of carbon mass concentrations between IMPROVE and CSN.
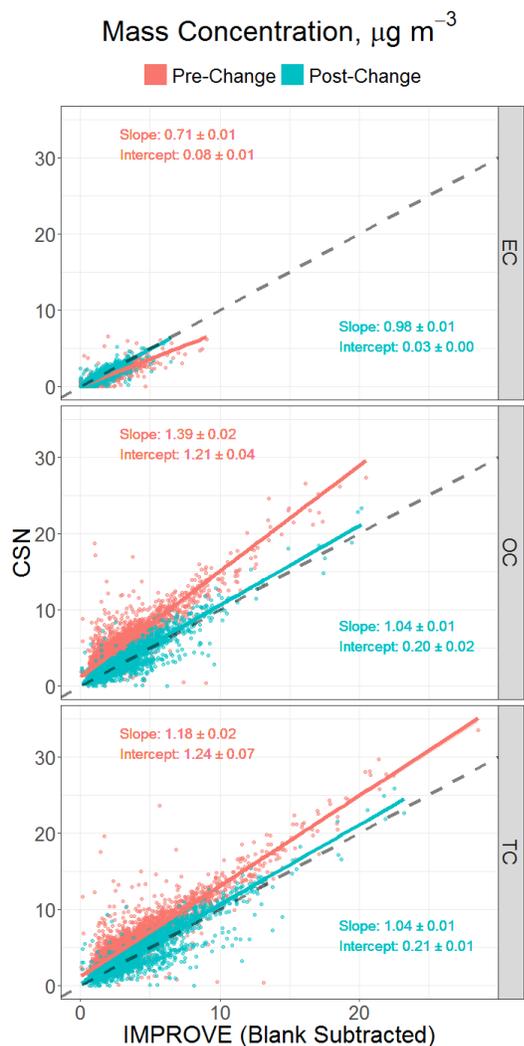
Example

Elemental and organic carbon are measured by thermal optical reflectance/transmittance (TOR and TOT, respectively) in the CSN and IMPROVE networks:

 (http://vista.cira.colostate.edu/Improve/)

Historically, CSN collected quartz filter samples using relatively low flow rate instruments (6.7 L/min) and analyzed them using the NIOSH TOT protocol. The IMPROVE network, on the other hand, collected samples at a higher flow rate (23 L/min) and analyzed them using the IMPROVE_A TOR protocol. Additionally, the IMPROVE network employed blank correction to its carbon measurements while CSN did not.

The difference in flow rates affected the retention of semi-volatile organic compounds while the difference in analytical methods affected the apportionment of carbon as either elemental or organic.

The combination of these differences led to systematic difference between collocated samples from the two networks, as shown in Figure 6. CSN altered its sampling and analysis methods over a period from 2007 to 2009 to better match the IMPROVE network, and since then the regression comparison has shown greater unity between the two networks. Note that the bifurcation in the post-change OC and TC plots are due to equipment issues at one of the collocated sites.

6.1.2. Multi-Parameter Trends

In the DART interface, time-series plots can be easily 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. These parameters are further explained in Section 7, ***Explanation of Terms***. Alternatively, an analyst may wish to view different combinations of parameters or different calculations based on local profiles. For example, an analyst may wish to calculate "soil" based on the local geochemistry rather than the general form available in DART. Another analyst may wish to sum the cation and anion species for their sites and inspect the ratio between them (i.e. $\sum cations / \sum anions$). One useful method for examining elemental data is referred to as enrichment factors (EFs). An EF is a double ratio of

elements where one element represents a specific source of interest (e.g. soil). In this case, the formula for an EF would be

$$EF_{target\ element} = \frac{\left[target\ element \middle/ soil\ element \right]_{measured}}{\left[target\ element \middle/ soil\ element \right]_{reference}}$$

where the reference ratio would be derived from a chemical profile of local soil. Generally, EFs greater than 10 are considered "enriched" with respect to the source of interest. If an analyst calculated an EF of 25 for copper with respect to titanium and an EF of 1 for magnesium with respect to titanium, that analyst could expect the titanium and magnesium to be related to resuspended soil while the copper is enriched due to some other source. Alternatively, if the analyst had a chemical profile of a specific source's emissions, then they could use that profile for the reference ratios (denominator). This way, all elements with an EF near one may be related to the specific source.

Example

EFs may be used to identify potential contamination issues. In Figure 7, EFs were calculated for chlorine, copper, nickel, and zinc with respect to titanium. Titanium is useful as it is not commonly emitted in significant quantities, making it a suitable choice for a crustal element signature. Chlorine may be routinely enriched at coastal sites due to sea salt influence, though not observed at this site. The zinc signature displays several events greater than 200 times enriched. If no other elements were enriched at these times, contamination may have occurred during shipping or analysis. Since copper and nickel also display a similar temporal profile, then the sample may be enriched in these metals. EF analysis should be performed with multiple crustal signature elements (e.g. magnesium, potassium, silicon, titanium) to increase confidence in interpretations.

Figure 7. Enrichment factors calculated for several elements. Chlorine is typically a natural signature for sea salt while copper, nickel, and zinc may be from anthropogenic sources when enriched.
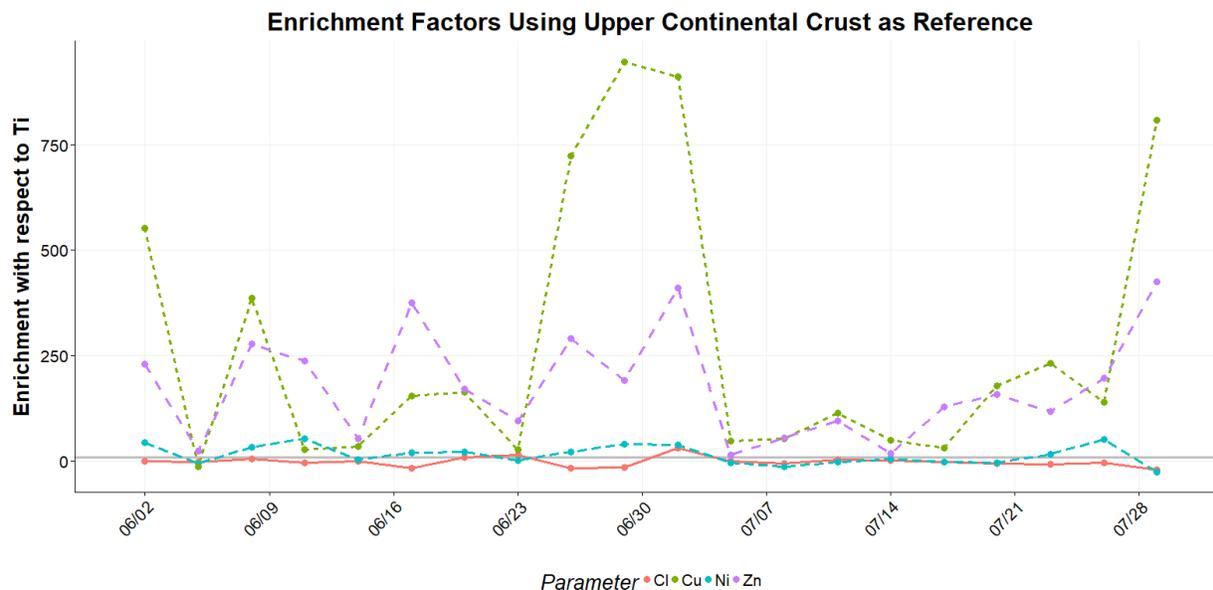
7. Explanation of Terms

In DART, some parameters have been provided that **<u>do not go on to AQS</u>**. UCD calculates and provides these parameters to aid in visualization of the dataset and validation. **These parameters do not need to be validated.** Calculated parameters are shown in Table 2.

Table 2. Calculated parameters included in 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 + Pb	Sum of elements measured by XRF. Calculated in DART.
00002	Ions	chloride + ammonium + sodium ion + potassium ion + nitrate + sulfate	Sum of ions measured by IC. Calculated in DART.
88339	Ammonium sulfate	4.125 * sulfur	This is a historical estimation of ammonium sulfate using XRF sulfur rather than measured sulfate.
88344	Ammonium	1.29 * nitrate	-

		nitrate	
88348	Soil	2.2*aluminum + 2.49*silicon + 1.63*calcium + 2.42*iron + 1.94*titanium	These are estimated from stoichiometric relationships of crustal soil oxides.
88350	Organic Carbon Mass (OMC)	1.4 * organic carbon	Organic carbon by TOR.
88401	Reconstructed Mass	ammonium sulfate + ammonium nitrate + soil + 1.8*chloride + elemental carbon + OMC	Elemental carbon by TOR.

Parameter 88502 (Acceptable $PM_{2.5}$ mass) from one site currently goes into AQS and needs to be validated in DART; however, Parameter 88501 ($PM_{2.5}$ Raw Data) **does not go to AQS** and **does not need validated in DART**. This mass value is retrieved from the AirNowTech web service and provided in DART to allow data validators to compare it to reconstructed mass from the CSN species. UCD cannot see what method is used to collect these data. Also, the data in the AirNowTech service is dynamic, so the value passed to DART may change between the time of AirNowTech collection and DART data delivery. The DART validator should download recent data from the AirNowTech when reviewing the data.

7.1. Field Blanks and Trip Blanks

Field blanks are scheduled in CSN for quality assurance and to calculate method detection limits. Trip blanks, however, are not regularly scheduled and occur if there is an issue with sampling. Since the handling of trip blanks is not consistent, they are not used in routine calculations. They are collected and analyzed but are not submitted to AQS.

7.2. Value Limits

As mentioned previously, certain limits exist either to understand typical parameter ranges or to determine whether or not a sample needs to be invalidated or have a qualifier flag applied. Table 3 lists the limits that are currently used for operational parameters such as flow rates, temperature, and pressure. The "AH" null flag is applied when samples are collected outside of the listed flow limits. Additionally, samples can be invalidated with the "AH" flag when the flow variability (Flow CV) is above 5% for the Met One SASS and 2% for the URG 3000N.

Table 3. Operational limits by instrument.

Instrument	URG 3000N	Met One SASS
Temperature, °C	-20 to 45	600 to 810
Pressure, mmHg	-30 to 50	600 to 810
Nominal Flow, L min^{-1}	22	6.7
Upper Limit, L min^{-1}	24.2	7.37
Lower Limit, L min^{-1}	19.8	6.03

Table 4 lists ranges for parameter comparisons. It should be noted that all limits are being re-evaluated with historical context. The outlier flag, "5", is only assigned for the comparison of reconstructed mass to measured gravimetric mass and when both 3 x sulfur/sulfate and potassium/potassium ion ratios are out of bounds. The bounds for the ion mass balance were used by the previous analytical laboratory contractor and have not been fully explored. The ion mass balance limits are not utilized until the re-evaluation has been completed.

Table 4. Historical limits for inter-parameter comparisons. Note that these values are currently being re-examined.

Ratio	Upper Limit	Lower Limit	"5" Flag Assigned
3 x sulfur / sulfate	1.8	0.667	Yes
Potassium / potassium ion	2	0.5	Yes
Gravimetric mass / reconstructed mass	2	0.5	Yes
Sum of anions / sum of cations	2.82	0.86	No

8. Further Information

Additional resources are available online for data 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://www.epa.gov/sites/production/files/2016-09/documents/dart_training_for_pams_and_csn.pdf

Any questions related to the DART interface can be directed to DART@sonomatech.com.

The Ambient Monitoring Technology Information Center (https://www.epa.gov/amtic) contains many useful documents concerning CSN methods, parameters, and network changes. More detailed discussions concerning the carbon methods and parameters can be found there. Another useful resource is the searchable list of AQS codes, found at:

https://www.epa.gov/aqs/aqs-code-list.

For direct assistance, please email CSNsupport@sonomatech.com. This email service will reach members of the EPA, UCD, and Sonoma Technology, Inc. Questions regarding data within DART should be directed to this address.

9. Appendix I – AQS Codes

Qualifier Code	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 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
TC	Component Check & Retention Time Standard	Null Data Qualifier
TS	Holding Time Or Transport Temperature Is Out Of Specs.	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
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