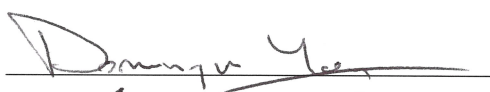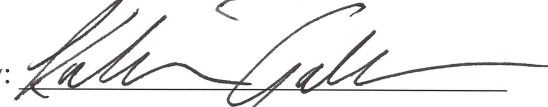# UCD CSN Technical Information #801B

CSN Data Processing

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

November 30, 2018
Version 1.1

Prepared By: _Dominique Yee_   Date: _11/28/2018_

Reviewed By: _____   Date: _11/28/2018_

Approved By: _____   Date: _11/28/18_

UCDAVIS
AIR QUALITY RESEARCH CENTER

DOCUMENT HISTORY

Date Modified	Initials	Section/s Modified	Brief Description of Modifications
11/30/18	NJS	All	Reworded for clarity, equations added, and updated name changes.

Table of Contents

List of Figures

List of Tables

1. **PURPOSE AND APPLICABILITY**

   The subject of this technical information (TI) document is processing the sampling and analytical data from the Chemical Speciation Network (CSN). The raw operational information from field samplers is combined with laboratory analytical results to generate concentrations, uncertainties, and detection limits.

2. **SUMMARY OF THE METHOD**

   The UCD analyst will use functions in the *crocker* software package to calculate final results and post them to the UCD CSN database. The analyst will also review any output messages for errors.

3. **DEFINITIONS**

   - **AQRC:** The Air Quality Research Center at the University of California, Davis.
   - **crocker:** A custom software package in the R language that contains the data processing code used to produce, check, and post the final results.
   - **CSN database:** A SQL Server database that is the central warehouse of CSN preliminary and final data at UC Davis.
   - **MDL:** Minimum detection limit.
   - **RCM:** Reconstructed Mass.

4. **HEALTH AND SAFETY WARNINGS**

   Not applicable.

5. **CAUTIONS**

   Not applicable.

6. **INTERFERENCES**

   Not applicable.

7. **PERSONNEL QUALIFICATIONS, DUTIES, AND TRAINING**

   The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to this project all have advanced training in database programming and database management. All have direct experience through recent involvement in designing and managing a similar database for IMPROVE.

8. PROCEDURAL STEPS

Data processing is performed using the *crocker* R package, which is developed and maintained by UCD specifically for data processing and operational monitoring of the CSN data. Data processing is performed by the UCD Data & Reporting Group on monthly batches of data (a calendar month of sample start dates). Processing occurs in two steps.

1. Calculate concentrations, uncertainties, and MDLs, and check for duplicates (Section 8.1).
2. Post results to the UCD CSN database (Section 8.2).

8.1 Calculate concentrations, uncertainties, and MDLs, and check for duplicates

Laboratory results for ions and carbon fractions are stored in the database as mass per filter and elements are stored as mass per cm^2. The *csn_calculate_* methods in the *crocker* package combine per filter analysis results with filter operational data (e.g., flow rates) and corresponding blanks to calculate concentrations, uncertainty estimates, and MDLs. The details and specific equations are provided in Section 4.

To calculate values for all measured and derived parameters, the operator (typically the data validation analyst) will open an R environment (such as RStudio) and run the following command[1]:

[month_data] <- crocker::csn_calculate_all([YYYY], [MM], 'production')

This command will calculate concentrations, uncertainties, and MDLs for all measured and derived parameters for the year *([YYYY])* and month *([MM])* and return them (in memory) to the variable *[month_data]*. The last argument in the command specifies that the calculations will use the "production" database (i.e., the CSN operational database).

Within the *csn_calculate_all* function, the concentrations, uncertainties, and MDLs are calculated for each of the three main groups of species (elements, ions, and carbon) as well as for operational parameters, calculate values such as sulfate, nitrate, soil, and OMC, and reconstructed mass. Following the calculations performed for each group, a check for duplicates is performed and processing will cease if duplicates are identified. As this duplicate check typically indicates a problem, the data validation analyst will investigate and work to resolve the issue. Duplicate issues can result from the same filter having more than one set of analysis results or because there appears to be more than one filter of the same filter type running on the same day at the same site. Once the issue(s) is resolved, the *csn_calculate_all* command will be re-run.

---

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

8.2 Post results to CSN database

After all duplicates have been resolved, the results can be uploaded to the UCD CSN database in preparation for Level 1 data validation. The results will be uploaded to both the test and production databases. The operator will execute the following commands:

[post] <- crocker::csn_post_results([month_data], server = 'test', delete.file = FALSE, AnalysisQcCode = 1, comment = ['Initial Posting'], replacingId = NULL, replacingQcCode = NULL)

[post] <- crocker::csn_post_results([month_data], 'production', delete.file = FALSE, AnalysisQcCode = 1, comment = ['Initial Posting'], replacingId = NULL, replacingQcCode = NULL)

This command appends the processed data to the analysis.Results table in the UCD CSN production database as an analysis set. It also inserts a record into the analysis.ResultsSets table that provides summary information for this set, including the comment and the *AnalysisQcCode*. *AnalysisQcCode = 1* is used for routine data. The command also creates a text file copy (csnAnalysisResults_[Date]_[Time].txt) for archive that is stored in \\CL-SQL\Production_DB_Inserts. During level 1 validation, the data may be modified and *csn_calculate_all* is run again and a new complete data set is to be posted to the database. The data analyst can use the *csn_post_results* command again, change the comment to record what the dataset relates to as well as changing the analysis QC code of the previously posted dataset(s) by including the data set ID of the previous posting (*replacingId*) and the analysis QC code (*replacingQcCode*) that should be associated with that data set.

9.   DATA PROCESSING EQUATIONS

Section 9.1-9.4 present the equations used to calculate aerosol concentrations and the associated uncertainty and MDL. All these calculations are performed by the *crocker* R package.

The mass of material on the filter is equal to the difference between the mass measured on the sample and the mass on the unused filter. The concentration is the mass on the filter divided by the sample volume.

9.1 Ions

Ions are measured on the nylon filter by ion chromatography. The ions measurement – chloride (Cl^-), nitrate (NO_3^-), sulfate (SO_4^{2-}), ammonium (NH_4^+), potassium (K^+), and sodium (Na^+) – are delivered as micrograms per filter in the ions data files from DRI. The ion concentration, MDL, and uncertainty in micrograms per cubic meter are calculated using Equations 1, 2, and 3 respectively.

$$C = \frac{A - B}{V} \qquad\qquad 1$$

Where,

C = ambient concentration ($\mu g/m^3$)

A = mass measured on sample (μg/filter)

B = artifact mass (μg/filter) = monthly median of field blanks (FB).

V = sample volume (m^3)

The MDL is calculated as the 95th percentile minus the median of monthly field blanks, with a minimum floor value from laboratory replicate analysis as shown in Equation 2.

$$MDL = \frac{max \begin{cases} P_{95} - med \\ MDLfloor \end{cases}}{V} \qquad 2$$

Where,

P_{95} = 95th percentile of at least 50 field blanks for the matching month, including nearby months if necessary (μg/filter)

med = median of at least 50 field blanks for the matching month, including nearby months if necessary (μg/filter)

MDLfloor = MDL floor value determined by taking the mean of repeated analysis of blanks (μg)

V = sample volume (m^3)

Uncertainty is reported with each concentration. The general model for the uncertainty is a quadratic sum of two components of uncertainty as shown in Equation 3.

$$\sigma(c) = \sqrt{(S_{FB})^2 + (fC)^2} \qquad 3$$

Where,

C = ambient concentration ($\mu g/m^3$)

f = Fractional uncertainty, as calculated in Equation 4-6 from five years of collocated precisions (cp).

$$Scaled\ Relative\ Difference\ (SRD) = \frac{(collocated - routine) / \sqrt{2}}{(collocated + routine) / 2} \qquad 4$$

$$Collocated\ Precision\ (cp) = \frac{(84th\ percentile\ of\ SRD) - (16the\ percentile\ of\ SRD)}{2} \qquad 5$$

$$Fractional\ Uncertainty\ (f) = 100 \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(cp)_i^2} \qquad 6$$

n = number of years used for calculation of fractional uncertainty

S_{FB} = Analytical uncertainty, as calculated in Equation 7.

$$S_{FB} = \frac{MDL}{1.6449}$$

7

The MDL floor and fraction uncertainty for each ion species given in Table 1. Note that no fractional uncertainty estimate was available for Chloride, and a value of 0.25 was assigned. These values are stored in the database.

Table 1. MDL floor and fractional uncertainty (f) for ions estimated from 2009-2014 collocated data. For the initial estimates shown here, the entire five-year period was used to calculate collocated precision rather than separated by year. This method was chosen initially to maximize the number of collocated pairs above the three times MDL threshold**.**

Species	MDLfloor	f
Chloride	0.104	0.25
Nitrate	0.067	0.076
Sulfate	0.123	0.049
Ammonium	0.019	0.071
Sodium	0.044	0.247
Potassium	0.587	0.126

9.2 Carbon

Carbon is measured by thermal optical reflectance (TOR) using a quartz filter. Carbon measurements are stored in the carbon file as micrograms per filter. For the eight carbon species, the primary source of fractional uncertainty is the separation into temperature ranges. This may be associated with temperature regulation, but it may also be from the inherent variability of the species involved. The concentration, MDL, and uncertainty in micrograms per cubic meter for the carbon fractions and sums – OC1, OC2, OC3, OC4, OPTR, OPTT, EC1, EC2, EC3, OCTR, ECTR, and TCTC – are calculated both with and without an artifact correction using Equations 8-14. For the versions without artifact correction (e.g., OC1_raw, ECTR_raw), B is 0. For the values with artifact correction (e.g., OC1, ECTR), B is the monthly median mass loading of field blanks across the network.

$$C = \frac{A - B}{V}$$

8

Where,

C = ambient concentration ($\mu g/m^3$)

A = mass measured on sample (μg/filter)

B = artifact mass (μg/filter) = monthly median of field blanks

V = sample volume (m^3)

The MDL corresponds to the maximum of the 95th percentile minus the median of monthly field blanks or a fixed floor value, as shown in Equation 9. The floor value is the analytical MDL developed by the carbon analysis lab. The values are listed in Table 2.

$$MDL = \frac{max \begin{cases} P_{95} - med \\ MDLfloor \end{cases}}{V} \qquad 9$$

Where,

P_{95} = 95th percentile of at least 50 field blanks for the matching month, including nearby months if necessary (µg/filter)

med = median of at least 50 field blanks for the matching month, including nearby months if necessary (µg/filter)

MDLfloor = MDL floor value determined by taking the mean of repeated analysis of blanks (µg)

Uncertainty is reported with each concentration. The general model for the uncertainty is a quadratic sum of two components of uncertainty as shown in Equation 7.

$$\sigma(c) = \sqrt{(S_{FB})^2 + (fC)^2} \qquad 10$$

Where,

C = ambient concentration (µg/m^3)

f = Fractional uncertainty, as calculated in Equation 11-13 from five years of collocated precisions (cp).

$$Scaled\ Relative\ Difference\ (SRD) = \frac{(collocated\ -\ routine) / \sqrt{2}}{(collocated\ +\ routine) / 2} \qquad 11$$

$$Collocated\ Precision\ (cp) = \frac{(84th\ percentile\ of\ SRD) - (16the\ percentile\ of\ SRD)}{2} \qquad 12$$

$$Fractional\ Uncertainty\ (f) = 100 \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(cp)_i^2} \qquad 13$$

n = number of years used for calculation of fractional uncertainty

S$_{FB}$ = Analytical uncertainty, as calculated in Equation 14.

$$S_{FB} = \frac{MDL}{1.6449} \qquad 14$$

Table 2. MDL floor and fractional uncertainty for the carbon species estimated from 2009-2014 collocated data. For the initial estimates shown here, the entire five-year period was used to

calculate collocated precision rather than separated by year. This method was chosen initially to maximize the number of collocated pairs above the three times MDL threshold**.**

Species	MDL floor	f
OC1	0.17	0.329
OC2	1.05	0.136
OC3	0.42	0.178
OC4	0.26	0.136
OP (OPTR)	0.23	0.251
OPTT	0.23	0.173
OCTR	1.43	0.116
OCTT	1.48	0.073
EC1	0.22	0.129
EC2	0.22	0.368
EC3	0.06	0.25
ECTR	0.37	0.155
ECTT	0.37	0.128
TCTC	1.48	0.25

9.3 Elements

Elements are measured using X-ray fluorescence (XRF) (PANalytical Epsilon 5) on PTFE filters from the SASS sampler. Because the XRF instrument reports areal densities, concentrations (*C*) are calculated using Equation 15.

$$C = \frac{AD \cdot a - L}{V}$$ 15

Where,

AD = areal density (μg/cm^2)

L = median mass loading of at least 50 field blanks for the matching month, including nearby months if necessary (μg/filter)

a = sample deposit area (cm^2)

V = sample volume (m^3)

The MDL corresponds to the maximum of the 95th percentile minus the median of monthly field blanks or a fixed floor value, as shown in Equation 16. The floor value is the analytical MDL developed by the carbon analysis lab. The values are listed in Table 2.

$$MDL = \frac{max \begin{cases} P_{95} - med \\ MDLfloor \end{cases}}{V} \qquad 16$$

Where,

P_{95} = 95th percentile of at least 50 field blanks for the matching month, including nearby months if necessary (µg/filter)

med = median of at least 50 field blanks for the matching month, including nearby months if necessary (µg/filter)

MDLfloor = MDL floor value determined by taking the mean of repeated analysis of blanks (µg)

Uncertainty is reported with each concentration. The general model for the uncertainty is a quadratic sum of two components of uncertainty as shown in Equation 17.

$$\sigma(c) = \sqrt{(S_{FB})^2 + (fC)^2} \qquad 17$$

Where,

C = ambient concentration (µg/m^3)

f = Fractional uncertainty, as calculated in Equation 18-20 from five years of collocated precisions (cp).

$$Scaled\ Relative\ Difference\ (SRD) = \frac{(collocated\ -\ routine)\ /\ \sqrt{2}}{(collocated\ +\ routine)\ /\ 2} \qquad 18$$

$$Collocated\ Precision\ (cp) = \frac{(84th\ percentile\ of\ SRD) - (16the\ percentile\ of\ SRD)}{2} \qquad 19$$

$$Fractional\ Uncertainty\ (f) = 100 \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(cp)_i^2} \qquad 20$$

n = number of years used for calculation of fractional uncertainty

S_{FB} = Analytical uncertainty, as calculated in Equation 21.

$$S_{FB} = \frac{MDL}{1.6449} \qquad 21$$

Table 3. MDL floor and fractional uncertainty for elemental species estimated from 2009-2014 collocated data. For the initial estimates shown here, the entire five-year period was used to calculate collocated precision rather than separated by year. This method was chosen initially to maximize the number of collocated pairs above the three times MDL threshold.

Parameter	MDLfloor	f
Na	0.858	0.164
Mg	0.397	0.245
Al	0.312	0.252
Si	0.133	0.152
P	0.02	0.179
S	0.036	0.062
Cl	0.02	0.342
K	0.021	0.106
Ca	0.054	0.168
Ti	0.031	0.174
V	0.013	0.128
Cr	0.026	0.389
Mn	0.062	0.154
Fe	0.17	0.17
Ni	0.018	0.178
Cu	0.11	0.269
Zn	0.03	0.123
Se	0.051	0.25
Br	0.044	0.15
Pb	0.118	0.185

9.4 Reconstructed Mass

Reconstructed mass is compared to mass at collocated continuous monitors during data validation. Reconstructed mass (RCM) concentration is calculated using Equation 22.

$$RCM = NHSO + NHNO + Soil + 1.8\,Chloride + ECTR + OMC \qquad 22$$

Where,

NHSO = Ammonium sulfate concentration (see 4.4.1)

NHNO = Ammonium nitrate concentration (see 4.4.2)

Soil = Soil concentration (see 4.4.3)

Chloride = Chloride concentration as measured by IC (Section 4.1)

ECTR = Total elemental carbon concentration by TOR (Section 4.2)

OMC = Concentration of organic mass by carbon (see 4.4.4)

For all of the terms in Equation 22, zero substitution is applied to negative values. The derived components of RCM and described in the sections below. The MDL for RCM is 0. Uncertainty is calculated as combination of the individual uncertainties via Equation 23.

$$\sigma_{RCM} = \sqrt{\sigma_{NHSO}^2 + \sigma_{NHNO}^2 + \sigma_{Soil}^2 + (1.8\,\sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2}$$

23

For Equation 23, each uncertainty term must be at least zero.

9.4.1 Ammonium Sulfate (NHSO)

Sulfur is predominantly present as sulfate in the atmosphere, generally as ammonium sulfate $(NH_4)_2SO_4$, though also as ammonium bisulfate $(NH_4)HSO_4$, sulfuric acid H_2SO_4, gypsum $CaSO_4 \cdot 2H_2O$, and, in marine areas, sodium sulfate $NaSO_4$. In many cases, the particle will include associated water, but this is omitted from the calculation. All sulfur is assumed to be present as ammonium sulfate, and calculate NHSO concentration, MDL, and uncertainty as:

$$C_{NHSO} = C_S * 4.125$$

24

$$MDL_{NHSO} = MDL_s * 4.125$$

25

$$\sigma_{NSHO} = \sigma_S * 4.125$$

26

9.4.2 Ammonium Nitrate (NHNO)

This is the total dry concentration associated with nitrate, assuming 100% neutralization by ammonium. The concentration, MDL, and uncertainty are derived from the nitrate ion measurement and calculated as:

$$C_{NHNO} = C_{nitrate} * 1.29$$

27

$$MDL_{NHnO} = MDL_{nitrate} * 1.29$$

28

$$\sigma_{NSNO} = \sigma_{nitrate} * 1.29$$

29

9.4.3 Soil

The soil component consists of the sum of the predominantly soil elements measured by XRF, multiplied by a coefficient to account for oxygen for the normal oxide forms $(Al_2O_3, SiO_2, CaO, K_2O, FeO, Fe_2O_3, TiO_2)$, and augmented by a factor to account for other compounds not included in the calculation, such as MgO, Na_2O, water, and CO_2. The following assumptions are made:

- Fe is split equally between FeO (oxide factor of 1.29) and Fe_2O_3 (oxide factor of 1.43), giving an overall Fe oxide factor of 1.36.

- Fine K has a non-soil component from smoke. Based on the K/Fe ratio for average sediment (*Handbook of Chemistry and Physics*) of 0.6, 0.6*Fe is used as a surrogate for soil K and then add the oxide factor for K_2O (oxide factor of 1.2) to get a total Fe factor of 0.72*Fe (0.6*1.2) for the potassium oxide in soil. This increases the factor for Fe from 1.36 to 2.08.
- The oxide forms of the soil elements account for 86% of average sediment; in order to obtain the total mass associated with soil, the final factors are divided by 0.86 (*Handbook of Chemistry and Physics*). The final equations for fine soil concentration, MDL, and uncertainty are:

$$C_{soil} = 2.2\,Al + 2.49\,Si + 1.63\,Ca + 2.42\,Fe + 1.94\,Ti \qquad 30$$

$$MDL_{soil} = max(2.2\,MDL_{Al}, 2.49\,MDL_{Si}, 1.63\,MDL_{Ca}, 2.42\,MDL_{Fe}, 1.94\,MDL_{Ti}) \qquad 31$$

$$\sigma_{soil} = \sqrt{(2.2\,\sigma_{Al})^2 + (2.49\,\sigma_{Si})^2 + (1.63\,\sigma_{Ca})^2 + (2.42\,\sigma_{Fe})^2 + (1.94\,\sigma_{Ti})^2} \qquad 32$$

For Equations 30-32, zero substitution is applied to any negative terms.

9.4.4 Organic Mass by Carbon (OMC)

To determine the total amount of organic mass associated with organic carbon, the ratio of organic mass to organic carbon is assumed to be 1.4.
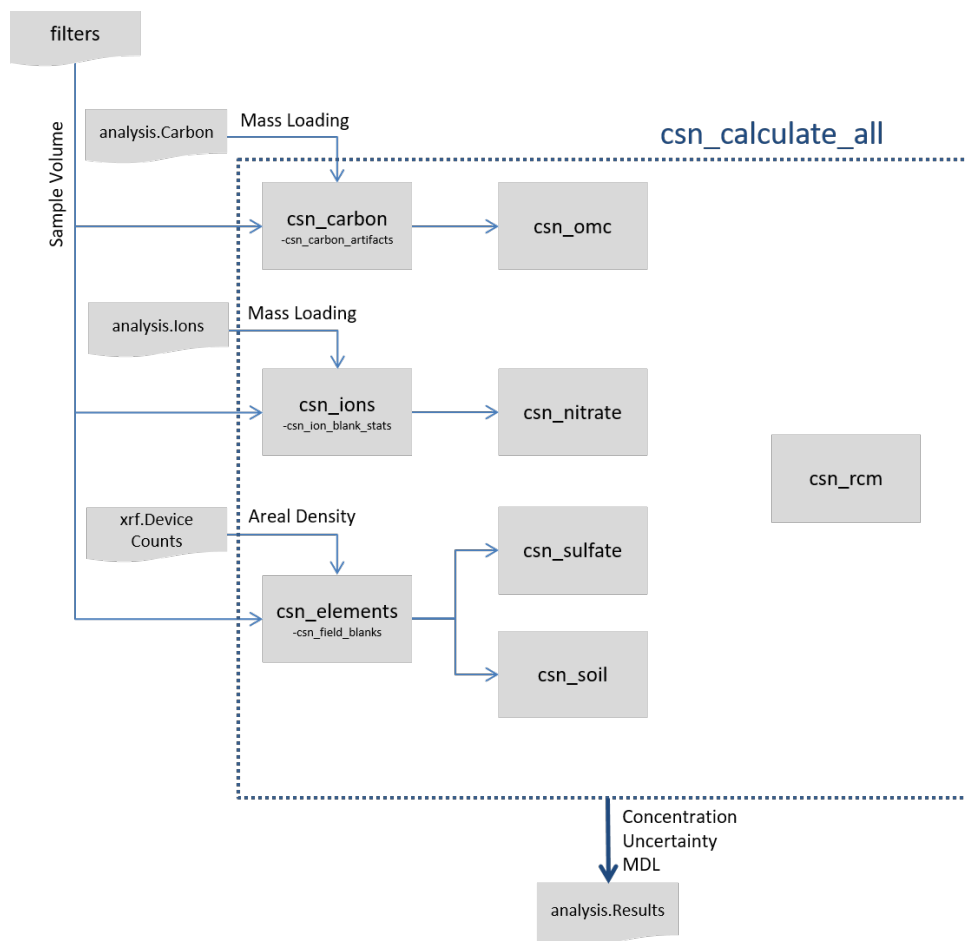
$$C_{OMC} = 1.4 * C_{OCTR} \qquad 33$$

$$MDL_{OMC} = 1.4 * MDL_{OCTR} \qquad 34$$

$$\sigma_{OMC} = 1.4 * \sigma_{OCTR} \qquad 35$$

10. DATA PROCESSING CODE

This section describes the flow of data through the data processing code used to calculate concentration, mdl, and uncertainty for all CSN parameters. Figure 1 outlines the flow of data from the filter and analysis specific database tables to final results. The wrapper function *csn_calculate_all* is the only function executed directly by the analyst (see 3.1); *csn_calculate_all* in turn calls several functions sequentially to calculate first measured and then derived concentrations. Source code for the functions shown in Figure 1 is stored in the UCD source repository.

Figure 1. Flow diagram of the processing code in *crocker::csn_calculate_all*. The raw sampling and analysis data are inputs to the process. Rectangles represent R functions and arrows indicate inputs and outputs.

11. EQUIPMENT AND SUPPLIES

The associated hardware and software used for CSN data ingest are described in the associated UCD SOP #801.

12. QUALITY ASSURANCE AND QUALITY CONTROL

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

13. REFERENCES
Not Applicable.