

UCD CSN Technical Instruction #801B

CSN Data Processing

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

*July 31, 2021
Version 1.4*

Prepared By: _____ Date: _____

Reviewed By: _____ Date: _____

Approved By: _____ Date: _____

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.
7/31/19	DEY, KAG	1,3,7,8,9,10	Changed process for carbon and ion analysis pathways. Wording changes for clarity.
7/31/20	KAG	5-13, 15	Added further description of analytical MDL. Defined process for updating analytical MDL and fractional uncertainty. Wording changes for clarity.
3/31/21	DEY, AMM	8, 9	Wording changes for clarity, added and updated equations for Reconstructed Mass and Soil MDLs.

TABLE OF CONTENTS

1. Purpose and Applicability.....	4
2. Summary of the Method	4
3. Definitions.....	4
4. Health and Safety Warnings	4
5. Cautions	4
6. Interferences.....	4
7. Personnel Qualifications	5
8. Procedural Steps.....	5
8.1 Calculate concentrations, uncertainties, and MDLs, and check for duplicates.....	5
8.2 Post results to CSN database.....	6
9. Data Processing Equations.....	6
9.1 Ions.....	6
9.2 Carbon	8
9.3 Elements.....	9
9.4 Reconstructed Mass.....	11
9.4.1 Ammonium Sulfate (NH ₄ SO ₄)	12
9.4.2 Ammonium Nitrate (NH ₄ NO ₃).....	12
9.4.3 Soil.....	12
9.4.4 Organic Mass by Carbon (OMC).....	13
10. Data Processing Code	14
11. Equipment and Supplies	14
12. Quality Assurance and Quality Control.....	15
13. References.....	15

LIST OF FIGURES

Figure 1. Flow diagram of the processing code in crocker::csn_calculate_all.....	14
--	----

1. PURPOSE AND APPLICABILITY

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

2. SUMMARY OF THE METHOD

The University of California, Davis (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

- **Chemical Speciation Network (CSN):** EPA's PM_{2.5} sampling network, with sites located principally in urban areas.
- **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 UCD.
- **Method Detection Limit (MDL):** A lower limit of detection specific to method of analysis and reported parameter.
- **RCM:** Reconstructed mass
- **Ion Chromatography (IC):** An analytical technique used to determine the concentration of ions.
- **Thermal Optical Analysis (TOA):** An analytical technique used to determine the concentration of carbon.
- **Energy Dispersive X-Ray Fluorescence (EDXRF):** An analytical technique used to determine the concentration of elements.

4. HEALTH AND SAFETY WARNINGS

Not applicable.

5. CAUTIONS

Not applicable.

6. INTERFERENCES

Not applicable.

7. PERSONNEL QUALIFICATIONS

The UCD Air Quality Research Center (AQRC) Data & Reporting Group staff assigned to tasks described in this document have advanced training in database programming and database management.

8. PROCEDURAL STEPS

Data processing is performed using the *crocker* R package, which is developed and maintained by UCD specifically for data processing, monitoring, and validation 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 cm². 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 9.

To calculate values for all measured and derived parameters, the analyst should open an R environment (such as RStudio) and run the following command¹:

```
[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 calculated values such as ammonium sulfate, ammonium nitrate, soil, organic mass by carbon, and reconstructed mass. Following the calculations, a check for duplicates is performed and processing will cease if duplicates are identified. When the duplicate check indicates a problem, the analyst will investigate and 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

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

on the same day at the same site. Once the issues are resolved, the *csn_calculate_all* command should be re-run.

8.2 Post results to CSN database

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

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

This command appends the processed data to the analysis.Results or analysis.CompositeResults 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. During level 1 validation, the data may be modified and *csn_calculate_all* is run again and a new complete data set is posted to the database. The analyst should use the *csn_post_results* command again, change the comment to record what the dataset relates to as well as change 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 through Section 9.4 present the equations used to calculate species concentrations and the associated uncertainties and MDLs. These calculations are performed by the *crocker* R package.

9.1 Ions

Ions are measured on the nylon filter by ion chromatography (IC). The ions measurement results – chloride (Cl⁻), nitrate (NO₃⁻), sulfate (SO₄²⁻), ammonium (NH₄⁺), potassium (K⁺), and sodium (Na⁺) – are delivered as micrograms per filter in the ions data files from RTI. The ion concentration, MDL, and uncertainty in micrograms per cubic meter are calculated using Equations 1, 2, and 3 respectively.

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

1

Where,

C = ambient concentration (µg/m³)

A = mass measured on sample (µg/filter)

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

V = sample volume (m³)

The MDL is calculated as the 95th percentile minus the median of monthly field blanks (Equation 2) and is dependent on sample volume. Since the MDL calculated here is determined using field blanks, it is expected to be higher than the analytical MDL which is calculated by the laboratory using laboratory blanks. As such, the analytical MDL is used as a minimum floor value for the unlikely case where the field blank MDL is lower than the analytical MDL. The field blank MDL is reported with each sample concentration and varies per month of reported data.

$$MDL = \frac{\max \left\{ \begin{array}{l} P_{95} - med \\ analytical\ MDL \end{array} \right.}{V} \quad 2$$

Where,

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

med = median of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{filter}$)

analytical MDL = analytical MDL value as reported by the ion chromatography laboratory (μg). Analytical MDL for each species is stored in the CSN database. The analytical MDLs are reviewed annually (corresponding with delivery of data for samples collected during January), and updated if needed.

V = sample volume (m^3)

Uncertainty is reported with each concentration and varies for each concentration reported (Equation 3).

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

Where,

S_{FB} = analytical uncertainty, calculated as shown in Equation 4.

$$S_{FB} = \frac{MDL}{1.6449} \quad 4$$

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

f = fractional uncertainty (Equation 6), calculated using scaled relative differences (SRD; Equation 5) between sample pairs collected at CSN collocated sites (using collocated data from the most recent two-year period, June through May). Fractional uncertainty for each species is stored in the CSN database. If the total number of valid collocated pairs over the two-year period is less than 60, a value of 0.25 is adopted as the fractional uncertainty. The fractional uncertainties are reviewed and updated annually (corresponding with delivery of data for samples collected during January).

$$\text{scaled relative difference (srd)} = \frac{(\text{collocated} - \text{routine})/\sqrt{2}}{(\text{collocated} - \text{routine})/2} \quad 5$$

$$\text{Fractional Uncertainty (f)} = \frac{(\text{84th percentile of srd}) - (\text{16th percentile of srd})}{2} \quad 6$$

9.2 Carbon

Carbon is measured on the quartz filter by thermal optical analysis (TOA). Carbon measurement results are reported from the laboratory 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 operationally defined 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 (e.g., OC1, ECTR) and without (e.g., OC1_raw, ECTR_raw) an artifact correction using Equations 7 through 11. When calculated without artifact correction, the B variable in Equation 7 is zero.

$$C = \frac{A - B}{V} \quad 7$$

Where,

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

A = mass measured on sample ($\mu\text{g}/\text{filter}$)

B = artifact mass ($\mu\text{g}/\text{filter}$) = 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 (Equation 8) and is dependent on sample volume. Since the MDL calculated here is determined using field blanks, it is expected to be higher than the analytical MDL which is calculated by the laboratory using laboratory blanks. As such, the analytical MDL is used as a minimum floor value for the unlikely case where the field blank MDL is lower than the analytical MDL. The field blank MDL is reported with each sample concentration and varies per month of reported data.

$$\text{MDL} = \frac{\max \left\{ \begin{array}{l} P_{95} - \text{med} \\ \text{analytical MDL} \end{array} \right.}{V} \quad 8$$

Where,

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

med = median of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{filter}$)

analytical MDL = analytical MDL value as reported by the thermal optical analysis laboratory (μg). Analytical MDL for each species is stored in the CSN database. The analytical MDLs are reviewed annually (corresponding with delivery of data for samples collected during January), and updated if needed.

Uncertainty is reported with each concentration and varies for each concentration reported (Equation 9).

$$\sigma(c) = \sqrt{(S_{FB})^2 + (fC)^2} \quad 9$$

Where,

S_{FB} = analytical uncertainty, calculated as (Equation 10).

$$S_{FB} = \frac{MDL}{1.6449} \quad 10$$

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

f = fractional uncertainty (Equation 12), calculated using scaled relative differences (SRD; Equation 11) between sample pairs collected at CSN collocated sites (using collocated data from the most recent two-year period, June through May). Fractional uncertainty for each species is stored in the CSN database. If the total number of valid collocated pairs over the two-year period is less than 60, a value of 0.25 is adopted as the fractional uncertainty. The fractional uncertainties are reviewed and updated annually (corresponding with delivery of data for samples collected during January).

$$\text{scaled relative difference (srd)} = \frac{(\text{collocated} - \text{routine})/\sqrt{2}}{(\text{collocated} - \text{routine})/2} \quad 11$$

$$\text{Fractional Uncertainty (f)} = \frac{(84\text{th percentile of srd}) - (16\text{th percentile of srd})}{2} \quad 12$$

9.3 Elements

Elements are measured on the PTFE filter using energy dispersive X-ray fluorescence (EDXRF). The element measurement results for 33 element species are delivered as areal densities in micrograms per centimeter squared. The element concentration, MDL, and uncertainty are calculated using Equation 13, 14, and 15, respectively.

$$C = \frac{AD \cdot a - L}{V} \quad 13$$

Where,

AD = areal density ($\mu\text{g}/\text{cm}^2$)

L = median mass loading of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{filter}$)

a = sample deposit area (cm^2)

V = sample volume (m^3)

The MDL is calculated as the 95th percentile minus the median of monthly field blanks (Equation 14) and is dependent on sample volume. Since the MDL calculated here is determined using field blanks, it is expected to be higher than the analytical MDL which is calculated by the laboratory using laboratory blanks. As such, the analytical MDL is used as a minimum floor value for the unlikely case where the field blank MDL is lower than the analytical MDL. The field blank MDL is reported with each sample concentration and varies per month of reported data.

$$MDL = \frac{\max \left\{ \begin{array}{l} P_{95} - med \\ analytical\ MDL \end{array} \right.}{V} \quad 14$$

Where,

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

med = median of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{filter}$)

analytical MDL = analytical MDL value as reported by the X-ray fluorescence laboratory (μg). Analytical MDL for each species is stored in the CSN database. The analytical MDLs are reviewed annually (corresponding with delivery of data for samples collected during January), and updated if needed.

Uncertainty is reported with each concentration and varies for each concentration reported (Equation 15).

$$\sigma(c) = \sqrt{(S_{FB})^2 + (fC)^2} \quad 15$$

Where,

S_{FB} = analytical uncertainty, calculated as (Equation 16).

$$S_{FB} = \frac{MDL}{1.6449} \quad 16$$

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

f = fractional uncertainty (Equation 18), calculated using scaled relative differences (SRD; Equation 17) between sample pairs collected at CSN collocated sites (using collocated data from the most recent two-year period, June through May). Fractional uncertainty for each species is stored in the CSN database. If the

total number of valid collocated pairs over the two-year period is less than 60, a value of 0.25 is adopted as the fractional uncertainty. The fractional uncertainties are reviewed and updated annually (corresponding with delivery of data for samples collected during January).

$$\text{scaled relative difference (srd)} = \frac{(\text{collocated} - \text{routine})/\sqrt{2}}{(\text{collocated} - \text{routine})/2} \quad 17$$

$$\text{Fractional Uncertainty (f)} = \frac{(\text{84th percentile of srd}) - (\text{16th percentile of srd})}{2} \quad 18$$

9.4 Reconstructed Mass

Reconstructed mass (RCM; Equation 19) is a composite variable estimated from chemical speciation measurements, and can be compared to gravimetric measurements to assess overall data quality. Gravimetric analysis is not routinely performed using CSN filters; thus, for comparison purposes data from collocated continuous monitors (available from AirNow Tech) are used for validation.

$$RCM = NHSO + NHNO + \text{Soil} + 1.8 \text{ Chloride} + ECTR + OMC \quad 19$$

Where,

NHSO = ammonium sulfate concentration (see 9.4.1)

NHNO = ammonium nitrate concentration (see 9.4.2)

Soil = soil concentration (see 9.4.3)

Chloride = chloride concentration as measured by IC (Section 9.1)

ECTR = elemental carbon concentration by TOA (Section 9.2)

OMC = concentration of organic mass by carbon (see 9.4.4)

For all of the terms in Equation 19, zero is substituted for negative values.

The MDL is calculated as the 95th percentile minus the median of monthly field blanks. If there are not at least 50 valid field blank events (i.e, valid field blanks for matching sites/dates for PTFE, nylon, and quartz filters), field blanks from previous months are also used in the calculation.

$$MDL_{RCM} = (P_{95} - med) \quad 20$$

Where,

P_{95} = 95th percentile of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{m}^3$)

med = median of at least 50 field blanks for the matching month, including nearby months if necessary ($\mu\text{g}/\text{m}^3$)

Uncertainty is calculated as the combination of the individual uncertainties (Equation 21).

$$\sigma_{RCM} = \sqrt{\sigma_{NHSO}^2 + \sigma_{NHNO}^2 + \sigma_{Soil}^2 + (1.8 \sigma_{Chloride})^2 + \sigma_{ECTR}^2 + \sigma_{OMC}^2} \quad 21$$

For Equation 21, 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 $(\text{NH}_4)_2\text{SO}_4$ (ammonium sulfate) though also as $(\text{NH}_4)\text{HSO}_4$ (ammonium bisulfate), H_2SO_4 (sulfuric acid), $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ (gypsum), and, in marine areas, NaSO_4 (sodium sulfate). In many cases, the particle will include associated water, but this is omitted from the calculation. Assuming all sulfur is present as ammonium sulfate (NHSO), the concentration, MDL, and uncertainty are derived from the sulfur elemental measurement and are calculated as:

$$C_{NHSO} = C_S * 4.125 \quad 22$$

$$MDL_{NHSO} = MDL_S * 4.125 \quad 23$$

$$\sigma_{NSHO} = \sigma_S * 4.125 \quad 24$$

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 are calculated as:

$$C_{NHNO} = C_{nitrate} * 1.29 \quad 25$$

$$MDL_{NHNO} = MDL_{nitrate} * 1.29 \quad 26$$

$$\sigma_{NSNO} = \sigma_{nitrate} * 1.29 \quad 27$$

9.4.3 Soil

The fine soil component consists of the sum of the predominantly soil elements measured by EDXRF, 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 Fe₂O₃ (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 of 0.6 for average sediment (*Handbook of Chemistry and Physics*), 0.6*Fe is used as a surrogate for soil K; accounting for the oxide factor for K₂O (oxide factor of 1.2), a total Fe factor of 0.72*Fe (0.6*1.2) is used 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:

$$C_{soil} = 2.2 Al + 2.49 Si + 1.63 Ca + 2.42 Fe + 1.94 Ti \quad 28$$

$$MDL_{Soil} = \frac{P_{95} - med}{V} \quad 29$$

Where,

P₉₅ = 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)

V = sample volume (m³)

Uncertainty is calculated as the combination of the individual uncertainties.

$$\sigma_{soil} = \sqrt{(2.2 \sigma_{Al})^2 + (2.49 \sigma_{Si})^2 + (1.63 \sigma_{Ca})^2 + (2.42 \sigma_{Fe})^2 + (1.94 \sigma_{Ti})^2} \quad 30$$

For Equations 28-30, 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.

$$C_{OMC} = 1.4 * C_{OCTR} \quad 31$$

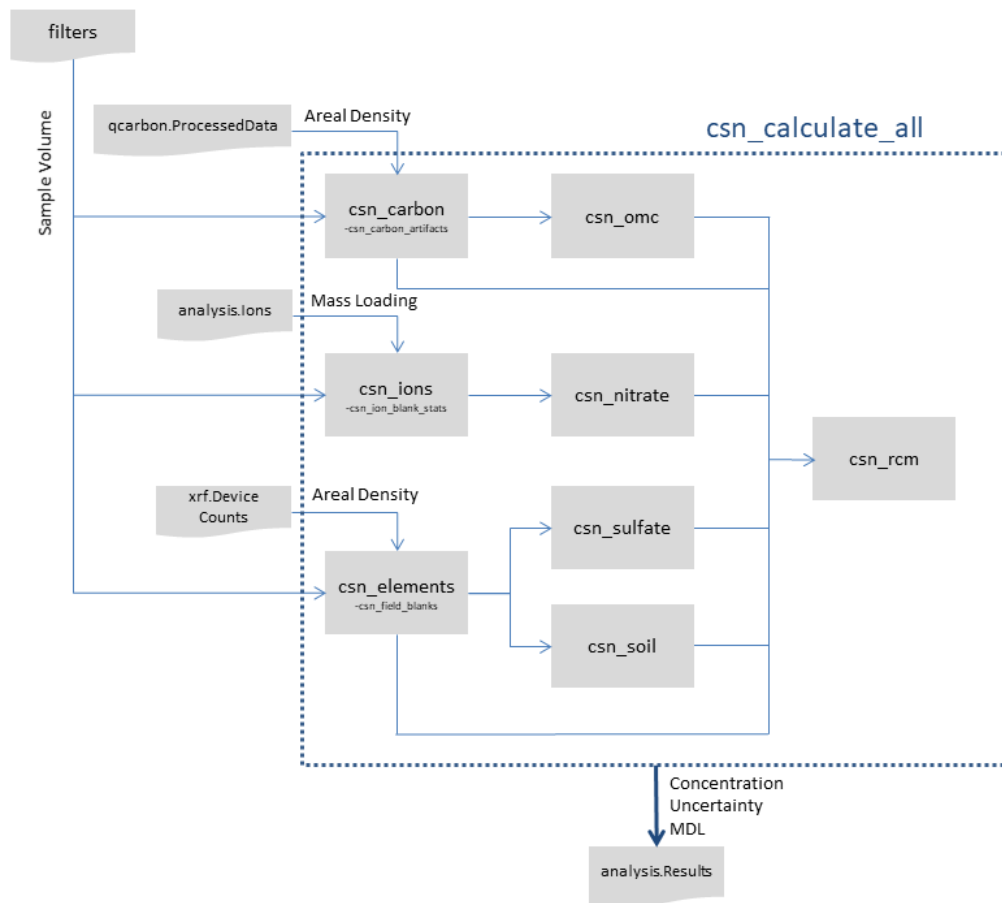
$$MDL_{OMC} = 1.4 * MDL_{OCTR} \quad 32$$

$$\sigma_{OMC} = 1.4 * \sigma_{OCTR} \quad 33$$

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` from the crocker R package is the only function executed directly by the analyst (see Section 8.1); `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. Gravimetric mass is not included; mass results are typically only available for specially designated sites.



11. EQUIPMENT AND SUPPLIES

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

12. QUALITY ASSURANCE AND QUALITY CONTROL

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

13. REFERENCES

UCD CSN SOP #801: Processing & Validating Raw Data