CABM SP2 analysis code documentation

## Introduction

The goal of this code package is to provide time series of refractory black carbon (rBC) mass and number concentration from sp2 measurements. All raw sp2 data and the instrument operating information is entered into a MySQL database. Time series data is then derived and also stored in the database.

All data processing code is written in Python 2.7. There are two relevant code libraries:

* SP2\_library – contains general code for sp2 analysis
* CABM\_sp2 – contains CABM specific code for sp2 analysis

A schematic of the general database and data analysis structure is given on the following page. The general workflow is given following that, and more detail on each step is given in Sections 1 through 7.

## Database and Data Analysis Structure

../../../Downloads/Copy%20of%20CABM%20SP2%20analysis%20-%20v3.

Blue = database tables separated by location

Yellow = single database table for all locations

## General Workflow

1. Check that the site information is in the sp2\_locations table. If this is a new location, enter the required information. (For more detail see Location and Instrument Information)
2. Check that the SP2 instrument information is in the sp2\_instrument\_info table. If this is a new instrument, enter the required information. (For more detail see Location and Instrument Information)
3. If there are new calibrations, enter the instrument ID, the location ID, the calibration date, the calibration material, and the calibrated channel into the sp2\_calibrations table. (For more detail see Calibration Data).
4. Use the ‘CABM\_add\_calibration\_points\_to\_db’ script to enter the calibration data point values into the sp2\_calibration\_points table. (For more detail see Calibration Data).
5. Use the ‘CABM\_fit\_and\_plot\_calibration’ script to fit a line to the calibration points, plot it for viewing, and, if this is a new calibration, update the sp2\_calibrations table with the fit parameters. (For more detail see Calibration Data).
6. Use the the CABM\_add\_config\_data\_to\_db script to read the SP2 configuration files and write the relevant parameters to the database. (For more detail see Configuration Files).
7. Use the CABM\_add\_hk\_data\_to\_db script to read the SP2 housekeeping files and write the relevant parameters to the sp2\_hk\_data\_locnX table. Here X indicates a specific location ID from the sp2\_locations table. (For more detail see Housekeeping Data).
8. Use the CABM\_add\_hk\_IDs\_to\_raw\_data\_table script to add housekeeping ids into the raw data table for use as foreign keys. (For more detail see Housekeeping Data).
9. Use the CABM\_add\_single\_particle\_data\_to\_db script to add incandescence data from single particles to the sp2\_single\_particle\_data\_locnX table. (For more detail see Raw SP2 Particle Data).
10. Enter any intervals that should be ignored for QC reasons into the sp2\_qc\_intervals\_locnX table. (For more detail see QC Checks).
11. Use the CABM\_compile\_time\_interval\_data script to populate the sp2\_time\_intervals\_locnX table. At this point the user must decide if they wish to extrapolate the calibration or only include particles within the calibration range. (For more detail see Time Series Data).
12. Use the The CABM\_compile\_interval\_distribution\_data script to calculate the fraction of the total rBC mass within the detection limits of the instrument, and write this fraction to the sp2\_time\_intervals\_locnX table. (For more detail see Time Series Data).

Note: All scripts can be run from the command line and have help files that can be accessed using the –h argument.

## Scripts for Plotting

There are several scripts for plotting. These can be useful for more in depth investigation of the data.

* The CABM\_fit\_and\_plot\_calibration script can be used to plot and fit a calibration (with either a quadratic or linear function). It also has the option to update the database with the new fit parameters.
* The CABM\_plot\_hk\_parameters script can be used to plot the sample flow, sheath flow, and yag power over a given time interval for a given instrument and location.
* The CABM\_plot\_raw\_sp2\_signals script can be used to view the raw signals from a single particle.
* The CABM\_QC\_color\_ratio script can be used to plot the Broadband/Narrowband incandescence signal ratio as a function of particle size over a given time interval for a given instrument and location. Note that this script can take a significant time to run, long intervals are ok, but may be slow to process.
* The CABM\_compile\_interval\_distribution\_data script has the option of plotting the mass distribution over a given time interval for a given instrument and location (the database will not be updated unless specified).
* The CABM\_plot\_timeseries script can be used to plot the mass and number concentration over a given time interval for a given instrument and location.

# Location and Instrument Information

This section concerns the basic information on the instruments and the deployment sites.

## Relevant module:

* None

## Relevant scripts:

* None

## Database tables:

* sp2\_instrument\_info
* sp2\_locations

## Details:

These tables are the top of the hierarchy; all other database tables will have a foreign key reference to an instrument (a particular SP2) and a location (one of the CABM sites). These tables are short but must be built manually.

The sp2\_locations table includes an id and a location name. The id is used in all further tables to link both tables and data to a location.

The sp2\_instrument\_info table has an entry for each instrument. Fields include, the instrument owner (in this case ECCC), the instrument number (e.g. 17,44,58, etc), the number of channels, the acquisition rate, the bytes per record, the minimum detectable signal, and the saturation limit. Note that there was a change in bytes per record for sp2 #17 when it moved from Whistler to East Trout Lake. This has been accounted for in the CABM\_add\_single\_particle\_data\_to\_db script.

# Calibration Data

This section concerns information on the SP2 incandescence calibrations. This calibration is performed using an external standard and relates incandescence peak height to refractory black carbon mass. Note that if Aquadag is used as the calibration standard the slope of the calibration curve is scaled by a factor of 1.43 to account for the SP2 bias toward Aquadag (See: Laborde, M., Mertes, P., Zieger, P., Dommen, J., Baltensperger, U., and Gysel, M.: Sensitivity of the Single Particle Soot Photometer to different black carbon types, Atmos. Meas. Tech., 5, 1031-1043, doi:10.5194/amt-5-1031-2012, 2012.).

## Relevant module:

* SP2\_calibration

## Relevant scripts:

* CABM\_add\_calibration\_points\_to\_db
* CABM\_fit\_and\_plot\_calibration

## Database tables:

* sp2\_calibrations
* sp2\_calibration\_points

The sp2\_calibrations table holds the information for each calibration. This includes the instrument ID, the location ID, the calibration date, the calibration material, the calibrated channel and information about the fit to the calibrations points.

The sp2\_calibration\_points table holds the individual calibration data points for each calibration. This table has a foreign key set as calibration\_ID to link the data points to a calibration in the sp2\_calibrations table.

## Details:

Calibration data is worked up and entered into excel files by the SP2 technician.

General calibration data is entered manually into the sp2\_calibrations table. This data includes the instrument ID, the calibration location ID, the calibration date, the calibration material and the channel that was calibrated.

The ‘CABM\_add\_calibration\_points\_to\_db’ script is used to enter the calibration data point values into the sp2\_calibration\_points table. This script takes in a calibration date, an instrument number, a calibrated channel (BBHG or BBLG) and the full path for a file containing the calibration data. The file with the calibration data should be a tab-delimited text file with a single line header and rows containing the mobility diameter followed by the signal.

For example:

calibration for SP2#58 at Alert on 20161113

mobility\_diameter signal

150 465.88

175 1006.3

200 1710.5

225 2492.4

250 3357.2

269 4046.2

300 5787.3

350 7385.2

400 9868.1

450 12607

The ‘CABM\_fit\_and\_plot\_calibration’ script is used to fit a line to the calibration points, plot it for viewing, and (optionally) update the sp2\_calibrations table with the fit parameters. This script takes in a date, a location, an instrument number, and a fit type (linear or quadratic). The optional –u argument can be used to update the sp2\_calibrations table. Note that the date does not have to be the date on which a calibration was performed. If another date is used, the relevant calibration for that date will be plotted. If the SP2 is an 8-channel instrument, both the low and high gain channels will be plotted.

# Configuration Files

## Relevant module:

* SP2\_configuration

## Relevant scripts:

* CABM\_add\_config\_data\_to\_db

## Database tables:

* sp2\_config\_parameters

The sp2\_config\_parameters table holds the data from the configuration files that is relevant when looking at mass and number concentration.

## Details:

The CABM\_add\_config\_data\_to\_db reads the SP2 configuration files and writes the relevant parameters to the database

There are two issues to worry about in the instrument configuration when looking at mass and number concentration.

1. The time sample factor, which is 1 out of how many minutes the instrument was recording data
2. The particle sample factor, which is 1 out of how many detected particles was saved to file

Issue number 1 is addressed when calculating time interval data. When looking at intervals we check to see if a particle has a long collection interval and ignore it if it does. For example, if we were recording data 1 out of every 10 minutes, the sampling interval would be 10 minutes long, compared to less than a second in most cases. We ignore this particle because this long sampling interval drastically dilutes the particle concentration.

Issue number 2 is also addressed when calculating time interval data. In this case we use the particle sample factor when calculating particle concentrations to correct the concentrations accordingly (i.e if only 1/10 particles was written to file our sample volume should be divided by 10).

# Housekeeping Data

## Relevant module:

* SP2\_housekeeping

## Relevant scripts:

* CABM\_add\_hk\_data\_to\_db
* CABM\_add\_hk\_IDs\_to\_raw\_data\_table
* CABM\_plot\_hk\_parameters

## Database tables:

* sp2\_hk\_data\_locnX
* sp2\_hk\_limits\_locnX

The sp2\_hk\_data\_locnX table holds housekeeping data from the .hk files. Parameters stored are the sample flow, sheath flow, yag power, and yag crystal temperature.

The sp2\_hk\_limits\_locnX table holds upper and lower limits for the housekeeping parameters. These are used to determine if the instrument was in good operational condition which is a quality control check for the data.

## Details:

The CABM\_add\_hk\_data\_to\_db script reads the SP2 housekeeping files and writes the relevant parameters to the sp2\_hk\_data\_locnX table. This script takes in dates to start and finish parsing hk data for, a location name, an instrument number, and the full path for the directory containing the daily raw data folders (or full path for the directory containing all files, if they are all in a single folder). It is also necessary to specify the column numbers from the raw .hk files that hold the parameters of interest (these change over time and from instrument to instrument). The columns of interest are ‘seconds past midnight’, ‘sample flow’, ‘yag power’, ‘sheath flow’, and ‘yag crystal temperature’. Optional arguments include a flag to set if all files are in a single folder (true for some older instruments), an input for time zone if the hk data is not in UTC, and an input for data interval (in seconds) if the data is recorded at something other than 1Hz.

It is much faster to add the hk ids after fully building the single particle table, so the CABM\_add\_hk\_IDs\_to\_raw\_data\_table script adds hk ids into the raw data table for use as foreign keys. This script takes in a start and end date for the ids to be entered, as well as an instrument location.

*Note:* For some older instruments (like sp2 #17) only time since midnight *local* time is recorded in the housekeeping files. Local time depends on what time zone the computer clock was set for, and it is not possible to know from the housekeeping files alone what time zone the housekeeping data is recorded in. For example, in Egbert in 2009-2010 SP2#17 was using PST (UTC-8). The time zone that the housekeeping data are written in can be determined by comparing the labview UTC timestamp and the short\_timestamp, both from a single particle record (i.e a binary record of a single ambient particle) since the particle records record both local and UTC time. Once we know these we can correct the housekeeping timestamps if necessary to make all timestamps in the database UTC.

# Raw SP2 Particle Data

For each particle detected by the SP2 a particle record is created and written to a .sp2b file in binary format. This section concerns parsing these binary records and storing the relevant incandescence information in the database.

## Relevant module:

* SP2\_raw\_data

## Relevant class:

* SP2\_particle\_record

## Relevant scripts:

* CABM\_add\_single\_particle\_data\_to\_db
* CABM\_plot\_raw\_sp2\_signals

## Database tables:

* sp2\_single\_particle\_data\_locnX

The sp2\_single\_particle\_data\_locnX tables hold the data for individual incandescent particles. This includes the instrument ID (refers to the sp2\_instrument\_info table), the location ID (refers to the sp2\_locations table), a housekeeping ID (refers to the sp2\_ sp2\_hk\_data\_locnX table), an .sp2b file name, a file index, the start time for the sampling interval in which that particle was recorded and the end time for the sampling interval in which that particle was recorded (the sampling interval is the time elapsed since the previous particle was detected). Also stored are the peak amplitudes of the broad-band high-gain incandescence signal and the broad-band low-gain incandescence signal and the positions of these peaks. The final stored information are the peak amplitudes of the narrow-band high-gain incandescence signal and the narrow-band low-gain incandescence signal.

## Details:

Single particle incandescence data is added to the database using the CABM\_add\_single\_particle\_data\_to\_db script. This script takes in a start date and end date for the records to be parsed. It also requires a location, an instrument number, and the full path for the directory containing the daily raw data folders. The daily raw data folders are named using the date (e.g. 20150101) and containing all the .sp2b files for that day. Some older instruments did not break the raw files into daily folders and for these cases the optional –s argument should be used to indicate that all the files are in a single folder and the path given should be the full path of the directory containing all of the files.

The CABM\_plot\_raw\_sp2\_signals script can be used to view a single particle record. This script takes in a location, an instrument number, the full path for the directory containing the .sp2b file of interest, and the record number to be viewed. It outputs an interactive plot of the recorded raw high and low gain scattering and incandescence signals.

*Note:* There was a change in the byte rate of SP2 #17 (from 2458 to 1498) when it was installed at East Trout Lake in 2013. A code snippet to correct for this was added to the CABM\_add\_single\_particle\_data\_to\_db script.

# QC Checks

## Relevant class:

* CABM\_SP2\_time\_interval (a subclass of SP2\_time\_interval)

## Relevant script:

* CABM\_QC\_color\_ratio

## Database tables:

* sp2\_hk\_limits\_locnX
* sp2\_qc\_intervals\_locnX
* sp2\_qc\_code\_definitions

## Details:

QC checks should be performed before the size distributions or time intervals are calculated. From an instrumental stand point these are:

1. housekeeping yag power (any apparent drops in yag power need to be confirmed by looking at the color ratio)
2. housekeeping sample flow

For the housekeeping yag power and the sample flow we use the sp2\_hk\_data\_locnX database table which has the 1 Hz housekeeping values for the yag power, sample air flow, sheath air flow, and yag crystal temperature. We determine reasonable upper and lower limits for each (eg. For a sample flow set point at 120 sccm reasonable limits are 105-135 sccm) then when retrieving individual particle records from the database, we can exclude particles collected during periods when the housekeeping parameters were outside of the set limits (an example is that when the instrument is started up, the flows fluctuate wildly for a few minutes and we can ignore particles collected within this time span). It’s important to note that the yag power in the housekeeping files is not a direct measure of the laser intensity but rather a measurement of light leaking from the cavity. Changes in detector alignment can cause the measured power to change while the laser itself remains constant. Therefore any apparent changes in yag power need to be confirmed by looking at the incandescent particle color ratio (details about the color ratio can be found in: The Detection Efficiency of the Single Particle Soot Photometer J. P. Schwarz, J. R. Spackman, R. S. Gao, A. E. Perring, E. Cross, T. B. Onasch, A. Ahern, W. Wrobel, P. Davidovits, J. Olfert, M. K. Dubey, C. Mazzoleni, and D. W. Fahey, Aerosol Science And Technology Vol. 44 , Iss. 8,2010)

For color ratio the CAMB\_QC\_color\_ratio script is run to plot the size dependent color ratio as a function of time. If any issues are spotted that period is examined in more detail and the periods to be excluded are entered into the sp2\_qc\_intervals\_locnX table with the code 1. QC code definitions are stored in the sp2\_qc\_code\_definitions table.

Site-based interferences (such as contamination from local sources) are also identified and periods to be excluded from analysis are entered into the sp2\_qc\_intervals\_locnX.

# Time Series Data

## Relevant module:

* CABM\_distribution

## Relevant class:

* CABM\_SP2\_time\_interval (a subclass of SP2\_time\_interval)

## Relevant scripts:

* CABM\_compile\_time\_interval\_data
* CABM\_compile\_interval\_distribution\_data

## Database tables:

* sp2\_time\_intervals\_locnX
* sp2\_time\_interval\_binned\_data\_locnX

The sp2\_time\_intervals\_locnX table holds the total mass, the mass uncertainty\*, the total number, and the total volume sampled for each time interval as well as the calibration, location, and instrument IDs. It also has a field for the fraction of the total mass distribution sampled by the instrument.

The sp2\_time\_interval\_binned\_data\_locnX table contains size binned data for each interval.

\* The uncertainty in total mass arises from the calibration uncertainty and from the uncertainty in the correction factor applied when using Aquadag as an external standard

## Details:

*CABM\_compile\_time\_interval\_data script*

The CABM\_compile\_time\_interval\_data script is used to populate the sp2\_time\_intervals\_locnX and the sp2\_time\_interval\_binned\_data\_locnX tables. For a specified period of interest, instrument, and location, this script integrates the rBC mass and number for fixed time intervals (default is 1 minute). The instrument location, instrument ID, the calibrations used, the interval start and stop times, the total interval mass, the total interval number and the volume of air sampled are all recorded in the sp2\_time\_intervals\_locnX table. In addition, the rBC measurements for the interval are binned (default bin width is 10nm) and the binned data is entered into the sp2\_time\_interval\_binned\_data\_locnX table.

The mass and number values for both tables are calculated for a fixed rBC diameter range which can be specified in three different ways:

1. The default behavior is to use the largest and smallest calibration masses as the upper and lower size limits. This ensures that only particles within the calibration range are considered. One drawback is that the bin limits can be a bit awkward (e.g. 63-73 nm vs 60-70 nm)
2. The user can specify upper and lower size limits using the optional –mn and/or –mx arguments. (Note: An electrical issue with SP2 #17 prior to 2012 gave anomalous signals at masses above ~240nm. There for for SP2#17 prior to 2012 an upper limit to the particle diameter should be set.)
3. The user can choose to extrapolate the calibration curve (see details below).

Note: The user can specify a lower size limit *and* extrapolate the calibration to larger sizes.

Note: The total mass and number values only include particles within the specified size range.

*Option: Extrapolating the calibration.*

In many cases the sp2 is not calibrated throughout its entire detection range. For example, the saturation limit might be at an rBC mass of 40fg, but the largest calibration mass is only 30fg. In this case to get the mass of larger particles, we must extrapolate the calibration curve. This is not done by default, but can be enabled with the –e argument in the CABM\_compile\_time\_interval\_data script. If the extrapolation isn’t done, the missing mass can be accounted for by the fit of the distribution, but in this method, over short intervals, the impact of individual large particles is lost. However, extrapolating outside the calibration curve is risky, especially if a quadratic fit is used. Ultimately the choice is made by the user after considering their needs and objectives. Note that if this option is enabled the calib\_extrapolated field in the sp2\_time\_intervals\_locnX table will be set to 1 (0 means no extrapolation was done)

*CABM\_compile\_interval\_distribution\_data script*

The sp2 has a limited size range over which it can detect rBC particles. For a 4 channel instrument such as #17, this is ~60-250nm volume equivalent diameter. For an 8-channel instrument it is ~60-800nm VED. Regardless of whether the calibration is extrapolated or not, smaller particles are lost in the baseline noise and larger ones saturate the detector. To account for mass outside the detection range, it is possible to fit the distribution of measured particles and estimate what fraction of the mass is missing. The CABM\_compile\_interval\_distribution\_data script does this by compiling the binned data from short intervals over a longer period so that a reliable size distribution can be achieved (over a 1 minute interval the size distribution is often very noisy, especially in clean sampling conditions). This compiled size distribution (default is 24 hours) is then fit and the fraction of the mass being measured is written to the sp2\_time\_intervals\_locnX table. A QC check is performed when the script is run and any unacceptable periods are left out of the calculation.