CABM SP2 analysis code documentation

# Introduction

Include general description of what code does and put in database map here

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

# Location and Instrument information

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

## Relevant module:

* None

## Relevant scripts:

* None

## Database tables:

* sp2\_instrument\_info
* sp2\_locations

## Process:

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.

# Calibration data

This section concerns information on the SP2 incandescence calibrations. This calibration is done using an external standard and relates incandescence peak height to refractory black carbon mass.

## 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.

## Process:

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:

mobility\_diameter signal #calibration is for SP2#58 at Alert 20161113

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 config files that is relevant when looking at mass and number concentration.

## Process:

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.

## Process:

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.

## Process:

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

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

* housekeeping yag power (any apparent drops in yag power need to be confirmed by looking at the color ratio)
* 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 (mass uncertainty), total number, and total volume sampled in a given 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.

## Process:

The CABM\_compile\_time\_interval\_data script is used to populate the sp2\_time\_intervals\_locnX table. It calculates interval data between a start time and end time, for a given instrument and location number. It has optional arguments to take in a size bin width (default is 10nm), and an interval length (default is 1 min). It also has an optional argument for extrapolating the calibration curve which is explained in more detail later on.

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. 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.

*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 should be accounted for by the fit of the distribution, but 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.