**CABM SP2 Data analysis**

The information in the following paragraphs is also shown schematically in Figure 1.

All sp2 data is stored in a relational database (MySQL). The code for the analysis is written in Python.

*Analysis summary:*

The SP2 data has six principle components. At the highest level are the instrument used to gather the data and the location that the instrument was deployed in. The sp2\_instrument\_info table contains an entry for each instrument and the sp2\_locations table contains an entry for each location in the network. All sp2 data is associated with both an instrument and a location.

The next four components are listed below. A database table is created and populated for each component. There is also a python module with the methods required for reading the data for each of these components from the raw files and adding them to the appropriate database table. Since the CABM dataset is so large, the sp2\_single\_particle\_data (individual records) and sp2\_hk\_data (housekeeping) tables have been separated by location.

* The individual particle records
  + When collecting data, the sp2 writes the full, time-dependent, detector signal for every channel to a binary file. For rBC mass and number analysis, only the peak amplitudes from the incandescence channels are needed. These are written to the database table for the high and low gain, and narrow and broad band, channels. Also required is the time that it took for the particle to be detected as this allows the sampled volume to be calculated.
* The instrument housekeeping data
  + These are the self-diagnostic data that the instrument collects. This information is used to exclude from the analysis any periods when the instrument was not functioning properly. The recorded sample flow rate is also used, in conjunction with the time it took to detect a particle, to calculate the sampled volume.
* The instrument configuration data
  + These are data on the operational configuration of the instrument. A common operational configuration is to record data from only a set fraction of the detected particles. For example, it is common to record data from only 1/10 detected particles. It is necessary to know about the skipped particles in order to accurately calculate a mass or number concentration.
* Calibration data
  + The sp2\_calibrations table contains information on each calibration performed. Each entry is associated with a location, an instrument, and a channel. It also includes a calibration date and a calibration material (e.g. Aquadag). The sp2\_calibration\_points table contains the individual values for mobility diameter and incandescence signal amplitude and each entry in this table is associated with a calibration. Fitting the points from a calibration allows the fit coefficients to be written to the sp2\_calibrations table. These are used to calculate the rBC mass in an individual particle.

Using all four components in the list above, the rBC mass and number concentration can be calculated for any desired time interval. Data for each interval is entered into the sp2\_interval\_data table. Again, since the CABM dataset is so large, the interval tables have been separated by location.

For any given interval, the rBC mass and number distribution can also be determined. The distribution data is written to sp2\_distribution\_data tables. Each entry in this table has an associated interval ID, upper and lower bin limits, and the rBC mass and number for that bin.

The final table is the sp2\_distribution\_fits table. For each distribution in the distributions table, a lognormal fit can be performed. The fit coefficients as well as the fraction of the fit distribution within the instrument detection limits is recorded in this table. Each entry is associated with a particular interval (typically a 24 hour interval) but can be used to correct the measured mass for shorter intervals within that time span.

****

Figure CABM SP2 data analysis and database structure