# SODA-2 Data Processing Software
NCAR/MMM
July 2013

## Introduction

The System for OAP Data Analysis (SODA) is a software package designed to process and view raw image data from Optical Array Probes (OAPs).  The version described here, SODA-2, is for use with modern probes such as those manufactured by DMT and SPEC, or probes using custom hardware interfaces such as the NCAR Fast-2DC.  SODA-2 supports a variety of data formats, including the native raw data formats from DMT, SPEC, SEA M-200/300, and NCAR/RAF.  Older probes, such as the PMS 2D-C and 2D-P, should be processed with the original SODA-1 package which can account for the more complicated buffer timing used with those instruments.

During initial processing SODA-2 creates particle distributions binned by size, area ratio, and aspect ratio, makes corrections for shattering and out-of-focus particles, organizes housekeeping data, and creates image links to locations in the raw data files.  All of this data is saved in a new file which can be used for further analysis. After processing, SODA-2 can be used to evaluate probe performance, compute cloud/precipitation parameters, and export data to image sequences or to netCDF files.  This manual describes the installation, features, and processing details for SODA-2.  For more information, please contact the developers or refer to the open-source code at https://github.com/abansemer/soda2.

## Installation

SODA-2 requires the Interactive Data Language (IDL) software package, either as a full IDL distribution or the freely available IDL Virtual Machine.

Using a full IDL distribution:
1. Get the latest version of the code from the SODA-2 repository:
   a. Using Git: "git clone https://github.com/abansemer/soda2"
   b. By download: Go to https://github.com/abansemer/soda2, download the zip file, and unzip into a directory on your local machine.
2. If using the IDL Desktop Environment, add the code directory to IDL's search path using the menus (Window/Preferences/IDL/Paths/Insert…).  If running IDL via command line modify the "!path" system variable:
       IDL> !path = !path + ':/my_programs/soda'
   This command can be run automatically by adding it to the IDL startup script.
3. Type "soda2" at the command prompt to start the processing software.

Using the IDL Virtual Machine:
1. Start the virtual machine.
2. Choose the file "soda2.sav" when prompted. The soda2.sav file is available by FTP at ftp.ucar.edu/pub/mmm/bansemer/soda.

## Data Processing

Type "soda2" at the IDL command prompt to start the processing software, or choose the file "soda2.sav" when prompted by the IDL Virtual Machine. The data processing window appears when the software is started. To begin processing data:

1. Under the "OAP Data" section, click on "Add file..." and select the raw OAP files to be processed. Multiple files can be selected using the [Shift] or [Ctrl] keys, but they should all be for the same instrument and for the same flight.
2. Enter the source of true air speed (TAS) for the flight, which is required for an accurate estimate of the probe's sample volume. This file can be in several formats:
   a. An ASCII file with time (UTC seconds) in the first column and TAS (meters/second) in the second column. Space, tab, or comma delimiters are accepted.
   b. An IDL .sav file, which should have a single structure named 'data' containing the variables 'time' (in UTC seconds) and 'tas' (in meters/second). The time and TAS records in this file should match what will be entered into the 'Processing Options' fields.

   Enter a fixed TAS if no other source is available.

3. Select processing options:
   - Project name: Enter any project identifier to be saved with the data.
   - Date: The flight date in format MMDDYYYY.
   - Start/Stop time: The time interval to process from the raw data. Enter in HHMMSS format. A smaller interval will save memory and disk space and speed processing time.
   - Probe: Select the probe to be processed from the dropdown list. If a new probe needs to be added modify the soda2_probespecs.pro

file.

- Adjust the bin endpoints if desired. Clicking the 'Default' button will load a recommended bin distribution based on the currently selected probe's resolution.
- Averaging time: Enter an averaging interval. Intervals shorter than one second are possible but will require more memory to store the particle distributions.
- Clock correction: If the data acquisition computer clock was not set correctly then a correction can be applied. The time entered will be added to the recorded time.
- Shatter correct: Apply a shattering correction based on particle interarrival times. The method is described in Field et al. (JTECH, 2006).
- All-in: Only process particles that are fully imaged and do not touch an edge of the array.
- Water Processing: Apply stricter roundness criteria to limit processing to particles that may be liquid water. Also apply the Korolev (JTECH, 2007) size correction.
- Stuck Bit Correct: Look for diodes that are either continuously on or continuously off and correct for the errors by using neighboring diodes.
- Area Size: Use equivalent area diameter as a sizing method. Please note that the effective sample area for this method is not well known, and not implemented in the software.

4. Select output options:
- IDL "sav" file: Select to save data in IDL's proprietary format. This file records all processing options, processed data, and housekeeping data. This file is also required to view data with the IDL browser or to export to other data formats. See the file format section of this document for detailed information about this file.
- Particle-by-Particle file: Creates an ASCII or netCDF particle-by-particle file with detailed information on each particle processed. This file may grow to be very large so a short time window defined by the start/stop time fields is recommended.
- Housekeeping: Some raw data files record end diode voltages, laser performance parameters, and other housekeeping data. This option will output a separate "sav" file containing these variables.
- Output directory: The location where the output files will be written.

5. Click "BEGIN PROCESSING" to process the data.

Processing will take several minutes to hours depending on the amount of data. Once completed, new files containing the processed data will be saved with the naming convention "date_starttime_probetype.dat", "date_starttime_probetype.txt", or "date_starttime_probetype.nc".

## Data Reprocessing

Settings from a previously processed ".dat" file can be loaded under the "File / Load Settings" menu option. The processing options can then be changed with the graphical interface, and the

data reprocessed.  The old file will be overwritten unless a different output directory is selected.

Alternately, the IDL files can be reprocessed via command-line or script.  Modifications are directly applied to the options structure (see file format information at the end of this document) and then reprocessed.  For example:

IDL> restore, 'myfile.dat'
IDL> data.op.rate=1                    ;Change the sampling rate
IDL> soda2_process_2d, data.op    ;Reprocess the data

## Browsing Processed Data

Select "Other Actions / Browse Data" from the main window to enter the data browser.  The browser may also be accessed directly from IDL command line by typing "soda2_browse".  Load a processed (*.dat) file under the "File / Load" menu to begin browsing.

*Navigating through the data:*
The first 3 tabs (Distributions, Particles, and Timing/Diodes) display data for a single time period.  To move forward in time, left-click anywhere on the main plot.  To move backward in time, right-click on the main plot.  The scroll wheel on a mouse may also be used to move forward or backward more rapidly.  A blue indicator line shows the current position in a small concentration plot at the bottom of the screen.  Left-click on this plot to directly access a new time period.  A new time may also be typed into the text box at the bottom-left corner of the window in either 'hhmmss' or seconds-from-midnight format.  Click the 'HMS' or 'SFM' label to toggle formats.

*Distributions tab:*
The default tab shows the normalized particle size distribution, the area ratio distribution, and a colorized contour composite of these two distributions.  Computed bulk values such as total number concentration, ice water content, and mean diameter for the current time period are displayed.

*Particles tab:*
The 'Particles' tab displays the images of the particles recorded for each time period.  *The original raw data files must be accessible in order to view this screen; the images are not saved in the processed file.*  Only images that fit on the screen are displayed.  To see more images click on the arrow buttons below the displayed images.

*Timing tab:*
This tab shows the interarrival time and diode histograms.  Ideally, the interarrival time plot should have a shape resembling a Poisson distribution.  The diode histogram the total number of shadows recorded during each time period.  If sufficient particles were recorded it should be

a flat line.  The distributions of interarrival time and diode histograms for the entire flight are also displayed in a red dashed line.

*Time series tab:*
This window displays time series plots of derived parameters and housekeeping data.  Two plotting windows are available, and the value to be plotted on each is changed with the drop-down menus.  The start and end times can be adjusted with the mouse by click-dragging a box on either the data plots or the reference plot at the bottom of the screen.  The green and red indicators on the reference plot show the current range.
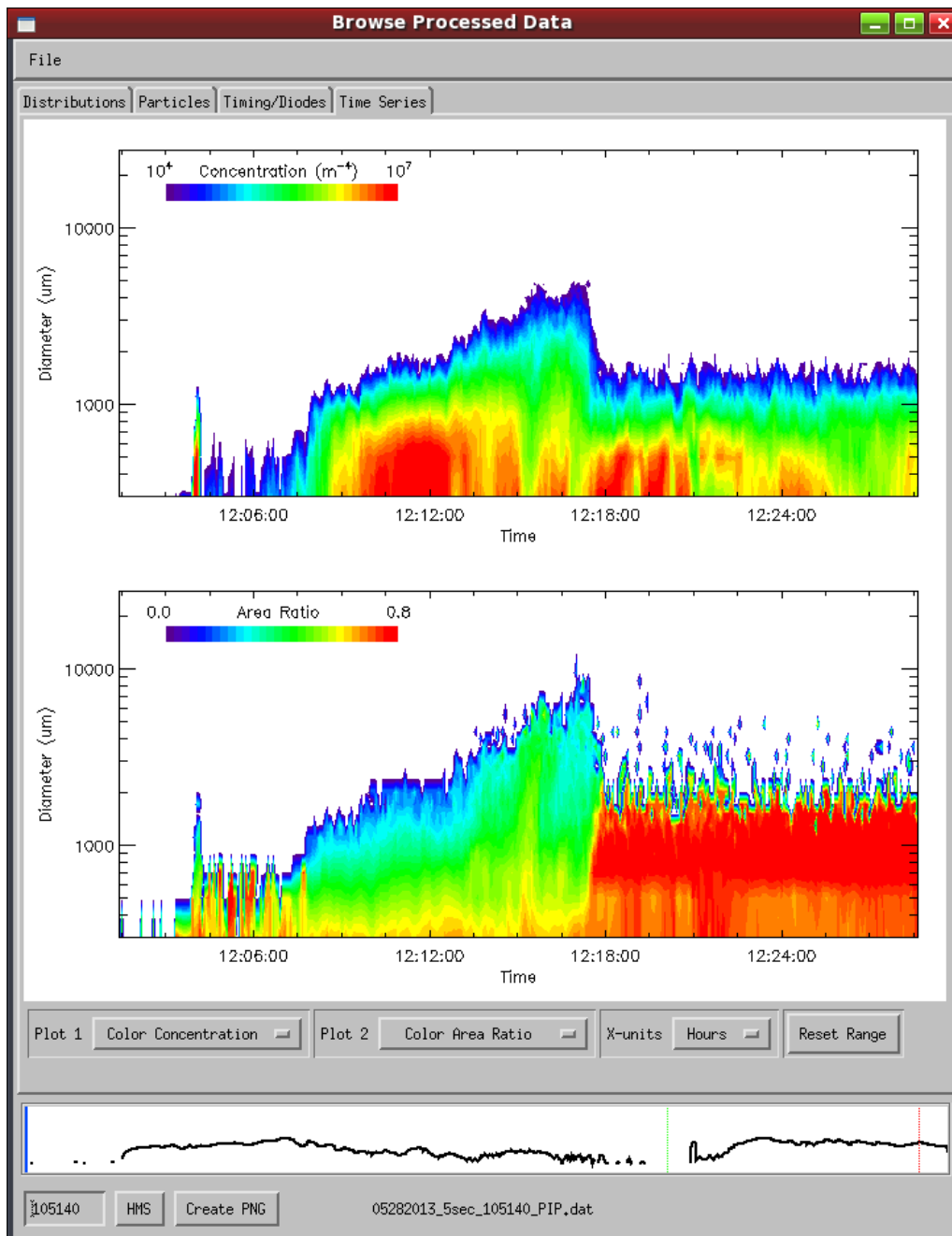
*Saving plots:*
Click the "Create PNG" button on any screen to save the current plot(s) to a PNG image. It will be saved in the directory where the processed file is located.

*File properties:*
The options used to process the file are displayed from the "File / Properties" menu

## Exporting Data

The processed data and images can be exported to netCDF (data) and PNG (images) for compatibility with Matlab, Python, or other software packages. Select "Other Actions / Export Data" from the main window to open the data export menu. Add processed (*.dat) files to the list for export to the netCDF format or to PNG images.

The netCDF files will contain particle size distributions, counts, interarrival time distributions, and a variety of computed parameters such as IWC, mean diameter, and total area. See the program soda2_export_ncdf.pro for more options.

PNG files each contain one minute of sample images, with one image buffer (roughly 1000 slices) shown for each time interval that was processed. Use the command-line version to output all images or to specify start and/or stop times, for example:
    IDL> soda2_imagedump, 'myfile.dat', /all, starttime=hms2sfm(130000)

## Processing Details

*Part 1:  Particle Sizing and Sample Area*

Particles can be measured by several methods, including circle-fit, sizing across the array (x-size) and sizing with the airflow (y-size).

The circle-fit method is the default sizing method. It fits the smallest possible circle around a particle image and uses the diameter of that circle as the diameter of the particle. This method is used for its computational efficiency, as well as its ability to produce a clean comparison of the area of particle to the area of the circle. This "area ratio" is used for subsequent particle rejection, roundness detection, and may also be used for computing such parameters as fall velocity and optical extinction.

The x-size and y-size methods measure the maximum difference between shaded pixels in their respective directions. X-size may be useful for spinning disc calibrations, or for any time where the probe's timing did not match the particle speed resulting in distorted images.

Under 'water' processing a sizing correction is applied following Korolev (2007). This correction is based on the size of the Poisson spot seen when imaging liquid particles, and indicates magnification of a particle due to its position in the depth of field. If a Poisson spot is detected its area is measured and compared to the area of the complete particle. The ratio of these two areas is used to find a correction factor, which reduces the size measurement to its expected pre-magnification value.

In all sizing methods, partially imaged particles which touch either or both ends of the diode array are allowed by default. The sample area of the probe is computed following the "reconstruction" method in Equation 17 of Heymsfield and Parrish (1978). If the user elects to reject partially imaged particles, the sample area is computed following Equation 4 of the same reference.

*Part 2: Shattering Corrections*

Large particles that impact on the forward surface of a probe arm can break into many pieces and then be imaged by the probe. This results in an overestimate of the concentration of small particles. Since these small particles appear in clusters, the time between neighboring particles, or interarrival time, may be used to detect suspected shattering events. SODA-2 corrects for shattering events using the method described in Field, et al. (2006).

*Part 3: Particle Rejection Criteria*

The particle rejection criteria in SODA-2 serve two purposes, to distinguish between "round" and "irregular" particles if water processing is enabled, and to remove image artifacts. Image artifact rejection is based on the area ratio. The rejection criteria details are as follows:

Under default processing particles are rejected if:
      Area ratio < 0.1
      Particle size outside of size-bin range

Under 'water' processing particles are rejected if:
      Area ratio < 0.4
      Area ratio < 0.5 for particles 10 pixels or larger
      Size greater than 6mm
      Corrected particle size outside of size-bin range

## Processed Data File Format

Processed data is saved in a raw data file using IDL's save/restore format. This file can be used directly for analysis beyond the capabilities of the Data Browser. Type "restore, xxxx_xxxx_xxx.dat" at the IDL command prompt to load the data into the IDL workspace. Once loaded, all data will be available in a structure named "data". The structure has a number of tags with processed information, and a sub-structure named "op" containing processing

options.  Descriptions of the data variables are in the table below:

| Variable | Description |
| --- | --- |
| OP | A sub-structure containing the processing options. |
| OP.FN | The original filenames entered into the GUI. |
| OP.DATE | Date string entered into the GUI. |
| OP.STARTTIME | Start time (UTC seconds). |
| OP.STOPTIME | Stop time (UTC seconds). |
| OP.FORMAT | Data acquisition format. |
| OP.SUBFORMAT | Data acquisition sub-format. |
| OP.PROBETYPE | Probe type (2DC, 2DP, etc.). |
| OP.RES | Probe resolution (microns). |
| OP.ENDBINS | Size bin endpoints (microns). |
| OP.ARENDBINS | Area ratio bin endpoints. |
| OP.RATE | Averaging interval. |
| OP.SMETHOD | Particle sizing method used ('fastcircle', 'xsize', etc.). |
| OP.PTH | IDL .sav file which contains true air speed data. |
| OP.PARTICLEFILE | Flag for creating a particle-by-particle file. |
| OP.INTTIME_REJECT | Flag for applying interarrival time rejection. |
| OP.RECONSTRUCT | Flag to allow particles that touch an edge of the array. |
| OP.STUCKBITS | Flag to turn on stuck bit detection and correction. |
| OP.WATER | Flag to use 'water' processing algorithm. |
| OP.FIXEDTAS | Fixed air speed to use if pthfile is unavailable. |
| OP.OUTDIR | Output directory. |
| OP.PROJECT | Project name entered in GUI. |
| OP.TIMEOFFSET | Filenames that pass data integrity test. |
| OP.ARMWIDTH | Probe arm width (cm). |
| OP.NUMDIODES | Number of diodes. |
| OP.GREYTHRESH | Threshold on which to size particles for grayscale probes. |
| OP.PROBEID | Probe ID for raw files that contain multiple probes. |
| TIME | Time in seconds from midnight UTC on the date specified in 'DATE'. |
| DATE_PROCESSED | Date and time of processing. |

| | |
|---|---|
| TAS | True airspeed used in concentration computation. |
| MIDBINS | Size bin mid-points. |
| SPEC1D | Counts per size bin in a [time, size bin] array. |
| SPEC2D | Counts per bin in a [time, size bin, area ratio bin] array. |
| CONC1D | Normalized particle concentration in a [time, size bin] array (#/m3/m). |
| ACTIVETIME | Probe activity time (seconds). |
| SA | Sample area of each size bin (m2) |
| INTSPEC_ALL | Counts per interarrival bin in a [time, interarrival bin] array for all (accepted+rejected) paritcles. |
| INTSPEC_ACCEPTED | Counts per interarrival bin in a [time, interarrival bin] array for accepted particles. |
| INTENDBINS | Interarrival bin end-points. |
| INTMIDBINS | Interarrival bin mid-points. |
| COUNT_ACCEPTED | The number of particles accepted. |
| COUNT_REJECTED | The number of particles rejected in a [time, reason] array. Particles may be rejected for the following reasons: |
| | 0: Unused |
| | 1: Area ratio too low. |
| | 2: Interarrival time below threshold. |
| | 3: Particle size out of size bin range. |
| | 4: Particle touches edge of array. |
| | 5: Unused |
| COUNT_MISSED | The number of particles that were not recorded. |
| CORR_FAC | Correction factor for interarrival time correction. |
| POISSON_FAC | Coefficients for the double-Poission interarrival time fit. |
| INTCUTOFF | Interarrival time threshold for accepted/rejected particles. |
| POINTER | Pointer to each buffer in the raw data files. |
| IND | Time index into which each buffer starts. |
| CURRENTFILE | File number for each buffer/pointer. |
| NUMBUFFSACCEPTED | Number of accepted buffers. |
| NUMBUFFSREJECTED | Number of rejected buffers. |
| DHIST | Detector shadow counts in a [time, n_diodes] array. |