SODA-2 Data Processing Software

NCAR/MMM
October 2023

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, are also supported but should be processed and verified 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, or the slide presentation available at this link.

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: issue the command **git clone https://github.com/abansemer/soda2**
    b. By download: Go to https://github.com/abansemer/soda2, click on the green 'Code' button, download the zip file, and unzip into a directory on your local machine.
2. If using the IDL Desktop Environment, add the SODA-2 directory to IDL's search path using the menus under *IDL/Settings/IDL/Paths/Insert*.
3. If running IDL via command line modify the ***!path*** system variable to include the SODA-2 directory:

    IDL> !path = !path + ':/my_programs/soda2'

This command can be run automatically by adding it to the IDL startup script.

4. Type ***soda2*** at the IDL command prompt to start the processing software.

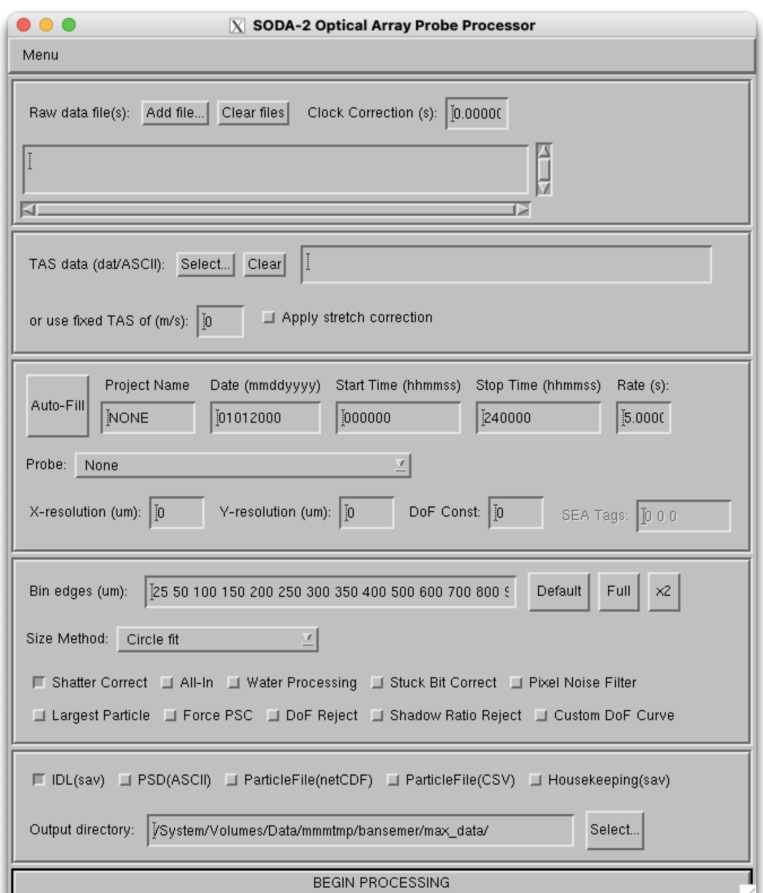
Using the IDL Virtual Machine:

1. Install the virtual machine. It is freely available as part of the trial IDL installation at https://www.nv5geospatialsoftware.com/
2. Start the virtual machine.
3. Choose the file *soda2.sav* when prompted. The *soda2.sav* file is available online here.

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. **Select raw data files:**
In the Raw Data section, click on **Add file...** and select the raw OAP files to be processed. These files are usually named *baseYYMMDDhhmmss.2DS* for SPEC instruments, *Imagefile_01CIP*.raw* for DMT instruments, or *YYYY-MM-DD-hhmm.sea* for data recorded on SEA data acquisition systems. 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. If the raw data require a time offset adjustment, enter the offset time in seconds in the **Clock Correction** box.



2. **Enter TAS:**
Enter the source of true air speed (TAS) for the flight in the **TAS Data** box, which is recommended for computing 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 Probe Options start/stop time fields.

Enter a fixed TAS if no other source is available. Select the **Apply stretch correction** option if the aircraft TAS and the probe slicing TAS were different, leading to stretched or compressed particles in the airflow direction.

3. **Select probe options:**
   a) Click **Auto-Fill** to interrogate the selected raw data files and automatically fill in the date and start/stop times. This also limits the probe selection list to instruments that have a compatible raw data format.
   b) Project name: Enter any project identifier to be saved with the data.
   c) Date: The flight date in format MMDDYYYY, if not entered correctly by **Auto-Fill**.
   d) Start/Stop time: The time interval to process from the raw data. Enter in HHMMSS format. A shorter time interval will save memory and disk space and speed processing time.
   e) Rate: Enter an averaging interval for the time series data. Intervals shorter than one second are possible but will require more memory to store the particle distributions.
   f) Probe: Select the probe to be processed from the dropdown list. If a new probe needs to be added modify the list in *soda2_probespecs.pro*.
   g) Make any necessary adjustments to the X-resolution, Y-resolution, or Depth-of-Field constant based on laboratory calibrations. The SEA tag numbers can also be adjusted here if they do not match the original configuration.

4. **Select processing options:**
   a) Adjust the bin edge values if desired. The **Default** button will load a recommended bin distribution based on the currently selected probe's resolution. The **Full** button will load a linear distribution of bins centered on the current X-resolution value, one bin for each element in the diode array. The **x2** button will do the same thing as the **Full** button, but doubles the number of bins to cover particle sizes up to twice the array width.
   b) Select a particle sizing method which will be used for the particle size distributions.
   c) Shatter correct: Apply a shattering correction based on particle interarrival times. The method is described in Field et al. (JTECH, 2006).
   d) All-in: Only process particles that are fully imaged and do not touch an edge of the array.
   e) Water Processing: Apply stricter roundness criteria to limit processing to particles that may be liquid water. Also apply the Korolev (JTECH, 2007) size correction.
   f) Stuck Bit Correct: Look for diodes that are either continuously on or continuously off and correct for the errors by using neighboring diodes.
   g) Pixel Noise Filter: Apply a filter to remove speckle noise, for CIP-Gray instruments only.
   h) Largest Particle: If multiple particles exist in a frame, only use the largest one.
   i) Force PSC: Apply the Korolev (2007) Poisson-spot size correction to all particles, regardless of shape or size.

j) DoF reject: For instruments that record a depth of field flag, reject particles where this flag is activated. On grayscale probes, reject particles that do not contain at least one pixel at the darkest available shadow level.
k) Shadow Ratio Reject: On grayscale probes or the SEA-1D2D, reject particles were fewer than half of the imaged diodes are at the darkest available shadow level.
l) Custom DoF Curve: Opens a dialog to load a file with a custom depth of field table instead of using the customary formula.

5. **Select output options:**
   a) 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.
   b) PSD file: Creates an ASCII file containing the particle size distribution results.
   c) ParticleFile: Creates an ASCII (comma separated values) 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.
   d) Housekeeping: Some raw data files record end diode voltages, laser performance parameters, and other housekeeping data. This option will create a separate IDL "sav" file containing these variables.
   e) Output directory: The directory where all output files will be written.

6. 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 *Menu/Browse Data* from the main SODA-2 window to load 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.

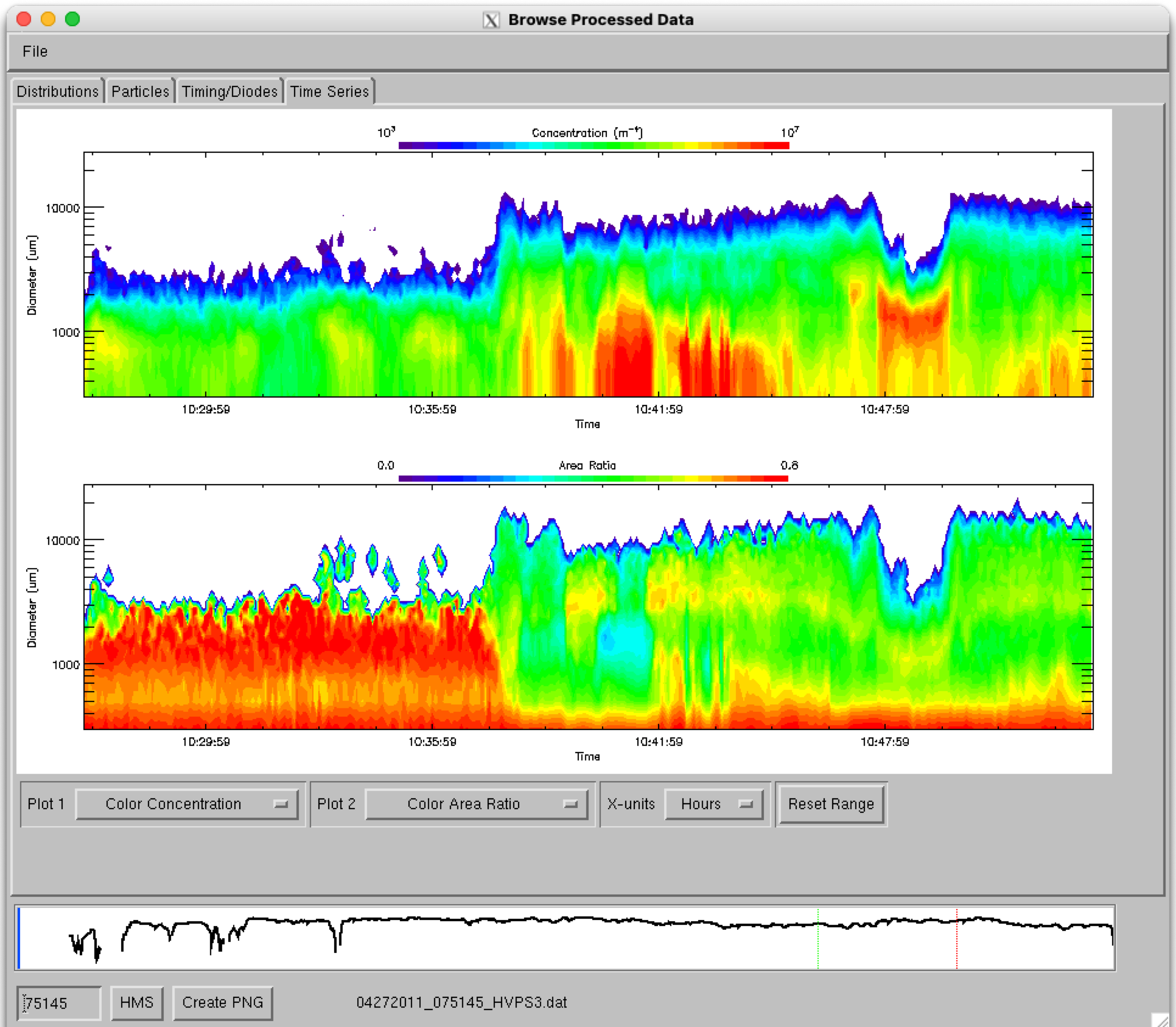
Distributions tab:
The default tab shows the normalized particle size distribution, the area ratio distribution, and a colorized contour composite of these two distributions. Computed bulk values such as total number concentration, ice water content, and mean diameter for the current time period are 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 (top panel) should have a shape resembling a Poisson distribution. The diode histogram (bottom panel) shows the total number of shadows recorded during each time period. If sufficient particles were recorded, then the histogram should be a relatively flat line. The distributions of interarrival time and diode histograms for the entire flight are also displayed with 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 or ASCII-CSV (size distributions) and PNG (images) for compatibility with Matlab, Python, or other software packages. Select *Menu/Export Data* from the main SODA-2 window to open the data export menu. Add processed (*.dat) files to the list for export.

The netCDF files will contain particle size distributions, counts, interarrival time distributions, and a variety of derived bulk 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

1. Particle Sizing and Sample Area:
Particles can be measured by several methods, including circle-fit, sizing across the array (x-size), sizing with the airflow (y-size), area equivalent sizing, and slice-width sizing (Lx).

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 distance 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 stretched or compress images in the airflow direction. Similarly, Lx sizing defines particle size by the maximum distance between shaded pixels on any individual slice of a particle (but not of the entire particle). This is used for situations where particles have a skewed appearance from transiting through the laser in an off-axis direction.

Area equivalent sizing defines particle size as the diameter of a circle which would have the same shadowed area as the particle image. 1D and 2D Emulation sizes replicate legacy instrument sizing methods such as the 260X and 260Y.

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 if the center of the particle is deemed to be within the array. The sample area of the probe is computed following the *center-in* method described in Heymsfield and Parrish (1978). If the user elects to reject partially imaged particles (*All-in* option), the sample area is computed following Equation 4 of the same reference.

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). This method requires at least 100 particles per time period, so it is recommended to use a sufficiently long averaging time (in the *Rate* box on the SODA-2 main screen) to ensure that enough particles are available to activate the correction.

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 is outside of size-bin range
- Depth of field criteria not met (if enabled by user with *DoF_reject* setting)
- Particle center is deemed to be outside the array
- Particle touches an edge of the array (if enabled by user with *All-in* setting)

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 is outside of size-bin range
- Depth of field criteria not met (if enabled by user with *DoF_reject* setting)
- Particle center is deemed to be outside the array
- Particle touches an edge of the array (if enabled by user with *All-in* setting)

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. Libraries are available for reading these files directly into Python (scipy.io.readsav). 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 *data.op* containing processing options. Descriptions of the data variables:

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.PROBEID	Probe ID for raw files that contain multiple probes
OP.SHORTNAME	Probe name used when constructing filenames
OP.RES	Probe resolution across the array (microns)
OP.ARMWIDTH	Distance between probe arms (cm)
OP.NUMDIODES	Number of diodes in the image array
OP.WAVELENGTH	Laser wavelength (m)
OP.ENDBINS	Size bin endpoints (microns)
OP.ARENDBINS	Area ratio bin endpoints (unitless)
OP.RATE	Averaging interval (seconds)
OP.SMETHOD	Particle sizing method used (‘fastcircle’, ‘xsize’, etc.)
OP.PTH	IDL .sav file or ASCII file which contains true air speed data
OP.PARTICLEFILE	Flag for creating an ASCII particle-by-particle file
OP.NCDFPARTICLEFILE	Flag for creating a netCDF particle-by-particle file
OP.INTTIME_REJECT	Flag for applying interarrival time rejection

OP.EAWMETHOD	Equivalent array width method ('centerin', 'allin')
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.GREYTHRESH	Threshold on which to size particles for grayscale probes
OP.IGNOREDEADTIME	Process without considering probe overload time
OP.ASCIIPSDFILE	Flag for creating an ASCII particle size distribution file
OP.CLUSTERTHRESH	Flag to enable a cluster-based shattering algorithm (experimental)
OP.SAVFILE	Flag for creating a .sav file
OP.RAKEFIX	Enable correction for particle raking
OP.KEEPLARGEST	Only measure the largest particle in a frame
OP.JUELICHFILTER	Remove speckle noise on CIP-Gray probes
OP.SEATAG	Tags for reading from SEA files [image_tag, tas_tag, elapsedtime_tag]
OP.YRES	Probe resolution in the airflow direction (microns)
OP.APPLY_PSC	Flag to apply Korolev (2007) size correction on all particles
OP.DOFCONST	Depth of field constant for computing sample area
OP.DOFREJECT	Flag for rejecting particles for instruments that record a depth of field flag (or grayscale instruments)
OP.STRETCHCORRECT	Enable stretch correction when there is a mismatch between the aircraft TAS and the probe slicing TAS
OP.STRICTCOUNTER	For CIP-Gray only, enable strict particle rejection when the particle counter goes awry
OP.ACTIVETIMEMISSED	Flag to compute active time from the number of missed

	particles (experimental)
OP.DIODERANGE	Lower and upper index of active diodes, for cropping the array of inactive diodes
OP.CUSTOMDOF	Custom depth of field values
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
PROBETAS	True airspeed used by the probe for slicing rate
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
SPEC2D_ASPR	Counts per bin in a [time, size bin, aspect 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 (seconds)
INTMIDBINS	Interarrival bin mid-points (seconds)
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: ‘Water’ criteria not met
	6: ‘Ice’ criteria not met
	7: Depth of field flag rejection
COUNT_MISSED	The number of particles that were not recorded.
MISSED_HIST	Histogram of the number of missed particles for each time period
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
HIST3D	Experimental
SPEC2D_ORIENTATION	Particle orientation counts in a [time, size bin, orientation bin] array. Orientation bins are 10-degrees each
ORIENTATION_INDEX	Orientation index in a [time, size bin] array
HOUSE	A substructure containing housekeeping data, when available
PBPSTARTINDEX	The index of the first particle in the particle-by-particle files for each time period

Particle-by-particle Data File Format

Particle-by-particle data variable descriptions and units are available directly in the header of the ASCII-CSV particle-by-particle files. They are also available in the variable attributes of netCDF particle-by-particle files. Use any netCDF browser or the command-line ***ncdump*** utility to view them.

VARIABLE	**DESCRIPTION**
TIME	UTC time [seconds]
PROBETIME	Unadjusted probe particle time [seconds]
BUFFERTIME	Buffer time [seconds]
RAWTIME	Raw time [slices or seconds]
REFTIME	Reference time for buffer matching [secnds]
INTTIME	Interarrival time from previous particle [seconds]
DIAM	Particle diameter from circle fit. No Poisson spot size corrections applied [microns]
XSIZE	X-size (across array). No Poisson spot size corrections applied [microns]
YSIZE	Y-size (along airflow). No Poisson spot size corrections applied [microns]
XEXTENT	Maximum x-extent (across array) for all individual slices. No Poisson spot size corrections applied [microns]
ONED	1-D emulation size. Number of latched pixels. No Poisson spot size corrections applied [microns]
TWOD	2-D emulation size. Slice with maximum number of shaded pixels. No Poisson spot size corrections applied [microns]
AREASIZE	Equivalent area size. No Poisson spot size corrections applied [microns]
AREARATIO	Area ratio [unitless]
AREARATIOFILLED	Area ratio with particle voids filled [unitless]
ASPECTRATIO	Aspect ratio [unitless]
AREA	Number of shaded pixels [pixels]

AREAFILLED	Number of shaded pixels including voids [pixels]
PERIMETERAREA	Number of shaded pixels on particle perimeter [pixels]
AREA75	Number of shaded pixels at the 75% (or grey level-3) shading [pixels]
XPOS	X-position of particle center (across array) [pixels]
YPOS	Y-position of particle center (along airflow) [pixels]
ALLIN	All-in flag (1=all-in) [unitless]
CENTERIN	Center-in flag (1=center-in) [unitless]
DOFFLAG	Depth of field flag from probe (1=accepted) [unitless]
EDGETOUCH	Edge touch (1=left 2=right 3=both) [unitless]
SIZECORRECTION	Size correction factor from Korolev 2007 (D_edge/D0). Use to adjust sizes in this file if necessary [unitless]
ZD	Z position from Korolev correction [microns]
MISSED	Missed particle count [number]
PROBETAS	True air speed for probe clock [m/s]
AIRCRAFTTAS	True air speed for aircraft (if available) [m/s]
OVERLOADFLAG	Overload flag [boolean]
PARTICLECOUNTER	Particle counter [number]
ORIENTATION	Particle orientation relative to array axis [degrees]
REJECTIONFLAG	Particle rejection code (see soda2_reject.pro) [unitless]