

HyperInSPACE

HyperInSPACE is designed to provide Hyperspectral In situ Support for the PACE mission by processing automated, above-water, hyperspectral ocean color radiometry collected on automated platforms with Satlantic HyperSAS SolarTracker instruments. Inclusion of additional instrument platforms is in progress.

Author: Dirk Aurin, USRA @ NASA Goddard Space Flight Center dirk.a.aurin@nasa.gov

Acknowledgements: Nathan Vandenberg (PySciDON; <https://ieeexplore.ieee.org/abstract/document/8121926>)

Version

1.0.b; While this version has been substantially updated from 1.0.a to (among other things) better accomodate a standard format into which data from above water radiometry suites other than HyperSAS/SolarTracker can be assimilated, the instrumentation available on the market today is highly varied. Therefore, we would appreciate any feedback from the community regarding the instruments and data formats you are interested in seeing implemented in HyperInSPACE in the future. —

NASA Goddard Space Flight Center (GSFC)
Software distribution policy for Public Domain Software

The HyperInSPACE code is in the public domain, available without fee for educational, research, non-commercial and commercial purposes. Users may distribute this code to third parties provided that this statement appears on all copies and that no charge is made for such copies.

NASA GSFC MAKES NO REPRESENTATION ABOUT THE SUITABILITY OF THE SOFTWARE FOR ANY PURPOSE. IT IS PROVIDED "AS IS" WITHOUT EXPRESS OR IMPLIED WARRANTY. NEITHER NASA GSFC NOR THE U.S. GOVERNMENT SHALL BE LIABLE FOR ANY DAMAGE SUFFERED BY THE USER OF THIS SOFTWARE.

Requirements and Installation

Requires Python 3.X is installed on a Linux, MacOS, or Windows computer. The Anaconda distribution is encouraged. A nice walkthrough can be found here: <https://youtu.be/YJC6ldI3hWk>. The Zhang (et al. 2017) sky/sunglint correction also requires Xarray (instructions here: <http://xarray.pydata.org/en/stable/installing.html>). Utilization of GMAO atmospheric models for use in the Zhang 2017 glint correction and filling in environmental conditions not otherwise provided in field logs will now require a user account on the NASA EARTHDATA server. New profiles can be created here: <https://urs.earthdata.nasa.gov/users/new>

Save this entire HyperInSPACE file repository to a convenient, sensibly named directory on your computer such as Users/yourname/HyperInSPACE/. When Main.py is launched for the first time, data and directories within this location will be updated as described below.

HyperInSPACE is a Main-View-Controller Python package that can be launched in several ways to run the Main.py module, such as by navigating to the program folder on the command line and typing the following command:

```
prompt$ python Main.py
```

The following folders will be created automatically when you first run the program:

- Config - Configuration and instrument files (by subdirectory - auto-created), SeaBASS header configuration files
- Data - Optional location for input and/or output data. Data delivered at the top level directory in the original distribution will be moved here when Main.py is run the first time.
- Logs - Most command line output messages generated during processing are captured for later reference in .log text files here
- Plots - A variety of optional plotting routines are included which create name-appropriate sub- directories (i.e. 'L1C_Anoms', 'L1D', 'L1E', 'L2', 'L2_Spectral_Filter')

(Note: Data, Plots, and Logs directories are not tracked on git.)

Guide

Main Window

The Main window appears once Main.py is launched. It has options to specify a configuration file, input/output directories, ancillary input files (i.e. environmental conditions from a SeaBASS file format), single-level processing, and multi-level processing. For batch processing, pop-up windows from failed (no output) files can be suppressed. No output files are often a normal result of quality control filtering.

The ‘New’ button allows creation of a new configuration file. ‘Edit’ allows editing the currently selected configuration file. ‘Delete’ is used to delete the currently selected configuration file and corresponding auto-created calibration directories (see Configuration). After creating a new configuration file, select it from the drop-down menu, and select ‘Edit’ to launch the Configuration module and GUI. Configuration files are saved in the ‘/Config’ subdirectory of HyperInSPACE.

The ‘Input...’ and ‘Output Data Parent Directory’ buttons allow optional selection of data directories from any mounted/mapped drive. Note that output data sub-directories (e.g. for processing levels) are also auto-created during processing as described below. The parent directory is the directory containing the sub-directories for processing levels (e.g. “/L1A/”, “/L1B/”, etc.) If no input or output data directories are selected, ‘/Data’ under the HyperInSPACE directory will be used by default as the parent data directory.

Ancillary Data files for environmental conditions used in L2 processing (including aerosol optical depth, cloud cover, salinity, water temperature, and wind speed) must be text files in SeaBASS format with columns for date, time, lat, and lon. See <https://seabass.gsfc.nasa.gov/> for a description of SeaBASS format. It is recommended that ancillary files are checked with the ‘FCHECK’ utility as described on the SeaBASS website. They will be interpreted using the included SB_support.py module from NASA/OBPG. In case environmental conditions were not logged in the field, or for filling in gaps in logged data, they will be retrieved from GMAO models as described below.

Configuration

Launch the configuration module and GUI (ConfigWindow.py) from the Main window by selecting a configuration file or creating a new one as described above. This file will be instrument-suite-specific, and is usually also deployment-specific according to which factory calibration files are needed, as well as how the instrument was configured on the platform or ship. Some cruises (e.g. moving between significantly different water types) will probably require multiple configurations to obtain the highest quality ocean color products at Level 2.

Calibration & Instrument Files:

In the ‘Configuration’ window, click ‘Add Calibration Files’ to add the calibration and instrument files (telemetry definition files) from the relevant extracted Satlantic ‘sip’ file (i.e. ‘.cal’ and ‘.tdf’ files). The calibration or instrument file can be selected using the drop-down menu. Enable (in the neighboring checkbox) only the files that correspond to the data you want to process with this configuration. You will need to know which .cal/.tdf files correspond to each sensor/instrument, and which represent light and dark shutter measurements. For example:

- SATMSG.tdf > SAS Solar Tracker status message string (Frame Type: Not Required)
- SATTHSUUUUA.tdf > Tilt-heading sensor (Frame Type: Not Required)
- SATNAVVVVA.tdf > Solar Tracker (Frame Type: Not Required)
- GPRMC_NMEAWWW.tdf > GPS (Frame Type: Not Required)
- SATPYR.tdf > Pyrometer (Frame Type: Not Required)
- HEDXXXAcal > Es (Frame Type: Dark)
- HSEXXXA.cal > Es (Frame Type: Light)
- HLDYYYYA.cal > Li (Frame Type: Dark)
- HSLYYYYA.cal > Li (Frame Type: Light)
- HLDZZZA.cal > Lt (Frame Type: Dark)
- HSLZZZA.cal > Lt (Frame Type: Light)

where UUUU, VVV, WWW, XXX, YYY, and ZZZ are the serial numbers of the individual instruments, which are followed where appropriate by factory calibration codes (A, B, C, D, E, ...).

Selections:

-Add Calibration Files - Allows loading calibration/instrument files (.cal/.tdf) into HyperInSPACE. Once loaded the drop-down box can be used to select the file to enable the instrument and set the frame type. -Enabled checkbox - Used to enable/disable loading the file in HyperInSPACE. -Frame Type - ShutterLight/ShutterDark/Not Required/LightAncCombined can be selected. This is used to specify shutter frame type (ShutterLight/ShutterDark) for dark correction.

For each calibration file:

Click 'Enable' to enable the calibration file. Select the frame type used for dark data correction, light data, or 'Not Required' for navigational and ancillary data. Each radiometer will require two calibration files (light and dark). Data from the GPS and SATNAV instruments is interpreted using the corresponding, required Telemetry Definition Files ('tdf').

As you create your new Configuration, cal/TDF files are copied from their selected locations into the /Config directory within an automatically created sub-directory named for the Configuration (i.e. KORUS.cfg results in /Config/KORUS_Calibration/[calibration & TDF files]).

Level 1A through Level 2 processing configurations are adjusted in the Configuration window. If you are reading this for the first time, the Configuration Window is a good reference to accompany the discussion below regarding processing. *The values set in the configuration file should be considered carefully, as they will depend on your viewing geometry and desired quality control thresholds. Do not use default values without consideration.* Level 1d includes a module that can be launched from the Configuration window to assist with data deglitching parameter selection ('Anomaly Analysis'). More details with citations and default setting descriptions are given below. A separate module to assist in the creation of SeaBASS output files is launched in Level 1E processing, and applied to L1E and L2 SeaBASS output as described below.

Click 'Save' or 'Save As' to save the configuration file. The configuration files are saved to the /Config directory under the HyperInSPACE main directory with a .cfg extension.

Processing Overview

It will be helpful to set your 'Input Data Parent' and 'Output Data Parent' directories from the Main window. As an example, one could use a cruise directory containing RAW HyperSAS data as the Input Parent Directory, and then create another directory named for the cruise to use as the Output Parent Directory when processing from L0. Once L0 processing is complete, change in the input directory to match the output, and files will be automatically sorted by processing level in the automatically created sub-directories. If not selected, these will default to the /Data directory within HyperInSPACE. Your Main window set-up (including configuration file, Input/Output directories, and Ancillary File) will be saved in Config/main.config and reopened the next time you launch Main.py.

Process the data by clicking on one of the buttons for single-level or multi-level processing. A file selection dialogue will appear. Multiple data files can be processed together (successively) by selecting them together in the GUI (Shift- and or Ctrl- click, or Ctrl-A for all). Input files will be checked for match to expected input level (e.g. L1A file input for for L1B processing). Multi-level processing works the same as single-level by processing each input raw file through all levels before moving on to the next raw file. However, it will only continue with a file if the preceding level was created immediately (within 1 minute) prior. In other words, if – due to changes in QA/QC parameterization – a file is entirely discarded at a given level, but an old file of the same name still exists in that directory, it will be ignored, and processing for that file will be terminated for higher levels.

Level 1A - Preprocessing

Process data from raw binary (Satlantic HyperSAS 'RAW' collections) to L1A (Hierarchical Data Format 5 'hdf'). Calibration files and the RawFileReader.py script allow for interpretation of raw data fields, which are read into HDF objects.

Solar Zenith Angle Filter: prescreens data for high SZA (low solar elevation) to exclude files which may have been collected post-dusk or pre-dawn from further processing. *Triggering the SZA threshold will skip the entire file, so don't be overly conservative with this selection.* Further screening for SZA min/max is available in L2 processing.

Default: 60 degrees (e.g. Brewin et al., 2016)

Level 1B

Process data from L1A to L1B. Factory calibrations are applied and data arranged in a standard HDF5 format.

L1B Format: *Data being processed from non-HyperSAS SOLARTRACKER systems should be formatted to match L1B in order to be successfully processed to L2 in HyperInSPACE.* An example of a L1B HDF file is provided in /Data for reference. Datasets are grouped by instrument and contain their data in an array called .data. For example, latitude data is stored in the group ‘GPS’ under LATPOS.data. Data should also contain ‘dtype’, a list of column headers strings (e.g. wavelength, see example file). The following datasets and attributes and groups are generally required:

Root level attributes:

- ‘CAL_FILE_NAMES’ - list of TDF and CAL file names
- ‘ES’/‘LI’/‘LT_UNITS’ - calibrated data units for Es, Li, and Lt
- ‘FILE_CREATION_TIME’ - DD-MMM-YYYY HH:MM:SS in UTC
- ‘PROCESSING_LEVEL’ - ‘1b’
- ‘WAVELENGTH UNITS’ - ‘nm’

Group id: ‘GPS’, attributes: ‘CalFileName’ (TDF file), ‘Frametag’ (‘\$GPRMC’), ‘InstrumentType’ (‘GPS’)

datasets: (where y is the number of samples in the file)

- ‘LATPOS’ - y-length vector of latitude, DDMM.MM (D-Degree(positive), M-Decimal minute)
- ‘LATHEMI’ - y-length vector of latitude hemisphere ‘N’ or ‘S’
- ‘LONPOS’ - y-length vector of longitude, DDDMM.MM (D-Degree(positive) , M-Decimal minute)
- ‘LONHEMI’ - y-length vector of longitude hemisphere ‘W’ or ‘E’
- ‘UTCPOS’ - y-length vector of UTC, HHMMSS.SS (no leading zeroes, e.g. 12:01:00AM is 100.0)

Group id: ‘ES_DARK’, attributes: ‘CalFileName’ (CAL file), ‘Frametag’ (‘SAT’+instrument code+serial number), ‘FrameType’ (‘ShutterDark’), ‘InstrumentType’ (‘Reference’)

datasets:

- ‘DATETAG’ - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- ‘TIMETAG2’ - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- ‘ES’ - Array x by y, where x is waveband, y is series. Calibrated Es dark shutter data

Group id: ‘ES_LIGHT’, attributes: ‘CalFileName’ (CAL file), ‘Frametag’ (‘SAT’+instrument code+serial number), ‘FrameType’ (‘ShutterLight’), ‘InstrumentType’ (‘Reference’)

datasets:

- ‘DATETAG’ - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- ‘TIMETAG2’ - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- ‘ES’ - Array x by y, where x is waveband, y is series. Calibrated Es light shutter data

Group id: ‘LI_DARK’, attributes: ‘CalFileName’ (CAL file), ‘Frametag’ (‘SAT’+instrument code+serial number), ‘FrameType’ (‘ShutterDark’), ‘InstrumentType’ (‘SAS’)

datasets:

- ‘DATETAG’ - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- ‘TIMETAG2’ - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- ‘LI’ - Array x by y, where x is waveband, y is series. Calibrated Li dark shutter data

Group id: ‘LI_LIGHT’, attributes: ‘CalFileName’ (CAL file), ‘Frametag’ (‘SAT’+instrument code+serial number), ‘FrameType’ (‘ShutterLight’), ‘InstrumentType’ (‘SAS’)

datasets:

- ‘DATETAG’ - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- ‘TIMETAG2’ - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- ‘LI’ - Array x by y, where x is waveband, y is series. Calibrated Li light shutter data

Group id: ‘LT_DARK’, attributes: ‘CalFileName’ (CAL file), ‘Frametag’ (‘SAT’+instrument code+serial number), ‘FrameType’ (‘ShutterDark’), ‘InstrumentType’ (‘SAS’)

datasets:

- ‘DATETAG’ - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- ‘TIMETAG2’ - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- ‘LT’ - Array x by y, where x is waveband, y is series. Calibrated Lt dark shutter data

Group id: 'LT_LIGHT', attributes: 'CalFileName' (CAL file), 'Frametag' ('SAT'+instrument code+serial number), 'FrameType' ('ShutterLight'), 'InstrumentType' ('SAS')

datasets:

- 'DATETAG' - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- 'TIMETAG2' - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- 'LT' - Array x by y, where x is waveband, y is series. Calibrated Lt light shutter data

Group id: 'SOLARTRACKER', attributes: 'CalFileName' (CAL file), 'Frametag' ('SATNAV'+instrument code+serial number), 'FrameType' ('SATNAVNNNN'), 'InstrumentType' ('SAS')

datasets:

- 'DATETAG' - Vector of length y, YYYYDOY.0, where DOY is the UTC sequential day of the year
- 'TIMETAG2' - Vector of length y, HHMMSSUUU UTC (hour, minute, second, millisecond)
- 'AZIMUTH' - Vector of length y, Solar Azimuth Angle (dtype 'SUN')
- 'ELEVATION' - Vector of length y, Solar Elevation (dtype 'SUN')
- 'HEADING' - Array 2 by y, True Sensor Azimuth (dtype 'SAS_TRUE'), Ship Heading (dtype 'SHIP_TRUE', optional)
- 'PITCH' - Vector of length y, Ship pitch
- 'ROLL' - Vector of length y, Ship roll

Level 1C

Process data from L1B to L1C. Data are filtered for vessel attitude (pitch and roll), viewing and solar geometry. *It should be noted that viewing geometry should conform to total radiance (Lt) measured at about 40 degrees from nadir, and sky radiance (Li) at about 40 degrees from zenith (Mobley 1999, Mueller et al. 2003 (NASA Protocols)).*

Rotator Home Angle: The offset between the neutral position of the SolarTracker unit and the bow of the ship. This /should/ be zero if the SAS Home Direction was set at the time of data collection as per Satlantic SAT-DN-635. Set to zero if using a non-Satlantic solar tracker and L1B contains the 'SAS_TRUE' data field.

Rotator Delay: Seconds of data discarded after a SAS rotation is detected.

Default: 60 seconds (Vandenberg 2016)

Pitch & Roll Filter (optional): Data outside these thresholds are discarded if this is enabled in the checkbox.

Default 5 degrees (IOCCG Draft Protocols; Zibordi et al. 2019; 2 deg “ideal” to 5 deg “upper limit”)

Relative Solar Azimuth Filter (optional): Relative azimuth angle in degrees between the viewing Li/Lt and the sun.

Default: 90-135 deg (Mobley 1999, Zhang et al. 2017); 135 deg (Mueller 2003 (NASA Protocols)); 90 deg unless certain of platform shadow (Zibordi et al. 2009, IOCCG Draft Protocols)

Level 1D

Process data from L1C to L1D. Light and dark data are screened for electronic noise (“deglitched” - see Anomaly Analysis), which are then removed from the data (optional, but advised). Shutter dark samples are then subtracted from shutter lights after dark data have been interpolated in time to match light data.

(e.g. Brewin et al. 2016, Sea-Bird/Satlantic 2017)

{To Do: Allow provisions for above water radiometer set-ups that do not have a dark shutter correction.} {To Do: discard dark data outside thresholds as hinted at in Zibordi 2009 (no actual threshold or methods given)}

Anomaly Analysis (optional)

Deglitching the data (which must follow L1C processing, as it is evaluated on a L1C file) is highly sensitive to the deglitching parameters described below, as well as environmental conditions and the variability of the radiometric data itself. Therefore, a separate module was developed to tune these parameters for individual files, instruments, and/or field campaigns. The tool is launched (after processing L1C files) by setting the parameters (windows and sigma factors described below) in the Configuration window, *SAVING THE CHANGES* to the Configuration file, and then running the tool on an example of L1B data using the Anomaly Analysis button and file dialog. Plots produced automatically in the /Plots/L1C_Anoms directory can be used to evaluate the choice of parameter values.

For each waveband of each sensor, and for both light and dark shutter measurements, the time series of radiometric data are low-pass filtered with a moving average using discrete linear convolution of two one dimensional sequences with adjustable window sizes. For darks, a stationary standard deviation anomaly (from the moving average) is used to assess whether data are

within an adjustable “sigma factor” multiplier within the window. For lights, a MOVING standard deviation anomaly (from the moving average of separately adjustable window size) is used to assess whether data are within a separately adjustable sigma. The low-band filter is passed over the data twice. First and last data points for light and dark data cannot be accurately filtered with this method, and are discarded.

Defaults: Dark Window 9, Light Window 11, Dark Sigma 2.7, Light Sigma 3.7 determined empirically from KORUS cruise (Abe et al. 2006, Chandola et al. 2009)

(API Reference: <https://docs.scipy.org/doc/numpy/reference/generated/numpy.convolve.html>)

‘Waveband interval to plot’ selects which waveband time-series plots will be generated (e.g. a value of 2 results in plots every ~7 nm for HyperOCR of resolution ~3.5 nm). Time-series plots of Es, Li, and Lt showing the results of the anomaly detection are saved to /Plots/L1C_Anoms. Data flagged for removal given the parameterizations chosen in the Configuration window are shown for the filter first pass (red box) and second pass (blue star). Review of these plots and adjustment of the parameters allow the user to optimize the low-pass filter for a given instrument and collection environment prior to L1E processing.

Level 1E

Process data from L1D to L1E. Interpolates radiometric data to common timestamps and wavebands, optionally generates spectral plots of Li, Lt, and Es, and optionally outputs text files (see ‘SeaBASS File and Header’ below) containing the data and metadata for submission to the SeaWiFS Bio-optical Archive and Storage System (SeaBASS; <https://seabass.gsfc.nasa.gov/>)

Each HyperOCR collects data at unique time intervals and requires interpolation for inter-instrument comparison. Satlantic ProSoft 7.7 software interpolates radiometric data between radiometers using the OCR with the fastest sampling rate (Sea-Bird 2017), but here we use the timestamp of the slowest-sampling radiometer (typically Lt) to minimize perturbations in interpolated data (i.e. interpolated data in HyperInSPACE are always closer in time to actual sampled data) (Brewin et al. 2016, Vandenberg 2017). Each HyperOCR radiometer collects data in a unique set of wavebands nominally at 3.3 nm resolution. For merging, they must be interpolated to common wavebands. Interpolating to a different (i.e. lower) spectral resolution is also an option. No extrapolation is calculated (i.e. interpolation is between the global minimum and maximum spectral range for all HyperOCRs). Spectral interpolation is linear by default, but has an option for univariate spline with a smoothing factor of 3 (see ProcessL1e.interpolateL1e in ProcessL1e.py). **(API: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.UnivariateSpline.html>)**

Note that only the specified datasets in each group will be interpolated and carried forward. For radiometers, this means that ancillary metadata such as SPEC_TEMP and THERMAL_RESP will be dropped at L1E and beyond. See ProcessL1e.py at Perform Time Interpolation comment.

Optional plots of Es, Li, and Lt of L1E data can be generated which show the temporal interpolation for each parameter and waveband to the slowest sampling radiometer timestamp. They are saved in /Plots/L3.

{To Do: Include selections to weight for various ocean color sensor bands (MODIS, VIIRS, SeaWiFS, Sentinel 3, etc.) using spectral response functions.} {To Do: Allow provision for above water radiometers that operate simultaneously and/or in the exact same wavebands.}

SeaBASS File and Header

To output SeaBASS formatted text files, check the box. A subfolder within the L3 directory will be created, and separate files made for Li, Lt, and Es hyperspectral data.

An eponymous, linked module allows the user to collect information from the data and the processing configuration (as defined in the Configuration window) into the SeaBASS files and their headers. The module is launched by selecting the New, Open, or Edit buttons in the Configuration window to create, select, or edit a SeaBASS header configuration file, which is automatically stored in the /Config directory with a .hdr extension. Instructions are given at the top of the SeaBASS Header window which launches when the Edit button is pressed. Within the SeaBASS Header window, the left column allows the user to input the fields required by SeaBASS. Calibration files (if they have been added at the time of creation) are auto-populated. In the right hand column, additional header fields can be auto-populated or updated from the Configuration window, additional comments can be added, and the lower fields are autopopulated from each data file as it is processed. To override auto-population of fields, enter the desired value here in the SeaBASS Header window, and save it.

When updating values in the Configuration window, be sure to apply those updates in the SeaBASS Header by editing the header, selecting the ‘Update from Config Window’ button, and saving the header file.

Level 2 Preliminary

Use of the Ruddick et al. 2006 and the Zhang et al. 2017 glint corrections require wind data, and Zhang (2017) also requires aerosol optical depth, salinity, and sea surface temperature. Since most field collections of above water radiometry are missing some or all of these ancillary parameters, this function allows the user to download model data from the NASA EARTHDATA server. These data are generated by the NASA Global Modeling and Assimilation Office (GMAO) as hourly, global ‘MERRA2’ HDF files at 0.5 deg (latitude) by 0.625 deg (longitude) resolution (<https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/>). Two files will be downloaded for each hour of data processed (total ~8.3 MB) and stored in /Data/Anc. Local ancillary data files will be reused for anywhere in the world, so it is not recommendable to clear this directory unless updated models are being released. Details for how these data are applied to above water radiometry are given below.

As of January 2020, access to these data requires a user login and password, which can be obtained here (<https://oceancolor.gsfc.nasa.gov/registration>). A link to register is also provided in the Configuration window. When the user selects ‘Download ancillary models’, pop-up windows will allow the user login and password to be entered.

Level 2

Process L1E to L2. Further quality control filters are applied to data, and data are averaged into optional time interval ensembles prior to calculating the remote sensing reflectance within each ensemble. While the use of ensembles is optional (set to 0 to avoid averaging), it is highly recommended, as it allows for the statistical analysis required for Percent Lt calculation (see below) within each ensemble, rather than %Lt across an entire collection.

Prior to ensemble binning, data may be filtered for **Maximum Wind Speed**.

Default 7 m/s (IOCCG Draft Protocols 2019; D’Alimonte pers.comm 2019); 10 m/s Mueller et al. 2003 (NASA Protocols); 15 m/s (Zibordi et al. 2009);

Solar Zenith Angle may be filtered for minimum and maximum values.

Default Min: 20 deg (Zhang et al 2017); Default Max: 60 deg (Brewin et al. 2016)

Spectral Outlier Filter may be applied to remove noisy data prior to binning. This simple filter examines only the spectrum of Es, Li, and Lt from 400 - 700 nm, above which the data are extremely noisy. Using the standard deviation of the normalized spectra for the entire sample ensemble, together with a multiplier to establish an envelope, spectra with data outside the envelop in any band are rejected. Currently, the arbitrary filter factors are 5.0 for Es, 8.0 for Li, and 3.0 for Lt. Results of spectral filtering are saved as spectral plots in ./Plots/L2_Spectral_Filter.

Meteorological flags based on (Wernand et al. 2002, Garaba et al. 2012, Vandenberg 2017) can be optionally applied to screen for undesirable data. Specifically, data are filtered for unusually low downwelling irradiance at 480 nm (default $2.0 \text{ uW cm}^{-2} \text{ nm}^{-1}$, for data likely to have been collected near dawn or dusk ($\text{Es}(470)/\text{Es}(680) < 1.0$), and for data likely to have high relative humidity or rain ($\text{Es}(720)/\text{Es}(370) < 1.095$).

Ensemble Interval can be set to the user’s requirements. Setting this to zero avoids temporal binning, using the common timestamps established in L1E. The ensemble period is used to calculate the spectra within the lowest X percent of Lt(780), the ensemble average spectra for Es, Li, and Lt, as well as variability in spectra within the ensemble, which is used in estimating sample uncertainty.

Percent Lt Calculation Data are optionally limited to the darkest X percent of Lt data at 780 nm within the sampling interval (if binning is performed; otherwise across the entire file) to minimize the effects of surface glitter. **Default: 5 % (Hooker et al. 2002, Zibordi et al. 2002, Hooker and Morel 2003); <10% (IOCCG Draft Protocols).**

Skyglint/Sunglint Correction (rho) The default value for sea-surface reflectance (**Rho_sky**, sometimes called the Fresnel “factor”) is set by default to 0.0256 based on (Mobley 1999, Mueller et al. 2003 (NASA Protocols)), which can be optionally adjusted for wind speed and cloud cover using the relationship found in (Ruddick et al. 2006) (i.e. $\text{Li}(750)/\text{Es}(750) < 0.05$ for clear skies, $\text{Rho_sky} = 0.0256 + 0.00039 * U + 0.000034 * U^2$, else $\text{Rho_sky} = 0.0256$). The default wind speed should be set by the user depending on in situ conditions for instances when the ancillary data and models are not available (more information is given below on how ancillary wind data are applied). This correction does not account for the spectral dependence (Lee et al. 2010, Gilerson et al. 2018) or polarization sensitivity (Harmel et al. 2012, Mobley 2015, Hieronymi 2016, D’Alimonte and Kajiyama 2016, Foster and Gilerson 2016, Gilerson et al. 2018) in **Rho_sky**. Uncertainty in rho is estimated from Ruddick et al. 2006 at +/- 0.003.

The other option provided for glint correction is based on Zhang et al. 2017. This model explicitly accounts for spectral dependence in rho, separates the glint contribution from the sky and the sun, and accounts for polarization. This approach requires knowledge of environmental conditions during sampling including: wind speed, aerosol optical depth, solar and sensor azimuth and zenith angles, water temperature and salinity. To accomodate these parameters, HyperInSPACE uses either the ancillary data file provided in the main window, GMAO models, or the default values set in the Configuration window as follows: field data ancillary files are screened for wind, water temperature, and salinity. These are each associated with the nearest timestamps of the radiometer suite to within one hour. Radiometer timestamps still lacking wind and aerosol data

will extract it from the GMAO models, if available. Otherwise, the default values set in the Configuration window will be used as a last resort.

{To Do: Include a bidirectional correction to L_w based on, e.g. Lee 2011, Zibordi 2009} *{To Do: Improve estimates (e.g. Zibordi et al. 2009)}*

Remote sensing reflectance is calculated as $R_{rs} = (L_t - \rho_{sky} * L_i) / E_s$ (e.g. (Mobley 1999, Mueller et al. 2003, Ruddick et al. 2006)). Normalized water leaving radiance (nL_w) is calculated as $R_{rs} * F_0$, where F_0 is the top of atmosphere incident radiation adjusted for the Earth-Sun distance on the day sampled.

Uncertainties in L_i , L_t , and E_s are estimated using the standard deviation of spectra used in the ensemble or collection average. Uncertainty in ρ is estimated at ± 0.003 from Ruddick et al. 2006. Uncertainty in R_{rs} and nL_w are estimated using propagation of uncertainties from L_i , L_t , E_s , and ρ assuming random, uncorrelated error.

Additional glint may be removed from the R_{rs} and nL_w by subtracting the value in the NIR from the entire spectrum (Mueller et al. 2003 (NASA Protocols)). This approach, however, assumes negligible water-leaving radiance in the 750-800 nm range (not true of turbid waters), and ignores the spectral dependence in sky glint, and **should therefore only be used in clear waters with caution**. Here, a minimum in $R_{rs}(750-800)$ or $nL_w(750-800)$ is found and subtracted from the entire spectrum.

To output SeaBASS formatted text files, check the box. A subfolder within the L4 directory will be created, and separated files made for L_i , L_t , E_s , and R_{rs} hyperspectral data.

References

- Abe, N., B. Zadrozny and J. Langford (2006). Outlier detection by active learning. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA, ACM: 504-509.
- Brewin, R. J. W., G. Dall’Omo, S. Pardo, V. van Dongen-Vogels and E. S. Boss (2016). “Underway spectrophotometry along the Atlantic Meridional Transect reveals high performance in satellite chlorophyll retrievals.” Remote Sensing of Environment 183: 82-97.
- Chandola, V., A. Banerjee and V. Kumar (2009). “Anomaly detection: A survey.” ACM Comput. Surv. 41(3): 1-58.
- D’Alimonte, D. and T. Kajiyama (2016). “Effects of light polarization and waves slope statistics on the reflectance factor of the sea surface.” Optics Express 24(8): 7922-7942.
- Foster, R. and A. Gilerson (2016). “Polarized transfer functions of the ocean surface for above-surface determination of the vector submarine light field.” Applied Optics 55(33): 9476-9494.
- Garaba, S. P., J. Schulz, M. R. Wernand and O. Zielinski (2012). “Sunglint Detection for Unmanned and Automated Platforms.” Sensors 12(9): 12545.
- Gilerson, A., C. Carrizo, R. Foster and T. Harmel (2018). “Variability of the reflectance coefficient of skylight from the ocean surface and its implications to ocean color.” Optics Express 26(8): 9615-9633.
- Harmel, T., A. Gilerson, A. Tonizzo, J. Chowdhary, A. Weidemann, R. Arnone and S. Ahmed (2012). “Polarization impacts on the water-leaving radiance retrieval from above-water radiometric measurements.” Applied Optics 51(35): 8324-8340.
- Hieronimi, M. (2016). “Polarized reflectance and transmittance distribution functions of the ocean surface.” Optics Express 24(14): A1045-A1068.
- Hooker, S. B., G. Lazin, G. Zibordi and S. McLean (2002). “An Evaluation of Above- and In-Water Methods for Determining Water-Leaving Radiances.” Journal of Atmospheric and Oceanic Technology 19(4): 486-515.
- Hooker, S. B. and A. Morel (2003). “Platform and Environmental Effects on Above-Water Determinations of Water-Leaving Radiances.” Journal of Atmospheric and Oceanic Technology 20(1): 187-205.
- Lee, Z., Y.-H. Ahn, C. Mobley and R. Arnone (2010). “Removal of surface-reflected light for the measurement of remote-sensing reflectance from an above-surface platform.” Optics Express 18(25): 26313-26324.
- Mobley, C. D. (1999). “Estimation of the remote-sensing reflectance from above-surface measurements.” Applied Optics 38(36): 7442-7455.
- Mobley, C. D. (2015). “Polarized reflectance and transmittance properties of windblown sea surfaces.” Applied Optics 54(15): 4828-4849.

- Mueller, J. L., A. Morel, R. Frouin, C. O. Davis, R. Arnone, K. L. Carder, Z. P. Lee, R. G. Steward, S. B. Hooker, C. D. Mobley, S. McLean, B. Holbert, M. Miller, C. Pietras, K. D. Knobelspiesse, G. S. Fargion, J. Porter and K. J. Voss (2003). Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volume III. Ocean Optics Protocols for Satellite Ocean Color Sensor Validation. J. L. Mueller. Greenbelt, MD, NASA Goddard Space Flight Center.
- Ruddick, K. G., V. De Cauwer, Y.-J. Park and G. Moore (2006). “Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters.” *Limnology and Oceanography* 51(2): 1167-1179.
- Scientific, S.-B. (2017). Prosoft 7.7 Product Manual SAT-DN-00228 Rev. K.
- Vandenberg, N., M. Costa, Y. Coady and T. Agbaje (2017). PySciDON: A python scientific framework for development of ocean network applications. 2017 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM).
- Wernand, M. R. (2002). Guidelines for (ship-borne) auto-monitoring of coastal ocean colour. Ocean Optics XVI, Sante Fe, NM, The Oceanography Society.
- Zhang, X., S. He, A. Shabani, P.-W. Zhai and K. Du (2017). “Spectral sea surface reflectance of skylight.” *Optics Express* 25(4): A1-A13.
- Zibordi, G., S. B. Hooker, J. F. Berthon and D. D’Alimonte (2002). “Autonomous Above-Water Radiance Measurements from an Offshore Platform: A Field Assessment Experiment.” *Journal of Atmospheric and Oceanic Technology* 19(5): 808-819.
- Zibordi, G., F. Mélin, J.-F. Berthon, B. Holben, I. Slutsker, D. Giles, D. D’Alimonte, D. Vandemark, H. Feng, G. Schuster, B. E. Fabbri, S. Kaitala and J. Seppälä (2009). “AERONET-OC: A Network for the Validation of Ocean Color Primary Products.” *Journal of Atmospheric and Oceanic Technology* 26(8): 1634-1651.
- Zibordi, G. and K. J. Voss (2019). Protocols for Satellite Ocean Color Data Validation (DRAFT). I. O. C. C. Group.