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 1.0.3 (see Changelog.md);

While this version has been substantially updated from 1.0.α 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 not only the general functionality of this version, but also the instruments and data formats you are interested in seeing implemented in HyperInSPACE in the future.

Requirements and Installation

Save this entire HyperInSPACE file repository (~2.5 GB) to a convenient, sensibly named directory on your computer such as yourPath/HyperInSPACE/. When Main.py is launched for the first time, data and directories will be updated as described below. No system files will be changed.

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. (If you are running a minimal, bootstrap distribution such as Miniconda, several additional packages may be needed including: scipy, matplotlib, pyqt5, h5py, fpdf, and requests.)

HDF5 data files will be read and written using the h5py module (2.9.0 at the time of writing) included in a standard Anaconda Python package. The Zhang (et al. 2017) sky/sunglint correction also requires Xarray, which requires installation (complete instructions here: http://xarray.pydata.org/en/stable/installing.html). To install Xarray with Anaconda:

```
prompt$ conda install xarray dask netCDF4 bottleneck
```

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 . The Pysolar module (https://anaconda.org/conda-forge/pysolar) must be installed unless the raw data includes solar geometries (e.g. SolarTracker with SATNAV and associated device file as described below). To install Pysolar with Anaconda :

```
prompt$ conda install -c conda-forge pysolar
```

To generate PDF reports for each file processed at Level-2, the package fpdf will need to be installed:

```
prompt$ pip install fpdf
```

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, Main view configuration file
- 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'). As with the Data, this path for outputting plots is optional and will be overwritten by choosing an alternate Data/Plots parent directory (see below).
- Data - This directory now comes unpacked in the distribution. By default, it contains only Pope & Fry/Smith & Baker water absorption properties, Thuillier and

satellite spectral response functions, banner images for the GUI, and the Zhang glint correction database. This is also the optional fallback location for input and/or output radiometry data, though setting up separate locations for field data is highly recommended (see below).

- Source - This new directory (which comes unpacked with the distribution) now holds the majority of the code.

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

Guide

Overview

1. Identify the cruise, relevant calibration files, and ancillary data files to be used in processing
2. Launch HyperInSPACE and set up the Main window for data directories and ancillary file
3. Create a new Configuration (or edit and existing Configuration)
4. Add and enable only *relevant* calibration and instrument files to the Configuration; there is no such thing as a standard instrument package
5. Choose appropriate processing parameters for L1A-L2 (do not depend on software defaults; there is no such thing as a standard data collection)
6. HDF files will be produced at each level of processing, plus optional SeaBASS text files for radiometry at L1E and L2. Plots can be produced at L1D, L1E, and L2. Processing logs and plots are aggregated into PDF Reports at L2 (covering all processing from RAW to L2) written to a dedicated Reports directory in the selected Output directory.

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 and relevant geometries in a SeaBASS file format), single-level processing, and multi-level processing. For batch processing, pop-up windows from "failed" (no output due to either corrupt raw binary data or stringent quality control filtering) files can be suppressed. Producing no output file at a given processing level is often a normal result of quality control filtering, so this option allows batches to continue uninterrupted.

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.

The 'Input...' and 'Output Data/Plots Parent Directory' buttons are fairly self explanatory and allow optional selection of data and directories from any mounted/mapped drive. Note that output data and plot 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' and 'Plots' under the HyperInSPACE directory structure will be used by default as the parent directories.

Ancillary Data files for environmental conditions and relevant geometries used in L2 processing 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. Optional data fields include station number, ship heading, relative sensor azimuth, aerosol optical depth, cloud cover, salinity, water temperature, and wind speed. An example ancillary file is included for use as a template. 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. The ancillary data file is optional (though strongly advised for adding wind speed at a minimum) provided the sensor suite is equipped with a SolarTracker or equivalent to supply the relevant sensor/solar geometries. If no SolarTracker-type instrument is present to report the relative sensor/solar geometries, the ancillary file must be provided with at least the ship heading and relative angle between the bow of the ship and the sensor azimuth as a function of time.

Configuration

Launch the configuration module and GUI (ConfigWindow.py) from the Main window by selecting/editing a configuration file or creating a new one. This file will be instrument-suite-specific, and is 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) may also require multiple configurations to obtain the highest quality ocean color products at Level 2. Sharp gradients in environmental conditions could also warrant multiple configurations for the same cruise (e.g. sharp changes in temperature may effect how data deglitching is parameterized, as described below).

Calibration & Instrument Files:

NOTE: IT IS IMPORTANT THAT THESE INSTRUCTIONS FOR SELECTING AND ACTIVATING CALIBRATION AND INSTRUMENT FILES ARE FOLLOWED CAREFULLY OR PROCESSING WILL FAIL

Note: You do not need to move/copy/paste your calibration and instrument files; HyperInSPACE will take care of that for you.

In the 'Configuration' window, click 'Add Calibration Files' to add the *relevant* calibration or instrument files (date-specific HyperOCR factory calibrations or ancillary instrument Telemetry Definition Files; i.e. the '.cal' and '.tdf' files). *Only add and enable those calibration and instrument files that are relevant to the cruise/package you wish to process (see below).* Each instrument you add here -- be it a radiometer or an external data instrument such as a tilt-heading sensor -- requires at least one .cal or .tdf file for raw binary data to be interpreted. Two .cal files are required in the case of radiometers calibrated seperately for shutter open (light) and shutter closed (dark) calibrations, as is typical with Satlantic/Seabird HyperOCRs. Instruments with no calibrations (e.g. GPS, SolarTracker, etc.) still require a Telemetry Definition File (.tdf) to be properly interpreted.

Adding new .cal and .tdf files will automatically copy these files from the directory you identify on your machine into the HyperInSPACE directory structure once the Configuration is saved.

The calibration or instrument file can now 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) (Note: Use of built-in flux-gate compass is extremely inadviseable on a steel ship or platform. Best practice is to use externally supplied heading data from the ship's NMEA datastream or from a seperate, external dual antenna GPS incorporated into the SolarTracker. DO NOT USE COURSE DATA FROM SINGLE GPS SYSTEMS.)

- 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)

- HLDYYYA.cal > Li (Frame Type: Dark)

- HSLYYYA.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 (usually A, B, C, etc. associated with the date of calibration) *Be sure to choose the factory calibration files appropriate to the date of data collection.*

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 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. in version 1.0.2 (which only fully supports Satlantic/Seabird HyperSAS) each radiometer will require two calibration files (light and dark). Data from the GPS and SATNAV instruments are interpreted using the corresponding Telemetry Definition Files ('.tdf').

Once you have created your new Configuration, CAL/TDF files are copied from their chosen locations into the /Config directory HyperInSPACE directory structure within an automatically created sub-directory named for the Configuration (i.e. a configuration named "KORUS" creates a KORUS.cfg configuration file in /Config and creates the /Config/KORUS_Calibration/directory with the chosen 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. 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'). Spectral filters are also plotted in L2 to help with filter parameterization factors. 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. If SeaBASS output is selected, a pop-up will remind the user to update the SeaBASS Header (the button in L1E) prior to saving the configuration file. The configuration file is 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 to use as the Output Parent Directory when processing from L0 (raw binary). Once L0 processing is complete, the user can change the input directory to match the output, and files will be automatically sorted by processing level in the automatically created sub-directories (i.e. the software automatically creates and looks for L1A, L1B, L1C, L1D, L1E, and L2 directories under the parent directory). If not selected, the Input/Output parent directories 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 using the Save button or upon closing the Main window, 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 (e.g. Shift- or Ctrl- click, or Ctrl-A for all, depending on your platform). 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 given 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.

Bug: When running in MacOS, the Open File dialog window remains frozen open during processing, and closes once complete. Suggestions welcome...

Bug: Very rarely, when running the program for the first time, the first RAW binary data file opened for processing is not read in properly. Processing will fail with the error message: [filename] does not match expected input level for outputing L2. The file will process properly if run a second time (assuming it is a healthy file). Cause unknown.

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, not just samples within the file, so do not be overly conservative with this selection, particularly for files collected over a long period.* Further screening for SZA min/max at a sample level 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 in HDF5 to match L1B in order to be successfully processed to L2 in HyperInSPACE. HyperSAS data with no SolarTracker can be processed from raw data, provided the relative geometries are included in the ancillary data file as described above and below.* An example of an L1B HDF file is provided in /Data for reference. Datasets are grouped by instrument and contain their data in an array referenced '.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. the wavelength of sample '304.37', or for many instruments simply 'NONE' if the data is already well described in the group name; 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' ('$GPRMC' or '$GPGGA'), '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)
- Several optional datasets including, 'COURSE', 'SPEED', etc.

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

*Note that HDF5 does not support Python datetime, so datetime is calculated for use at each level using Datetag and Timetag2.

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 lenght y, Ship pitch
- 'ROLL' - Vector of lenght y, Ship roll

Level 1C

Process data from L1B to L1C. Data are filtered for vessel attitude (pitch, roll, and yaw when available), 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)**). Unlike other approaches, HyperInSPACE eliminates data flagged for problematic pitch/roll, yaw, and solar/sensor geometries *prior* to deglitching the time series (L1D), thus increasing the relative sensitivity of deglitching for the removal of non-environmental anomalies.

SolarTracker: Select when using the SolarTracker unit. In this case sensor and solor geometry data will come from the SolarTracker (i.e. SATNAV**.tdf). If deselected, solar geometries will be calculated from GPS time and position with Pysolar, while sensor azimuth (i.e. ship heading and sensor offset) must either be provided in the ancillary data or (eventually) from other data inputs. Currently, if SolarTracker is unchecked, the Ancillary file chosen in the Main Window will be read in, subset for the relevant dates/times, saved in the ANCILLARY_NOTRACKER group, and carried forward to subsequent levels (i.e. the file will not need to be read in again at L2). If the ancillary data file is very large (e.g. for a whole cruise at high temporal resolution), this process of reading in the text file and subsetting it to the radiometry file can be slow.

Rotator Home Angle Offset: Generally 0. This is the offset between the neutral position of the radiometer suite and the bow of the ship. This *should* be zero if the SAS Home Direction was set at the time of data collection in the SolarTracker as per Satlantic SAT-DN-635. If no SolarTracker was used, the offset can be set here if stable, or in the ancillary data file if changeable in time. Without SolarTracker, L1C processing will require at a minimum ship heading data in the ancillary file. Then the offset can be given in the ancillary file (dynamic) or set here in the GUI (static). *Note: as SeaBASS does not have a field for this angle between the instrument and the bow of the ship, the field "relaz" (normally reserved for the relative azimuth between the instrument and the sun) can is used in this version of the software.*

Rotator Delay: Seconds of data discarded after a SolarTracker rotation is detected. Set to 0 to ignore. Not an option without SolarTracker. **Default: 60 seconds (Vandenberg 2017)**

Pitch & Roll Filter (optional): Data outside these thresholds are discarded if this is enabled in the checkbox. Not currently an option without SolarTracker. *{To Do: see what other accelerometer data are being collected and accommodate.}* **Default: 5 degrees (IOCCG Draft Protocols; Zibordi et al. 2019; 2 deg "ideal" to 5 deg "upper limit")**

Absolute Rotator Angle Filter (optional): Angles relative to the SolarTracker neutral angle beyond which data will be excluded due to obstructions blocking the field of view. These are generally set in the SolarTracker software when initialized for a given platform. Not an option without SolarTracker. **Default: -40 to +40 (arbitrary)**

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 light frames 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)} Currently, spectra with anomalies in any band are deleted in their entirety, but this might be overkill -- certainly it is very conservative. It may be sufficient to set the anomalous values to NaNs, and only delete the entire spectrum if more than, say, 25% of wavebands are anomalous.

Anomaly Analysis (optional)

Deglitching the data (which must follow after L1C processing, as it is evaluated on a L1C file) is highly sensitive to the deglitching parameters described below, as well as some environmental conditions not controlled for in L1C 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 and conditions. A sharp temperature change or heavy vibration, for example, could change the optimal deglitching parameterization. The tool is launched (after processing L1C files) by setting the parameters (windows and sigma factors described below) in the Configuration window and then pressing the Anomaly Analysis button to select an example L1C file. Plots produced automatically in the [output_directory]/Plots/L1C_Anoms directory can be used to evaluate and adjust 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 over time using discrete linear convolution of two one dimensional sequences with adjustable window sizes. For darks, a *STATIONARY*** standard deviation anomaly *(from the moving average in time) 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 example data on the 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 6.6 nm for HyperOCR of resolution 3.3 nm). Time-series plots of Es, Li, and Lt showing the results of the anomaly detection are saved to [output_directory]/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 filters 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 temporal plots of Li, Lt, and Es, and ancillary data to show how data were interpolated. L1E also 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 perterbations 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: only the datasets specified in ProcessL1e.py in each group will be interpolated and carried forward. For radiometers, this means that ancillary instrument data such as SPEC_TEMP and THERMAL_RESP will be dropped at L1E and beyond. See ProcessL1e.py at Perform Time Intepolation comment.

Optional plots of Es, Li, and Lt of L1E data can be generated which show the temporal interpolation for each parameter and each waveband to the slowest sampling radiometer timestamp. They are saved in [output_directory]/Plots/L1E. Plotting is time and memory intensive, and so may not be adviseable for processing an entire cruise.

{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 SeaBASS subfolder within the L1E directory will be created, and separate files generated 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 Edit/Update button in the Configuration window. A SeaBASS header configuration file is

automatically stored in the /Config directory with the name of the Configuration and a .hdr extension. Instructions are given at the top of the SeaBASS Header window, which launches when the Edit/Update 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, the HyperInSPACE parameterizations defined in the Configurations window can be updated in the SeaBASS header using the 'Update from Config Window' button. Additional comments can be added in the second field, and the lower fields are autopopulated from each data file as it is processed. To override auto-population of the lower fields in the right column, 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. A reminder pop-up is given when saving the Configuration if SeaBASSS output is selected for L1E or L2.

Level 2

Process L1E to L2. Further quality control filters are applied to data, and data are then averaged within optional time interval ensembles prior to calculating the remote sensing reflectance within each ensemble. A typical field collection file for the HyperSAS SolarTracker is one hour, and the optimal ensemble periods within that hour will depend on how rapidly conditions and water-types are changing, as well as the instrument sampling rate. 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 (radiance acceptance fraction; see below) within each ensemble, rather than %Lt across an entire (e.g. one hour) collection, and it also improves radiometric uncertainty estimation.

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 spectra of Es, Li, and Lt from 400 - 700 nm, above which the data are noisy in the HyperOCRs. Using the standard deviation of the normalized spectra for the entire sample ensemble, together with a multiplier to establish an "envelope" of acceptable values, 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 [output_directory]/Plots/L2_Spectral_Filter. The filter can be optimized by studying these plots for various parameterizations of the filter.

Meteorological flags based on **(Ruddick et al. 2006, Mobley, 1999, Wernand et al. 2002, Garaba et al. 2012, Vandenberg 2017)** can be optionally applied to screen for undesirable data. Specifically, data are filtered for cloud cover, unusually low downwelling irradiance at **480 nm < default 2.0 uW cm^-2 nm^-1** for data likely to have been collected near dawn or dusk, or **(Es(470)/Es(680) < 1.0)**, and for data likely to have high relative humidity or rain (**Es(720)/Es(370) < 1.095**). Cloud screening (**Li(750)/Es(750) >= 0.05**) is optional and not well parameterized. Clear skies are approximately 0.02 (Mobley 1999) and fully overcast are of order 0.3 (Ruddick et al. 2006). However, the Ruddick skyglint correction (below) can partially compensate for clear versus cloudy skies, so to avoid eliminating non-clear skies prior to glint correction, set this high (e.g. 1.0). Further investigation with automated sky photography for cloud cover is warranted.

Ensembles Extract Cruise Stations can be selected if station information is provided in the ancillary data file identified in the Main window. If selected, only data collected on station will be processed, and the output data/plot files will have the station number appended to their names. At current writing, stations must be numeric, not string-type. If this option is deselected, all automated data (underway and on station) will be included in the ensemble processing.

Ensemble Interval can be set to the user's requirements depending on sampling conditions and instrument rate (**default 300 sec**). Setting this to zero avoids temporal bin-averaging, preserving the common timestamps established in L1E. Processing the data without ensemble averages can be very slow, as the reflectances are calculated for each spectrum collected (i.e. nominally every 3.3 seconds of data for HyperSAS). The ensemble period is used to process the spectra within the lowest percentile of Lt(780) as defined/set below. The ensemble average spectra for Es, Li, and Lt is calculated, as well as variability in spectra within the ensemble, which is used to help estimate sample uncertainty.

Percent Lt Calculation Data are optionally limited to the darkest percentile 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 from capillary waves. The percentile chosen is sensitive to the sampling rate. The 5% default recommended in Hooker et al. 2002 was devised for a multispectral system with rapid sampling rate. **Default: 5 % (Hooker et al. 2002, Zibordi et al. 2002, Hooker and Morel 2003); <10% (IOCCG Draft Protocols)**.

Skyglint/Sunglint Correction (rho) 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 anchillary parameters, an embedded 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 for one hour of field data) and stored in /Data/Anc. Global ancillary data files from GMAO will be reused, so it is not recommended to clear this directory unless updated models are being released by GMAO. 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 to enter a login and password. Once this has been done once, canceling the login pop-up dialog will force the program to use the current configuration (i.e. it is only necessary to re-enter the password if it has changed.)

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. Li(750)/Es(750)< 0.05 for clear skies, Rho_sky = 0.0256 + 0.00039* U + 0.000034* U^2, else Rho_sky = 0.0256). The default wind speed (U) 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 third 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 in the skylight term. 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: Work to optimize the processing of Zhang rho. Currently, this proceeds far more slowly in Python than Matlab. See notes in ZhangRho.py for gen_vec_quad and prob_reflection near L165} {To Do: Include additional skylight/skyglint corrections, such as Groetsch et al. 2017/2020} {To Do: Include a bidirectional correction to Lw based on, e.g. Lee 2011, Zibordi 2009 (for starters; ultimately the BRDF will need to be better parameterized for all conditions and water types.)} {To Do: Improve uncertainty estimates (e.g. Zibordi et al. 2009). The uncertainty in rho within the Zhang model is not well constrained.}

Remote sensing reflectance is calculated as Rrs = (Lt - rho_sky* Li) / Es (e.g. **(Mobley 1999, Mueller et al. 2003, Ruddick et al. 2006)**). Normalized water leaving radiance (nLw) is calculated as Rrs*F0, where F0 is the top of atmosphere incident radiation adjusted for the Earth-Sun distance on the day sampled.

Uncertainties in Li, Lt, and Es are estimated using the standard deviation of spectra in the ensemble or full-file average. Uncertainty in rho is estimated at +/- 0.003

from Ruddick et al. 2006. Uncertainty in Rrs and nLw are estimated using propagation of uncertainties from Li, Lt, Es, and rho assuming random, uncorrelated error.

Additional glint may be removed from the Rrs and nLw by subtracting the value in the NIR from the entire spectrum **(Mueller et al. 2003 (NASA Protocols))**. This approach, however, assumes neglible 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 the clearest waters and with caution** . Here, a minimum in Rrs(750-800) or nLw(750-800) is found and subtracted from the entire spectrum.

An alternate NIR residual correction can be applied based on **Ruddick et al. 2005, Ruddick et al. 2006**. This utilizes the spectral shape in water leaving reflectances in the NIR to estimate the residual glint correction for turbid waters with NIR reflectances from about 0.0001 to 0.03

Negative reflectances can be removed as follows: any spectrum with any negative reflectances between 380 nm and 700 nm is removed from the record entirely. Negative reflectances outside of this range (e.g. noisy data deeper in the NIR) are set to 0.

Spectral wavebands for a few satellite ocean color sensors can be optionally calculated using their spectral weighting functions. These will be included with the hyperspectral output in the L2 HDF files. Spectral response functions are applied to convolve the (ir)radiances prior to calculating reflectances. **(Burgghoff et al. 2020)**.

Plots of processed L2 data from each radiometer and calculated reflectances can be created and stored in [output_directory]/Plots/L2. Uncertainties are shown for each spectrum as shaded regions, and satellite bands (if selected) are superimposed on the hyperspectral data.

Select the "Derived L2 Ocean Color Products" button to choose, calculate, and plot derived biochemical and inherent optical properties using a variety of ocean color algorithms. Algorithms largely mirror those available in SeaDAS with a few additions. They include OC3M, PIC, POC, Kd490, iPAR, GIOP, QAA, and the Average Visible Wavelength (Vandermuellen et al. 2020) and GOCAD-based CDOM/Sg/DOC algorithms (Aurin et al. 2018).

To output SeaBASS formatted text files, check the box. A subfolder within the L2 directory will be created, and separate text files will be made for Li, Lt, Es, and Rrs hyperspectral data and satellite bands, if selected. Set-up for the SeaBASS header is managed with the 'Edit/Update SeaBASS Header' in the L1E configuration.

Upon completion of L2 processing for each file, a PDF summary report will be produced and saved in [output_directory]/Reports. This contains metadata, processing parameters, processing logs, and plots of QA analysis, radiometry, and derived ocean color products.

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, Association for Computing Machinery: 504–509.
- Brewin, R. J. W., G. Dall'Olmo, 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.
- Burggraaff, O. (2020). "Biases from incorrect reflectance convolution." Optics Express 28(9): 13801-13816.
- Chandola, V., A. Banerjee and V. Kumar (2009). "Anomaly detection: A survey." ACM Comput. Surv. 41(3): Article 15.
- 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.
- Hieronymi, 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., V. De Cauwer and B. Van Mol (2005). Use of the near infrared similarity reflectance spectrum for the quality control of remote sensing data, SPIE.
- 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.
- 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 AND OCEAN COLOR. Ocean Optics XVI. S. Ackleson and C. Trees. Santa Fe, NM, USA.
- 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., K. J. Voss, B. Johnson and J. L. Mueller (2019). Protocols for Satellite Ocean Colour Data Validation: In Situ Optical Radiometry. IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation. IOCCG. Dartmouth, NS, Canada, IOCCG.