

# Working with SGLI Level-2 data

## Contents

<b>Working with SGLI Level-2 data</b> .....	<b>1</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>2 Evaluating SGLI Level-2 data</b> .....	<b>1</b>
<b>3 Finding match-ups in SGLI Level-2 data</b> .....	<b>7</b>
<b>4 SGLI data in JASMES (*.nc) format</b> .....	<b>9</b>
<b>5 Case study evaluating SGLI Chla in the California Current area</b> .....	<b>10</b>
<b>6 SGLI Level-2 NWLR data</b> .....	<b>12</b>
<b>7 Evaluating SGLI Level-3 data</b> .....	<b>15</b>

## 1 Introduction

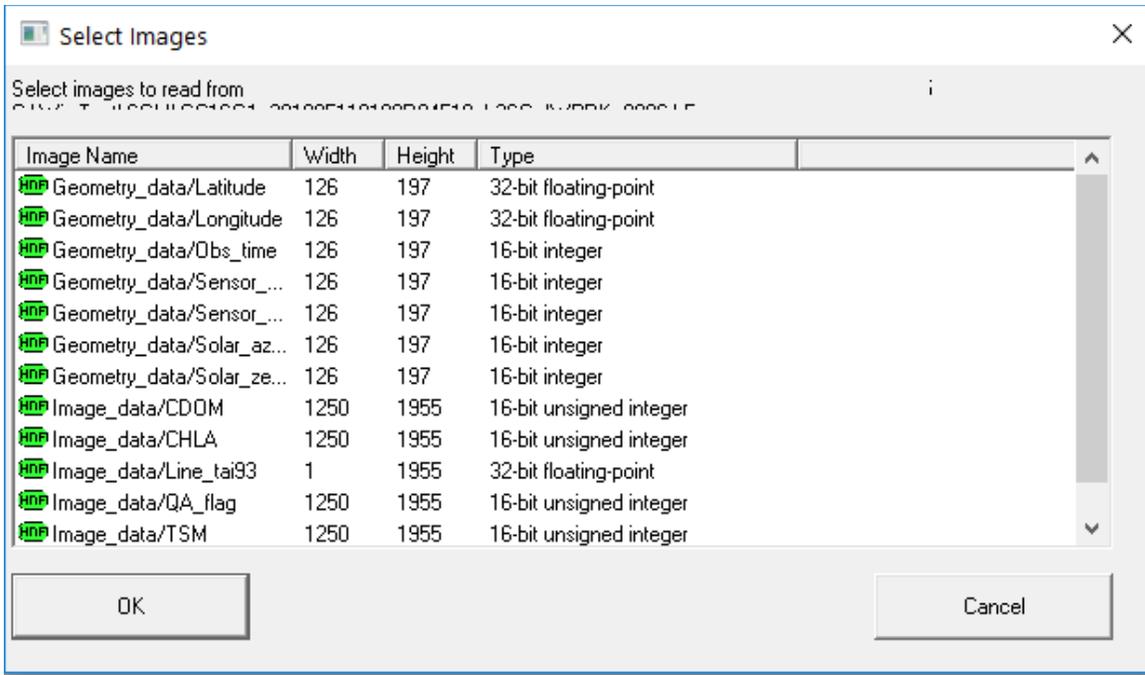
SGLI (second generation GLI) is the new color sensor on the JAXA GCOM-C satellite launched in December 2017 ([http://global.jaxa.jp/projects/sat/gcom\\_c/](http://global.jaxa.jp/projects/sat/gcom_c/)).

Departing from the common format used for most satellite datasets, i.e. that of netCDF, SGLI data are distributed in the HDF5 (\*.h5) format. Support for HDF5 and SGLI has been added starting from WIM version 11 and is being extended to WAM programs.

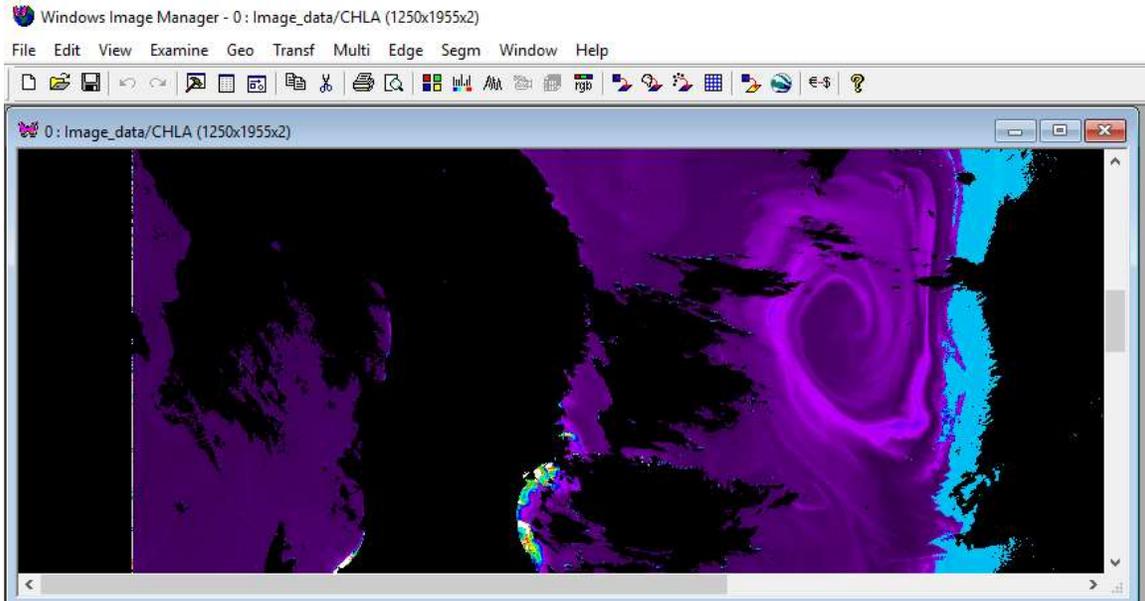
## 2 Evaluating SGLI Level-2 data

SGLI data are distributed in the HDF5 (\*.h5) format and can be read with WIM version 11 and later. You can read the SGLI datasets with the regular *File – Open – HDF5 file (\*.h5)* menu option. However, it is much faster to read them simply by clicking (double-clicking) on the .h5 file in Windows Explorer. To do that you need to associate WIM with the .h5 file extension – right-click on any .h5 file name and select *Properties*. Then select *Open with:* and find WIM.exe in *Program Files (x86)* directory. Then double-click (click) on the file name and – BE CAREFUL! - select that you read the file in WIM with *Open – HDF5 file (\*.h5)* and not as some other file type. After doing that you can always just click (or double-click) on a .h5 file and see the available the datasets that you can load. As a demonstration, we will read a 1 km resolution file GC1SG1\_201805110109R04510\_L2SG\_IWPRK\_0006.h5. A corresponding 250-m resolution file is named GC1SG1\_201805110109R04510\_L2SG\_IWPRQ\_0006.h5, i.e. “Q” instead of “K”.

When I click (double-click) on the file, I see the following available datasets:



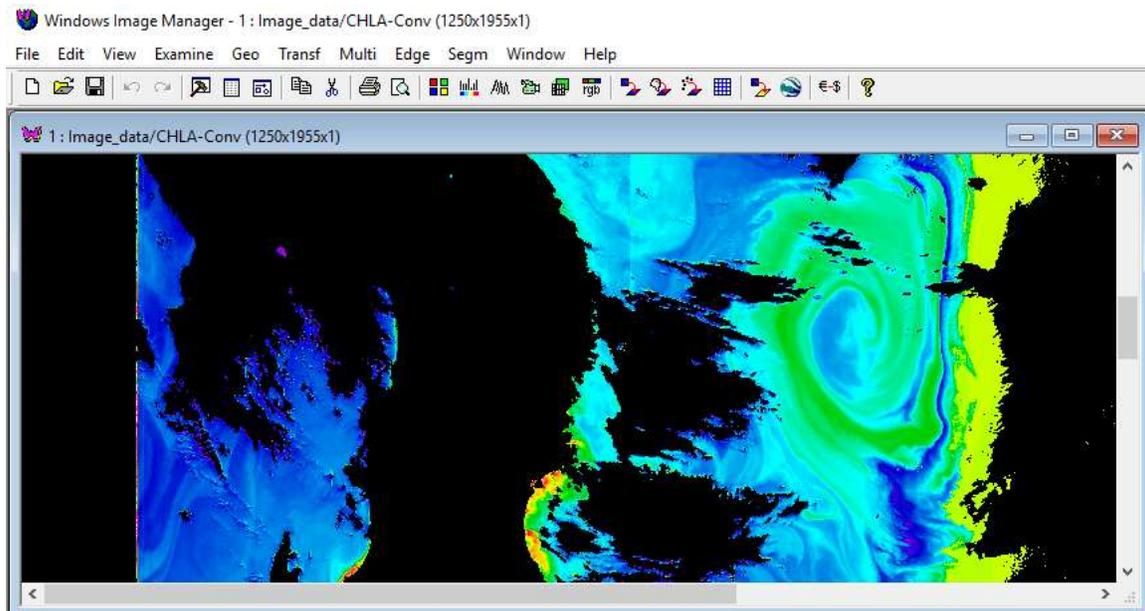
You can recognize *Latitude*, *Longitude*, *CDOM*, *CHLA* and others. Notice that these names are preceded by a “group” or “directory” name like *Geometry\_data/* or *Image\_data/*. Now select *CHLA* and we will see something like that:



This is Chla concentration in linear scaling. You can see the pixel values in both scaled integers and as Chla concentration (in brackets) by right-clicking on the image but to better visualize the

contrast we can convert the image to Log-scaling using the *Euro-\$* tool and *Log-Chl, mg/m3*. We will get something like that:

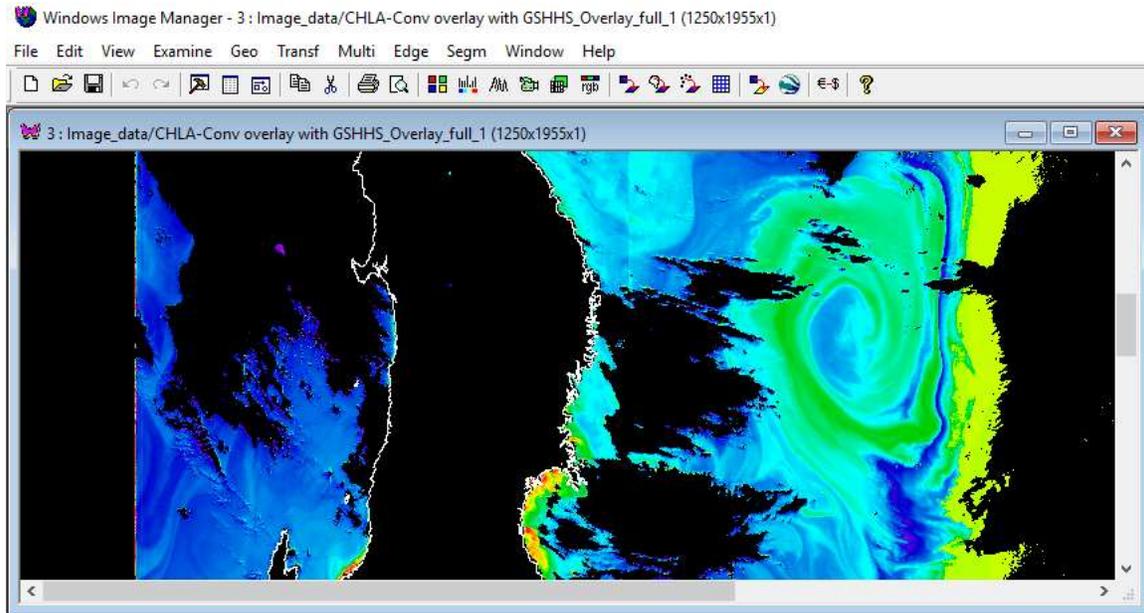




To check that we have the correct geo-location, we generate a coastlines image with *Geo-Get Map Overlay* with *coast\_full.b*, *Background Value* of 0 and *Foreground Value* of 255:

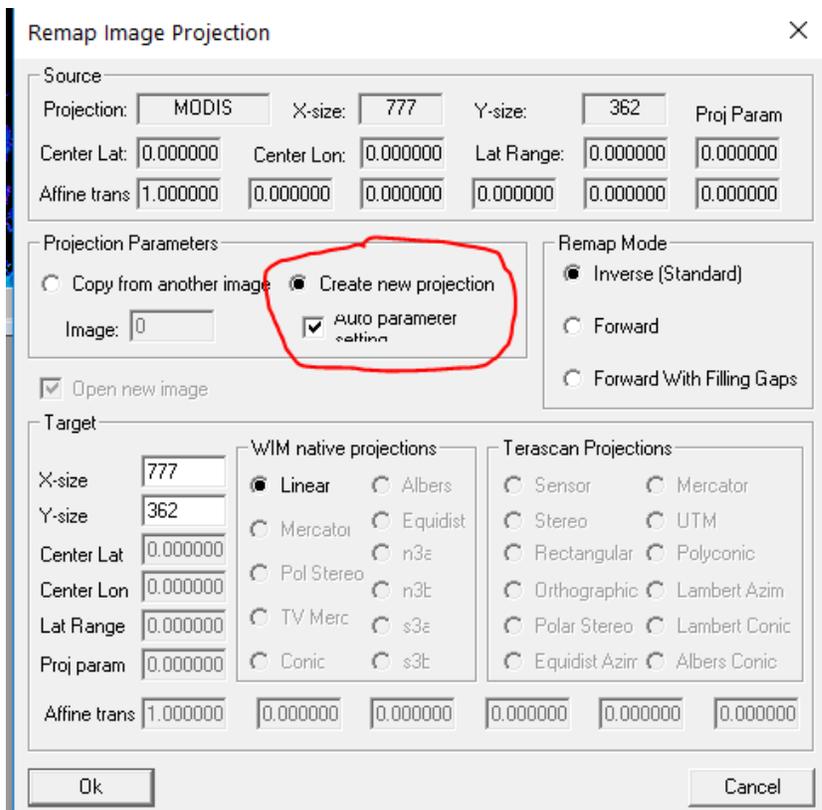


We now switch back (click) to the previous image and overlay the coastlines image with the Overlay tool . The coastlines match very well with the image, i.e. the geo-location is correct.

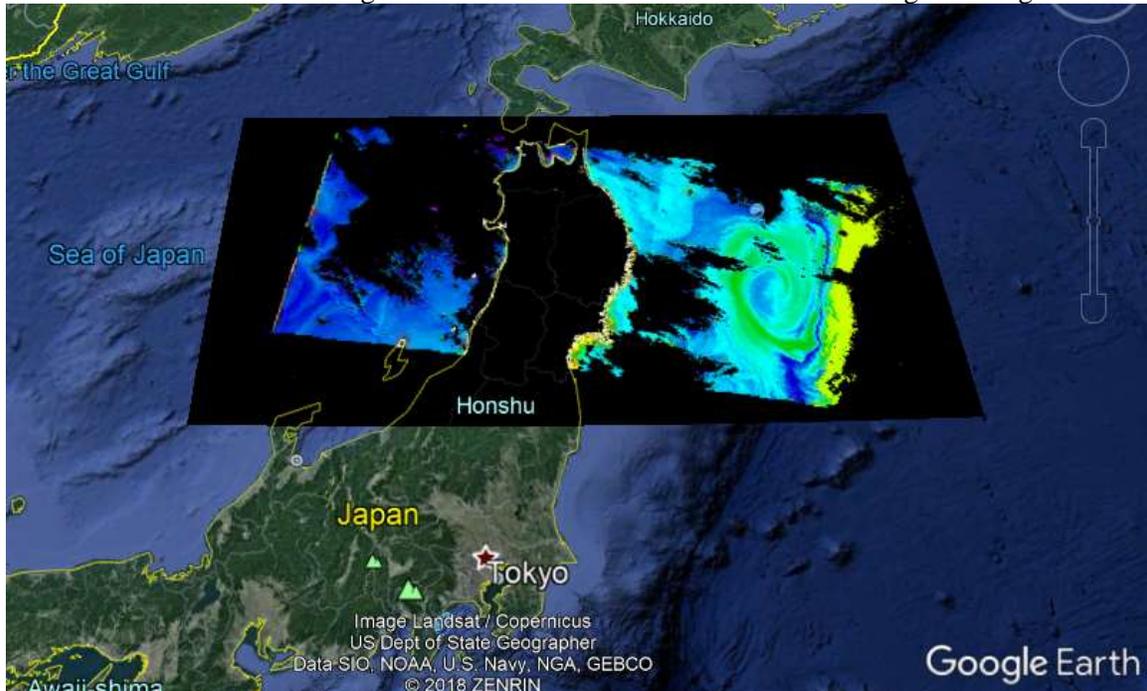


We can even view this image in Google Earth but for that the image has to be in *Linear* projection. To remap the image into a *Linear* projection we cut out the interesting part with the

Scissors tool on Toolbar  and then remap to itself using *Geo-Remap Projection* or  icon on Toolbar and then using the option *Create new projection*:



After that we click on the Google Earth icon  on Toolbar and see the image in Google Earth:



As you can see, we have a nice eddy seen by the Chla field east of the Japanese island Honshu. The yellow area east of the eddy seems to be too high and probably has some quality issues. We investigate that by looking at the quality flags and an image called *Image\_data/QA\_flag* in the same file *GC1SG1\_201805110109R04510\_L2SG\_IWPRK\_0006.h5*. Just click (or double-click) on the file name and select *Image\_data/QA\_flag*. After loading the image, stretch the colors with menu *View-LUT stretch*. Now scroll down and you should see something like that:



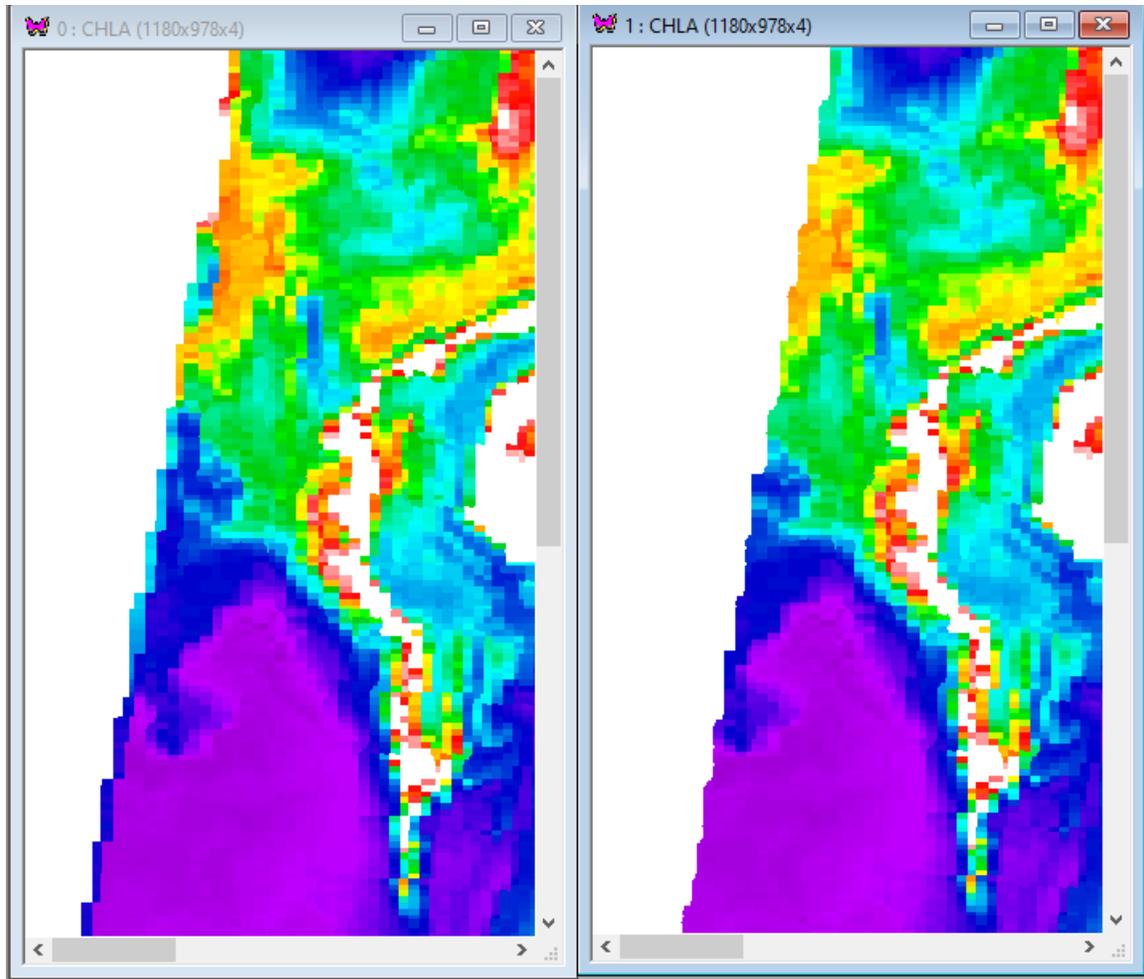
Now right-click on the different colors and you can see which flags are set (on). For example, the yellow area in the Chla image seems to have the *sunlint* flags set. Over land we have 3 lags set: *land*, *nodata*, *sunlint*. The white ellipsoidal area has *sunlint* and *shallow*. Some of the flags do not seem to be correct and there may be an error in the processing or reading the flags. This will be probably improved in the next processing.

As most WIM/WAM programs work with HDF4 (.hdf) files and not with HDF5 (.h5) files, it may be needed to save SGLI data to HDF4. You can save SGLI Level-2 data with correct geo-location by loading the images and saving in *HDF with Lat/Lon arrays* (\*.hdf). Do not save the *Geometry\_data/Latitude* and *Geometry\_data/Longitude* datasets as they will be saved automatically! Also, do not save the datasets of different sizes in the same file as the corresponding *Latitude* and *Longitude* arrays will be created and saved only for the first dataset! When saving SGLI Level-3 mapped datasets, save them as *HDF SDS* (\*.hdf) and NOT as *HDF with Lat/Lon arrays* (\*.hdf) as the image has been already mapped and has a projection and does not need separate *Latitude* and *Longitude* datasets. Saving those would only waste a lot of disk space and make geo-location slower.

Routine processing of Level-2 data should be done with command line programs like *wam\_l2\_map*. See a separate manual for [that](#). Here we show just a specific problem that seems to be causing at edges of SGI data swath. Therefore we use a special option to block 90 pixels at both sides of the data swath. The following is a sample command to extract CHLA from one SGLI Level-2 files and remap to a common map:

```
wam_l2_map F:\CAL\2021\SGLI\IWPRK\Good\GC1SG1_202106091816*.h5  
E:\Projections\California\Linear\ptConc_122.1_118.9w_32.9_35.1n_land_250m.hdf  
flags=1111100000001110 xpos=228 ypos=20 var=Image_data/CHLA dilate=0 swath=90
```

Note the option *swath=90* that blocks 90 pixels on both sides. The figure below shows a zoomed image with obviously bad pixels along the left side (left image) and after removal of 90 edge pixels (right image). Both images have been zoomed to better show the swath edge effects.



### 3 Finding match-ups in SGLI Level-2 data

For validating SGLI level-2 data we want to compare those values with *in situ* measurements. We can use a WAM command line programs `wam_match_l2` and `wam_read_match`. As currently I am not sure if the flags in `Image_data/QA_flag` are correct or we are not reading them correctly, we currently cannot use the flags. Let's assume that we have *in situ* data in Excel and have converted them to a CSV file called `points.csv` that looks like that in Excel:

Lon	Lat	Date	Time	Station	CHLA
142.462	40.168	5/10/2018	10:00	Sta1	0.194
143.56	39.085	5/10/2018	11:00	Sta2	1.05
141.011	38.031	5/10/2018	12:00	Sta3	11.074
141.192	38.362	5/10/2018	13:00	Sta4	34.458

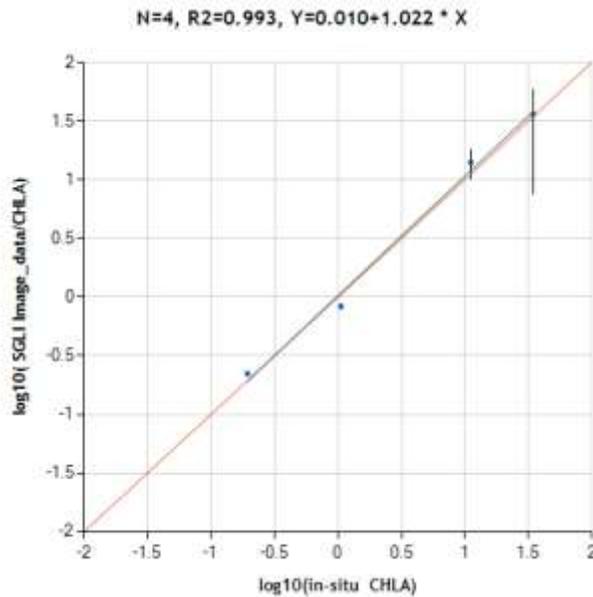
We have CHLA measurements from 4 stations. Note that *Longitude* and *Latitude* are both in decimal degrees and *Lon* is before *Lat*. Note that I am using *Date* is in the US format (MM/DD/YYYY) but other forms can be used as well. Time is always in GMT (UTC). These *in situ* values are actually fake as I manually picked pixel values in WIM with `Examine-Point Save`. `wam_match_l2` has many options which you can see by typing the name of the program in the command window. Here we use simplified options and ignore the flags:

```
wam_match_l2 points.csv Level2\GC1SG1*.h5 Image_data/CHLA maxDiffDays=5 noflags=yes
```

As input we specify *points.csv* (in situ data), matching .h5 files is the Level2 directory. *Image\_data/CHLA* is the name of the dataset that we are using, *maxDiffDays=5* specifies 5 days as the maximum time difference between the station (point) and satellite data and *noflags=yes* makes us ignore the flags dataset. The default for *maxDiffDays* is 30 days. The idea is to use a relatively long time interval (e.g. 5 days) here, collect at least some match-ups and later you can narrow the maximum time lag and quickly find match-ups with smaller time difference. As a result, we get a file *points\_SGLI.csv* with the match-ups. We can plot a scatter plot of the matchups with the following command:

```
wam_read_match points_sgli.csv all=yes maxDiffDays=1 plotNthcolumn=5 xMin=-2 xMax=2 hasFlags=no
```

Here we specify the matchups file *points\_sgli.csv*, force to plot all points irrespective of pixel to pixel variance (*all=yes*), use maximum time difference of 1 day (*maxDiffDays=1*), use column 5 (in situ CHLA, *plotNthcolumn=5*) as the X-axis, use log-scale limits -2 to 2 (i.e. 0.01 to 100 in linear units) and use no flags. You can suppress the min-max lines with *minMaxLine=no*. The scatter plots looks like that:



Of course, the plot looks almost perfect as I picked the *in situ* points from the same image (actually from the lower resolution version of it). There is substantial pixel to pixel variability in the highest CHLA region (the vertical min-max line is plotted for 3x3 pixel region) but almost no variability at lower CHLA levels. Note that if we don't use the *all=yes* option then the point with highest CHLA would be excluded due to high pixel to pixel variability. Note: this is a preliminary evaluation not using quality flags.

## 4 SGLI data in JASMES (\*.nc) format

While most SGLI data are in HDF5 (.h5) files, some of the data are being released after remapping in the more easily accessible netCDF4 files (.nc). Both global 5 km and local Japan 250 m data are available at

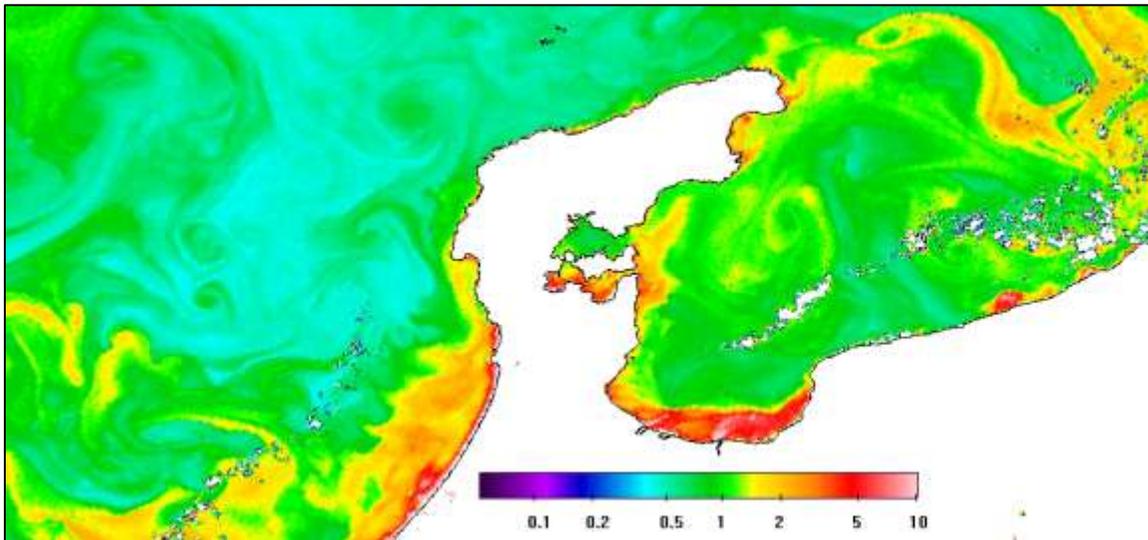
[https://www.eorc.jaxa.jp/JASMES/SGLI STD/about\\_sglistd.html](https://www.eorc.jaxa.jp/JASMES/SGLI STD/about_sglistd.html)

The Japan 250 m data are mapped on the following grid: Size of pixel (same as grid interval): 0.05 degree; Latitude and longitude of four corners of the global image (center of each corner pixel): Upper left: Lon: 49.99875 deg.E, Lat: 123.00125 deg.N, Lower right: Lon: 24.00125 deg.E, Lat: 149.99875 deg.N. WIM version 11.21 (April 2010) has made it easy to read these 250 m Japan area files and automatically retrieve geo-location. Geo-location and other coefficients are read from the attributes:

### Attributes of CHLA-Conv

Attribute Name	Owner	Type	Data
scale_factor	<GLOBAL>	float32	0.001600
add_offset	<GLOBAL>	float32	0.000000
Long_name	<GLOBAL>	char8	Chllorophyll-a concentration
Spatial_resolution	<GLOBAL>	float32	250.000000
Lower_left_latitude	<GLOBAL>	float32	30.000000
Lower_left_longitude	<GLOBAL>	float32	135.000000
Lower_right_latitude	<GLOBAL>	float32	30.000000
Lower_right_longitude	<GLOBAL>	float32	145.000000
Upper_left_latitude	<GLOBAL>	float32	40.000000
Upper_left_longitude	<GLOBAL>	float32	135.000000
Upper_right_latitude	<GLOBAL>	float32	40.000000
Upper_right_longitude	<GLOBAL>	float32	145.000000

As the 250 Japan area images are big (4000 x 4000 pixels), we show a small part of it after converting to *Log-Chl* scaling and overlaying the coastline (with pixel value 1). Below is a part of a CHL image in GC1SG1\_201903050119004900\_L2MG\_IWPRQ\_1000\_CHLA\_11.nc, i.e. 5-Mar-2019, showing Toyama bay and neighboring areas on Honshu Island.



## 5 Case study evaluating SGLI Chla in the California Current area

The above evaluation of SGLI Level-2 data was done with imaginary field data. Here we evaluate the SGLI CHL algorithm with real in situ data collected in the California Current area. We used in situ Chla data from 1036 stations in 2018 and 2019 against Level-2 SGLI datasets in IWPRQ files. We obtained 224 match-ups within  $\pm 1$  day and 27 match-ups within  $\pm 3$  hours. The locations of these match-up stations and the scatter plots are shown below.

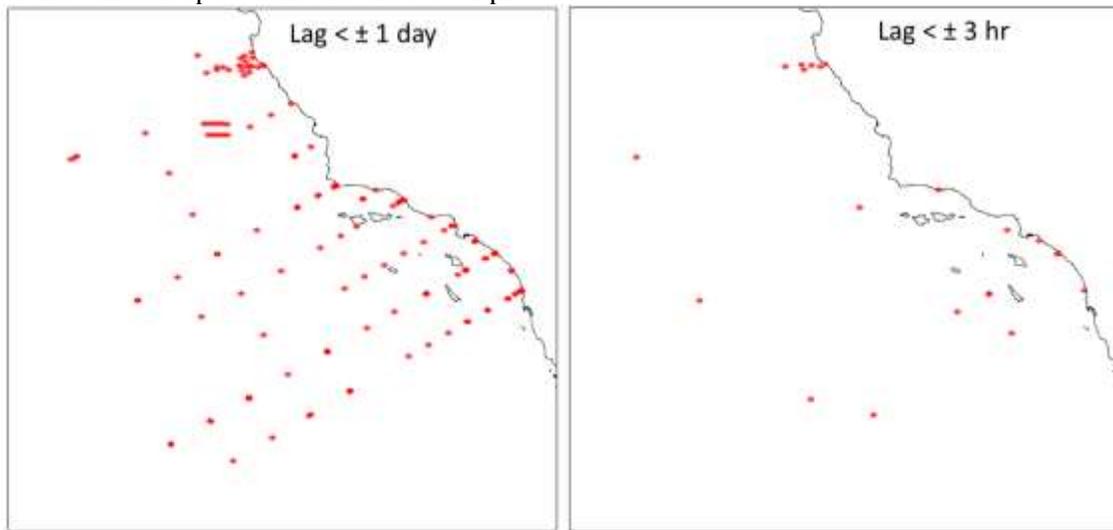


Fig. Locations of CHLA match-ups within  $\pm 1$  day (224 match-ups, left panel) and  $\pm 3$  hours (27 match-ups, right panel) between SGLI Level-2 IWPRQ data and in situ CHLA during 2018-2019.

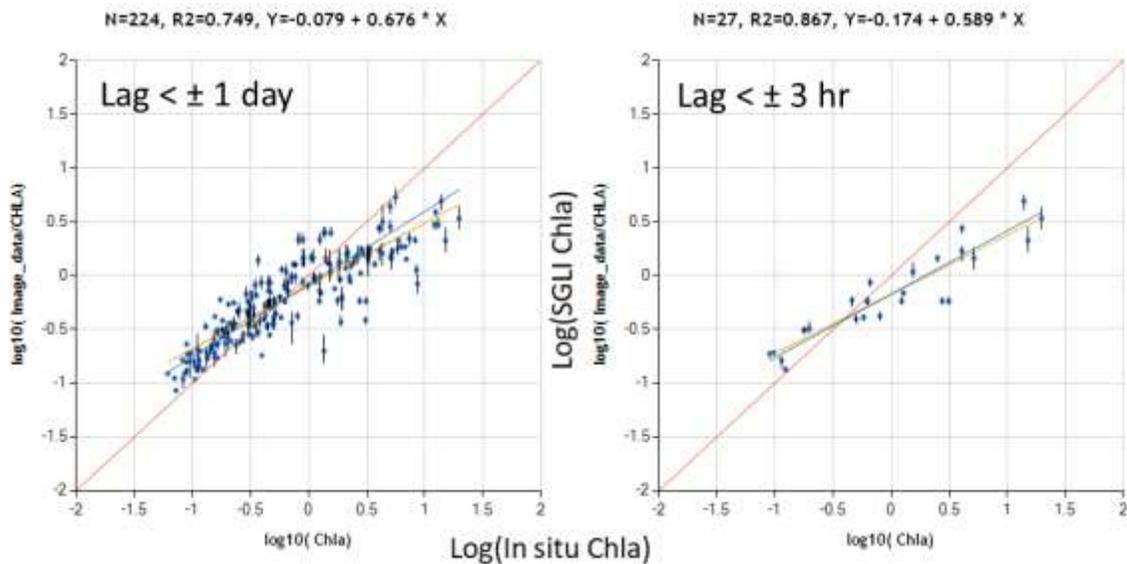


Fig. Scatter plots in log10 scales of SGLI CHLA in IWPRQ against in situ surface CHLA value within  $\pm 1$  day (left panel) and within  $\pm 3$  hours (right panel). The point along the vertical axis

represents the mean and the vertical line represents the range for each 3x3 pixel window of SGLI IWPRQ data.

The commands to produce these results are the following. First we find the matchups separately for each year (2018 and 2019) within  $\pm 5$  days. We then manually merge the CSV files from the different years.

```
wam_match_l2 top.csv IWPRK\G*.h5 Image_data/CHLA maxDiffDays=5 noflags=yes
```

We then make plots of within  $\pm 1$  days and  $\pm 3$  hours (0.125 days), selecting only those matchups with at least 5 valid pixels out of 9 (`nValidFrom=5`). This eliminates some matchups at cloud edges. Using the `saveSelected=yes` option saves the selected subsets of matchup to make spatial plots on a map image manually using WIM and *Geo-Get Vector Objects*.

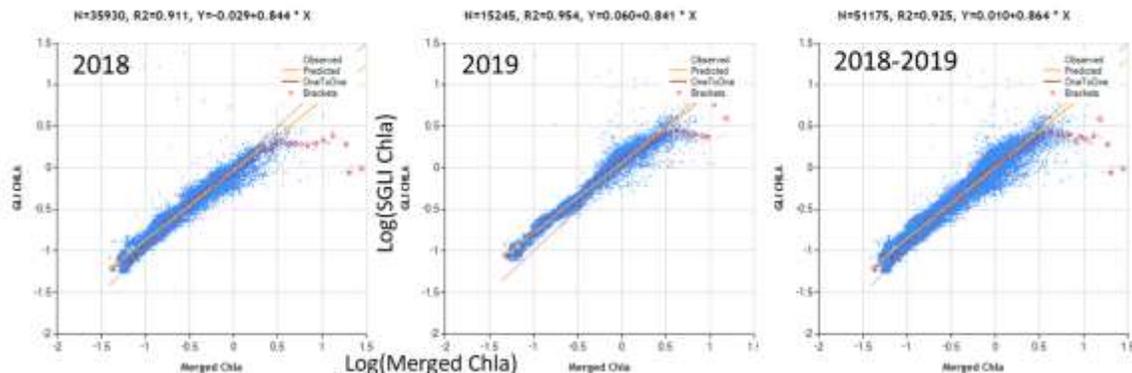
```
wam_read_match top_SGLI.csv maxDiffDays=1 plotNthcolumn=13 xMin=-2 xMax=2 yMin=-2 yMax=2 hasFlags=no maxDiffDays=1 nValidFrom=5 saveSelected=yes
```

```
wam_read_match top_SGLI.csv maxDiffDays=1 plotNthcolumn=13 xMin=-2 xMax=2 yMin=-2 yMax=2 hasFlags=no maxDiffDays=0.125 nValidFrom=5 saveSelected=yes
```

We confirm that the accuracy of SGLI CHLA estimates is similar to that of NASA and NOAA sensors but we detect a significant concentration-dependent bias. The SGLI standard CHLA values over-estimate in situ values at low CHLA and under-estimate in situ values at high CHLA. The  $R^2$  of 0.867 in 3 hr match-ups is already quite good and eliminating some match-ups is not likely to have a major influence. The reduced resolution (1-km, IWPRK) data have similar match-ups but slightly (insignificantly) lower  $R^2$ .

We can also evaluate SGLI CHL directly against other sensors using `wam_pixelwise_match`. To do that we need to map the daily L2 data to a common map (e.g. with `wam_l2_map`). We compare SGLI CHL against merged Chla from a number of other sensors (MODISA, MODIST, VIIRS\_SNPP, VIIRS-JPSS1, OLCI-A, OLCI-B). Be careful not to match too many (e.g. millions of pixels) as that may stall the plotting routine and use a mask file that has pixels from low to high concentrations. I use a mask file that looks like a strip from coast to offshore. To plot SGLI CHL against merged Chla we use a command like that:

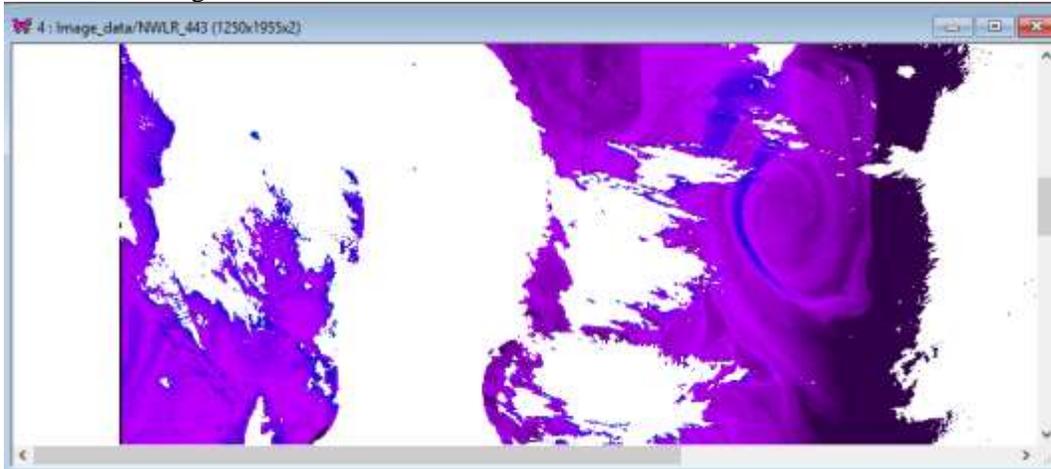
```
wam_pixelwise_match C2018_chl_day\C*annot.hdf G2018_chl_day\G*annot.hdf mask.hdf xMin=-1.5 xMax=1.5 yMin=-1.5 yMax=1.5 median1=yes median2=yes
```



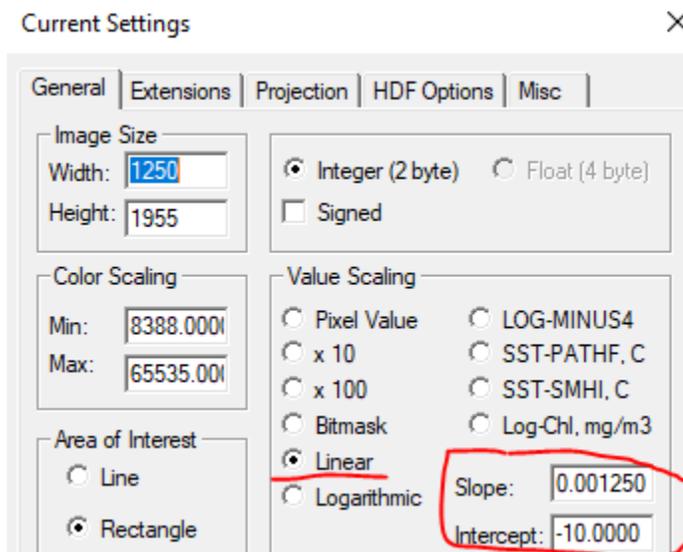
The plot above shows pixelwise scatter of same day CHL of SGLI versus the merged Chla of other sensors. The SGLI standard CHLA values over-estimate compared to other sensors at low CHLA. There may be some under-estimation at high CHLA but there is considerable scatter.

## 6 SGLI Level-2 NWLR data

You can work with SGLI level-2 *Normalized Water Leaving Radiance* (NWLR) data but keep in mind that there are currently (October, 2018) some issues with the flags (*Image\_data/QA\_flag*). When you load any NWLR image from a *\*NWLRK\*.h5* or *\*NWLRQ\*.h5* file and a part of the *NWLR\_443* image looks like that:



You can check the settings that look like that:



The image has 2-byte unscaled integer pixels with *Linear* scaling (slope=0.00125 and intercept=-10).

If you look at the *Attributes* with the  icon, you will see:

Offset	Image_da...	float32	-10.000000
Slope	Image_da...	float32	0.001250
Rrs_slope	Image_da...	float32	0.000001
Rrs_offset	Image_da...	float32	-0.005373
Rrs_unit	Image_da...	char8	sr <sup>-1</sup>
Data_description	Image_da...	char8	Normalized Water Leaving Radiance(NWLR) at ...
Unit	Image_da...	char8	W/m <sup>2</sup> /sr/um

As you see the WIM *Intercept* is called *Offset* in the attributes. The attributes also give you the coefficients to convert NWLR to Rrs.

We will now find NWLR match-ups using the same *Points.csv* file that we used for CHLA. We run *wam\_match\_l2* with the following arguments:

```
wam_match_l2 points.csv Level2\GC1SG*NWLR*.h5 Image_data/NWLR_380
Image_data/NWLR_412 Image_data/NWLR_443 maxDiffDays=1
```

Here we have chosen *maxDiffDays=1* and 3 bands (*NWLR\_380*, *NWLR\_412*, *NWLR\_443*) but we can add all the bands present (also bands at 490, 530, 565, 670 nm).

This is using default *QA\_flag* values that are currently set like that (1 means that the flag makes the pixel invalid, 0 makes the flag to be ignored):

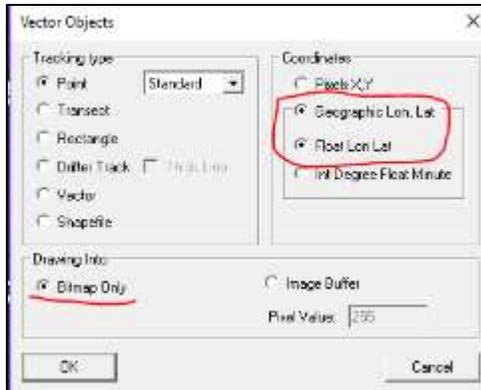
```
Value,Name
1,DATAMISS
1,LAND
1,ATMFAIL
1,CLDICE
1,CLDAFFCTD
0,STRAYLIGHT
0,HIGLINT
0,MODGLINT
0,HISOLZ
0,HISENZ
0,HITUAU
0,EPOUT
0,OVERITER
0,NEGNLW
0,HIGHWS
0,SPARE
```

You can select your own “bad” flags that make a pixel invalid, save in a text file like the list above and use the following command:

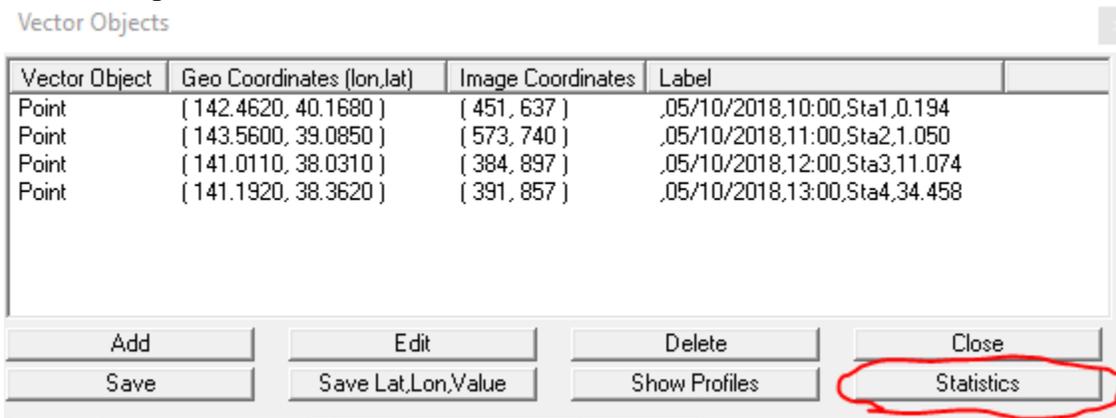
```
wam_match_l2 points.csv Level2\GC1SG*NWLR*.h5 Image_data/NWLR_380
Image_data/NWLR_412 Image_data/NWLR_443 maxDiffDays=1 badflags=myflags.csv
```

Here *myflags.csv* has the list of your flags, i.e. numbers of 1 or 0 in each line. Using the option *noflags=yes* ignored the flags in the *QA\_flag* dataset completely.

You can see the locations of the match-up points on the image if you have the image open and load the *points.csv* or *points\_SGLI.csv* with *Geo-Get Vector Objects*:

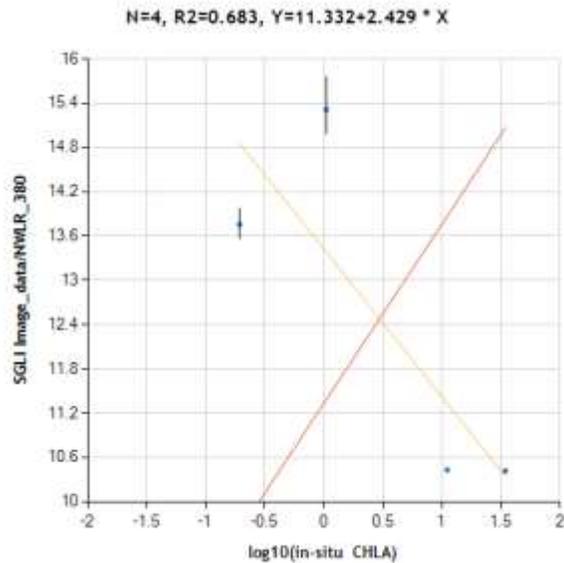


If you click on the object in the *Vector Objects* table, it starts to blink in the image. You can also do a simple statistics for a 3x3 pixel neighborhood by selecting an object (point) in the table and then selecting *Statistics* button:



*maxDiffDays=X* specifies X days as the maximum time difference with in situ data. It is recommended to use a longer X (more inclusive) and make your exclusions later with *wam\_read\_match*. For example, we can use following command:

```
wam_read_match points_sgli.csv all=yes maxDiffDays=1 plotNthcolumn=5 xMin=-2 xMax=2
logY=no yMin=10 yMax=16
```



Here we have specified the matchups file *points\_sgli.csv*, use maximum time difference of 1 day, use column 5 (in situ CHLA) as the X-axis, use log-scale limits -2 to 2, but linear Y axis from 10 to 16. Of course, it does not make much sense to plot NWLR\_380 against in situ CHLA. Also, keep in mind that NWLR values are currently not using the offset of -10.

## 7 Evaluating SGLI Level-3 data

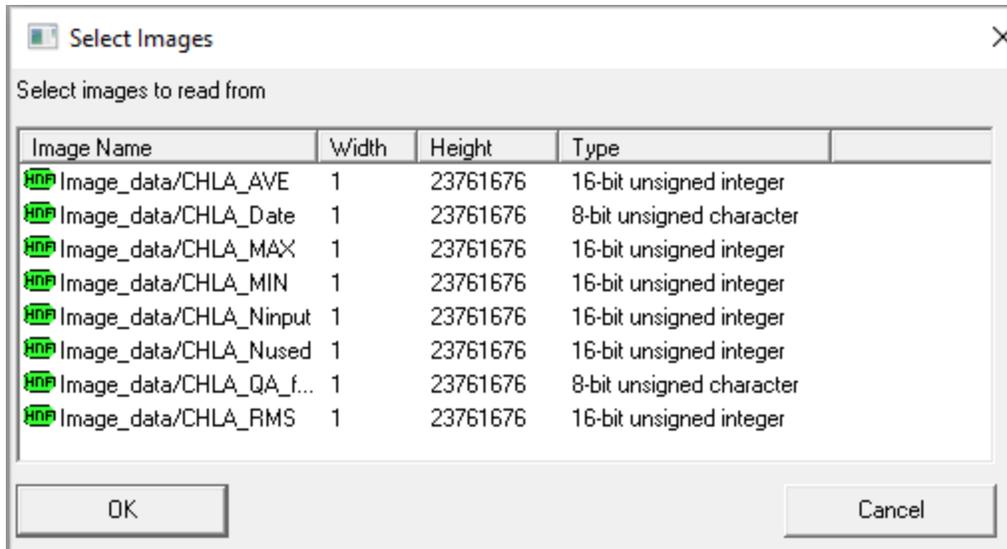
Working with SGLI Level-3 (i.e. binned and possibly mapped) data is beyond this manual but we include a few notes here because in a previous version of WIM (11.10) the projection was not correctly detected.

Let's look at some sample files:

GC1SG1\_20180701D01D\_X0000\_3**BSG**\_CHLAF\_0006.h5

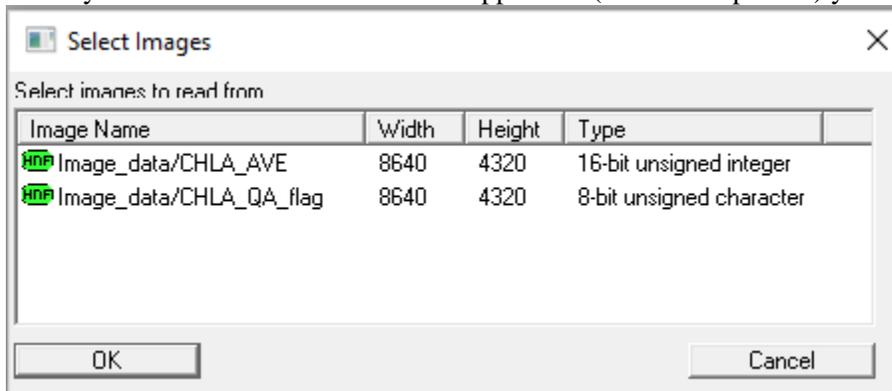
GC1SG1\_20180701D01M\_D0000\_3**MSG**\_CHLAF\_0100.h5

Note the bold B in the first file name and the bold M in the 2<sup>nd</sup>. The “**B**” means binned and the “**M**” means mapped. You can try to open the binned (not mapped) CHLA file and you can see the contents:

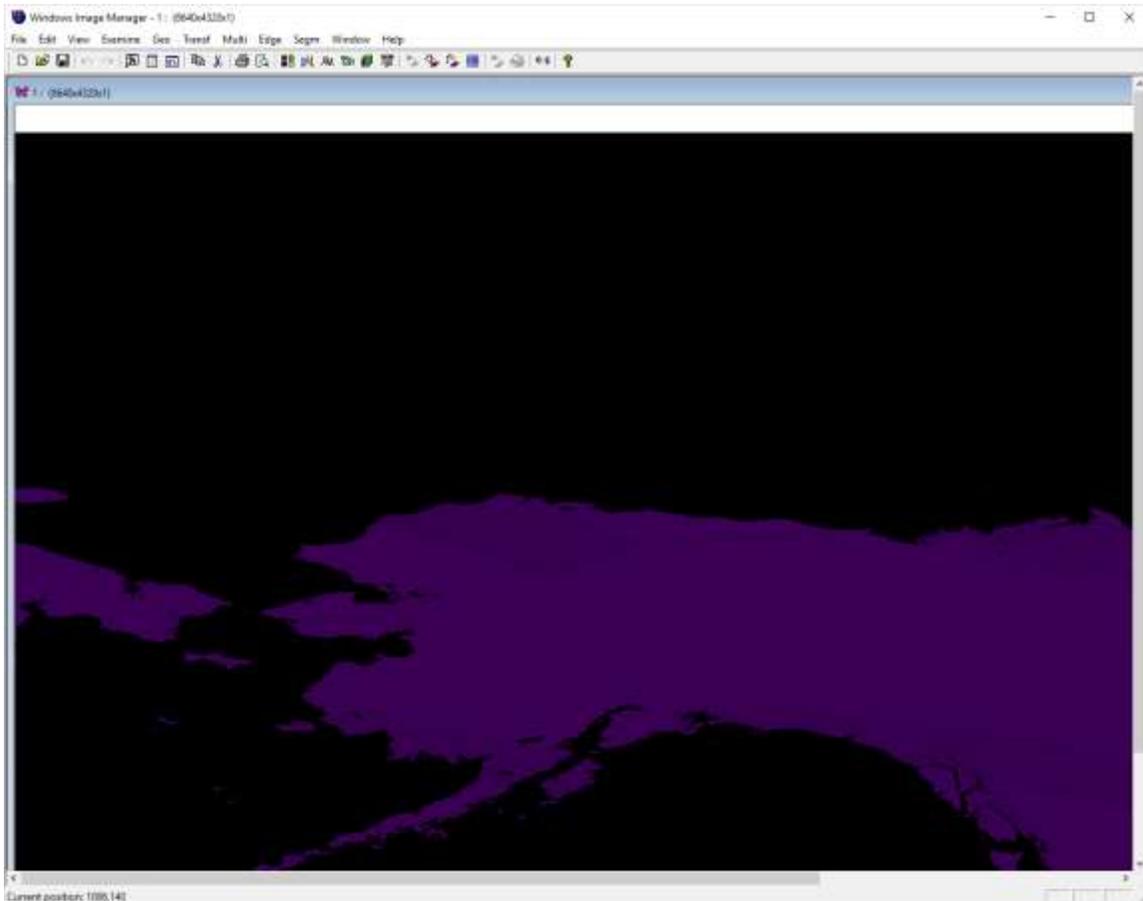


You see several vectors (width=1) but you cannot visualize them easily as they are not presented as matrices.

When you click on the file name of a mapped file (the 2<sup>nd</sup> sample file) you will see the contents:



Here you see the averaged dataset (*CHLA\_AVE*) and the flags (*CHLA\_QA\_flag*), both with a size of 8640 x 4320. These are the mapped datasets with the nominal ~4.5 km spatial resolution at the equator. Select both of them and you will see something like that:



You can recognize the coastline of Alaska in the *QA\_flag* image. Right-clicking on the image shows that ocean has pixel value 0, land has 3, white areas (no data) have value 255. You can

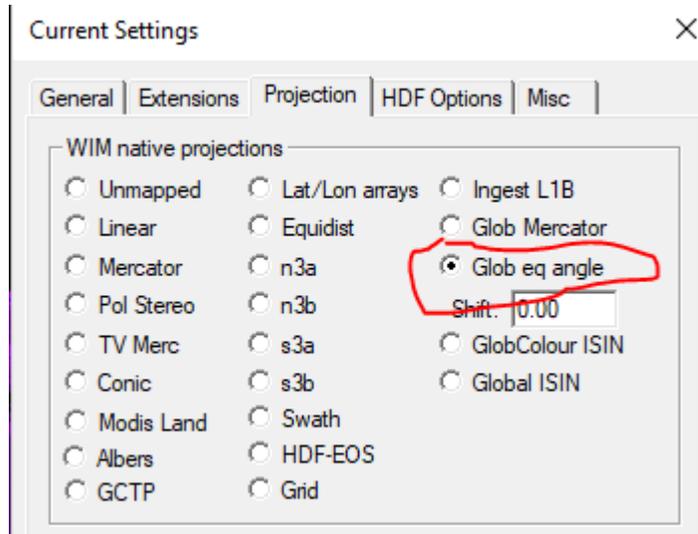
click on the Attributes icon  on the Toolbar and see the attributes:

Attributes of GC1SG1\_20180701D01M\_D0000\_3MSG\_CHLAF\_0100.h5

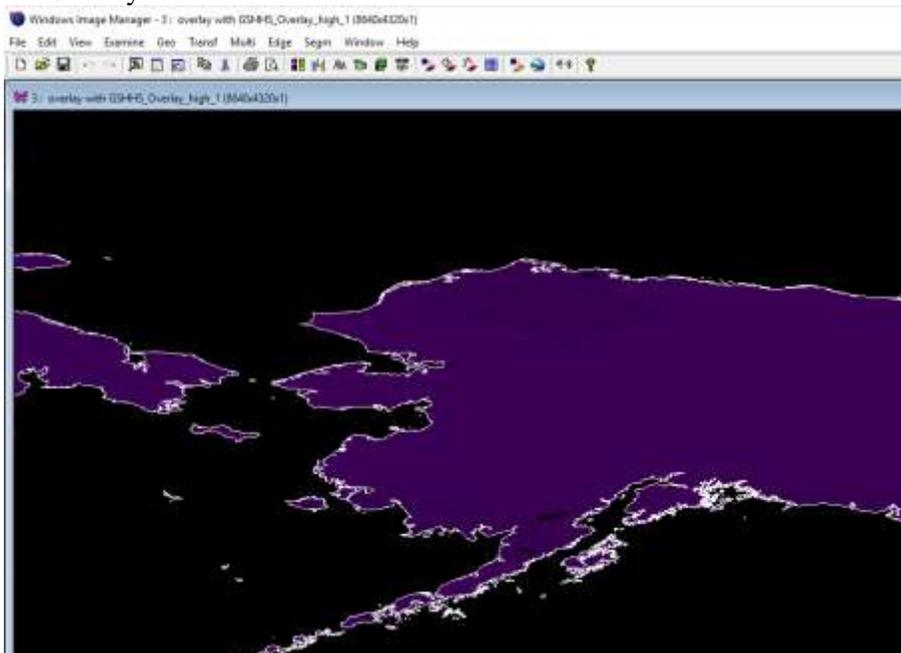
Attribute Name	Owner	Type	Data
Algorithm_developer	<GLOBAL>	char8	Japan Aerospace Exploration Agency (JAXA)
Dataset_description	<GLOBAL>	char8	Level-3 daily temporal binning file of CHLA
Product_file_name	<GLOBAL>	char8	GC1SG1_20180701D01M_D0000_3MSG_CHL...
Product_level	<GLOBAL>	char8	Level-3
Product_name	<GLOBAL>	char8	Level-3 binned map product
Product_version	<GLOBAL>	char8	0100
Satellite	<GLOBAL>	char8	Global Change Observation Mission - Climate (G...
Sensor	<GLOBAL>	char8	Second-generation Global Imager (SGLI)
Algorithm_version	<GLOBAL>	char8	1.00
Parameter_version	<GLOBAL>	char8	001.00
Image_start_time	<GLOBAL>	char8	20180701 00:07:36.910
Image_end_time	<GLOBAL>	char8	20180710 00:03:27.779
Grid_interval	Image_da...	float32	0.041667
Grid_interval_unit	Image_da...	char8	deg
Image_projection	Image_da...	char8	Equal Rectangular projection
Number_of_lines	Image_da...	int32	4320
Number_of_pixels	Image_da...	int16	8640

Buttons: Add, Edit, Delete, Copy To..., Close

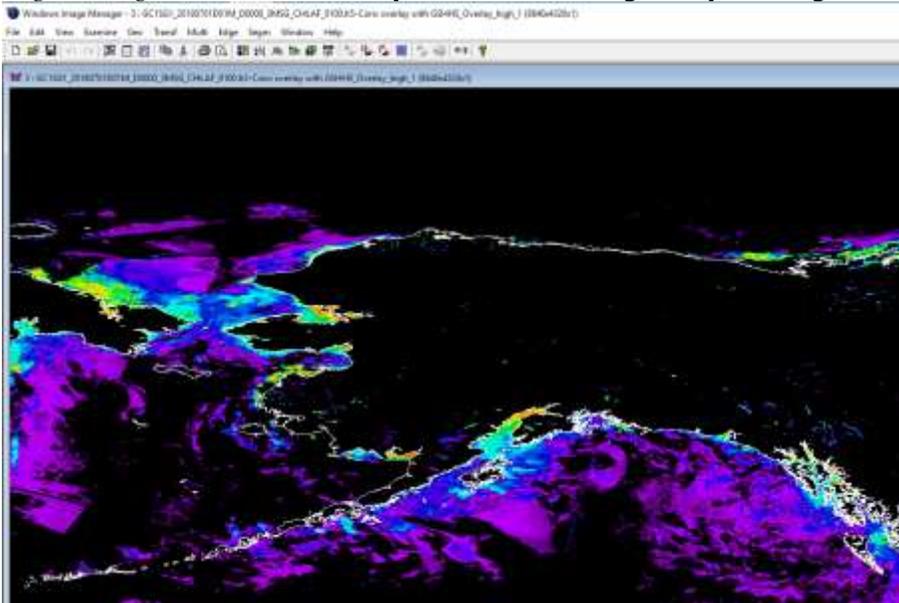
Right-clicking on the image shows the pixel values and their longitude and latitude. You can check the projection type in *Settings* (the Hammer icon), choose *Projection*, and see that *Glob. eq. angle* (*Global Equal Angle* has been selected which is the same as *Equal Rectangular projection*, see the attributes):



After doing that you can see the longitude and latitude by right-clicking on the image. You can also create the coastlines image by selecting the Map Overlay icon  on Toolbar and then picking a coastlines file, e.g. *coast\_high.b* and selecting Background value of 0 and foreground value of 255. Next click back on the source image (e.g. *QA\_flag*) and overlay the coastlines with the Overlay tool  :



You can see that the coastlines match the image very well. You can use the same coastlines image to overlay the *CHLA\_AVE* image but it is better to convert it first to log-scaling with  and *Log-Chl, mg/m3*. After the overlay the coastlines image and you will get something like that:



Automatic detection of the projection (geo-location) will be added in the next version. Multiple command-line tools will be also extended to these .h5 files that allow operations like compositing, remapping, match-up finding, etc.