Validation of satellite data with WIM/WAM – Japan East Sea

Contents

1 Introduction

Satellite measurements are often very indirect measurements of the variable that we want to estimate. For example, we use top of the atmosphere radiance to estimate the concentration of chlorophyll-a (Chl-a) in the water. It is known that about 90-95% of the radiance (photons) reaching the sensor at the top of the atmosphere is backscattered from the atmosphere and represent noise for estimating ocean parameters. Only 5-10% of the radiance at the top of the atmosphere can carry information about the water after being affected by absorption and scattering in the water column. Estimates of water properties by radiance at the top of the atmosphere have to be extremely accurate in order to prevent small errors in radiance measurements becoming big errors in estimated in-water properties. Satellite measurements of radiance are affected by errors in the calibration, atmospheric correction, estimated sea surface effects, adjacency effects, etc. In contrast to variable like Chl-a concentration, primary production (PP, mg C m^{-2} d^{-1}), is even more difficult to estimate remotely as it is a rate (mg C m^{-2} d^{-1}) and not a concentration, and the PP algorithm uses multiple inputs (e.g. Chl-a, PAR, etc) and model relationships to estimate the rate of PP integrated through the water column. In conclusion, satellite measurements are essential to make regional and global estimates of various important variables but often have large errors and need validation, i.e. comparison with in situ measurements. Comparison with in situ point measurements is often called a *match-up* analysis. WIM/WAM provides tools for validation and match-up analysis.

2 Prerequisites

We assume that you are familiar with the basics of WIM and of the Windows command prompt. If not, please check out the WIM and WAM manuals. We also assume that you have copied the *Images* and *Sat* folders from the DVD to the root of your C drive (e.g. *C:\Images* and *C:\Sat*.

Let's assume that we had a cruise in the Japan East Sea and collected surface Chl-a data measured with both the fluorometric and the HPLC methods. The data are in a spreadsheet *C:\Sat\SeaWiFS\L2\JES\jes9906_Chl_surf.csv*. This is a comma delimited text file (*.csv). As comma is used instead of a decimal point in many regions, there is a version of the same data in tab-delimited format in *C:\Sat\SeaWiFS\L2\JES\jes9906_Chl_surf_tab.csv*. Please open the spreadsheet file and familiarize yourself with the structure of the file:

Lon Lat DateGMT TimeGMT Cruise Station Cast Chla tChlHPLC

Note that *Longitude* (in decimal degrees) is first, followed by *Latitude* (in decimal degrees). Date and Time are both in GMT (UTC) and Date is in the US format MM/DD/YYYY. Time is in a simple 24 hour format. Conversion between date and time formats is often a problem. After the *Lon*, *Lat*, *Date* and *Time* we have the *Cruise*, *Station* and *Cast* columns. In this case the Cast column is not used but sometimes you may have multiple casts per station. The measured data are in *Chla* (fluorometric Chl-a) and *tChlHPLC* (total Chl-a from HPLC). Our task is to find matching satellite data to these 39 stations of surface Chl-a.

3 Using WIM interactively

3.1 Right-click

The easiest and most primitive way of a match-up analysis is to load an image into WIM and use the right click of the mouse to look up pixel values (see exercise 3.2 in *Practical Exercises with WIM and WAM*). For example, load some of the images in C:\Sat\SeaWiFS\L2\JES\Monthly and try to match the latitude/longitude of a station in the *jes9906_Chl_surf.csv* spreadsheet to a pixel (point) in the image. As you can see, it is extremely time consuming and error-prone.

3.2 Using Geo-Get Vector Objects

A little more advanced method for finding match-ups is to use the *Geo-Get Vector Objects* menu option in WIM (exercise 3.10 in *Practical Exercises with WIM and WAM*).

- Load the composited SeaWiFS Chl-a image *S1999156_S1999181_chl_comp_annot.hdf* in *C:\Sat\SeaWiFS\L2\JES\Monthly*. Use menu option *Geo-Get Vector Objects-Point* (*Bitmap Only*, *Geographic Lon*, *Lat*, *Float Lon Lat*) to load the station data from *jes9906_Chl_surf.csv*. If you don't get any match-ups then there is something wrong. For example, WIM may assume that you have *Longitude* first and *Latitude* second. Use the *Settings* (the Hammer icon on the Toolbar)-*Misc* and uncheck the *Lat first* checkbox.
- You should get a something like Fig.1 with small colored circles showing the stations on the image.
- Select a point in the *Vector objects* table and see which one is blinking. Select *Statistics* for the selected point. You need to make sure that the valid range is properly specified. For a log-Chl scaled byte image the proper range is Min = 0.011, Max = 64. The statistics is calculated for a 3 x 3 pixel window centered at the point. Note the values of ***N (in-range*** = number of pixels in the valid range) and ***N (out-range*** = number of pixels **out** of the valid range). As we

are using a monthly composite, the image is fairly good and has few missing pixels due to clouds.

- You can select one or more (or all) the points and save all the statistics into a new file, e.g. *test.csv* with the *Save Lat,Lon,Value* button (**not** the *Save* button that saves it in a HDF format). Load the saved *test.csv* file into a text editor, e.g. *Notepad*, or MS *Excel* and verify the file format. The header line has the names of the columns:

Lon Lat Value N_Valid N_Invalid Min Max Mean Median

Reading it into *Excel* is a bit tricky as there is space between *Lon* and *Lat* and tab between the other columns.

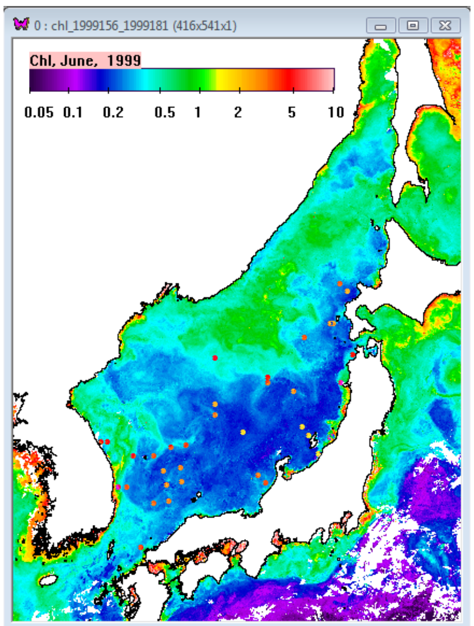


Fig. 1. Picking stations from an image with *Geo-Get Vector Objects*.

Using *Geo-Get Vector Objects* interactively in WIM works fine if you have a number of in situ values that you want to match with a **single** satellite image. In reality, you typically have many satellite images and the closest in time may be cloudy, so that and you need to match your in situ points with many other images. Which image to pick for each of the points becomes a difficult problem to solve as you need to consider (1) is the image area corresponding to the point clear; (2) which match-up image to pick (going forward in time or back in time) if you have more than one clear scene.

4 Using WAM GUI applications

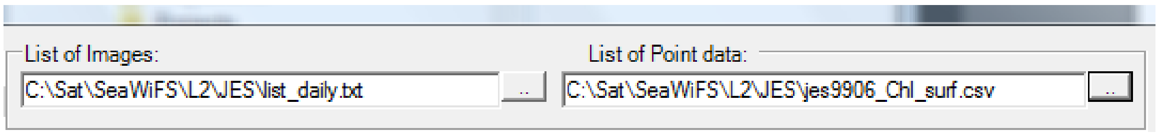
4.1 Using wam_match

Please see exercise 4.2 on *wam_match* in *Practical Exercises with WIM and WAM*. In *wam_match* you use a ***List of images*** (to specify the image files that you want to use), a **List of Point data** with longitude, latitude, date and time (e.g. *match.pnt*). The match-up output can be saved in a CSV file (e.g. *jes9906_Chl_surf.csv*). You can select the maximum time lag allowed (e.g. 28 hr), the size of the pixel window to consider (e.g. 3 x 3), the minimum number of valid pixels required (e.g. 5). You can then visualize the match-up points in a X-Y scatter plot, select, examine and eliminate individual match-up points, save in various formats.

You can run wam_match with the default setup and repeat exercise 4.2 in [*Practical Exercises with WIM and WAM*]. Before you can use *wam_match* with our JES data you need to create the list of images. We assume that your CSV file with station data is in *C:\Sat\SeaWiFS\L2\JES* and the satellite data you want to use are in *C:\Sat\SeaWiFS\L2\JES\Daily*. Use the following commands to create a list of images *list_daily.txt*:

```
cd C:\Sat\SeaWiFS\L2\JES
dir /b /s C:\Sat\SeaWiFS\L2\JES\Daily\*annotated.hdf > list_daily.txt
```

Now select the proper *List of Images* and *List of Point* data, e.g.



You should get output like Fig. 2.

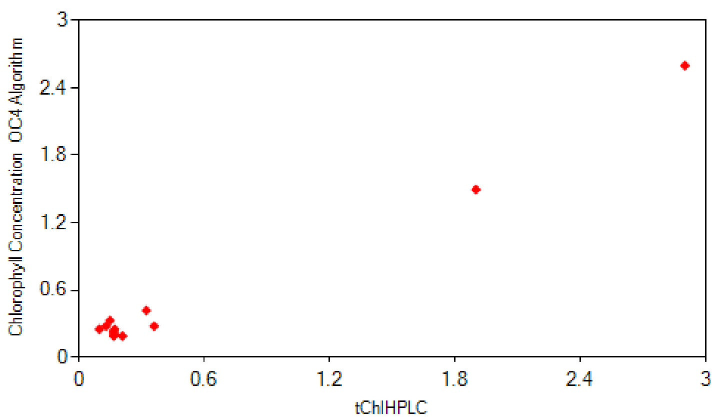


Fig. 2. Output from *wam_match* using mapped daily images and total HPLC Chl-a as X variable.

As you can see, you obtained 15 match-ups with time difference 28 hours or less. The problem is that you don't know what is the best time difference limit to use. If you set the limit too small then you may not get any match-ups but if you set the time limit too high then you may get multiple match-ups for a station and they may be too late or early to have a value in a changing environment.

5 Using WAM command line applications

5.1 Finding match-ups with wam_match_nearest

While *wam_match* uses a list of images and a preset time limit for the difference between the satellite image time and the point time and finds all match-ups within these limits, the command line program *wam_match_nearest* uses all matching image files in a path and, for each point, finds the nearest image with at least 3 valid pixels within the 3 x 3 pixel window. If an image has less than 3 valid points within the 3 x 3 pixel window centered at the point then it jumps to the next nearest image in time and if that one does not either have enough valid pixels, to the next nearest in time. This process continues until the image with enough valid pixels is found or, if no image within 30 days has enough valid pixels, it gives up and skips to the next point. It is possible to change many of these threshold values.

We use *wam_match_nearest* on the same data as *wam_match* with the following commands:

```
cd C:\Sat\SeaWiFS\L2\JES
wam_match_nearest jes9906_Chl_surf.csv Daily\S*.hdf
```

When the command finishes, it has found 39 match-ups for 39 points. Remember that we only used 36 images.

The output is saved in *jes9906_Chl_surf.csv_out.csv* (note that *_out.csv* has been added to the *jes9906_Chl_surf.csv* file name). Rename it to *jes9906_Chl_surf.csv_out_Daily4km.csv* to specify that we used daily 4-km images. We will produce match-ups for the same point file with other images.

You can visualize the output by loading the CSV file into *wam_match* with button *Load from CSV*.

Set *Select X Variable*: *tChlHPLC* and you get the following:

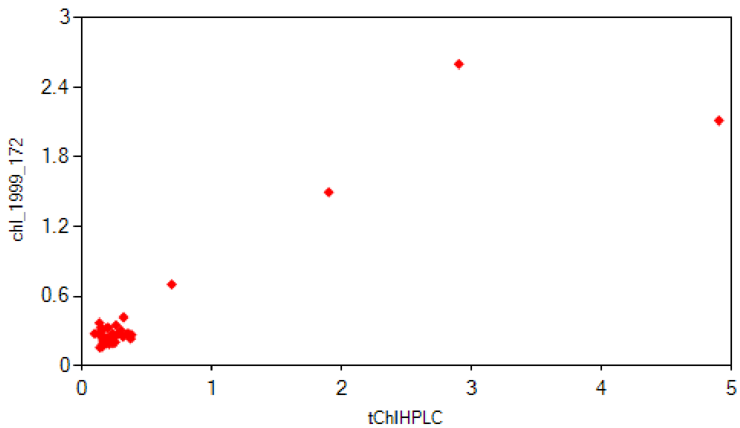


Fig. 3. Output from *wam_match_nearest* loaded into *wam_match* showing *tChlHPLC* as X variable and Chl-a from SeaWiFS mapped 4-km data as Y-variable.

As seen in Fig. 3, we have a positive correlation between in situ and satellite Chl-a values but mostly due to the few high values. At low Chl-a there is mostly a scatter without much correlation.

Don't forget to rename the previous output as otherwise it will be over-written. In the previous match-up we used 4-km daily data that were specially produced from Level-2 data. Global SeaWiFS data are available at 9-km resolution. Now issue the following command:

```
wam_match_nearest jes9906_Chl_surf.csv C:\Sat\SeaWiFS\L3\Daily\CHL_9\S*.hdf
```

Rename the output file *jes9906_Chl_surf.csv_out.csv* to *jes9906_Chl_surf.csv_out_Daily9km.csv* to keep it separate from the earlier output. In this case we used daily global Level-3 images at 9-km resolution and we found match-up points for all 39 stations. We had more input images but at lower spatial resolution (~9 km). Now visualize the output CSV file with *wam_match* with the *Load from CSV* button.

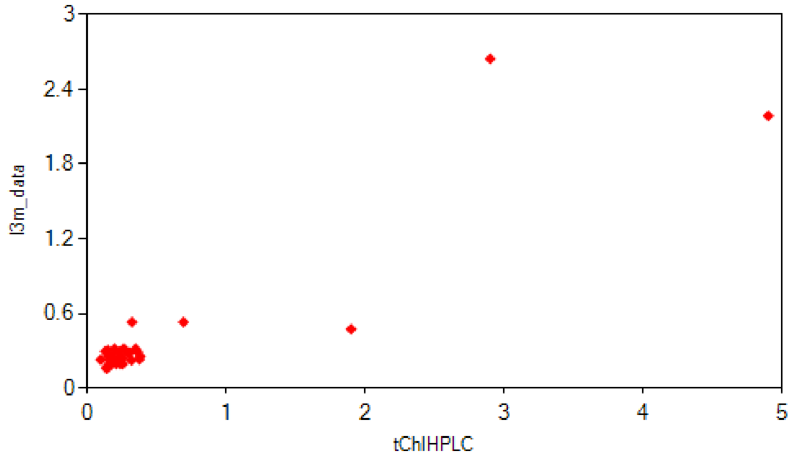


Fig. 4. Output from *wam_match_nearest* loaded into *wam_match* showing *tChlHPLC* as X variable and Chl-a from SeaWiFS global Level-3 data as Y-variable.

As you can see, the correlation has become worse because of using lower resolution global data. In most cases the spatial variability has patchiness, especially near the coasts, at rather small scales (smaller than the 9-km resolution of the Level-3 data) and therefore better to make match-ups with the satellite data of highest resolution possible. We now use another command to make match-ups with unmapped or Level-2 data that provide the best satellite data (unmodified by mapping).

```
wam_match_l2 jes9906_Chl_surf.csv C:\Sat\SeaWiFS\L2\JES\1999\S1999\MLAC\0\S*.hdf
```

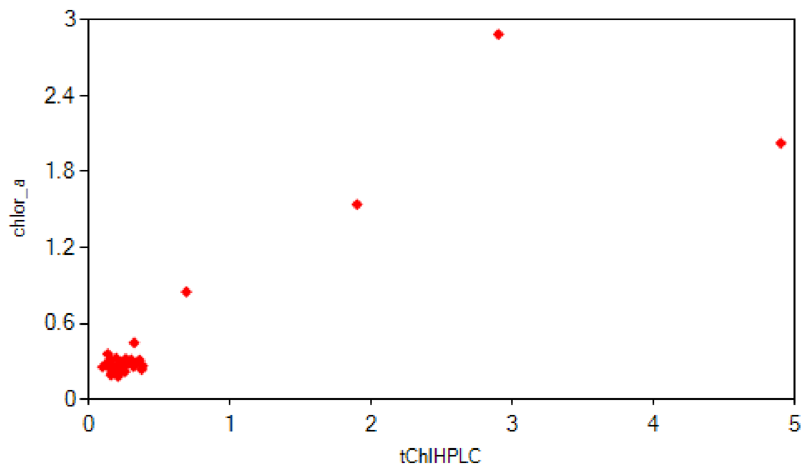


Fig. 5. Output from *wam_match_l2* loaded into *wam_match* showing *tChlHPLC* as X variable and Chl-a from SeaWiFS global Level-3 data as Y-variable.

Rename the output file *jes9906_Chl_surf_SEAWIFS.csv* to *jes9906_Chl_surf_SEAWIFS_MLAC.csv* to specify that we used MLAC L2 data and not GAC. In this case we used individual Level-2 files at the nominal 1-km resolution (at nadir, lower at swath edges) and we found match-up points for all 39 stations. Now visualize the output CSV file with *wam_match* with the *Load from CSV* button. As you can see in Fig. 5, in this case the correlation in match-ups is not substantially better than with lower resolution data. Also, we run *wam_match_l2* with default arguments. You can see all the possible options in *wam_match_l2* if you type the name of the command without any arguments. For example, using *maxdays*=5 would limit the maximum time difference to 5 days. By default, it keeps looking for a match-up until 30 days of time difference. Also, by default *wam_match_l2* uses the standard *chlor_a* that is in the L2-file. It can also calculate Chl-a using other algorithms. A big difference between L2 and L3 files is that L2 files have a set of variables called L2-flags that report all kinds of potential problems. Open the output file and try to guess the meaning of the L2-flags reported for each of the 9 match-up pixels.

Finally, you can load all the output CSV files into Excel and make plots. It is a common practice of making those plot as *Log-Log* plots instead of the linear *X-Y scatter* plots (due to the log-normal distribution of Chl-a). It is best to make new log10-transformed columns and make plots with these, rather than use the log-option in the scatter plot.