

Validation of satellite data with WIM/WAM

Contents

Validation of satellite data with WIM/WAM.....	1
1 Introduction	1
2 Prerequisites	2
3 Using WIM interactively.....	2
3.1 Right-click	2
3.2 Using Geo-Get Vector Objects.....	2
4 Using WAM GUI applications.....	3
4.1 Using wam_match.....	3
5 Using WAM command line applications	3
5.1 Finding match-ups with wam_match_nearest.....	3
5.2 Finding NPP match-ups with wam_npp_point.....	7

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 scattering or absorption in the water column. It is clear that 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 Chl-a, and those measurements are affected by errors in the calibration, atmospheric correction, estimated sea surface effects, adjacency effects, etc. Another important variable, primary production (PP, $\text{mg C m}^{-2} \text{d}^{-1}$), is even more difficult to estimate remotely as (1) it is a rate ($\text{mg C m}^{-2} \text{d}^{-1}$) and not a concentration, and (2) the PP algorithm uses an estimate of Chl-a with a set of other measured or modeled variables run through a model of PP 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 several tools for validation and match-up analysis.

2 Prerequisites

We assume that you are familiar with the basics of WIM and of the command line programs. If not, please check out the WIM and WAM manuals. We also assume that you have copied the *Images* folder from the DVD to your *Program Files\Wimsoft* folder. You also need SeaWiFS daily Level-3 data from the `\Sat\SeaWiFS\L3\daily` either from DVD or copied to your hard disk (e.g. to `C:\Sat`).

3 Using WIM interactively

3.1 Right-click

The easiest and most primitive way of a match-up analysis is using the mouse and the right click. Please see exercise 3.2 in [Practical Exercises with WIM and WAM](#). Of course, it is time consuming and error-prone to match the latitude and longitude of your in situ sample with the latitude and longitude of a pixel on the image and to manually pick pixel values.

3.2 Using Geo-Get Vector Objects

The next in the level of complexity and power is to use the *Geo-Get Vector Objects* menu option in WIM. Please see exercise 3.10 in [Practical Exercises with WIM and WAM](#).

- Load the composited SeaWiFS Chl image of the Baja California area *composite.hdf* in `Images\SeaWiFS\baja_2000_april`. Load a sample point file *ime9801.pnt* in `Images\SeaWiFS\baja_2000_april` with *Geo-Get Vector Objects-Point (Bitmap Only, Geographic Lon, Lat, Float Lon Lat)*. These are actual stations from a Mexican cruise IMECCAL 9801.
- Select a point in the *Vector objects* table and see which one is blinking. Select *Statistics* for the selected point. The statistics is calculated for a 3 x 3 pixel window centered at the point. Note the values of *Nin* (= number of pixels in the valid range) and *Nout* (= number of pixels **out** of the valid range).
- You can select one or more (or all) the points and save all the statistics into a new file, e.g. *test.pnt* with the *Save Lat,Lon,Value* button. Note that the valid data range is not available for changing in this operation and is assumed to be set previously. You can change the valid data range in *Examine - Statistics*. Now load the saved file into a text editor, e.g. *Notepad*, or MS *Excel* and verify the file format. Do **not** use the *Save* button that saves in an HDF file (it is not an image file that WIM can read).

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. Very often the satellite image closest in time is cloudy and you need to match your in situ points with 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 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** (*list_mapped_C.txt*), a **List of Point data** with longitude, latitude, date and time (*match.pnt*). The match-up output can be saved in a CSV file (e.g. *wam_match.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.

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 the image has less 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.

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

```
cd %WIMSOFT%  
wam_match_nearest match.csv Images\SeaWiFS\baja_2000_april\S*scaled.hdf
```

We assume here that folder *Images* is in *C:\Program Files\Wimsoft* which has been pointed to by the Environment Variable *%WIMSOFT%*. As “*Program Files*” has a space in it, we cannot use directly *C:\Program Files\Wimsoft\Images/* in our command line as it would be taken as 2 separate arguments. You can use the following command from any directory (assuming that you copied the *Images* folder to your *Wimsoft* folder):

```
wam_match_nearest match.csv %WIMSOFT%\Images\SeaWiFS\baja_2000_april\S*scaled.hdf
```

When the command finishes, it has found 45 match-ups for 66 points. Remember that we only used 15 *S*scaled.hdf* images.

The output is saved in *match.pnt_out.csv* (note that *_out.csv* has been added to the Point file name). Rename it to *match.pnt_out_LOCAL.csv* as we will produce another file from the same point file using global Level-3 images.

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

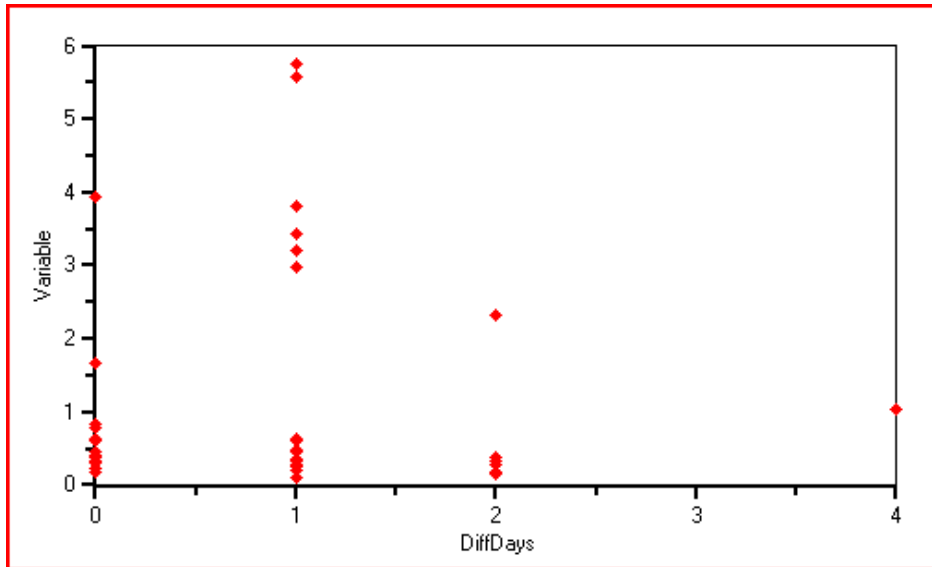


Fig. 1. Output from *wam_match_nearest* loaded into *wam_match* showing *DiffDays*.

As seen in Fig. 1, most match-ups have the time difference of either 0 or 1 days while the maximum was 4 days. Now *Select X Variable: AvgOfChla* and you get the following:

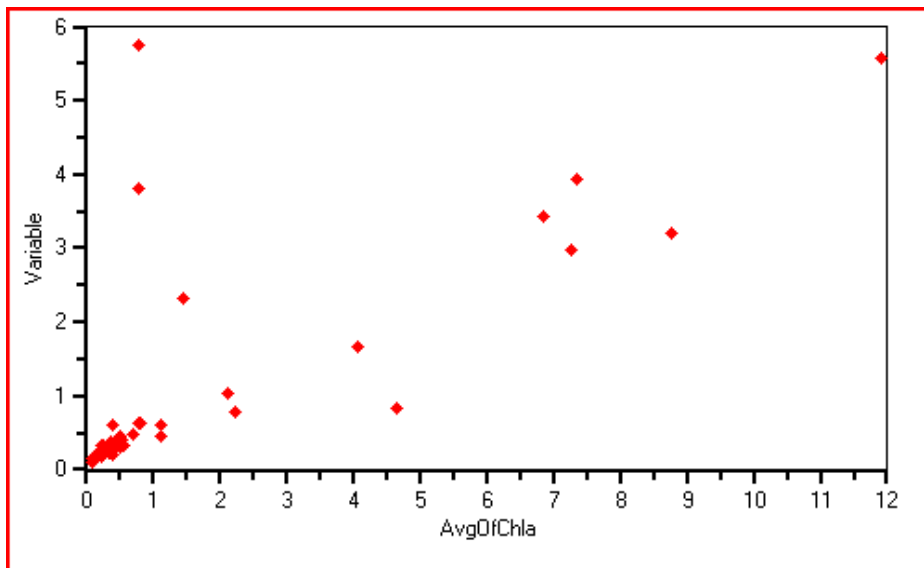


Fig. 2. Output from *wam_match_nearest* loaded into *wam_match* showing *AvgOfChla* as X variable.

As seen in Fig. 2, we have a nice linear relationship between in situ and satellite Chl-a values with a few outliers. You can try to figure out the reason for these outliers but it appears that the outliers are **not** due to the longer time difference (3 or 4 days). The worst outlier may have been caused by errors near cloud edges that is shown by the small number of valid pixels ($N_{in}=5$, $N_{out}=4$) but this is not the case for all outliers. If you exclude (by double-clicking) the 2 worst outliers then the resulting relationship looks

pretty good ($r^2 = 0.907$) but the satellite data is under-estimating the in situ data by about 2 times. The images that we used here (i.e. Images\SeaWiFS\baja_2000_april\S*scaled.hdf) were derived from Level-2 (unmapped) 1-km data mapped to a *Linear* projection in 2000. SeaWiFS data has been reprocessed several times since then. Level-3 data are more easily available as 4-km or 9-km MODIS data or as 9-km SeaWiFS data. We will therefore make another match-up analysis using global mapped Level-3 9-km data from SeaWiFS (MODIS data was not available in 2000). Don't forget to rename the previous output (using images in the *baja_2000_april* folder) to *match.pnt_out_LOCAL.csv* as otherwise it will be over-written. Now issue the following command:

```
wam_match_nearest match.pnt C:\Sat\SeaWiFS\L3\daily\S*_9
```

You get this:

Using Point file match.pnt

Area of interest size: 3 x 3

Required minimum number of valid pixels in area of interest: 3

Required maximum difference in days between point and satellite pass: 30

31 files found in C:\Projects\WIM_Add\Sat\SeaWiFS\L3\daily\

66 points read from match.pnt

Header: Lon,Lat,Date,Time,Cruise,Station,AvgOfChla

Point 1, year = 2000, Julian day = 98, station 93.26.7

Nearest image: S2000098.L3m_DAY_CHLO_9, diff days: 0

Nearest image: S2000098.L3m_DAY_CHLO_9, diff days: 0

0 pixels valid in S2000098.L3m_DAY_CHLO_9

0 pixels valid in S2000099.L3m_DAY_CHLO_9

0 pixels valid in S2000097.L3m_DAY_CHLO_9

0 pixels valid in S2000096.L3m_DAY_CHLO_9

0 pixels valid in S2000100.L3m_DAY_CHLO_9

0 pixels valid in S2000101.L3m_DAY_CHLO_9

0 pixels valid in S2000095.L3m_DAY_CHLO_9

Using S2000102.L3m_DAY_CHLO_9, valid: 3, diff days: 4

.....

Point 65, year = 2000, Julian day = 113, station 77.51

Nearest image: S2000113.L3m_DAY_CHLO_9, diff days: 0

Using S2000113.L3m_DAY_CHLO_9, valid: 9, diff days: 0

Point 66, year = 2000, Julian day = 113, station 77.49

Nearest image: S2000113.L3m_DAY_CHLO_9, diff days: 0

Using S2000113.L3m_DAY_CHLO_9, valid: 3, diff days: 0

=====

66 valid stations saved in match.pnt_out.csv

As you see, for the first station the nearest image with enough valid pixels has a time difference of 4 days. For the last 2 points (65 and 66) valid match-up pixels were found for the same day Chl-a image (diff days: 0). In this case we used daily global Level-3 images at 9-km resolution and we found match-up points for all 66 stations. We had more input images but at lower spatial resolution (~9 km). Rename the output file to *match.pnt_out_GLOBAL.csv* to keep it separate from the match-up using “local” resolution images. Now load the output CSV file with *wam_match* with the *Load from CSV* button.

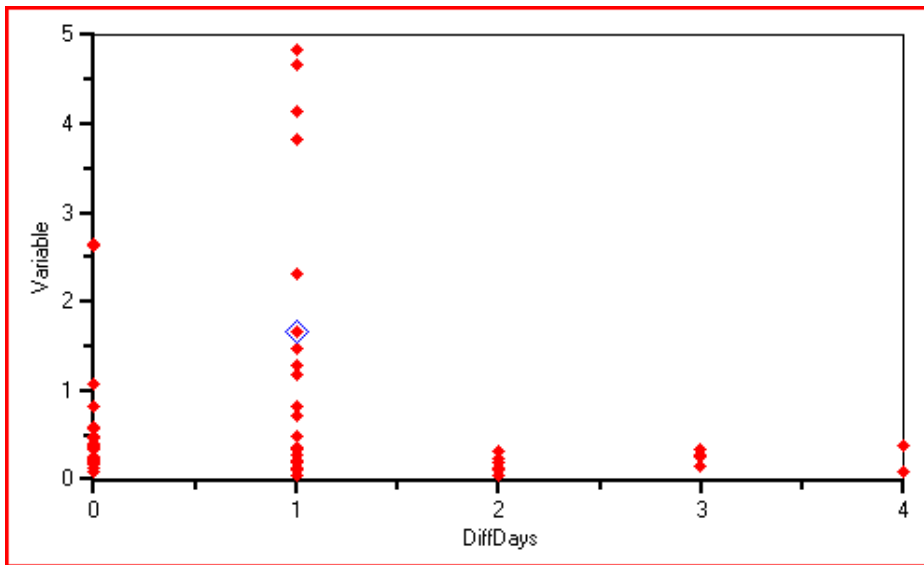


Fig. 3. Output from *wam_match_nearest* using global Level-3 data; loaded into *wam_match*; showing *DiffDays*.

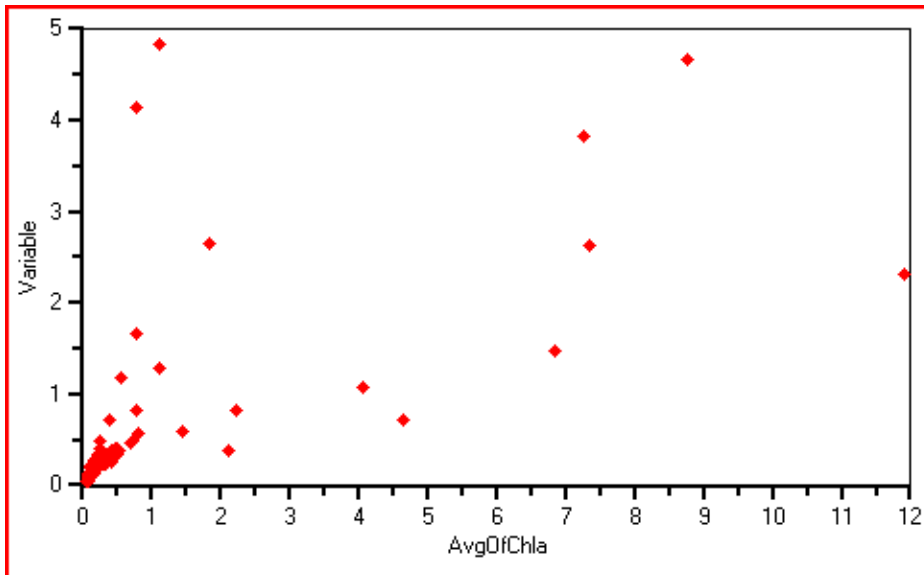


Fig. 4. Output from *wam_match_nearest* loaded into *wam_match* showing AvgOfChla as X variable and Chl-a from SeaWiFS global Level-3 data as Y-variable.

As you can see, we have more scatter and lower r^2 when using global Level-3 data compared to the “local” 1-km data. It seems that in this case the spatial variability had rather small scales (smaller than the 9-km resolution of the Level-3 data) and therefore we have better correspondence with the higher resolution (“local”) images from Level-2 data. Using the “local” data you may reach a conclusion that SeaWiFS Chl-a estimate has a very good (high r^2) relationship with in situ Chl-a but at least at higher concentrations the satellite-derived values are about 2 times lower than the in situ values. You can load both 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).

5.2 Finding NPP match-ups with *wam_npp_point*

The idea of *wam_match_nearest* of finding the closest valid pixels in time is being used in *wam_npp_point* for validating net primary production measurements. Satellite estimates of NPP are made using not one but several concurrent input images (e.g., of Chl-a, PAR and SST for VGPM and other models or even 5 images for the CbPM model). Therefore, the NPP images are not required for doing the match-up analysis of NPP as the pixel values are calculated in real time. You do need to have concurrent component images for the NPP model. The best matching images are found by the software (*wam_npp_point*) given the paths to these images. If the nearest matching images have no valid pixels in the 3 x 3 pixel window then the next nearest image in time is used. The process continues until at least 3 valid pixels are found or the time difference is over the limit (30 days). This process is applied to all input images. Please see a separate document [Tutorial Primary Productivity.pdf](#) in Course\4.