Validation of satellite data

- Please see *Course**3**Tutorial_Validation.pdf* on DVD or
- http://www.wimsoft.com/Tutorial_Validation.pdf
- Satellite measurements are often very indirect measurements of the variable that we want to estimate. For example, top of the atmosphere radiance is used to estimate the chlorophyll-a concentration (Chl-a) in the water. Only 5-10% of the radiance at the top of the atmosphere originates from the water column and can carry information about the water after being affected by scattering or absorption.
- Another important variable, primary production (PP, mg C m^{-2} d^{-1}), is even more difficult to estimate remotely as (1) it is a rate (mg C m^{-2} 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 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.
- A related task is to compare in situ measurements of any kind (e.g. fish catch, whale abundance, seabird abundance) with any satellite data (e.g. SST, Chl-a)

Levels of validation - 1

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.

Levels of validation - 2

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 IMECOCAL 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).

Levels of validation - 2

2. Using Geo-Get Vector Objects, cont.

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

Levels of validation - 3

3. ***Using wam_match :*** If you have a set of points and a set of images

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.csv*). 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.

Levels of validation - 4

4. Using WAM command line applications

4.1. 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, and **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.

Levels of validation - 4

4.1. wam_match_nearest, cont.

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

cd %WIMSOFT%

•**As "*Program Files*" has a space in it, we cannot use directly *C:\Program Files\Wimsoft\Images* in command line as it would be taken as 2 separate arguments.**

wam_match_nearest match.csv

Images\SeaWiFS\baja_2000_april\S*scaled.hdf (in 1 line)

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

Levels of validation - 4

4.1. wam_match_nearest, cont.

match.csv is a CSV file with columns separated by tab and dates (MM/DD/YYYY) and numbers in English-US format.

In many other cultures comma is used instead of decimal point and dates have DD/MM/YYYY format. Use match_pt.csv (separator = tab) or match_pt_br.csv (Portuguese-Brazil; separator = semicolon) if you have different settings. Change in Control Panel – Regional and Language Options.

wam_match_nearest match_pt_br.csv

Images\SeaWiFS\baja_2000_april\S*scaled.hdf (in 1 line)

Load output file into Excel or text editor, explain the column meanings!

SYear EYear SDay EDay Image TimeDiff_Days VarName Pointvalue Nin Nout Min Max Mean StDev Median

Make a plot in Excel in X-Y scatter format of *Median* vs *AvgOfChla*; the same in Log-Log scale.

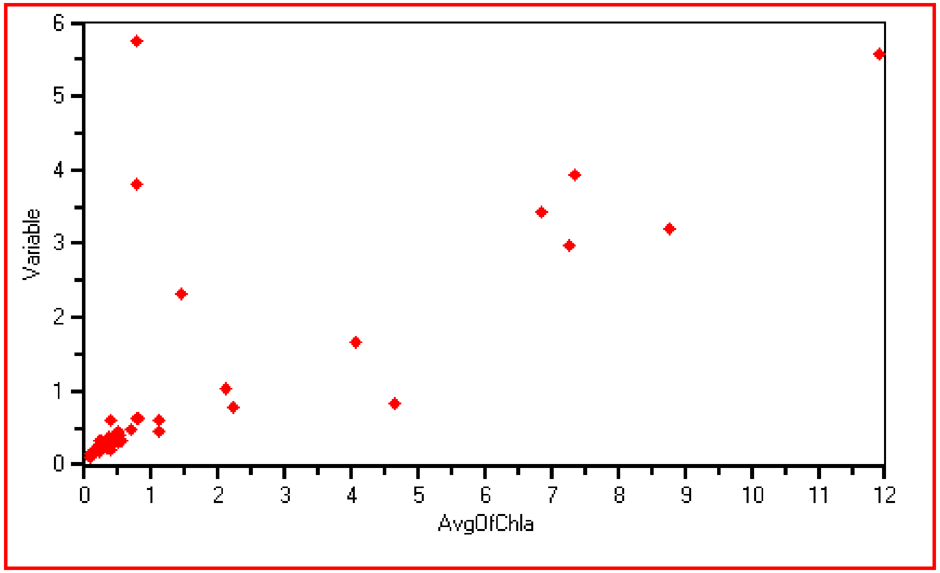
Levels of validation - 4

4.1. wam_match_nearest, cont.

Finds 45 match-ups for 66 points using 15 *S*scaled.hdf* images.

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

Visualize output by with ***wam_match*** with button ***Load from CSV****, use* ***AvgOfChl*** *in* ***Select X variable***.



You can eliminate outliers and increase the r^2, e.g. by eliminating the 2 worst outliers get $r^2 = 0.907$

Levels of validation - 4

4.1. wam_match_nearest, cont.

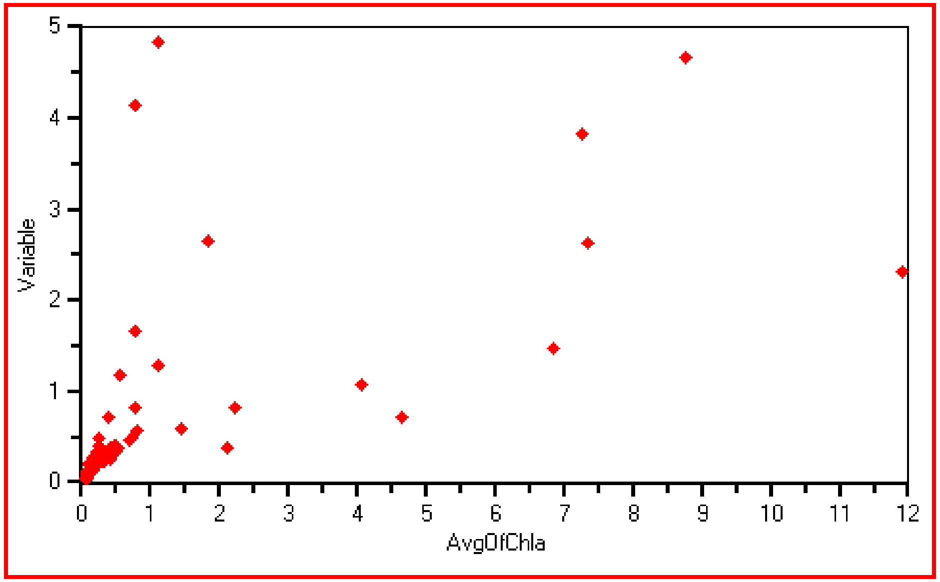
Match with Level-3 images:

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

Finds 66 valid stations, saved in match.csv_out.csv

Rename output file to *match.csv_out_GLOBAL.csv*

Visualize output by with *wam_match* with button *Load from CSV*.



Levels of validation - 4

4.1. wam_match_nearest, cont.

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).

Do you have in situ data of your own? If yes, try to compare with satellite data using the WIM/WAM validation tools.

Levels of validation - 4

4. Using WAM command line applications

4.2. 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. 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.