Variance and Anomaly Analysis with WIM/WAM

1 Introduction

Analysis of temporal variance of image data provides important clues on the functioning of the ecosystem. For example, the annual (seasonal) cycle, the long-term trends as well as anomalies from these are all important variables. This kind of satellite information is also useful background information for planning the locations of in situ measurements. Calculating anomalies is a powerful method of change detection in time series. In this tutorial we will use related techniques available in the WIM/WAM software and apply these to MODIS Aqua 4-km CHLO data in the California Current region.

2 Prerequisites

We assume that you are familiar with the basics of WIM and of the command line programs, i.e. how to open the command window, change directory, issue a command. If not, please check out the basic WIM and WAM manuals. We also assume that you have a set of images that you can use. In our example we use standard mapped (4 km) 8-day chlorophyll-a (CHLO_4) images from MODIS Aqua but with a few modifications you can use other datasets.

3 Selecting the standard map and area of interest

It may be easiest to apply our programs to the global 4-km images but these images are very big (8640 x 4320 pixels at 2 bytes per pixel ~ 75 MB) and some operations may either take too long or not work due to insufficient memory. Therefore, it is better to pick a subarea of interest. In addition to picking our standard area we can also pick our standard map projection. Here we keep the standard equal angle projection of the source images (and the 4-km resolution) of the standard mapped images and just cut out our area of interest. You can choose another area and/or remap the global images to a different map projection.

It is **VERY IMPORTANT** to keep all files **ORGANIZED** in separate folders. You can use your own folder names but I recommend the following. I create a top level folder C:\SatProjects\CAL\Variance for my project. Under that folder I have subfolders ***Chlo_4*** (for subsets of the chlorophyll-a 4-km images), ***Chlo_anomaly*** (for the chl-a anomalies), ***Chlo_means*** (for the monthly means), ***Chlo_Variance***. I recommend using the same structure:

- C:\SatProjects\CAL\Variance
    - Chlo_4
    - Chlo_annual
    - Chlo_anomaly

- Chlo_means
- Chlo_statist

Before you start running WAM programs you need to pick your area of interest and a suitable target map projection. If you have your area of interest and target map from a previous exercise then you can use that. You can generate a new target projection with the WIM *File-New* menu selection picking your latitude and longitude range and the pixel size (resolution):

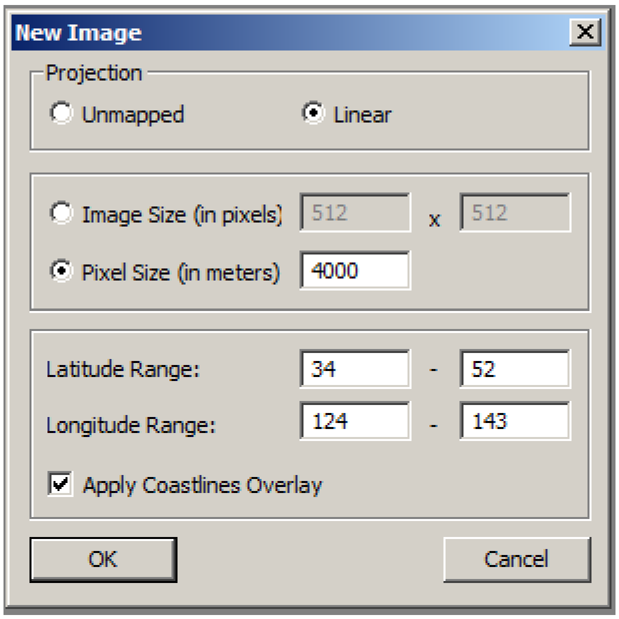


Fig. 1. File-New dialog for creating a new image in WIM.

However, here I am NOT using that method and I just cut out a suitable rectangle of a sample 4-km image with WIM (drag with the left mouse button, then use the Scissors icon to cut the subset). I save the cut image as *C:\SatProjects\CAL\Variance\target_4km.hdf*. As *wam_series* is using the latitude and longitude range to cut subset images, I find the latitude-longitude range of my subset image with the following command line:

```
wam_xy2ll   target_4km.hdf
Upper Left (lat, lon): 50.3333333333333 -132.375
Lower Right (lat, lon): 21.2083333333333      -104.916666666667
```

4 Creating a series of images with *wam_series*

Here we use *wam_series* to cut out image subsets from global images. Please try first with a few sample images! After you are satisfied with the output of a few images, run the same through all images. You can see below that I use to match a single image first, save as HDF and PNG, convert to byte using the *Log-Chl* scaling. The subset image created with *wam_series* may be slightly different in size from the image manually cut with WIM. That's because the latitude and longitude corners are calculated with certain approximation. That is important to notice when you are trying to overlay images as images need to be of the same size when doing overlay. If you create a linear projection target map with the New Image command (figure above) then you need to use "Remap to" with the target map name in the *wam_series* dialog (below).

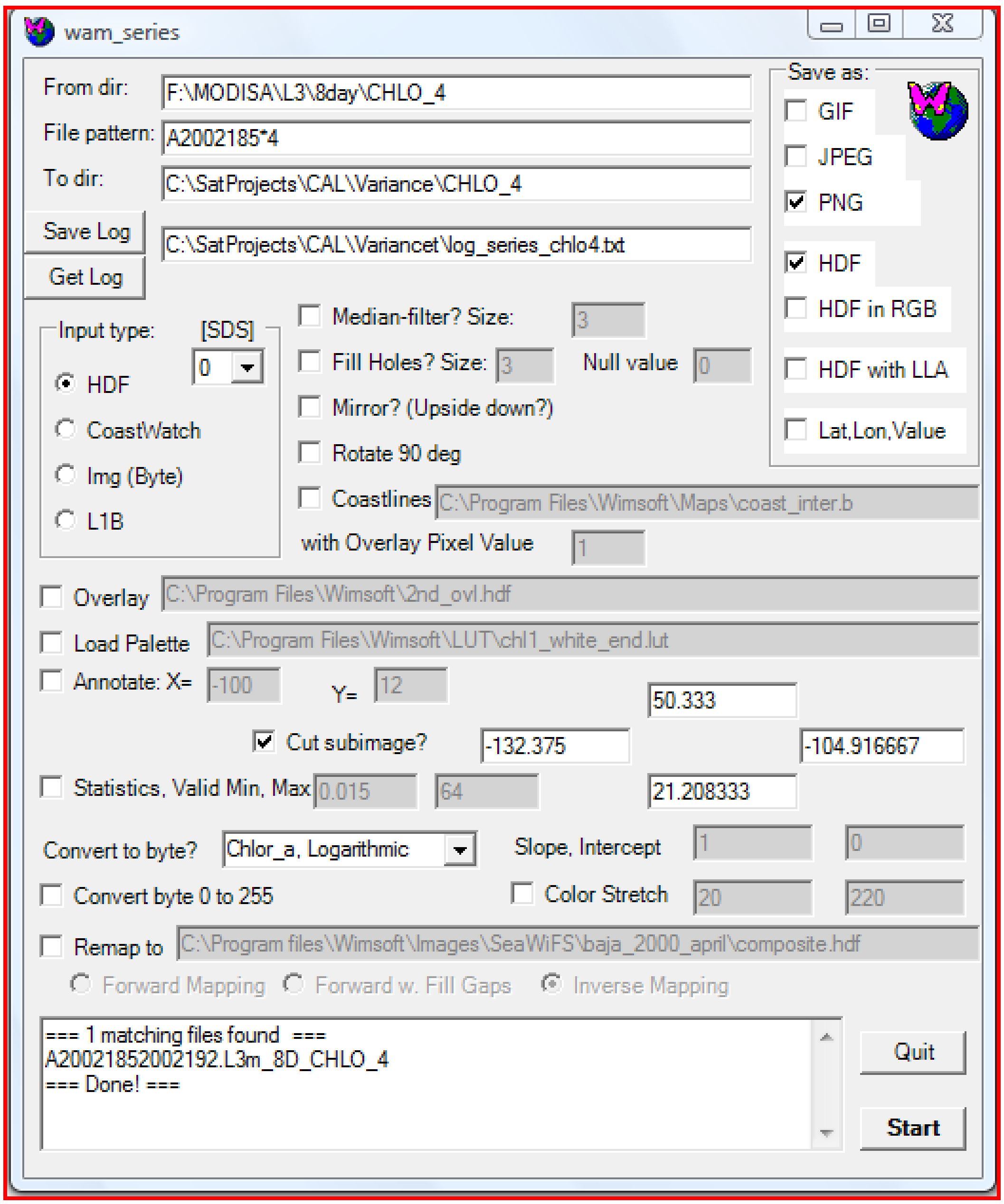


Fig. 2. Screen shot of *wam_series* for creating a test subset of the series of subset images.

When I am satisfied with the output, I do the same with all the images and save just the HDF files and not the PNG files.

5 Calculating overall mean and variance

In this step we calculate the overall mean as well as overall variance and the number of valid pixels. I issue the following command:

```
wam_variance CHLO_4\A*cut.hdf
```

I get the following response:

```
Saved mean in CHLO_4\..\Means.hdf
Saved Valid pixel count in CHLO_4\..\ValidCounts.hdf
Saved SD in CHLO_4\..\SD.hdf
Done with 262 images!
```

The output files are saved one level up from the source files. I move them to a folder *Chlo_statist*. They contain data on the overall mean, SD and the count of valid pixels.

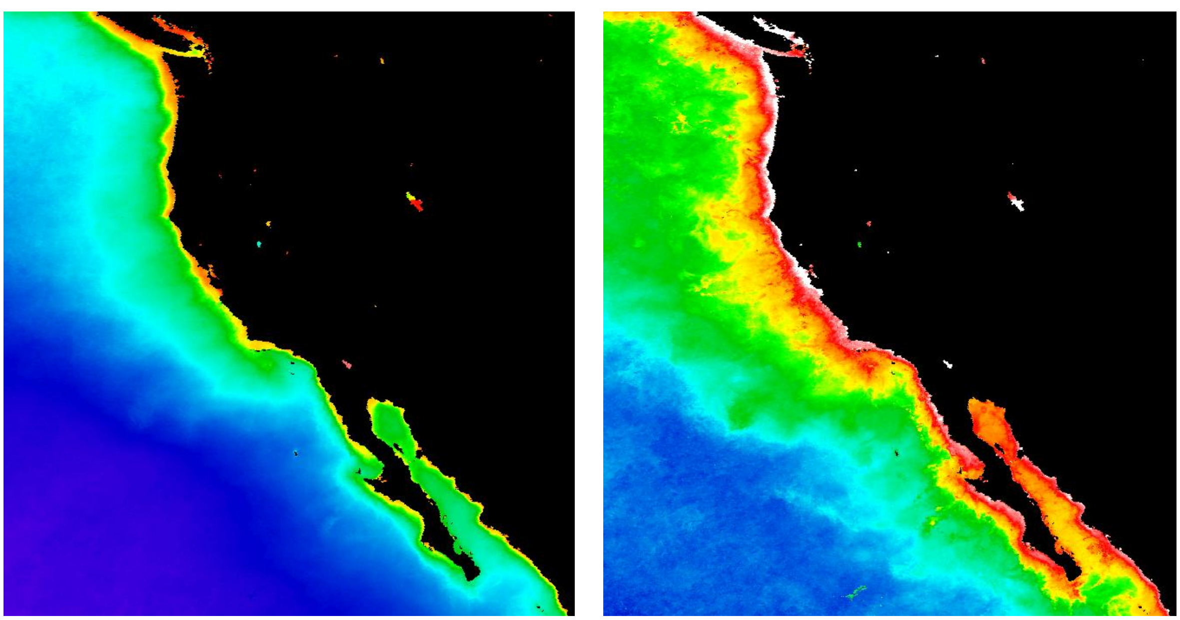


Fig. 3. Overall mean CHLO (left panel) and the SD (right panel).

It is clear that the areas with the highest mean Chl-a have also the highest variance (SD).

6 Calculating the mean annual cycle and the anomalies

The overall standard deviation can be partitioned into variability due to the mean seasonal cycle and the deviations from the mean seasonal cycle. The deviations from some kind of mean are called anomalies. In the next step I create the mean seasonal cycle and the anomalies with a single command.

```
wam_anomaly CHLO_4\A*cut.hdf 46 anomaly5.lut true
```

As output of this command I get 2 files with the *Means* and *Valid Counts*, respectively. I move them into a folder *Chlo_means*. The *Means* file has 46 individual datasets that correspond to the mean annual (seasonal) cycle of 46 eight-day periods and the *Valid Counts* file has the corresponding number of valid pixels in each 8-day period.

7 Extracting the mean annual cycle

The 46 individual images are all in the same file. In order to plot the mean annual cycle or use the *wam_variance* again and calculate the variance due to the annual (seasonal) cycle I extract the individual 8-day means and store them in individual files:

```
wam_overlay Chlo_means\A_cut.hdf_Means.hdf
```

As output of this command I get 46 HDF files and 46 JPG files that show the annual cycle. You can use the JPG files to create a movie loop of the annual cycle (e.g. with with the Babarosa GIF Animator).

I will also want to plot the annual cycle in the following 10 stations:

LonDec	LatDec	Station
-124	30.42	90.120
-121.99	31.42	90.90
-120.64	32.09	90.70
-119.48	32.65	90.53
-118.39	33.19	90.37
-120.03	34.28	81.8.46.9
-120.8	34.32	80.55
-121.84	33.82	80.70
-122.53	33.48	80.80
-123.91	32.82	80.100

Please not that Longitude (decimal degrees) is the 1st column, Latitude (decimal degrees) is the 2nd columns and the station name (text) is the 3rd column. This station list is in a CSV file *CCE_cardinal_stations.csv*.

I now create a list of images that I want to use with the following command:

dir /b /s Chlo_means*overlaid*.hdf > list_annual.txt

The list has the full path of filenames, e.g.

C:\SatProjects\CAL\Variance\Chlo_means\A_cut_overlaid_01.hdf_Means_overlaid_01.hdf

…….

I am now ready to run *wam_statist* with the list of images and a list of stations.

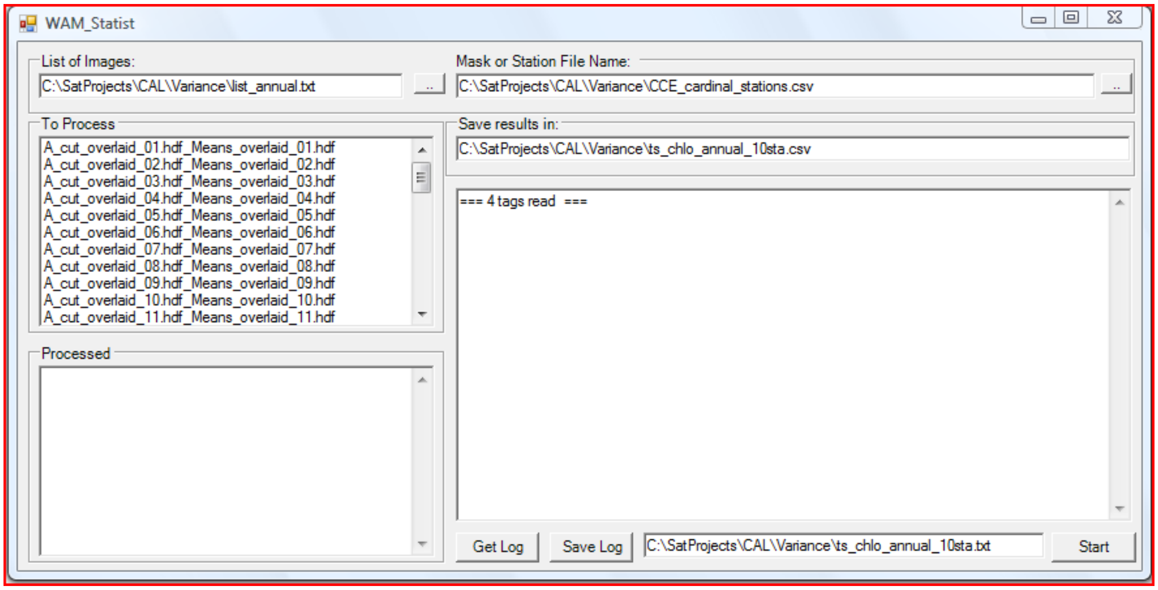


Fig. 4. Screen shot of *wam_statist* before finding the annual cycle for 10 stations.

As output, *wam_series* created a CSV file *ts_chlo_annual_10sta.csv* that has the statistics and all individual values for 3 x 3 pixel neighborhood centered at the station. However, the sequence of lines is: image file 1 with all the stations, image file 2 with all the stations, etc. This sequence is not convenient to plot without sorting. It is possible to sort in Excel but we use a special command for that:

sortstas ts_chlo_annual_10sta.csv

This command creates a sorted CSV file *ts_chlo_annual_10sta_sorted.csv* that we read into Excel for plotting. In order to plot the annual cycle in Excel, we need a time axis. You can notice that the SYear (=Start Year), SDay (=Start Day), EYear (=End Year), EDay (=End Day) are not useful in this file as they are for the whole time series of images and not for each 8-day period. We therefore create a new YearDay time axis by starting with day 4 (middle of the 1st 8-day period) and add 8 in each next step. The resulting mean time series of Chl-a in 3 x 3 pixel windows centered at the10 stations are shown below. Notice that this is the mean annual cycle for 2002-2008 (2002 and 2008 are partial years). Remember that you created the overall Mean cycle with the *wam_anomaly* command and later extracted individual images. If you want to create time series of the individual years or the whole time series then you can do it with *wam_statist* using a list of images that has all the original file names. Plots in Fig. 10 show that the annual cycle is very different in the high-Chl-a zone (near the coast) and low-Chl-a zone (offshore). While in the coastal zone the maximum is during days 100-150 (April-May), the offshore areas have a minimum during that period and a maximum in January or in December (Fig. 10B).

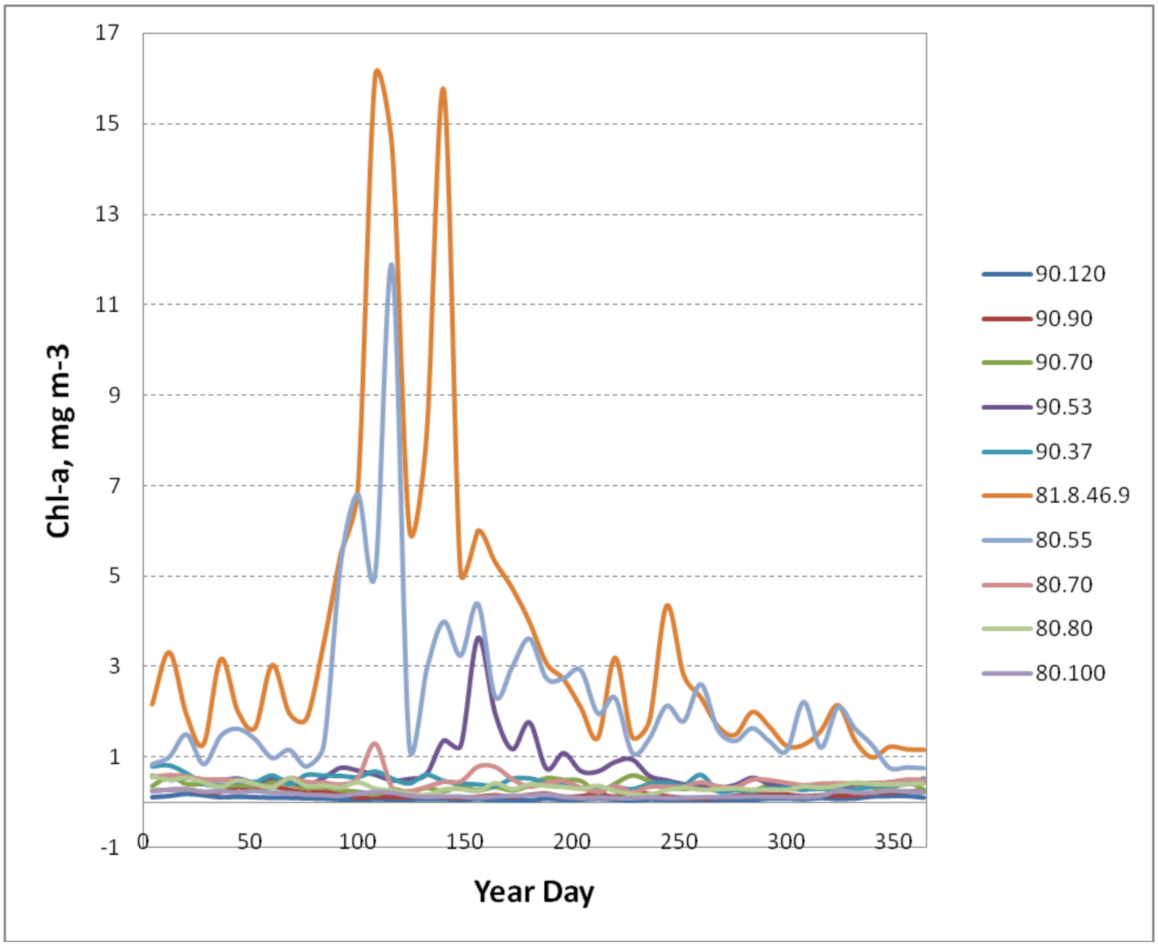


Fig. 5A. Mean Chl-a annual cycle in 3 x 3 pixel windows centered at the10 stations.

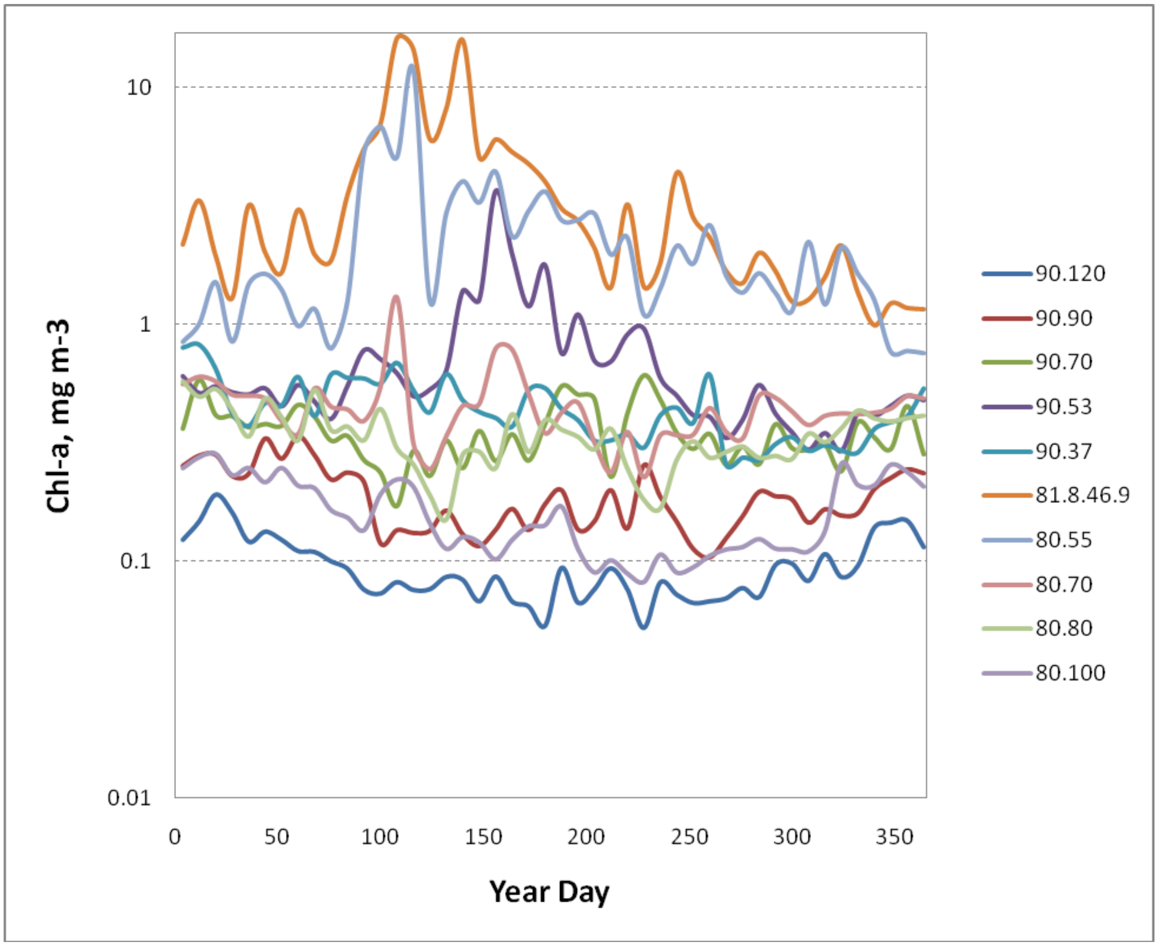
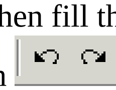


Fig. 5B. Mean Chl-a annual cycle in 3 x 3 pixel windows centered at the10 stations in Log10 scale.

8 Creating an overlay image

I will now create an overlay for nicer visualization using any of the sample (cut) images with *Geo-Get Map Overlay-coast_full.b*, background=0, foreground=1. I then fill the land with pixel value 255 (with *Edit-Draw*). If paint spills out, I undue with , close the gap with the paintbrush tool and try again. I save the overlay image under the same name as the original target map file *target_4km.hdf*.

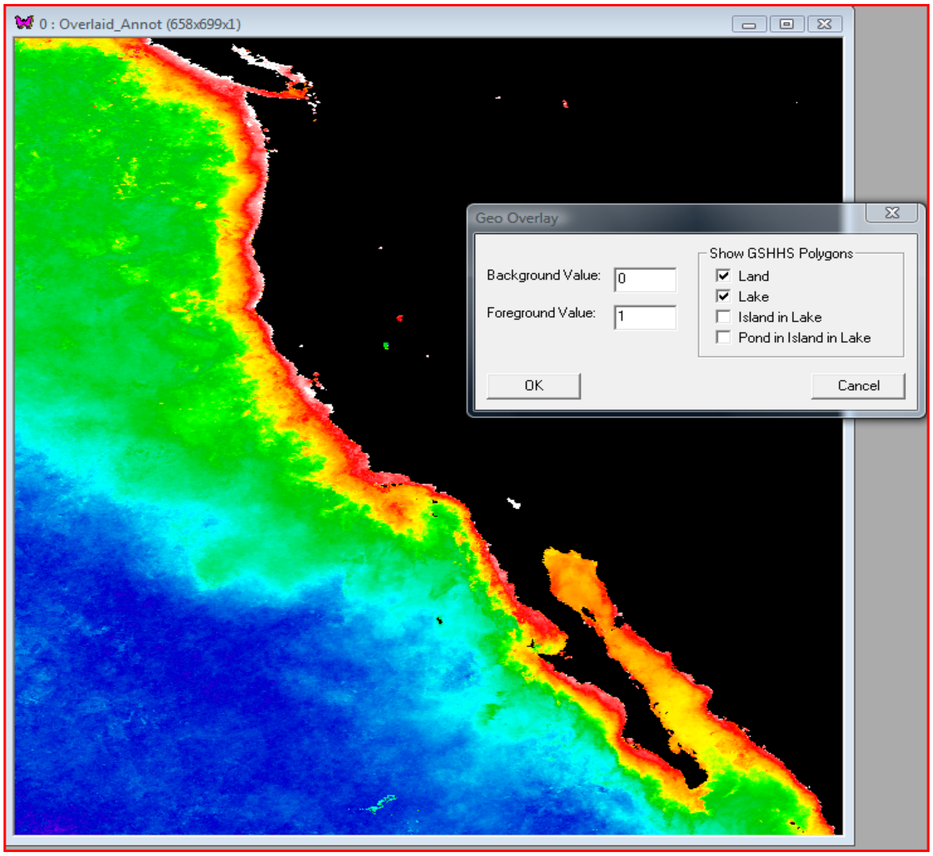


Fig. 6. Creating the coastline image with *Geo-Get Map Overlay*.

Fig. 7. Overlay image with coastlines (=pixel value 1) and land (pixel value 255).

9 Variance due to mean annual cycle

I will now calculate the variance due to the mean seasonal cycle (based on 8-day composites):

```
wam_variance Chlo_means\A*overlaid*.hdf
```

The response is:

```
Saved mean in Chlo_means\..\Means.hdf
Saved Valid pixel count in Chlo_means\..\ValidCounts.hdf
Saved SD in Chlo_means\..\SD.hdf
Done with 46 images!
```

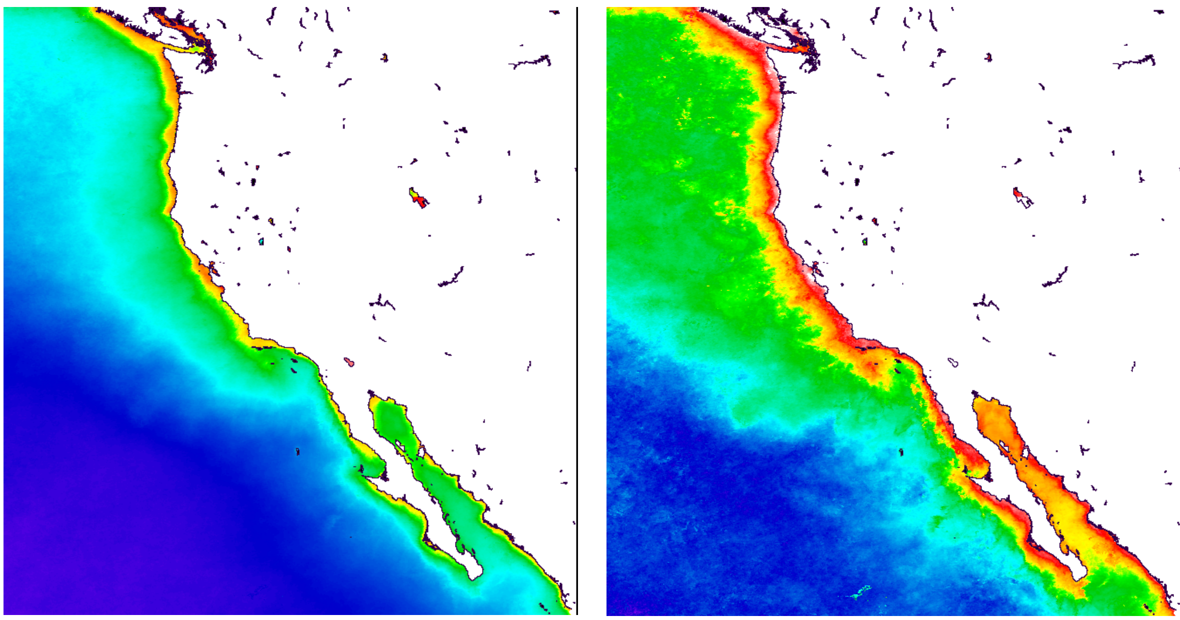


Fig. 7. Mean Chl-a of the individual 8-day mean composites (left) and the SD (right) of the annual cycle. Both very similar to the overall mean and the overall SD, respectively.

I have put the newly created overlay file on top of these Mean and SD image with the *Multi-Overlay Image* menu option. I move these files into a separate folder *Chlo_annual*.

10 Variance due to deviations from the mean annual cycle

I can now calculate the variance due to the anomalies from the mean annual cycle (based on 8-day composites):

wam_variance Chlo_anomaly\A*anomaly.hdf

The response is:

Saved mean in Chlo_anomaly\..\A_anomaly.hdf_Means.hdf

Max of valid counts per pixel = 262, > 255

Keeping ValidCounts image as Float32 instead of Byte

Saved Valid pixel count in Chlo_anomaly\..\A_anomaly.hdf_ValidCounts.hdf

Saved SD in Chlo_anomaly\..\A_anomaly.hdf_SD.hdf

Done with 262 images!

The mean image of the anomalies must be close to zero if using difference anomaly but with Chl-a we use anomaly that is the ratio of the current value to the mean value. Therefore the mean of the anomalies is a value that is close to 1. The SD of the anomalies is a *Float32* image. I stretch the scaling with *View-LUT stretch* and may also pick a different palette file with *File-Lookup table-Load LUT-chl1_white_end.lut*. I can then convert this to *Log-Chl* scaling and put the overlay image on top of that. The result is below:

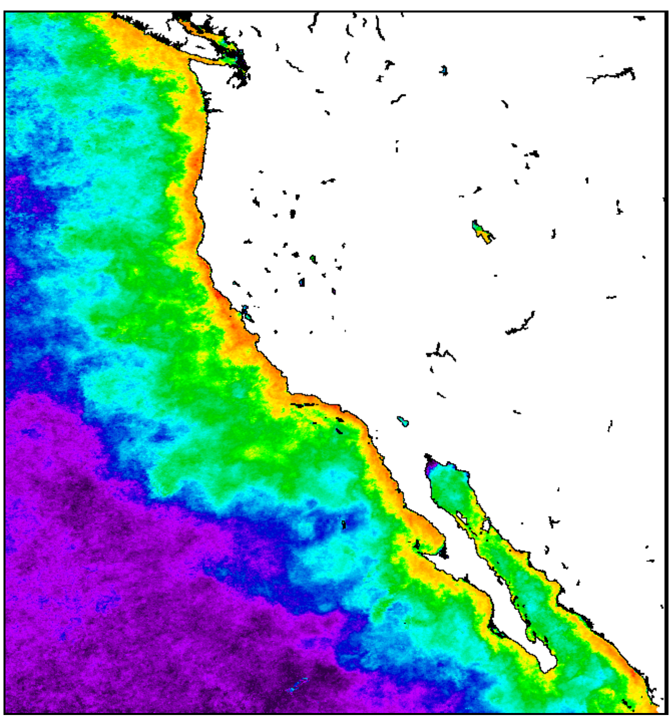


Fig. 8. Mean Chl-a anomaly relative to the mean annual cycle.

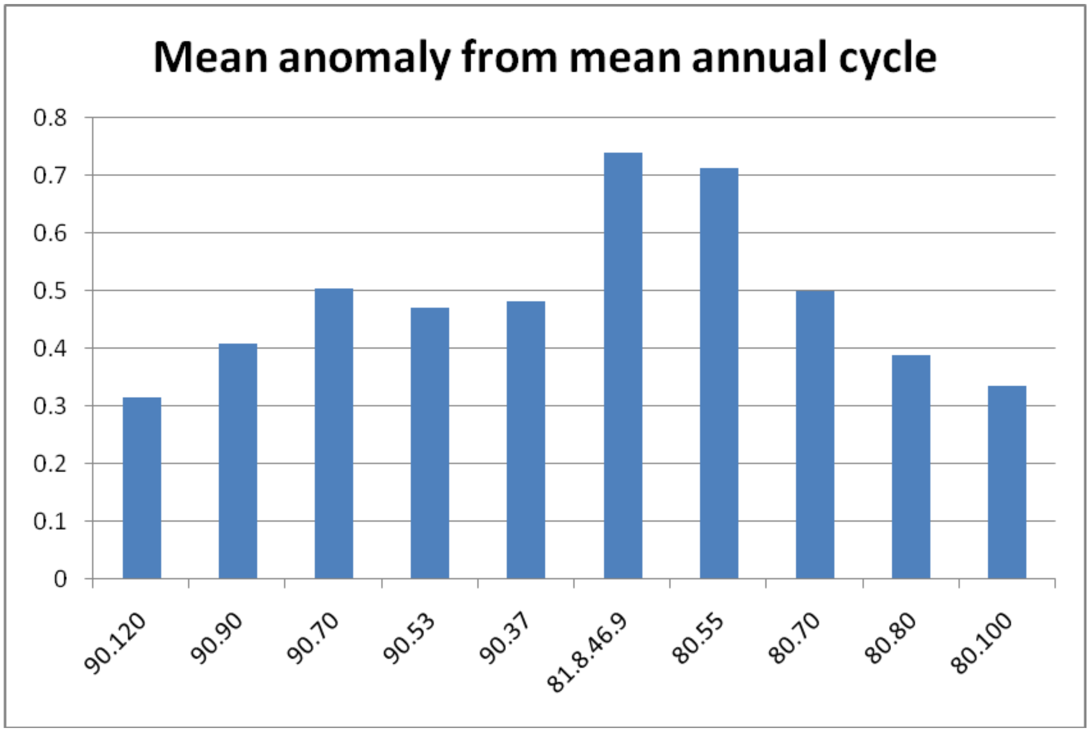


Fig. 9. Mean Chl-a anomaly relative to the mean annual cycle at 10 selected stations.

We can extract pixel values and statistics of 3 x 3 pixel neighborhoods from the mena anomaly image (Fig. 8) with the Geo-Get Vector Objects-Point. We then select all the objects and save. The saved file can be read into Excel. The mean anomaly corresponding to the 10 stations is shown in Fig. 9. We have to remember that anomaly for Chl is calculated as the ratio of the actual value to the mean value. Therefore the interpretation of the mean anomaly is tricky. However, it shows the expected high anomalies correspond to high Chl-a.

A conclusion of this exercise is that Chl-a variability is well correlated with its mean concentration. Areas with high mean values near the coast and in the coastal filaments are also areas with the most seasonal and non-seasonal variability. Offshore areas have both low Chl-a and low Chl-a variability.