
INDICATORS OF COASTAL WATER QUALITY

MAPPING GLOBAL CHANGES IN CHLOROPHYLL-A
CONCENTRATION USING SEAWIFS DATA

Data, Methods and Results

November 2009

This document and the associated data sets were prepared by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University under NASA contract NNG08HZ11C for the continued operation of the Socioeconomic Data and Applications Center (SEDAC).

For more information or to download the data, visit:
<http://sedac.ciesin.columbia.edu/es/seawifs.html>

CONTENTS

Introduction.....	3
Description of distributed data.....	5
Methodology.....	6
Chlorophyll Raw Data.....	6
Ancillary Data.....	7
Coastline Buffer.....	7
Exclusive Economic Zones (EEZ).....	7
Political information.....	8
Sequence grid and centroids.....	8
Step by step instructions.....	9
Data Analyses.....	10
Deriving Chlorophyll-a concentrations.....	10
Indicators of change.....	10
Notes on regression analyses.....	11
Calculation steps in Excel.....	12
Analysis Scenarios.....	13
Scenario 1: Sub-setting based on a Polygon buffer.....	13
Scenario 2: Interactive selection of arbitrary areas.....	16
References.....	16
Figures and tables.....	17
Global and Country Statistics.....	17
Selected imagery.....	23

Acknowledgements: CIESIN/SEDAC would like to acknowledge the advice of Dr. Ajit Subramaniam of the Marine Biology Department of the Lamont-Doherty Earth Observatory of Columbia University in the development of these indicators. The data were developed by Steffen Foerster, Malanding Jaiteh, and Alex de Sherbinin.

INTRODUCTION

The eutrophication of coastal waters is a growing problem in many countries owing to significant discharges of raw or processed sewage, agrochemical runoff, and even runoff from common lawn fertilizers into coastal waters. Nitrogen and phosphorus loading contributes to the growth of phytoplankton whose subsequent decomposition reduces oxygen availability to benthic sea creatures – fish, shell fish, and crustaceans. Changes to nutrient loadings can also change the phytoplankton species composition and diversity. In extreme cases eutrophication can lead to hypoxia – oxygen depleted “dead zones” – and harmful algal blooms (HABs) (Figure 2). Unless the amount of nutrients flowing into coastal waters is reduced, the spatial extent and severity of eutrophication is expected to grow. A twenty year analysis (1980-2000) of Coastal Zone Color Scanner (CZCS) and Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data found dramatic global increases in chlorophyll concentrations, particularly in the southern hemisphere (Antoine 2005).

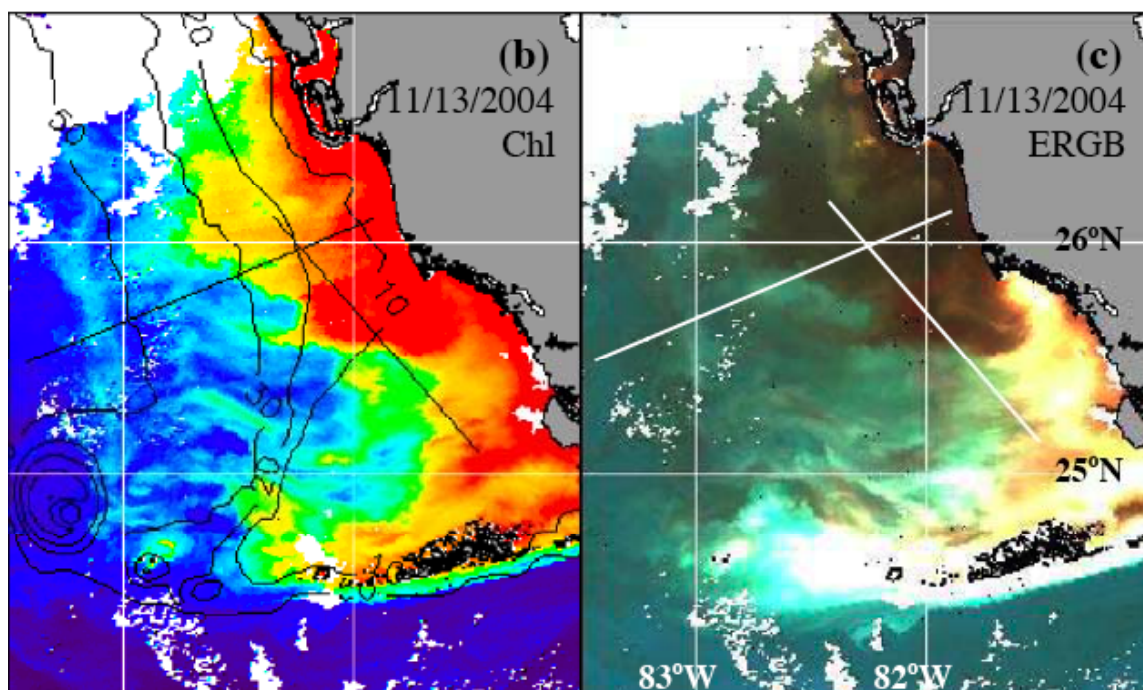


Figure 1: SeaWiFS captured these images of a harmful algal bloom off the coast of Florida coast on November 13, 2004 (see inset upper right). The left image shows the chlorophyll concentration while the right image is a pseudotrue color composite showing dark waters that are very likely HAB south of Tampa Bay. Note that the toxicity of algal blooms cannot be observed using remote sensing, but areas of algae that may be HABs can be identified and corroborated by local observation at sea level. Sources: Hu et al. 2005 and <http://eobglossary.gsfc.nasa.gov/study/redtide/>

More wide-spread and severe eutrophication, when combined with rising sea surface temperatures due to global warming, is likely to increase the incidence of hypoxia.

To better understand the extent of coastal eutrophication, we used global SeaWiFs data from 1997 to 2007 for a time series analyses to identify areas with significant changes in chlorophyll-a concentration as an indicator of algae biomass. For defining coastal zones, we excluded the first ten kilometers of coastal waters because of the potential for bottom-reflectance or suspended sediments affecting the satellite measurements in close proximity to the coast (Lavender et al. 2004). We limited the extent of coastal zones to 100 km offshore as an arbitrary cut-off above which impacts from land based sources on oceanic eutrophication are unlikely. By analyzing trends over a ten year period, we attempted to identify areas with improving, declining, and stable coastal water quality that can provide guidance for decision making in the context of environmental management.

DESCRIPTION OF DISTRIBUTED DATA

Table 1: LIST AND DESCRIPTION OF DATA DISTRIBUTED

Name	Code	Description
Yearly grids	yearly	Folder. Contains yearly gridded chlorophyll-a concentrations (in ng/m ³) clipped to near-coastal zone (10-100 km offshore) in ArcGIS GRID (integer) and GeoTIFF formats
Ancillary data	ancillary	Folder. Contains spreadsheets of ancillary data for each grid cell covered in the global time series grid, including: yearly chlorophyll-a concentrations, percent change, significance of change, lat long coordinates of grid cell center points, and GRIDCODE identifier; data provided in Excel 2007 format; in addition to country information, each grid cell is assigned to a region, using UN classification of region memberships ¹
	ancillary by country	Because of the large number of rows in the global dataset (>450,000) that cannot be handled by older spreadsheet software, for compatibility purposes we provide ancillary in multiple smaller tables in CSV format
Sequence grid	sequence	A sequence grid of global extent
Centroids	centroids.shp	Center points of global sequence grid cells with lat-long information added
Buffer	buffers.shp	Buffer of near-coastal waters for each country (restricted to its EEZ extent) in shapefile format

¹ Available at: <http://unstats.un.org/unsd/methods/m49/m49regin.htm>

METHODOLOGY

CHLOROPHYLL RAW DATA

Yearly averages of chlorophyll-a concentrations (in ng/m³)² are based on annual composites³ of SeaWiFS satellite data provided by the SeaWiFS Project, NASA/Goddard Space Flight Center and GeoEye. Data were downloaded from the Level-3 Browser at:

<http://seadas.gsfc.nasa.gov/cgi/level3.pl> in the form of HDF files at a resolution of approximately 1° x 1° on an Equidistant Cylindrical Map projection (about 9x9 km at the equator). A detailed documentation of the Level 3 global composite imagery can be accessed at http://disc.gsfc.nasa.gov/guides/GSFC/guide/SeaWiFS_L3_Guide.gd.shtml

HDF files were converted to ESRI GRID format using the Marine Geospatial Ecology Toolbox developed by Roberts et al (2007), version 0.4. After successful installation, the toolbox can be added to the set of toolboxes from the python[version_number]\Lib\site-packages\GeoEco\ArcGIS Toolbox folder. For parameters needed to convert into ESRI GRID, see Figure 2.

² Measures are in nanograms/m³ instead of milligrams/m³ in order to preserve the data as integer rather than floating point grids.

³ **Note:** For regions with heavy seasonal cloud cover, there is a potential bias in the use of annual composites in that these composites make use of every cloud-free valid pixel in a year. Thus, less cloudy periods will be over-represented in the sample. If chlorophyll-a abundance is correlated with cloud cover in a given region – and it often is (e.g., coastal upwelling which brings phytoplankton to the surface in the Pacific Northwest also creates a cloud layer) – then the presence of clouds creates a sampling bias. This product does not address this issue, but in the future we plan to develop a product that does by using monthly composites instead of annual composites.

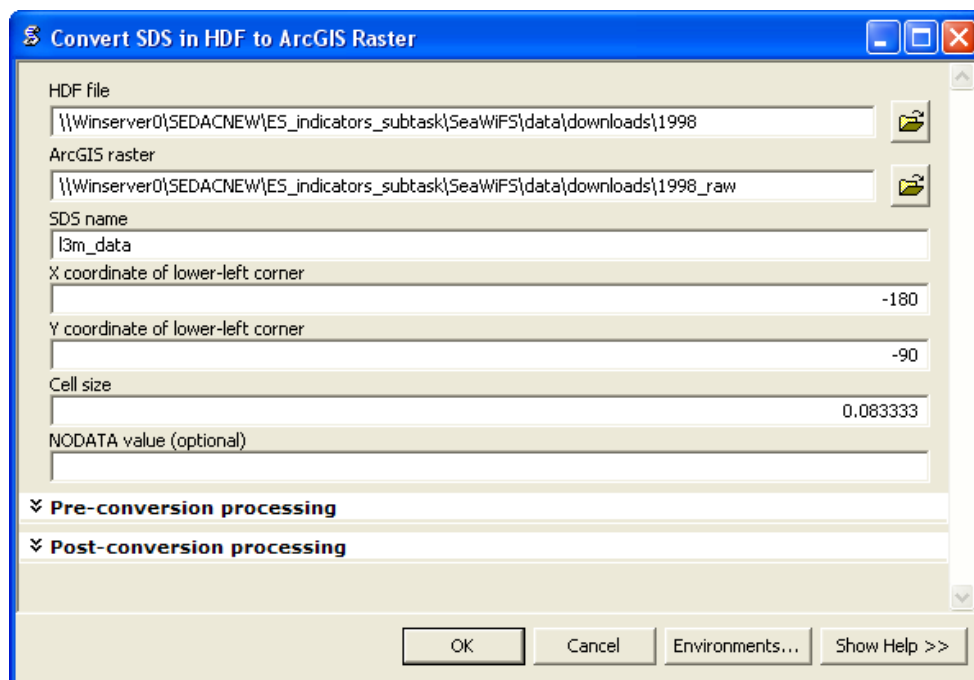


Figure 2: Screenshot of HDF-to-raster conversion interface of GEOECO toolbox

ANCILLARY DATA

COASTLINE BUFFER

For defining coastal zones to be used in our analyses, we created coastline buffers stretching from 10 to 100 km off the coast using the ESRI coastline layer.

EXCLUSIVE ECONOMIC ZONES (EEZ)

Coastline buffers were clipped by country and Exclusive Economic Zone (EEZ). A layer of EEZ boundaries was obtained from the VLIZ Maritime Boundaries Geodatabase provided by the Flanders Marine Institute, Belgium. According to metadata information available for the EEZ layer, it incorporated the GSHHS coastline layer from NGDS, downloaded in 2005 (version not specified, possibly version 1.3, 1999).

When EEZ boundaries were closer than 100 km to the coastline of a given country, we clipped the coastline buffer according to the EEZ, to limit country-wide analysis to that portion of the near coastal waters over which a country has jurisdiction. Due to the coarse nature of the coastline layer,

certain small island states are underrepresented because only bigger islands are contained in the ESRI layer. Thus, small island states such as Cocos island, Easter Islands and smaller islands were completely left out of the analysis.

POLITICAL INFORMATION

Both the EEZ and coastline layer contained country information. In certain instances the exact country names (spellings, abbreviations) differed between the two layers and had to be manually matched. In such cases we maintained the country name as provided in the EEZ layer.

Country buffers were intersected with the EEZ layer to maintain political information and use only that portion of a country's coastline buffer that was under its jurisdiction. All country buffers were then merged into a global coastline buffer layer (BUFFER) that was used for clipping global SeaWiFS grids.

SEQUENCE GRID AND CENTROIDS

To merge raster data representing yearly averages into a time series grid while maintaining attribute information of spatial location for each grid cell, we first created a global sequence grid (1° x 1° on an Equidistant Cylindrical Map projection) that assigns consecutive numbers to all grid cells. Using this sequence grid for merging different raster datasets (see below) allows for easy spatial matching of grid cells with the results from time series analyses computed outside the GIS environment (in Excel or statistical packages).

Similarly, we created a global fishnet of the same cell size (1 degree) that was then turned into a layer of centroids (center points) with the same identifying GRIDCODE as used in the sequence grid. Although not necessary for a time series analyses based on a country's EEZ extent (country information is part of the attribute table of the time series grid), using centroids greatly facilitates the interactive selection of any arbitrary area of interest within and across different EEZ zones for separate analysis.

STEP BY STEP INSTRUCTIONS

The following are step-by-step instructions for creating a global time series of chlorophyll concentrations from global raw SeaWiFS grids and the ancillary data described above:

1. Clip the raw grids (yearly CHLORO averages) using the global coastal buffer layer.
 - a. **IMPORTANT:** To avoid grid cell shift during calculations, set ENVIRONMENT in SPATIAL ANALYST options to RAW GRID extent and cell size to 0.083333.
2. Combine the clipped grids with the sequence grid (SEQ_GRID). This will be the time series grid of the raw data for the selected buffer zones.
3. Clip centroids layer (CENTROIDS) using the coastline buffer
4. Join the centroid selection from step 3 to the time series grid created at step 1 based on the GRIDCODE field
5. Intersect centroid selection from step 4 with BUFFER to associate political information with each grid cell of the time series grid
6. Extract attribute table and do analyses in Excel or other Stats package, taking care to maintain the original GRIDCODE field
7. For display purposes, in ArcMap, link the time series from step 1 with analysis results based on the GRIDCODE field

DATA ANALYSES

DERIVING CHLOROPHYLL-A CONCENTRATIONS

Radiance counts of raw level-3 data were converted into chlorophyll concentrations using the scaling equation provided as part of the HDF header file:

$$C = \text{Base}^{((\text{Slope} \times \text{l3m_data}) + \text{Intercept})}$$

Where C is the chlorophyll-a concentration in mg/cubic meter, base=10, slope=0.000058, l3m_data is the dataset in the HDF file containing the raw values to convert, and intercept=-2. To allow storage of concentrations in integer grid format, we multiplied concentrations by 1,000 to obtain ng/cubic meter (see footnote 2 for the reason for this transformation).

The header file can be extracted with the GeoEco toolbox (Data Management\HDF Files\Extract Header\Extract HDF Header).

INDICATORS OF CHANGE

Simple statistical measures describing the change of mean chlorophyll-a concentrations in a given area can easily be achieved via standard summarizing functions in Excel or any statistical package. However, given the setup of the data table (Figure 3), with grid cells as rows and repeated measurements as columns, regression analyses of temporal change require the calculation of regression equations and levels of significance for each grid cell separately. To calculate indicator statistics for each grid cell, we used array functions in Microsoft Excel 2007 (see next section). To facilitate the use of this dataset out of the box, we included the following statistics in the ancillary data table:

1. Slope of a linear regression as an indicator of the steepness of change over time [SLOPE]
2. Significance level of the linear regression of chlorophyll concentration on year [P]
3. Percent change in the average 2005-2007 value compared to the average 1998-2000 value [CHANGE]

4. Percent change over time where linear regression was significant, i.e. $p < 0.05$ [SCHANGE]

	A	B	C	D	E	F	G	H	I	J
1	GRIDCODE	EEZ	Country	Sovereign	Region	x1998	x1999	x2000	x2001	x2002
2	2304881	Croatian Exclusive Economic Zone	Croatia	Croatia	Europe	1.429434	0.89236	0.916761	1.348578	1.046767
3	2304882	Croatian Exclusive Economic Zone	Croatia	Croatia	Europe	1.626702	1.076252	1.032467	1.850453	1.286305
4	2309200	Croatian Exclusive Economic Zone	Croatia	Croatia	Europe	1.22904	0.808712	0.871164	1.056457	0.86687

J	K	L	M	N	O	P	Q	R	S	T
x2002	x2003	x2004	x2005	x2006	x2007	SLOPE	R2	CHANGE	SCHANGE	p
1.046767	0.739197	0.932941	0.720772	0.650771	0.676379	-0.06669	0.554133	-36.7643	-36.7643	<0.05
1.286305	0.909687	1.097883	0.893672	0.833048	0.963842	-0.06665	0.360483	-27.9716	-999999	n.s.
0.86687	0.607516	0.801937	0.710166	0.642566	0.580083	-0.05352	0.630802	-33.5555	-33.5555	<0.01

Figure 3: Setup of data table

NOTES ON REGRESSION ANALYSES

To obtain significance levels of our linear regression for each row of data (i.e. grid cell) we calculated t-values using the formula:

$$(1) \quad t = \frac{B}{SE(B)}$$

where B is the slope of the regression line and S.E. the standard error of the slope. The latter is calculated as:

$$(2) \quad SE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} / \sqrt{\sum (x_i - \bar{x})^2}$$

where y_i is the value of the dependent variable for observation i , \hat{y}_i is estimated value of the dependent variable for observation i , x_i is the observed value of the independent variable for observation i , \bar{x} is the mean of the independent variable, and n is the number of observations.

CALCULATION STEPS IN EXCEL

Using the data layout as shown in **Figure 3**, the following are the calculation steps and formula used to obtain indicator statistics in Excel (version 2003 and above).

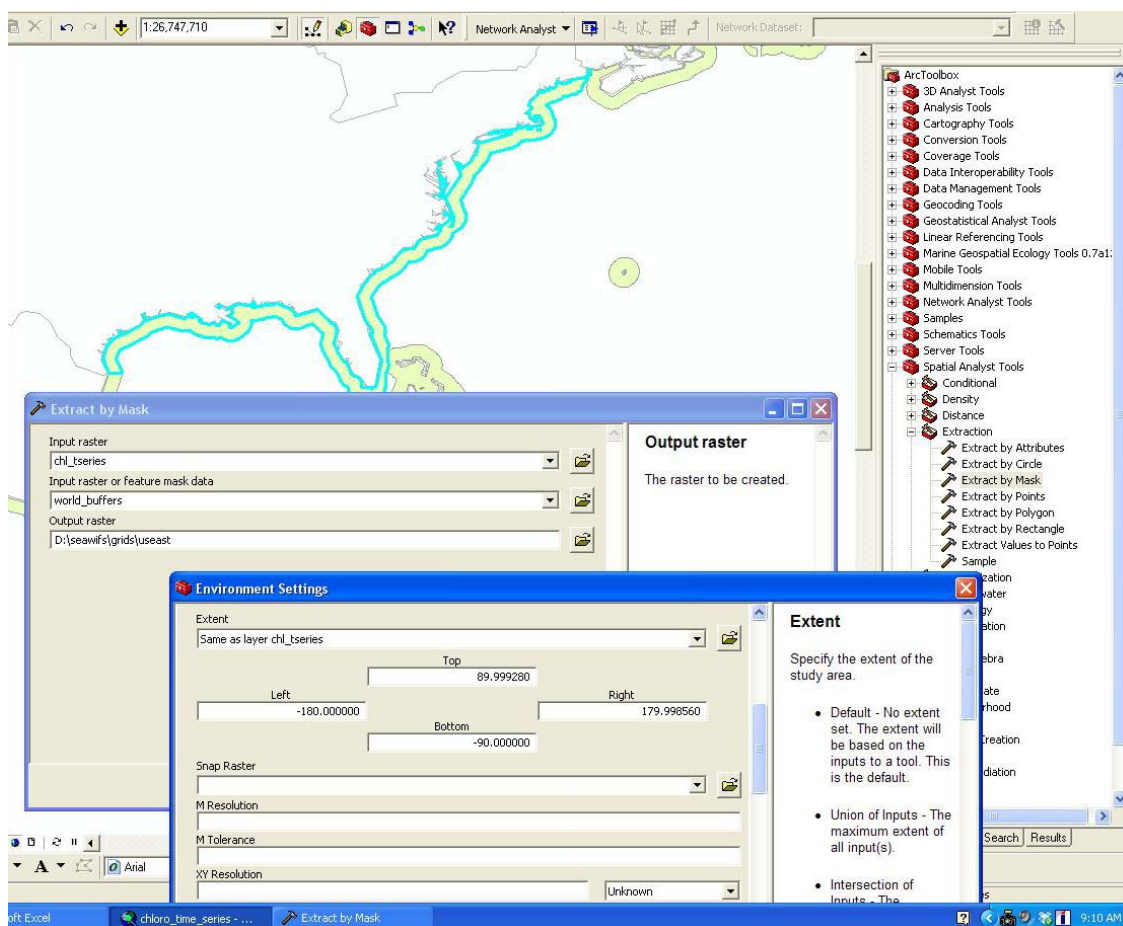
1. Calculate slope of linear regression line using the function LINEST, e.g.
`=LINEST(D2:M2,D1:M1,,FALSE)`
2. Calculate sum of squared differences between known y-values and predicted y-values based on the fitted linear regression line (nominator in equation 2 above)
 - a. `{=SUM((D2:M2-TREND(D2:M2,D1:M1,D1:M1,TRUE))^2/8)}`
 - i. TREND calculates the predicted y-values for a set of new x-values based on least squares method. In this case, the new x-values are the same as the known x-values, as we want to know the residuals from the predicted line for each year
 - ii. TRUE assures that constant b (in regression line $y=b + mx$) is calculated normally (i.e. intercept is included in regression)
 - iii. Making this an array function enables us to automatically sum these differences for every year in one destination cell, instead of having to create a new set of 10 columns and then adding the differences in another separate column. To make this an array function, press CTRL+SHIFT+ENTER instead of enter after typing the above equation
3. Calculate sum of squares for known x-values (years): same formula as above, but this time we will have one value for all rows (as x-values don't change, so the value from row 2 is copied and pasted into all other rows).
4. Calculate square roots of calculated summed differences for x and all y
5. Calculate the standard error of the slope regression line by dividing SQRT Y by SQRT X
6. Calculate t-value: Slope (from step 1) divided by its standard error (step 5)
7. Look up t-value cutoffs using Excel's TINV function: `=TINV('probability level', 'df')`.
 - a. For calculating 2-tailed t-values of a regression: $DF=N-2$
8. Bin the t-values from step 6 into significance levels with the following formula:
`=IF(AND(ABS(Z2)>2.306,ABS(Z2)<3.355),"<0.05",IF(AND(ABS(Z2)>3.355,ABS(Z2)<5.041),"<0.01",IF(ABS(Z2)>5.041,"<0.001","n.s.")))`

ANALYSIS SCENARIOS

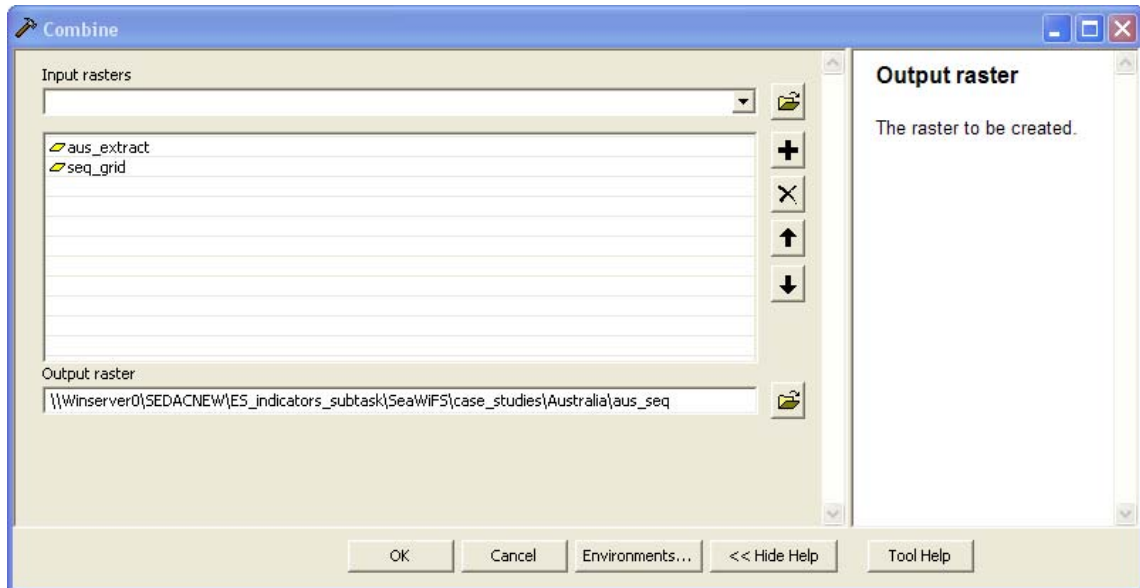
The following are two scenarios for how to extract time series data for specific regions of interest from the provided global grids (yearly or time series)

SCENARIO 1: SUB-SETTING BASED ON A POLYGON BUFFER

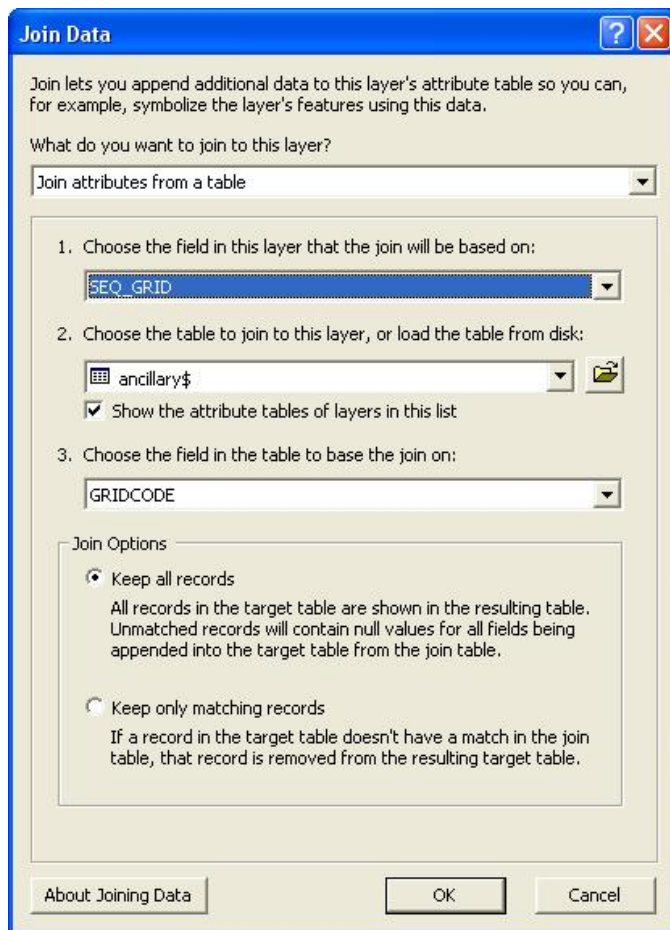
1. Add a layer containing the buffer to an ArcMap document containing a global grid (time series grid or a selected yearly grid)
2. Select the desired buffer (the example shown uses one of the coastal buffers we used to create the dataset)
3. Using Spatial Analyst Tools, **EXTRACT** the global grid using the selection (Extract by Mask) – in the **EXTRACT BY MASK** dialog box, set **EXTENT** in the **ENVIRONMENT** to extent of global grid to avoid grid cell shift during calculation



- Combine the extracted raster with the sequence grid (SPATIAL ANALYST TOOLS – LOCAL – COMBINE)



- Load the ancillary data file (ANCILLARY.CSV) into ArcGIS as a source table.
- Join the resulting grid to the appropriate ANCILLARY dataset, matching the SEQ_GRID field to GRIDCODE field.



7. EXPORT attribute table for further analyses and/or EXPORT data to make additional attribute information permanent

SCENARIO 2: INTERACTIVE SELECTION OF ARBITRARY AREAS

1. Add CENTROID layer to ArcMap document containing the raw data
2. Select and subset centroids covering the area of interest, copy selection to a new layer
3. Join the centroid selection to the appropriate ancillary data and export attribute data

REFERENCES

- Antoine, D. 2005. Recent interdecadal evolution (1980-2000) of the phytoplankton biomass in the global ocean. *La Lettre du Changement Global (Lettre PIGB-PMRC-France)*. May 2005, No. 19 : 22-32.
- Hu, C., F.E. Muller-Karger, C. Taylor, K.L. Carder, C. Kelble, E. Johns, C.A. Heil. 2005. Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters. *Remote Sensing of the Environment*. 97: 311-21.
- Lavender, S.J., Pinkerton, M.H., Froidefond, J.M., Morales, J., Aiken, J., Moore, J., 2004. G.F. SeaWiFS validation in European coastal waters using optical and bio-geochemical measurements. *International Journal of Remote Sensing* 25 (7-8), 1481-1488.
- Roberts, J. J., B. N. Best, and P. N. Halpin. 2007. Marine Geospatial Ecology Tools. Available online: <http://mgel.env.duke.edu/tools>.
- VLIZ (2006). Maritime Boundaries Geodatabase Version 1.0. Available online at <http://www.vliz.be/vmdcdata/marbound>.

FIGURES AND TABLES

GLOBAL AND COUNTRY STATISTICS

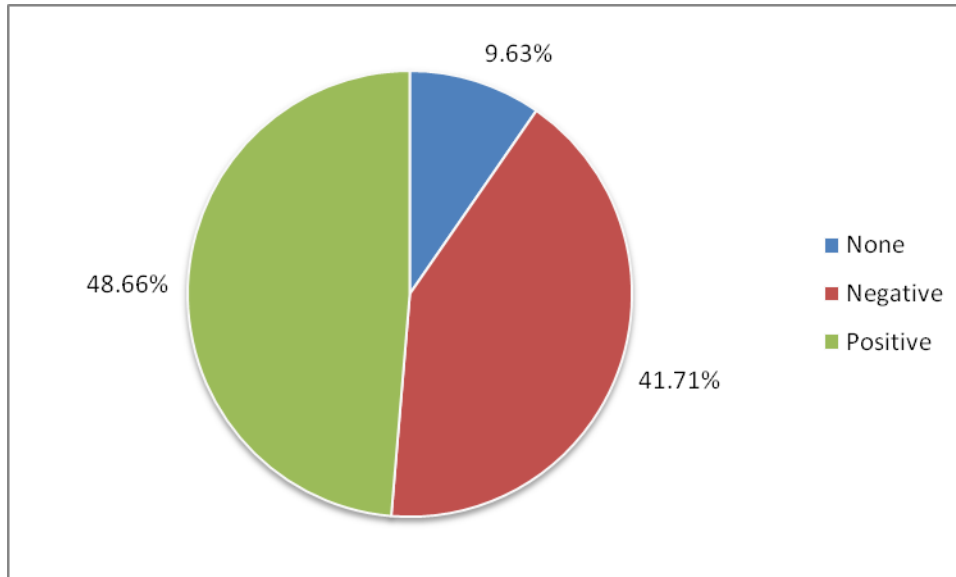


Figure 4: Percent of countries where significant trends in chlorophyll concentrations could be identified in at least some proportion of their coastal waters.

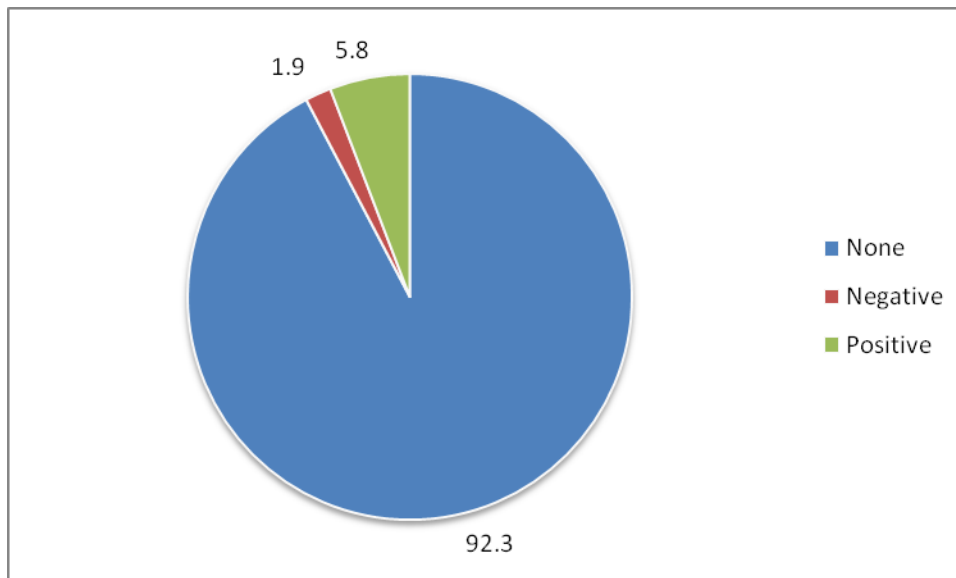


Figure 5: Percent of all grid cells covered by our global near-coastal-zone buffer (10-100 km off shore) where trends in chlorophyll concentrations between 1998 and 2007 were either absent, positive, or negative

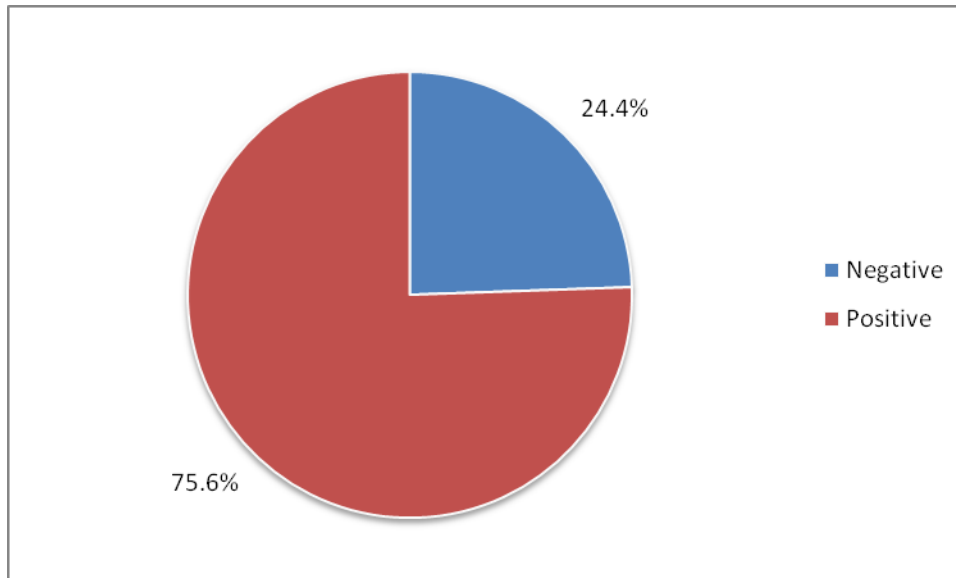


Figure 6: Percent of grid cells with significant trends in chlorophyll concentrations between 1998 and 2007 that showed positive versus negative change

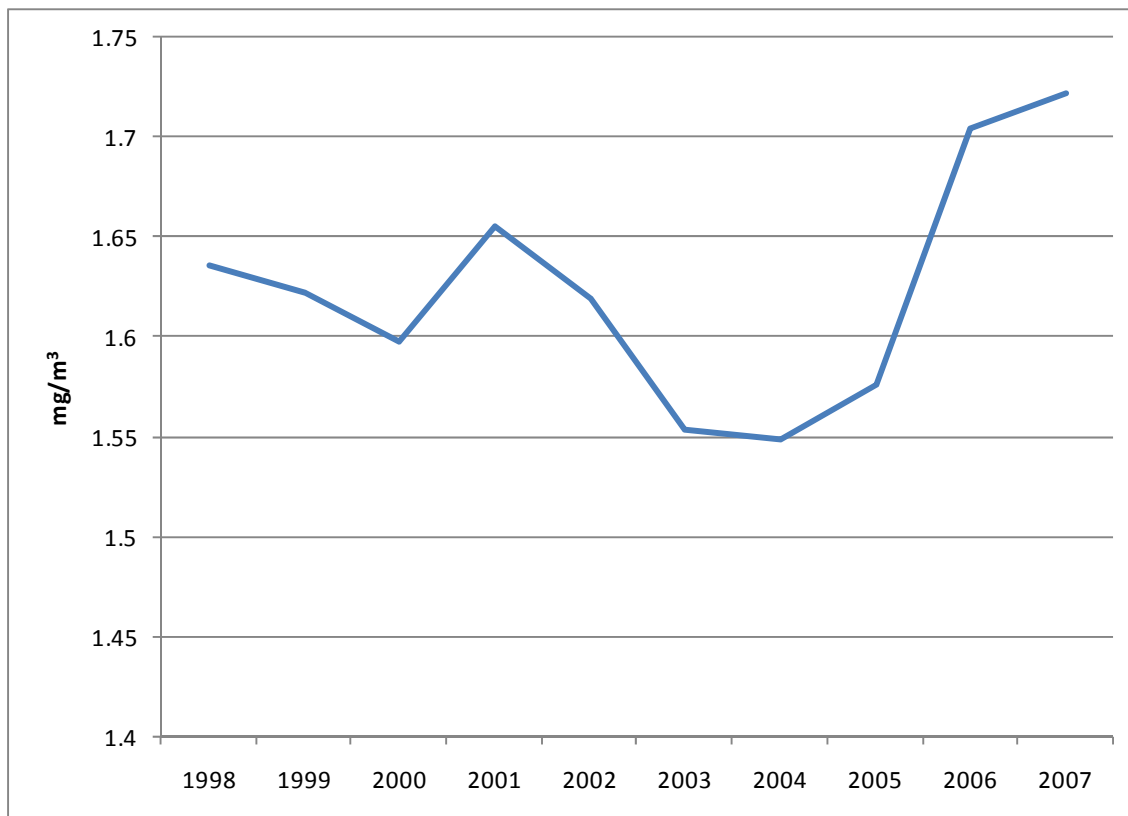


Figure 7: Change in average annual global chlorophyll-a concentrations in near-coastal waters between 1998 and 2007

Table 2: Country/territorial offshore waters in which chlorophyll-a concentrations showed a statistically significant positive trend line for all annual average observations from 1998 to 2007, the corresponding increase in percent (average of 2005-2007 compared to 1998-2000 values), and average coastal concentrations (mg/cubic meter) in the first and last year of the time series [note: in some cases the difference between start and end values is within the error bars of sensor accuracy and should be interpreted with caution]

Country/Territory	% Change	1998	2007
Cuba	15.3	0.44	0.50
Eritrea	28.2	1.36	1.68
Fiji	4.1	0.16	0.18
Finland	18.6	8.28	10.62
Iran	12.0	1.20	1.51
Kuwait	23.2	2.42	2.60
Peru	31.0	1.88	3.03
Qatar	16.3	1.22	1.53
Saudi Arabia	16.7	0.52	0.62
Sudan	27.8	0.30	0.36
United Arab Emirates	18.7	1.24	1.63

Table 3: Country/territorial offshore waters in which chlorophyll-a concentrations showed a statistically significant negative trend line for all annual average observations from 1998 to 2007, the corresponding decrease in percent (average of 2005-2007 compared to 1998-2000 values), and average coastal concentrations (mg/cubic meter) in the first and last year of the time series [note: in some cases the difference between start and end values is within the error bars of sensor accuracy and should be interpreted with caution]

Country/Territory	% Change	1998	2007
Croatia	-14.9	0.35	0.26
French Polynesia	-16.8	0.07	0.06
Georgia	-18.0	1.29	1.08
Hawaii	-5.7	0.08	0.08
India	-12.5	1.74	1.57
Italy	-12.5	0.31	0.26
Jamaica	-7.1	0.13	0.12
Prince Edward Islands	-27.2	0.24	0.18
Saint Kitts and Nevis	-12.3	0.14	0.12
Turkey	-24.0	0.93	0.67

Table 4: Coastal zones where significant **increases** in chlorophyll concentrations occurred in more than 10 percent of the area, and the mean percent of change in the affected grid cells

Country	% of Area	Mean Change in %
United Arab Emirates	65.0	28.2
Qatar	49.5	23.3
Eritrea	49.2	40.9
Bahrain	46.3	23.1
Kuwait	40.5	41.5
Sudan	31.8	40.0
Iran	23.9	29.1
Estonia	23.8	36.1
Saudi Arabia	21.1	26.7
Finland	14.3	38.0
Peru	14.3	76.4
China	13.9	14.5
Germany	13.8	33.3
Yemen	12.7	8.5
Indonesia	11.8	20.6
Norway	11.6	8.4
Tonga	11.6	26.3
Chile	11.5	60.8
Fiji	11.5	15.5
New Zealand	10.9	41.5
Morocco	10.9	28.1

Table 5: Coastal zones where significant **decreases** in chlorophyll concentrations occurred in more than 10 percent of area, and the mean percent of change in the affected grid cells

Country	% of Area	Mean Change in %
Saint Kitts and Nevis	75.9	-13.4
French Polynesia	58.5	-18.6
Georgia	46.1	-23.8
Turkey	32.8	-25.7
Cape Verde	32.2	-50.4
Jamaica	31.7	-12.5
Montserrat	29.6	-13.0
Micronesia	28.7	-24.9
Cook Islands	27.6	-14.8
Tokelau	22.8	-23.4
Hawaii	22.1	-10.1
Antigua and Barbuda	21.5	-11.8
Heard and McDonald Islands	20.5	-25.7
Prince Edward Islands	20.1	-33.2
Anguilla	20.0	-14.8
Libya	17.9	-10.9
Croatia	17.8	-19.5
Belize	16.7	-12.1
India	16.3	-28.5
Cuba	15.5	-0.7
Puerto Rico and Virgin Islands	14.7	-17.2
Guinea Bissau	14.6	-11.4
Andaman and Nicobar	14.2	-20.4
South Georgia and the South Sandwich Islands	12.2	-54.4
British Virgin Islands	11.5	-16.2
Denmark	10.5	-24.7

SELECTED IMAGERY

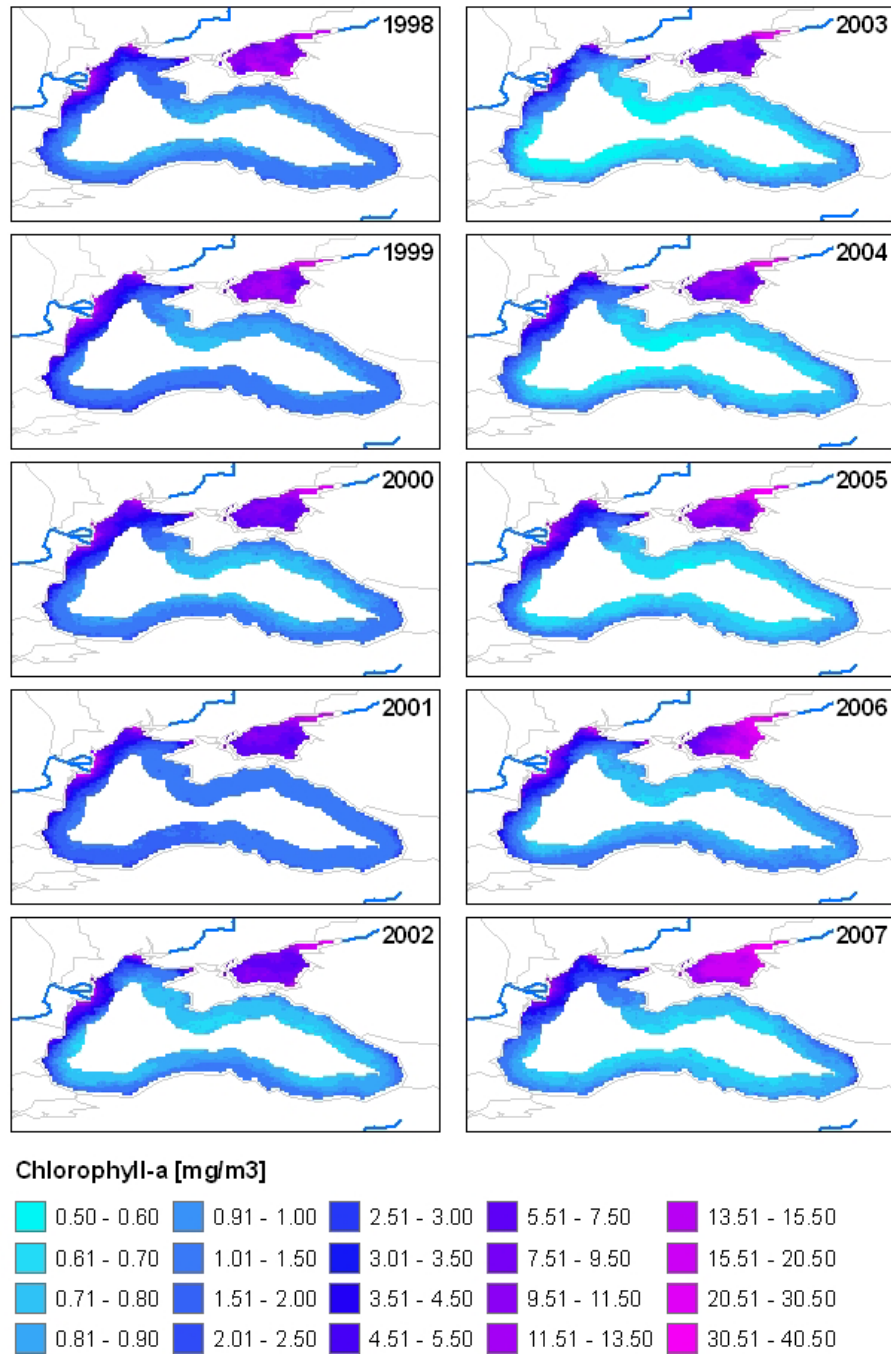


Figure 8: Time series of chlorophyll-a concentration in the black sea