

GRIMAS: GRIDDING AND ARCHIVING OF SATELLITE-DERIVED OCEAN COLOUR DATA FOR ANY REGION ON EARTH

Quinten Vanhellemont, Bouchra Nechad, Kevin Ruddick

Management Unit of the North Sea Mathematical Models (MUMM), Royal Belgian Institute of Natural Sciences (RBINS), Gulledele 100, 1200 Brussels, Belgium,
e-mail: quinten.vanhellemont@mumm.ac.be

ABSTRACT

A tool for GRIDDING of MARine Satellite data (GRIMAS) was created to archive available remote sensing data for any region on earth. GRIMAS is a set of flexible image processing software that can process large numbers of MODIS and MERIS images in order to create a gridded dataset of chlorophyll *a* concentration (chl), of sea surface temperature (sst – MODIS only) and of total suspended matter concentration (tsm). The software focuses on flexibility; a new region can be configured by defining the bounding box in latitude and longitude. The software can be easily adapted to produce new and custom products. Additional sensors or (georeferenced) datasets can be added if an appropriate reader is available or implemented. In a multi-year archive of remotely sensed parameters, oceanographic features are apparent. Such archives can be very rapidly generated and used in support of a wide variety of ocean related disciplines.

Keywords: *ocean colour, chlorophyll concentration, total suspended matter concentration, sea surface temperature, MODIS, MERIS, data archive, time-series*

1 INTRODUCTION

The Moderate Resolution Imaging Spectroradiometer (MODIS) and the Medium Resolution Imaging Spectroradiometer (MERIS) are multi-spectral sensors on board EOS-PM ('Aqua') and ENVISAT, both of which are polar orbiting satellites launched in 2002. Since then, they have been systematically recording our planet's land and oceans. Oceanographic parameters related to ocean colour can be derived from the spectral bands of these sensors, such as chlorophyll *a* and total suspended matter concentrations. The imagery that has been collected by these sensors can provide an extensive and unprecedented cover for coastal and oceanic regions, both regarding temporal and spatial aspects. In this paper we present a set of processing tools that can process and grid the data in this multi-year image set, in order to make the data more accessible to oceanographic researchers.

In the BELCOLOUR project (<http://www.mumm.ac.be/BELCOLOUR/EN/index.php>) an archive of chlorophyll *a*, total suspended matter and sea surface temperature maps was generated from SeaWiFS, MODIS-Aqua and MERIS data for several (nested) regions in the North Sea area (NOS, 48.5° N – 60° N, 4° W – 9° E). Recently, a remote sensing data archive and a near real-time data service was needed for the JELLYFOR project. The JELLYFOR project aims to set up a jellyfish forecasting system based on *in situ* and remote sensing data for a Spanish and an Irish region. Therefore, a new, generalized, and improved tool for GRIDDING of Marine Satellite data (GRIMAS) was created in Interactive Data Language (IDL, ITT Visual Information Solutions), based on BELCOLOUR software. In order to download, process and upload near real-time data, a wrapper script that uses the GRIMAS tool was written in Perl. Because of the generalized approach, the software allows the user to process large amounts of remote sensing data for any region on earth. The processing of the imagery is rapid, and when a gridded dataset is available, extraction of time-series and multi-

temporal products is trivial.

A number of tools exists that offer similar functionality to import, reproject and export satellite data. The SeaDAS package from the Ocean Color group at NASA/GODDARD (<http://oceancolor.gsfc.nasa.gov/>) offers a very complete set of tools to process and analyse ocean colour data from various sensors, including data display, batch processing and mapping capabilities. SeaDAS offers processing from low level satellite data, with custom atmospheric corrections and L2 output products. Map projection and multi-temporal binning options are available. The BEAM toolbox (<http://www.brockmann-consult.de/cms/web/beam/>) includes excellent display capabilities and offers reprojecting, subsetting and band maths, and also has a multi-temporal binning processor. Giovanni (<http://disc.sci.gsfc.nasa.gov/giovanni/>) offers time-series extraction and multi-temporal composite generation, but is limited to the reduced resolution (temporal and spatial) of the used input data. The GLOBCOLOUR project (<http://www.globcolour.info/index.html>) offers gridded datasets, with merging of data from several sensors (MODIS, MERIS, SeaWiFS) with a resolution of 4 km for the global dataset, and 2 km for the European dataset.

Batch processing in SeaDAS and BEAM requires good knowledge of the applications, and additional scripting is required. GRIMAS complements these applications with configuration free batch processing and reprojection to a standard grid. It is especially suited to process large numbers of standard L2 files and generate maps and gridded NetCDF files (<http://www.unidata.ucar.edu/software/netcdf/>). In addition to MERIS N1 L1B and L2 files, GRIMAS can import MODIS HDF L1B and L2 files, either standard files from the Ocean Color website or custom L2 files generated with SeaDAS. Any of the datasets from the source files can be added to the output grid files. Analysis and mapping tools for the gridded output files are also included. In contrast to, for example, GIOVANNI, non-standard products based on L1B/L2 datasets can easily be generated and included in the output files. With appropriate data readers, more datasets can be reprojected on the same grid, including past and future sensors. GRIMAS offers easy to use output products (NetCDF) which can be imported in GIS applications.

2 METHODS

2.1 REGION DEFINITION AND GRID CALCULATION

Region definition is as simple as defining four bounding coordinates in an ASCII configuration file, the northern (ϕ_1), southern (ϕ_0), eastern (λ_1) and western (λ_0) borders of the region. A custom equirectangular projection or 'grid' is calculated per region for several standard resolutions, based on the sensor resolutions, e.g. 1000 m for MERIS reduced resolution and MODIS data and 250 m for MERIS full resolution and MODIS high resolution data. The number of grid cells in longitude (x) and latitude (y) is calculated so that the pixel resolution at the centre latitude of the region corresponds to the required grid resolution. The central latitude is chosen as a reference, in order to reduce the error of using a non-equal-area projection. The distance covered by one degree of longitude at this latitude is given by $^\circ\lambda$ in equation [1] and the distance covered by one degree of latitude is given by $^\circ\phi$ in equation [2].

$$^\circ\lambda = 2 \cdot \pi \cdot R \cdot \cos((\phi_0 + \phi_1)/2 \cdot \pi/180^\circ)/360^\circ \quad [1]$$

$$^\circ\phi = 2\pi \cdot R/360^\circ \quad [2]$$

with R the earth's radius (used value: 6378.1370 km). The calculated distances are in kilometre by default and thus have to be scaled to get different grid resolutions. The number of grid cells (n_s and n_l) along the X and Y axes (longitudes and latitudes respectively) is

calculated following equations [3] and [4], rounded to the nearest integer value.

$$ns = \text{round}(\lambda \cdot |\lambda_0 - \lambda_1|) \quad [3]$$

$$nl = \text{round}(\varphi \cdot |\varphi_0 - \varphi_1|) \quad [4]$$

Datasets are reprojected to this custom grid using a nearest-neighbour resampling method, written in IDL. The nearest neighbour approach is chosen not to blur oceanic features and fronts between water masses, and to avoid mixing of water pixels with land or cloud/haze pixels.

2.3 SATELLITE DATA

Satellite remote sensing data is available in several processing levels. L0 files contain the raw sensor data. L1 files contain calibration, geolocation and top of atmosphere radiances. L2 files contain so called data products, such as (atmospherically corrected) marine reflectances and chlorophyll *a* concentration together with geolocation and quality flags. L3 files contain multi-temporal L2 products. In GRIMAS, L1 scenes are used only to create top of atmosphere ‘true colour’ maps, and will not be discussed further. For each L2 scene that is processed, the necessary datasets are imported, quality controlled and reprojected to the standard grid. L3 files can be generated from the gridded archive or on the fly.

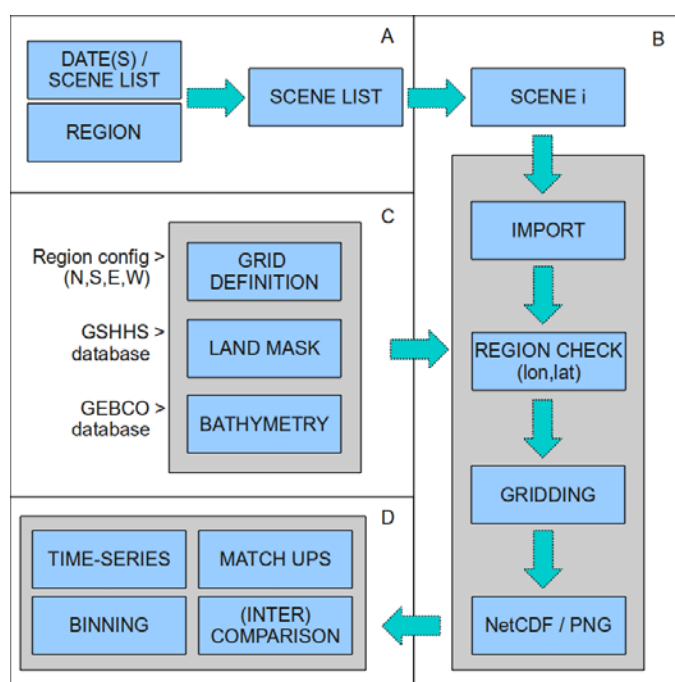


Figure 1. An overview of the GRIMAS processing chain. A) Required inputs are either the region and date(s), with the satellite images (MODIS/MERIS standard or custom L2) stored in a standardized manner, or the region and a list of files. B) Each scene is imported and checked for region cover, gridded and exported if necessary. C) If required, the grid is calculated, and land mask and bathymetry files are generated. D) A set of tools facilitates extraction of information, such as time-series and multi-temporal bins, from the archive of exported NetCDF files.

Some of the parameters are considered mature and are supplied by the space agencies. Other, more experimental or custom products, such as turbidity and total suspended matter concentration (Nechad et al., 2010), are computed from the included marine reflectances. Sea surface temperature can also be estimated with MODIS, and the dataset is available in the standard OceanColor distributed files.

2.4 MODIS L2 DATA

The MODIS L2 data is stored in the HDF format, and the appropriate IDL functions are used for accessing the files. The same format is used for SeaWiFS data, and a similar approach to the one described below can also be used to process SeaWiFS data. The current MODIS processing version is R2009.1 for data before, and R2010.0 for data after 01/01/2009. No or

only minor changes to GRIMAS' data reader are anticipated with future processor updates. Detailed information on the MODIS L2 products and flags is available on the OceanColor website (for example here: http://oceancolor.gsfc.nasa.gov/DOCS/Ocean_Level-2_Data_Products.pdf).

For each file the longitude and latitude datasets are read and a check is performed whether the target region is within the file. If it is not, the file is skipped, else a number of datasets is imported: (1) l2_flags, 32 bit quality flags, set during the processing from L1 to L2, (2), aot_869, aerosol optical thickness at 869 nm and (3) Rrs_xxx, Remote sensing reflectances at wavelengths 412, 443, and 448 nm. These datasets are interpreted to create an eight-bit set of scene flags (sc_flags), (described in table 1).

ID	Flag name	Flag use	Datasets or flags used for	
			(a) MODIS	(b) MERIS
1	NODATA	pixel out of scene limits	set for pixels outside scene	!= water, land, cloud
2	CLOUD	clouds/haze	== cloud	== cloud
3	GLINT	uncorrected sun glint	== hisunglint	== high_glint
4	HIANG	high sun/sat angles	== hisenz, hisolz	sun zenith > 75 sat zenith > 60
5	ATMFAIL	failure of atmospheric correction	== atmfail, maxaeriter, hilt, hisunglint aot_869 < 0 Rrs_412, 443, 448 < 0	not used
6	STLIGHT	stray light effect	== stlight	not used
7	LAND	land pixel	== land + GSHHS (see 2.6)	== land + GSHHS (see 2.6)
8	TURBID	turbid water	== turbid	== case_2s
9–16	SPARE	spare flag		

Table 1: GRIMAS Scene flags dataset definitions for MODIS and MERIS. '===' : if at least one of the named level 2 flags is set, '!=' if the named level 2 flag is not set. MODIS flag names follow the naming convention of NASA's Ocean Color team.

Additional datasets are imported according to the desired products in the gridded files. For the standard set of parameters, chlorophyll (chl) and total suspended matter (tsm [Nechad et al., 2010]) concentration and sea surface temperature (sst), the required datasets are: (1) chlor_a, chlorophyll a concentration, computed with OC3, (2) Rrs_667, remote sensing reflectance at 667 nm, to compute tsm and mask chl in turbid waters and (3) sst, sea surface temperature. A confidence check is performed for each of the required products and this information is saved in a dataset flags (ds_flags) dataset. As the number of datasets is dynamic, the dataset flag definition can change per file. Flag values are stored as attributes to the datasets. For the standard datasets (chl, tsm, sst) table 2a gives dataset flags definition.

Regarding product quality, the algorithm of Nechad et al. (2010), used to compute tsm from MODIS reflectance data, was calibrated in turbid waters and might provide inaccurate results

for clearer waters. The algorithm can be easily adapted with knowledge of regional specific inherent optical properties (SIOPs). The MODIS chlorophyll *a* dataset is known to be less reliable in turbid waters, therefore an additional quality masking is applied in these waters (Park et al., 2010, paragraph 2.3).

2.5 MERIS L2 DATA

MERIS scenes are stored in the ENVISAT N1 format. The datasets are imported from the scenes using an IDL routine, written in accordance with the ENVISAT file format definitions (available at http://earth.esa.int/pub/ESA_DOC/ENVISAT/Vol11_Meris_5b.pdf). L2 files from the MERIS processor version IPF 5.05 ('second reprocessing') are used. The third reprocessing was completed mid 2011, and data dissemination will follow later, and also here only minor changes to GRIMAS' data reader are anticipated.

Dataset	Description	Confidence flag set using	
		(a) MODIS	(b) MERIS
CHL	Chlorophyll <i>a</i> concentration	== 1, 2, 5, 6, 7 (table 1) == chlfail QC of Park et al., 2010	== 1, 2, 7 (table 1) == PCD_17
TSM	Total suspended matter concentration	== 1, 2, 5, 6, 7 (table 1) Rrs_667 < 0	== 1, 2, 7 (table 1) == PCD_17
SST	Sea surface temperature	== 1, 2, 5, 7 (table 1) == sstfail	

Table 2: GRIMAS Dataset confidence definitions for a file with the two or three 'standard products' from a) MODIS: chl, tsm, sst and b) MERIS: chl, tsm. '==' : if at least one of the named scene or level 2 flags is set, '!=' if the named scene or level 2 flag is not set.

For each file the longitude and latitude tie point grids are read and a check is performed whether the target region is inside the file. If it is not, the file is skipped, else the tie point grids are interpolated to the full scene and a number of datasets is imported, to create the scene flags dataset (see table 1b): (1) l2_flags, 32 bit quality flags, set during the processing from L1 to L2 and (2) sun_zenith and view_zenith, tie point grids, also interpolated to full scene. For the default MERIS products (chl and tsm) the following datasets are imported: (1) algal_2, the neural network chlorophyll concentration product and (2) total_susp, the neural network suspended matter concentration product (Schiller and Doerffer, 1999). Confidence datasets are generated, based mainly on the associated product confidence flags (PCD) see table 2b for details. More information on these flags and datasets can be found in the detailed MERIS L2 processing document (available at:

http://earth.esa.int/pub/ESA_DOC/ENVISAT/MERIS/MERIS_DPML2_i7r2A_re-issued.pdf)

2.6 ADDITIONAL INPUT DATA

Because of the generalised approach, there is no limit to the spatial data that can be processed, as long as it is georeferenced and an appropriate reader is available or implemented. Depending on the application different datasets may be interesting, e.g. wind or other meteorological data, hydrographic climatologies such as salinity and temperature, or validation data for hydrodynamic or ecosystem models. The full resolution version (0.04 km) of NOAA's Global Self-consistent, Hierarchical, High-resolution Shoreline Database

(GSHHS, Wessel et al., 1996) is used to extract land masks to additionally mask the land data on the processed grid, see table 1. When required, e.g. for exporting maps with isobaths, the GEBCO 0.5° or 1° bathymetry database is accessed to read water depths (available from https://www.bodc.ac.uk/data/online_delivery/gebco/).

2.7 OUTPUT DATA

Gridded scenes can be exported to NetCDF files or mapped to PNG files, with customizable annotations (e.g. scale bar, colour bar, gridlines, isobaths). In the NetCDF files, longitudes and latitudes are stored from west to east and from south to north in two single precision floating point vectors. These vectors give the geolocation for the centre of any (x,y) point in the dataset arrays. The datasets are stored in a floating point array of ns by nl dimensions. Two flagging datasets are included in the file, one containing general scene flags (sc_flags) and one containing dataset confidence flags (ds_flags). Depending on the number of flags, eight, sixteen or thirty-two bit datasets are used (default: sixteen bit).

3 RESULTS

3.1 GRID IMPLICATIONS

The grid definition makes using the archive very easy, as one vector for longitudes and one for latitudes is enough to specify the geolocation for all the files of the same grid. This facilitates powerful queries such as extracting of a multi-year time-series for a given point, or makes finding match-ups trivial. However, since the gridding uses an equirectangular projection centred along the central latitude, pixel sizes between longitudes shrink or expand when moving away from this latitude. Therefore the used grid is less suitable for large regions as the differences in pixel size between the first and last row of data become too large. The pixel size in these rows also does not compare well to the resolution of the sensor. Moreover, the true sensor resolution is greatly reduced at the edges of an image. An equal-area projection, such as a sinusoidal projection, will be considered for future development.

3.2 REMOTE SENSING DATA ARCHIVE

For the three regions in the JELLYFOR project (table 3), all MODIS Aqua L2 files from NASA's OceanColor website (<http://oceancolor.gsfc.nasa.gov/>) and all MERIS reduced resolution (RR) L2 files available on ESA's Mercator application (<http://merci-srv.eo.esa.int/merci/>) were ordered and processed for 2003-2010, the eight full years in which both sensors were active.

Region	Catalan Sea (CTL)	North Irish coast (NOI)	Belgian coast (BCZ)
ϕ_0, ϕ_1	0.5000°E, 3.4970°E	8.8900°W, 5.3539°W	1.8000°E, 3.9964°E
ns ($\Delta\phi$)	1004 (~0.0030)	906 (~0.0030)	612 (~0.0036)
λ_0, λ_1	40.0000°N, 42.4977°N	54.2500°N, 55.6078°N	50.8500°N, 51.7978°N
nl ($\Delta\lambda$)	1113 (~0.0022)	606 (~0.0022)	423 (~0.0022)

Table 3: Overview of the JELLYFOR regions and the automatically generated 250 m grid. The grid step is the range divided by the number of cells minus one, but this computation is unnecessary as latitude and longitude vectors are included in each file.

An incredible amount of information can be extracted from an archive of almost ten years of remote sensing data, including for example time-series for every location within the region and monthly and climatological average maps. The arithmetic mean is used when creating climatologies (IOCCG, 2004, p.38). Scenes are first combined per day (daily bin), before averaging over longer periods. If MODIS data from multiple overpasses is available, the best quality data is selected based on viewing angle. In quarterly climatologies of total suspended matter in the BCZ region (figure 2), there is a good agreement between MODIS and MERIS. Slightly lower concentrations (note the logarithmic scale) are observed offshore in the MODIS dataset spring and summer (Q2,Q3). Increased resuspension of sediments during the windier winter months causes higher suspended matter concentrations (Q1,Q4), lowest concentrations are observed during summer (Q2,Q3).

Climatological mean maps can also be useful to identify different water masses. A clear difference in temperature can be observed between the stratifying North-East Atlantic and well-mixed Irish Sea water masses (figure 3). This distinction can also be made in the chlorophyll *a* and total suspended matter datasets.

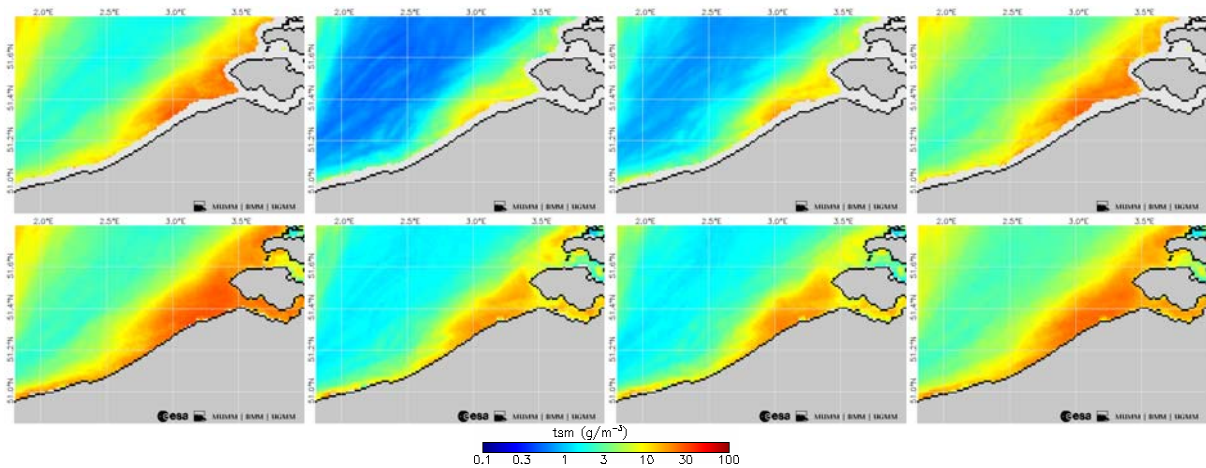


Figure 2: Quarterly climatologies (left to right: Q1-Q4) of tsm from MODIS (top row) and MERIS (bottom row) for the BCZ region. Near-coast pixels are excluded from the MODIS products to avoid the stray light effect.

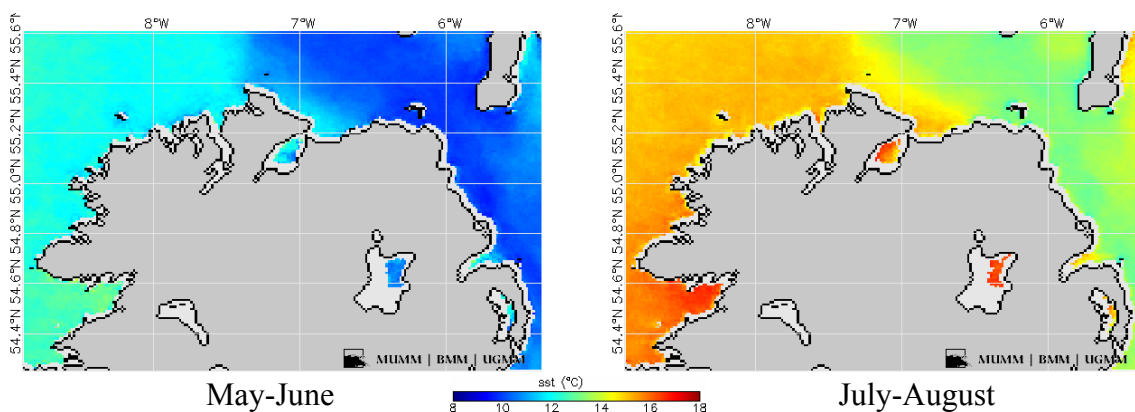


Figure 3: Two bimonthly climatologies of MODIS derived sst for the NOI region. Two different water masses can be observed: the stratifying North-East Atlantic waters in the west and the colder, well-mixed Irish Sea in the east.

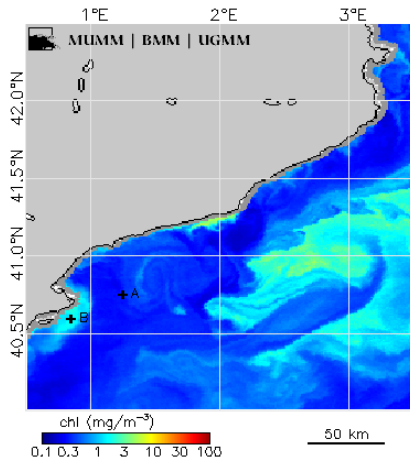


Figure 4: MODIS Aqua chl dataset for 10/04/2010 for the CTL region, with the locations of two points for which time-series were extracted, one in and one about 50 km from the plume of the Ebro river. A large phytoplankton bloom can be seen in the east.

Time-series can be extracted for any point in the region; a time-series for two points in the CTL region (figures 4 and 5) show that higher concentrations of chl and tsm are found in closer to the mouth of the Ebro river. Lowest chl and tsm values are observed in summer, corresponding to high water temperatures, and vice versa in winter.

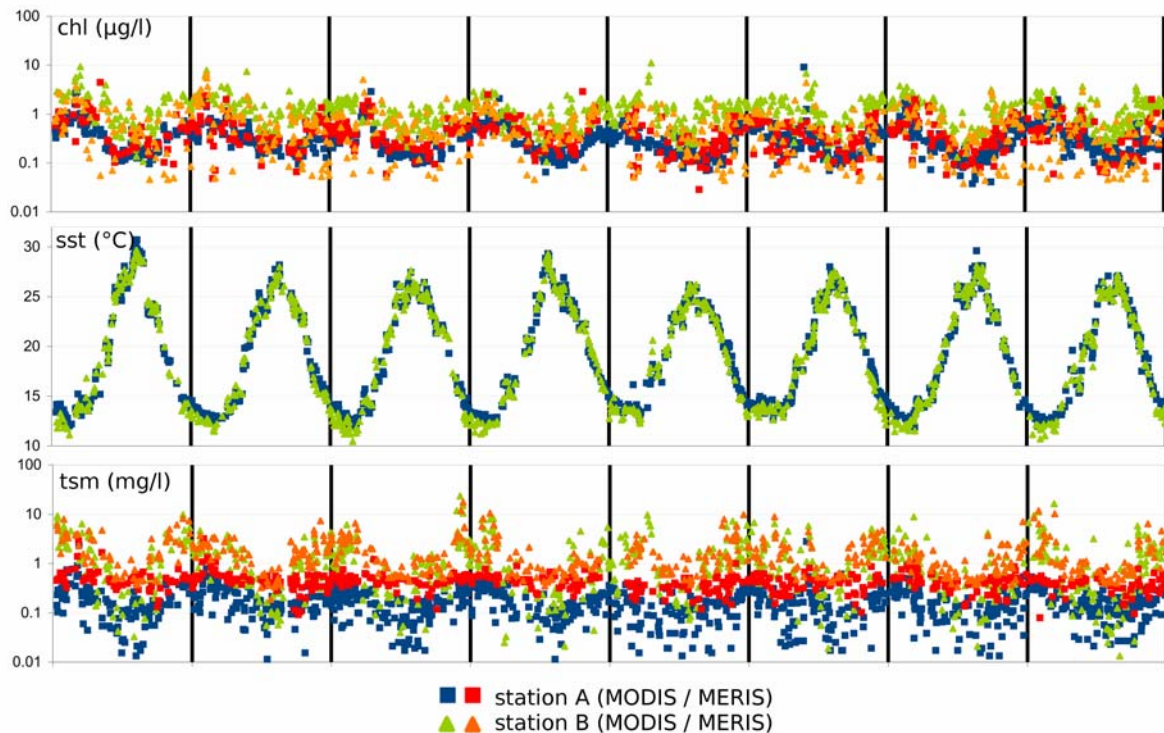


Figure 5: Eight year (2003–2010) time-series of chl and tsm from MODIS and MERIS and sst from MODIS for two locations in the CTL region (A: 40.75° N, 1.25° E and B: 40.60° N, 0.85° E, see figure 4). Values represented are the median of a 5 by 5 kernels over the point, extracted only if 13 or more of the kernel pixels are valid.

4 CONCLUSION

The GRIMAS software for extracting and gridding of the datasets is generic, highly automated and flexible, so that a multi-year archive of remote sensing data can readily be created for any region on earth. GRIMAS aims in making the available ocean colour data more accessible, and can also be used to process MODIS and MERIS (both full and reduced resolution) data on a near real-time basis. A long-term remote sensing dataset is a useful tool

for understanding the oceanography of any region, be it a well-studied or a relatively unknown one. Using monthly composites and time-series, the inter-annual changes and the evolution throughout the year can be analysed. The dataset can be used for a wide range of applications in marine biology, sediment transport, coastal management, etc. Some products are only valid for open ocean (case 1) situations, and others, such as the tsm product from Nechad *et al.* (2010) are calibrated for turbid waters. For each new region, a validation or recalibration should be performed for optimization of results. This recalibration can be easily added to GRIMAS.

ACKNOWLEDGMENTS

This work was funded by the STEREO Programme of the Belgian Federal Science Policy Office in the framework of the JELLYFOR-BE project (SR/37/135). NASA and ESA are acknowledged for distributing MODIS and MERIS products, respectively.

REFERENCES

- IOCCG (2004). Guide to the Creation and Use of Ocean-Colour, Level-3, Binned Data Products. Antoine, D. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 4, IOCCG, Dartmouth, Canada.
- B. Nechad, K.G. Ruddick, Y. Park. (2010), Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters, *Remote Sensing of Environment* 114: 854-866
- Park, Y-P., Ruddick, K., Lacroix, G. (2010), Detection of algal blooms in European waters based on satellite chlorophyll data from MERIS and MODIS, *International Journal of Remote Sensing*, Volume 31 Issue 24
- Schiller, H., and R. Doerffer (1999), Neural network for emulation of an inverse model-operational derivation of case II properties from MERIS data, *Int. J. Remote Sens.* 20, 1735–1746
- Wessel, P., and W. H. F. Smith, (1996) A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res.*, 101, #B4, pp. 8741-8743