The Nansen Environmental and Remote Sensing Center

a non-profit
research institute affiliated with
the University of Bergen

Thormøhlensgate 47,
N-5006 Bergen
Norway

NERSC Technical Report no. 296

VALIDATION AND IMPROVEMENT OF PROCESSING ALGORITHMS FOR RETRIEVAL OF WATER QUALITY PARAMETERS IN NORWEGIAN COASTAL WATERS

SUPPORITING SERVICES DEVELOPED UNDER THE
NORWEGIAN SPACE CENTRE SATHAV
AND
ESA MARCOAST PROJECTS

by

Lasse H. Pettersson, Anton Korosov,
Dmitry Pozdnyakov and Yunfei Wang
NERSC/NIERSC
Kai Sørensen and Are Folkestad,
NIVA

Bergen, January 16, 2009

Nansen Environmental and Remote Sensing Center
Thormøhlensgate 47
N-5006 Bergen - NORWAY
Phone: +47 55 20 58 00 Fax. +47 55 20 58 01
E-mail: administrasjon@nersc.no Web.site: http://www.nersc.no

a non-profit environmental research center affiliated with the university of Bergen

REPORT

TITLE Validation and improvement of processing algorithms for retrieval of water quality parameters in Norwegian Coastal waters.	**REPORT No.** Technical report no. 296
CLIENT Norwegian Space Centre (NRS) European Space Agency (ESA)	**CONTRACT** NRS SatHav ESA GSE MarCoast
CONTACT PERSONS Terje Wahl, NRS Gordon Campbell, ESA	**AVAILABILITY** OPEN
AUTHORS Lasse H. Pettersson, Anton Korosov, Dmitry Pozdnyakov, Yunfei Wang, Kai Sørensen (NIVA), Are Folkestad (NIVA)	**DATE** 22nd December, 2008
APPROVAL *Lasse H. Pettersson, Leading scientist*	 *Ola M. Johannessen, Director*

TABLE OF CONTENT

1 INTRODUCTION 1

2 THE OPERATIONAL PROCESSING SCHEME AT NERSC 2

2.1 Generall script `cron_HAB.bash` 3

2.2 Level 2 processing 4

2.3 Level 3 processing 5

3 MARCOAST VALIDATION RESULTS 2007 10

3.1 Remote sensing data 10

3.2 The independent validation data 11

3.2.1 Chlorophyll-a validation 11

3.2.2 Total Suspended Matter (TSM) 13

3.3 Comparison of satellite and ground truth data 13

3.4 Validation results 14

3.4.1 Chlorophyll-a products – HPLC samples 14

3.4.2 Chlorophyll-a products – Fluorescence ferry box data 16

3.4.3 Total Suspended Matter 19

3.4.4 The Fluorescence Line Height product 20

3.5 Assessment of the validation results 23

4 SUGGESTIONS FOR IMPROVEMENT OF SERVICE AND PRODUCTS 23

4.1 Application of advanced atmospheric correction 23

4.2 Calculation of chlorophyll-a concentrations from the FLH product 24

5 CONCLUSIONS 24

1 Introduction

Validation of satellite EO data products with *in situ* observations is a challenge and difficult task however needed in order to get confidence in and improve the algorithms used for retrieval of geo-bio-chemical information form satellite EO sensors. Specific validation protocols have been developed and dedicated validation campaigns have been executed in cooperation between the oceanographic research community and the space agencies for performing validation studies of various satellite based EO geo-bio-chemical products and sensors.

The unique spatial and temporal coverage of the satellite EO data, integrating the surface signal over e.g. a one or more km^2, and the *in situ* sampling providing "point" observations, but some times instrument and method specific, accurate measurement of one or more parameter based on sampling of often a "few litres" of water at the surface or at some specific depth. The very different nature of these types of environmental observations implies that a one-to-one comparison cannot be expected in such validation investigations. Also the natural small- to meso-scale variability of the oceans, and in particular in the coastal waters, is within these spatial and temporal scales of variation and similar in magnitude as expected to be found between satellite and *in situ* observations. The fact that the observed parameters often are not exactly the same and needs to be converted to the same units also introduces errors in the validation exercise. Additionally, also *in situ* measurements of the "same parameter" is pending on the measurement protocol or instrumentation used, which may cause additional errors. However, under these conditions it is still viable to do a validation comparison between satellite and *in situ* measurements of water quality and algae bloom events as part of the MarCoast service portfolio.

The MarCoast Water quality (WQ) and Harmful algae bloom (HAB) monitoring service operated by the Nansen Center (http://HAB.nersc.no) is based on precursor services developed and operated for "the extended Norwegian" coastal waters, started with using SeaWiFS and AVHRR satellite EO data in 1998. Accordingly, the supplied information has been gradually been developed to exploit the capabilities and to meet the users expectations and needs. The provided EO based products (Table 1) are based on both standard processing provided by the space agencies and regionally tuned algorithms developed by the Nansen Centers in Bergen and St. Petersburg as the service provider.

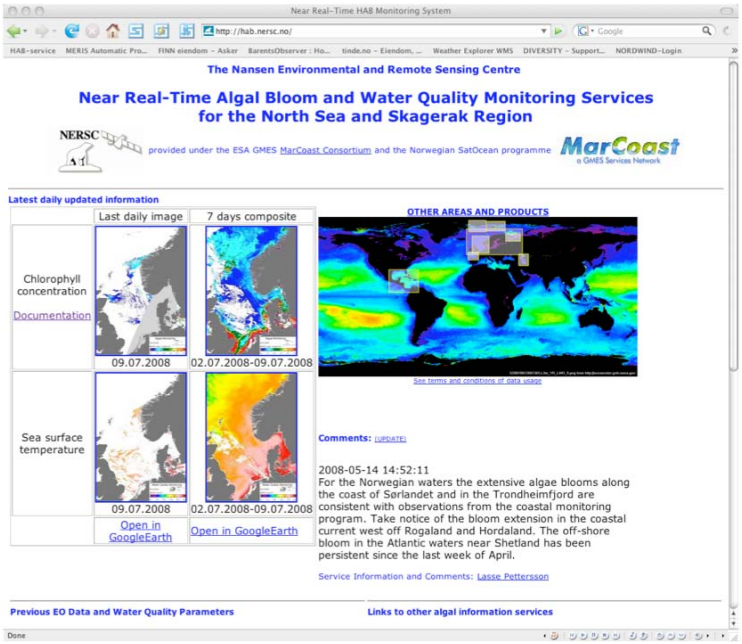

Figure 1: The access portal of the NERSC MarCoast web site (http://HAB.nersc.no) for monitoring water quality in the North Sea and Norwegian coastal waters. The map indicated the regions in which the service concept has been implemented.

Products name	Product	Algorithm	Validation data
chl case 1 ESA	Chlorophyll-a	ESA (case I)	Chl-a flurometric and HPLC samples
chl case II ESA	Chlorophyll-a	ESA (case 2)	Chl-a flurometric and HPLC samples
tsm ESA	Total Suspended Matter	ESA	Turbidity data available
cdom ESA	Colored Dissolved Organic Matter	ESA	No Yellow substance measurements
chl NERSC/NIERSC	Chlorophyll-a	NERSC/NIERSC	Chl-a flurometric and HPLC samples
tsm NERSC/NIERSC	Total Suspended Matter	NERSC/NIERSC	Turbidity data available
cdom NERSC/NIERSC	Colored Dissolved Organic Matter	NERSC/NIERSC	No yellow substance measurements
mse NERSC/NIERSC	Mean square error (difference between measured and reconstructed Rrsw spectra – a quality estimate)	NERSC/NIERSC	N/A
flh – NERSC	Phytoplankton concentration, i.e. Chlorophyll-a, measured by the Florescence Line Height.	NERSC	Chl-a flurometric and HPLC samples
RGB NERSC/NIERSC	Pseudo color image constructed from 2nd (442 nm), 5th (559 nm) and 6th (619) MERIS bands.	NERSC/NIERSC	N/A
sst	Sea Surface Temperature from MODIS	NASA	Validated standard product

Table 1: A summary of the NERSC MarCoast products generated by the service, their basis and the output parameter identification. All products are generated daily as one-day composites (1-2 passes) and 7-days running average as well as monthly once per calendar month. The validation parameters, provided by the NIVA FerryBox system, are indicated in the right column.

2 The operational processing scheme at NERSC

The service production chain for the information generation is based on near real-time ftp-access to the ESA on-line MERIS rolling archive in Kiruna and MODIS data from a NASA ftp-server. The rolling archive is daily and automatically checked for new relevant EO data updates within the geographic area of operation(s) for the service and subsequently new MERIS data are downloaded at a given time interval (every PM). These standard MERIS Level-2 data are locally stored at NERSC for further processing and generation of the service products. During 2008 the software routines for the processing and hardware storage capacities have been improved for the MarCoast WQ & HAB Service, in order to make the service performance more robust and that the routine processing can be more efficient and with less human interaction in case of irregularities. A modular based solution has been developed, with a structure that makes it re-usable for new geographical areas and for introduction of new products.

The data management and processing procedures for the "Water Quality Monitoring Service for the North Sea" at NERSC is organized along three parallel production lines for the

generation of the different service products (Table 1). Each of these the lines are associated with their EO data source used and algorithms applied (MERIS / MODIS; standard algorithms / '"in-house" algorithms).

The data processing is carried out at two levels (Figure 2) according to generally accepted scheme of remote sensing data processing (presented in Table 2). Level 2 processing includes import of N1 and HDF files, transformation of the images into the selected map projection, calculation of additional products (Kd490) and saving in internal format. Data in internal format is assumed to be of Level 2, sublevel C (L2C data). Programs for L2 processing are sensor specific – for each sensor there is a separate l2c Matlab function.

L3 processing includes: A) binning of acquired satellite images into daily, weekly and monthly composites; B) visualization of the averaged products as JPG and PNG files.

Similar services are implemented @ NERSC for different geographical regions (e.g. Central American Waters) and they are managed jointly at NERSC.

Level 1	TOA radiance measured by a satellite
Level 2	Normalized water leaving radiance (reflectance) calculated after atmospheric correction; geo-physical products (i.e. chlorophyll concentrations)
Level 3	Binned data (1, 7, 30 days averages of satellite images)

Table 2: General scheme of level for remote sensing data processing.

2.1 GENERALL SCRIPT CRON_HAB.BASH

The processing is launched from the most general shell script `cron_HAB.bash` called by a cron-daemon every day at 19:00 LNT. The script `cron_HAB.bash` launches simple Python scripts for downloading from the ESA and NASA On-line Archive and further processing of the satellite data taken for the time being for eight different geographical regions, including the North Sea:

- `get_all.py` (download MERIS, MODIS and AATSR images taken over all regions of interest);
- `make_northsea.py` (process MERIS, MODIS data over the North Sea);
- `make_sam.py` (process MERIS, MODIS, AATSR data over the CAW area);
- etc…

Standard and error output of these scripts is redirected to separate log-files with self explanatory names located in home directory of the sat-user @ NERSC (those monitoring and maintaining the services @ NERSC). Such organization of region-by-region processing makes it easier to identify and debug the code and fix the problems that may arise on a daily basis.

Level	Steps	Data format	Layer 3 scripts
Download	MERIS, MODIS, AATSR on FTP (ESA rolling archive, NASA FTP subscription)	N1 HDF	
	Data download		`get_all.py` `mer_download.py` `aqu_download.py`
	MERIS, MODIS, AATSR on NERSC server	N1 HDF	
L2C	Reading N1; Reprojection; Caluclating additional products; Application of advanced algorithms; Saving;		`l2.py` `l2c_mer.m` `l2c_aqu.m` `boreali.m`
	L2C data (algal_1, algal_2, total_susp, yellow_subs, kd490, sst_aqu, lmchl, lmtsm, lmdoc, lmmse)	MAT	
L3	Binning input data into 1, 7 and 30 day averages		`l3.py` `l3a.m`
	L3A data (averaged producs)	MAT	
	Visualization of averaged products		`l3.py` `l3b.m`
	L3B data visualized maps with concentration distributions	JPG PNG	
Dissemination	WEB-visualization		`index.php`
	Data on the screen and web-map for end users	HTML KML	

Figure 2: General scheme of data processing at NERSC.

2.2 LEVEL 2 PROCESSING

Downloaded satellite images are first converted into L2C format: by three Python and corresponding Matlab function:

```
l2c_mer("NorthSea"); %for MERIS import
l2c_aqu("NorthSea "); %for MODIS/Aqua import
```

These functions read XML configuration files where the following options are listed: names of the products to be read, desired map projection type and parameters; location of N1 and HDF files and L2C data in the file system.

After that each images is saved in Matlab format with internal structure equal for all satellites and sensors. Such files include: names of products, grids with products, grids with values of latitude and longitude. Names are also given to the files according to internal conventions: XXXYYYYMMDD_hhmmss.mat, where XXX – sensor specific prefix ('MER' for MERIS, 'AQU' for MODIS/Aqua, 'ATS' for AATSR, etc), YYYY – year, MM – month, DD – day, hh – hour, mm – minute, ss – second.

Parameters for the listed functions are the key-names of the regions. For processing data over CAW area the parameter should be equal to "SAM". In fact, these functions are region-universal and with other parameters ("NorthSea") are launched for processing data over the North Sea.

Advanced processing is realized in the Matlab-based software package Bio-Optical REtrieval Algorithm (BOREALi). The BOREALi code is based on the advanced algorithm for processing remote sensing data in the visible described in more details below. The software package consist of several Matlab functions and an executable file:

- `borealli(L2C_file_name, config_file_name)` – interface function;
- `setsat` – setting default constants for each of the input satellite products. These constants are different for SeaWIFS, MODIS and MERIS.
- `getini` – reading configuration file;
- `homodel` – reading the file with available hydro-optical models and development of the model specific for the processed satellite data;
- `masknflag` – application of quality assurance techniques, masking and flagging of erroneous input data;
- `cppcore` – launching the executable file for fast processing of remote sensing data with multivariate optimization algorithms;
- `save2mat` – saving of the results into L2C format.
- `lm.exe` – executable file where the Levenberg-Marquardt multivariate optimization technique is realized. Executable file is used instead of a Matlab-function due to higher efficiency.

2.3 LEVEL 3 PROCESSING

The main advantage of the level-by-level processing is that we need only one general L3 Python function for data from any sensor/satellite - `l3.py` . This function launches two Matlab function: `l3a`(<region> <sensor> <input folder> <output folder>) and `l3b`(<region> <sensor> <input folder> <input file>) for averaging of input data and for visualization. The parameters for the functions define the options that should be used while averaging or visualization. These options are saved in a XML configuration file and include: list of products to be averaged, color scheme, legend and scaling limits for visualization.

Processing schemes are illustrated in details respectively in Figure 3 for the standard MERIS processing, in Figure 4 for advanced MERIS processing and in Figure 5 for the MODIS SST processing.

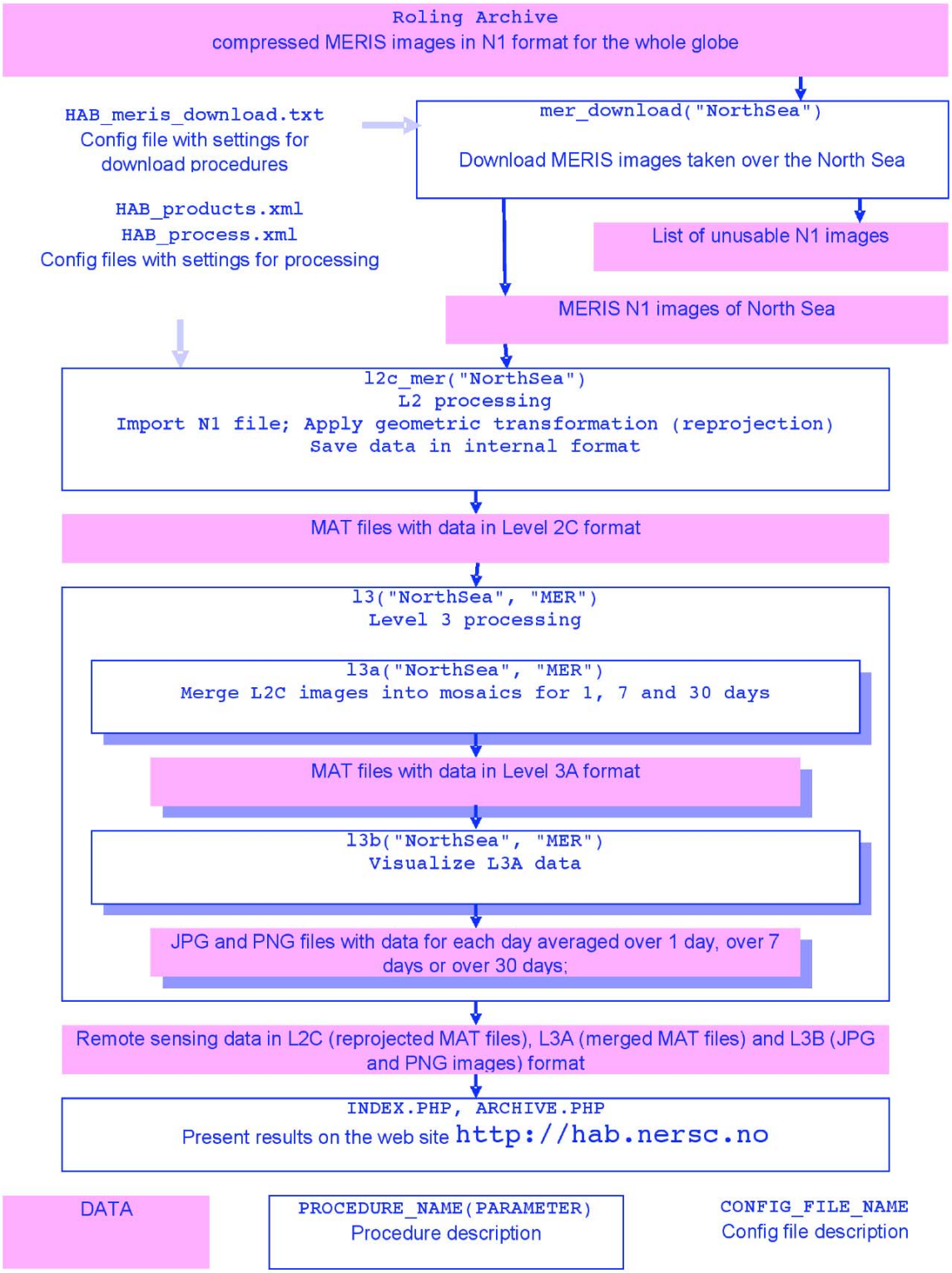

Figure 3: The scheme of the standard MERIS processing.

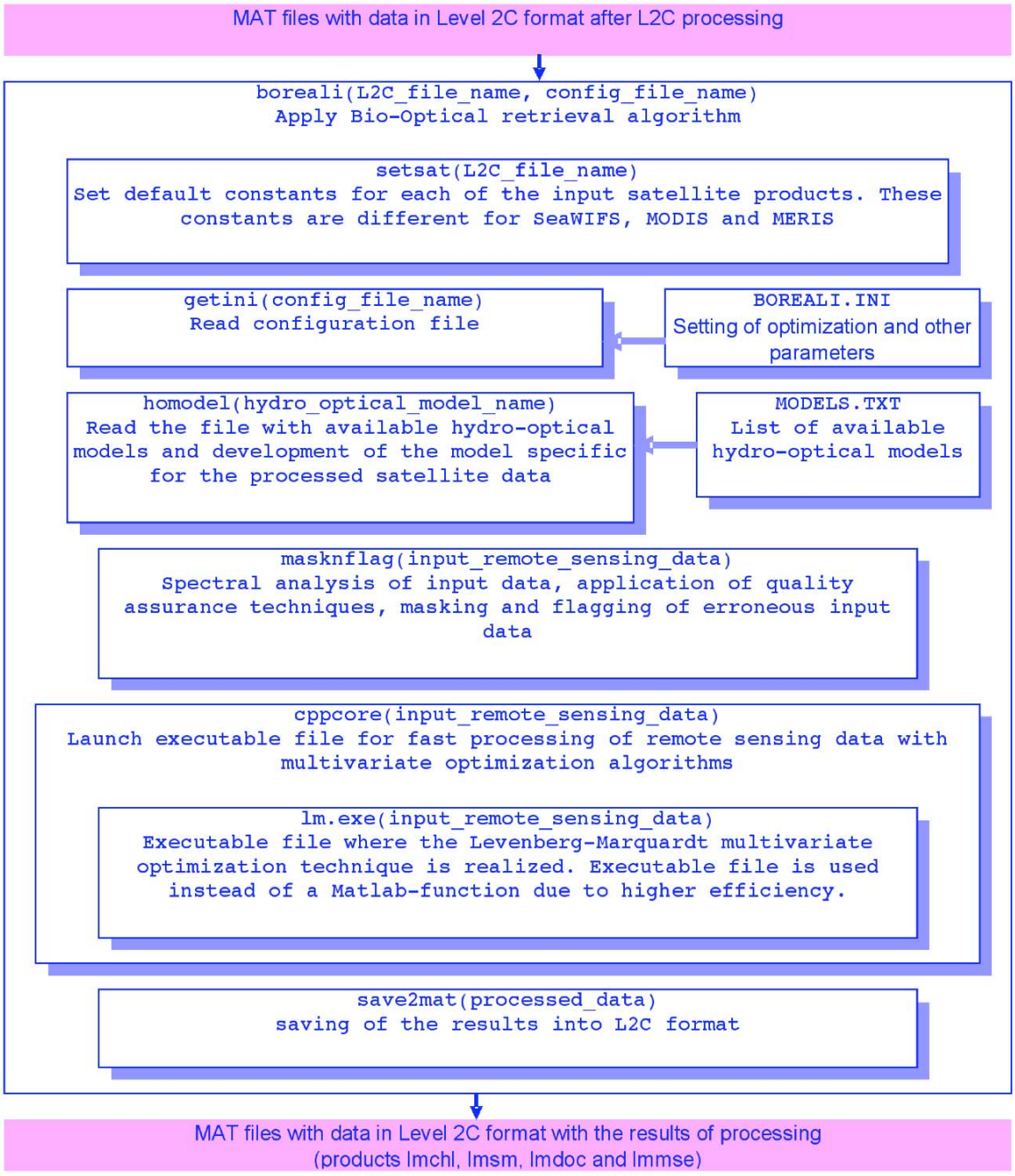

Figure 4: The scheme of the advanced MERIS processing.

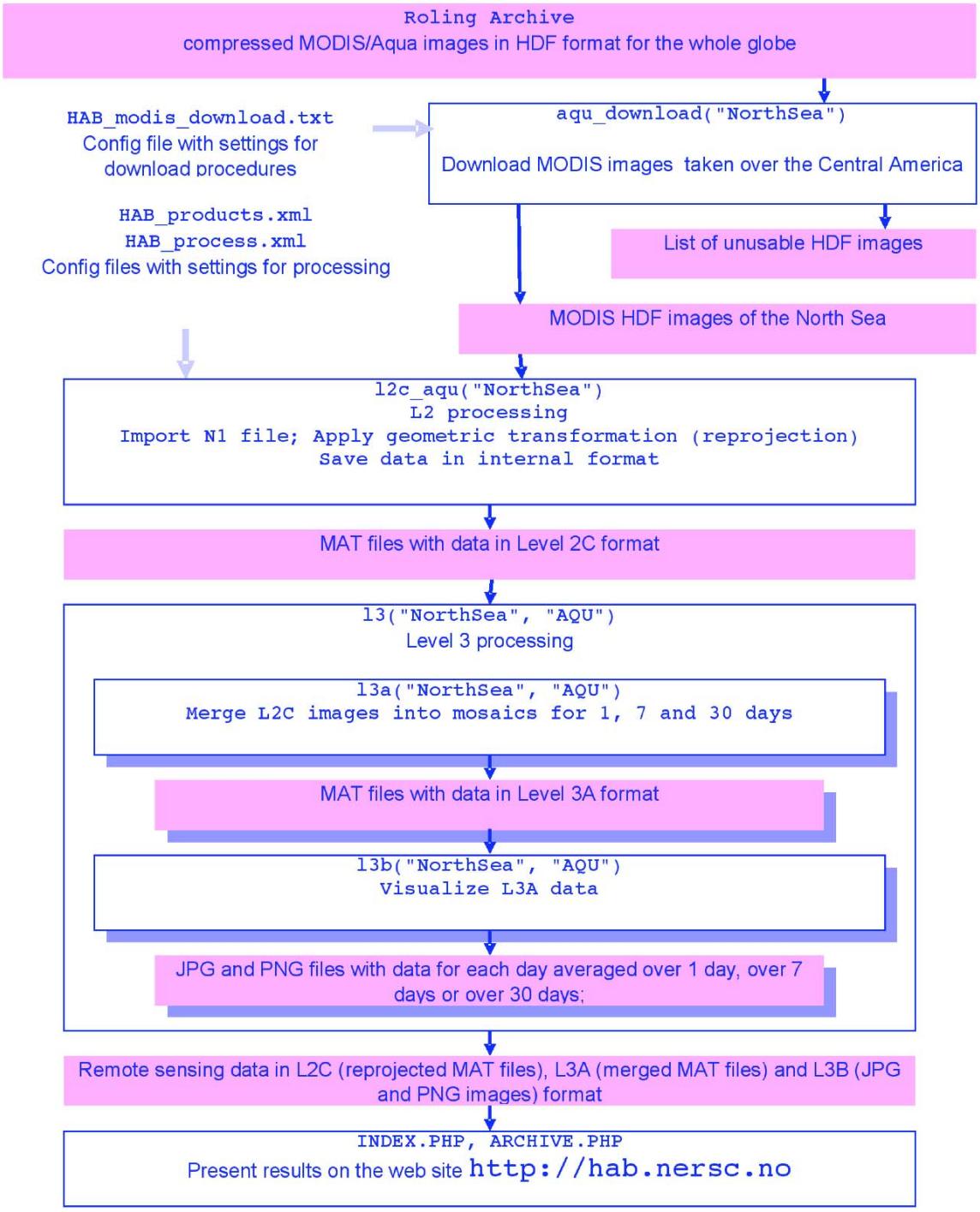

Figure 5: The scheme of the standard MODIS processing @ NERSC.

The standard ESA MERIS chlorophyll-a products for respectively Case-1 and -2 waters (algal_1 and algal_2), total suspended matter (total_susp) and CDOM (yellow_subs) are extracted from each satellite pass covering fully or partly the geographical area monitored by the service.

Additionally an RGB true-color image is routinely generated from the MERIS spectral channels 6, 5 and 2.

A new product Kd490 (diffuse attenuation coefficient at 490 nm, 1/m) was lately introduced to the monitoring system. Kd490 values are calculated as

$$Kd490 = 0.0166 + 0.08349 * algal\_10.63303$$

A regional "in-house" developed algorithm (Pozdnyakov et al, 2005[1]) is used for retrieval of a set of regional specific data products for the above listed bio-geophysical parameters (chlorophyll-a, total suspended matter and CDOM, with prefix "lm" in Table 1). These "regional" products are limited to the initial core area in North Sea and southern Norwegian coastal waters, where the current algorithm has its validity and the global ESA products are more questionable (see Section 3.4).

Sea surface temperature (SST) is generated from MODIS data obtained from an on-line NASA ftp-site. The product retrieval is based on well-defined and -proven processing algorithms and is accordingly not a target for validation. For each area the data are made avialble through a web-based graphical inter face at http://HAB.nersc.no (Figure 6).

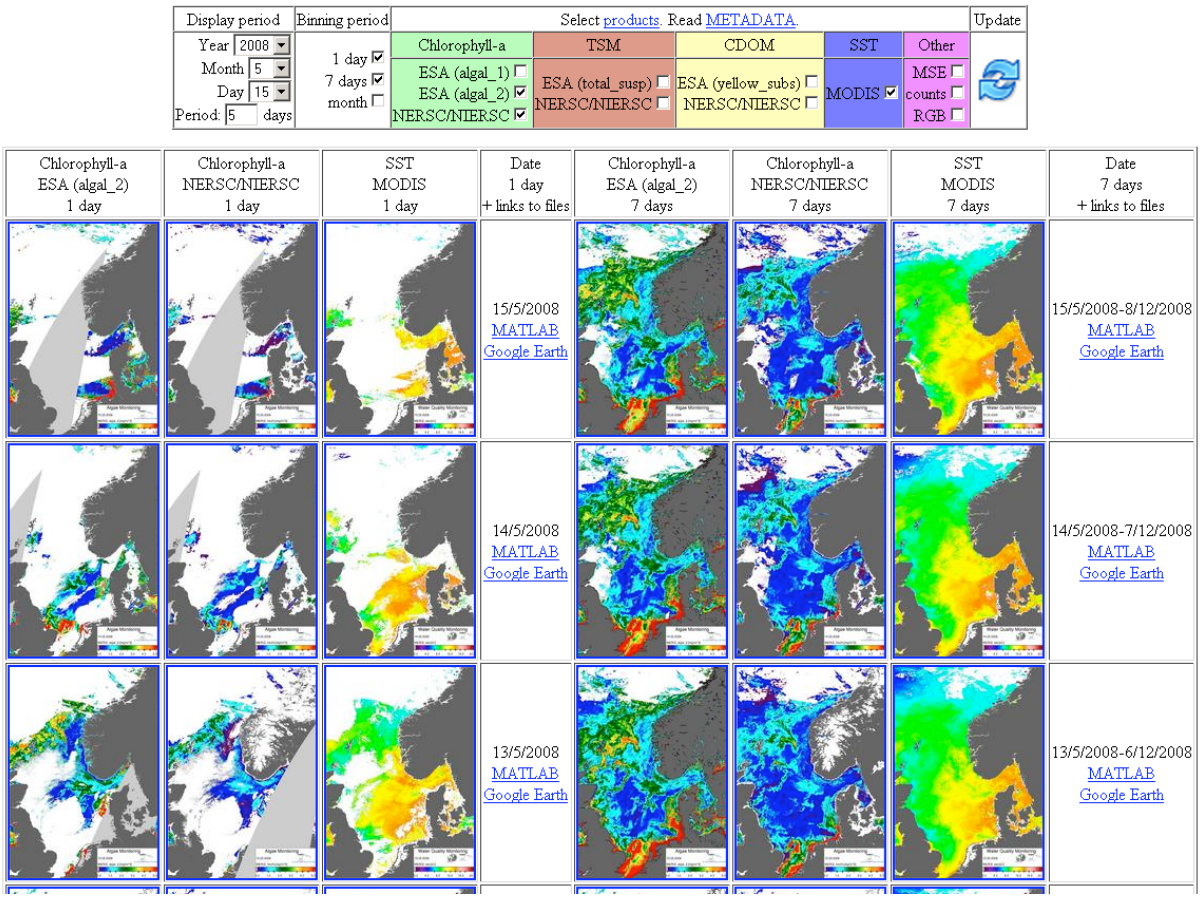

Figure 6: A selection of 1-day and 7-days binned information products for 3 days in May 2008 offered by the MarCoast North Sea service @ http://HAB.nersc.no. Products shown are; the ESA case 2 chlorophyll-a (algal_2), the regional NERSC/NIERSC chl-a (lmchl) and the MODIS sea surface temperature (SST). The image examples are binned for, respectively, 1-day and 7-day periods. The user can interactively select the information products for display and download the images in GeoTIFF, KML for Google or MatLab formats.

[1] Pozdnyakov D.V, A.A. Korosov, H. Grassl and L.H. Pettersson, 2005: An advanced algorithm for operational retrieval of water quality parameters from satellite data in the visible. Int. J. Remote Sensing, Vol 26, no 12, pp. 2669-2687, DOI:10.1080/01431160500044697

3 MarCoast validation results 2007

3.1 REMOTE SENSING DATA

Based on the water leaving reflectance data derived from the MERIS satellite EO sensor the following remote sensing products are provided by the NERSC MarCoast Service and delivered to our SLA users in near real-time at http://HAB.nersc.no (Figure 6):

- ***algal_1***: An algal pigment index for estimation of chlorophyll-a (CHL-a) concentrations from a MERIS image using the standard ESA algorithm for oceanic Case I waters. The algorithm is based on a blue-green band ratio and valid for oceanic waters. Unit: mg/m^3.
- ***algal_2***: An algal retrieval algorithm for estimation of chlorophyll-a (CHL-a) concentrations from a MERIS image using the standard ESA algorithm for "coastal" Case II waters. The algorithm is based on a neural network method and is valid for coastal waters where sediments particles and yellow substances might be present. Unit: mg/m^3.
- ***lmchl***: A regionally tuned algal retrieval algorithm for estimation of chlorophyll concentrations calculated from a MERIS image, using the advanced NERSC/ NIERSC algorithm for optically complex waters with application of a hydro-optical model developed specifically for the North Sea. Unit: mg/m^3.
- ***flh***: A measurement of the algal fluorescence signal in the red spectral region. The algorithm is based on the Chl-a fluorescence signal at 681,25 nm with reference bands at respectively 705 and 665 nm. Unit: sr^{-1}.
- ***total_susp***: An algorithm for estimation of the concentration of total suspended matter (TSM) calculated from a MERIS image using standard ESA algorithm for Case II waters. The algorithm is based on a neural network method and assumes a linear relation between the content of particles and their properties for scattering of light. Unit: g/m^3.
- ***lmsm***: A regionally tuned algal retrieval algorithm for estimation of concentration of total suspended matter calculated from a MERIS image using the advanced NERSC/NIERSC algorithm for optically complex waters with application of a hydro-optical model developed specifically for the North Sea. Unit: g/m^3.
- ***yellow_subs***: A neural network based algorithm similar to the one for the *algal_2* and *total_susp* retrieval. The MERIS retrieved yellow substance is defined as the colored dissolved organic material (CDOM an YS) and the bleached particle absorption. Unit: m^{-1}.
- ***lmdocm***: A regionally tuned algal retrieval algorithm for estimation of concentration of colored dissolved organic matter calculated from a MERIS image using the advanced NERSC/NIERSC algorithm for optically complex waters with application of a hydro-optical model developed specifically for the North Sea. Unit: m^{-1}.

As described in Section 2 remote sensing data are downloaded in Level-2 (L2) from the ESA and NASA ftp-data servers and processed, stored and provided at two levels: L2 and L3. The L2 data products are provided on a one-day image-by-image basis, while L3 data is provided at day-by-day basis (including one or more satellite pass during one day over the area of investigations) as well as weekly (7-days rolling) and monthly-binned data products. A polar orbit satellite, such as ENVISAT, passes several times over the North Sea per day and the MERIS sensor records several images from the region. Each of these images is considered as a separate L2 data product, but images are also binned to create daily, weekly and monthly

mosaics and averages – the L3 products. In the validation study only L2 data are used for comparison with ground truth data.

Each acquired MERIS image is converted into internal format and stored in a separate file with a conventional filename (see Section 2). Afterwards it is accessible for further processing, visualization and comparison with *in situ* data. Storage of MERIS images represents a database of remote sensing data. Special routines were developed for extracting pixel-by-pixel data from satellite images for a required date and location. Several pixel locations, e.g. 3x3 pixels, might be binned in the extraction process and for use in the validation.

3.2 THE INDEPENDENT VALIDATION DATA

The independent validation of the Norwegian HAB and WQ services operated by NERSC and its satellite EO data products is based on utilization of FerryBox data for the Colour Line ferry between Oslo, Norway and Hirtshals in Denmark, provided by NIVA [Sørensen, 2006][2]. The transect between Norway and Denmark is covered twice per 24 hrs. with typically a night-time transect from Oslo to Hirthals and a day-time returning to Oslo (see Figure 7). The MERIS satellite pass is typically around 10-11AM, implying that the vessel is "en route" during the satellite pass and that the typical time difference for the night-time transect may be up to 5-12 hours between the satellite pass and the FerryBox validation data collection in the Skagerrak waters. The measurements acquired with the FerryBox system are done analyzing water from the onboard intake at 3,5 meters depth. The measurements includes near time and continuous records of temperature, salinity, algae fluorescence and particles. Additionally waters samples are collected at given locations and preserved for further analysis in the laboratory, these include HPLC measurements of chlorophyll-a concentrations and TSM. The validation data used in the MarCoast validation has been quality assured by NIVA and used by NERSC for validation vs. the satellite based EO products generated as a part of the MarCoast service portfolio. For details on the sampling, preparations of the analysis and its validity it is referred to Sørensen [2006].

3.2.1 Chlorophyll-a validation

The *in vivo* or *in situ* measured chlorophyll-a fluorescence is strongly coupled to the biochemistry of the phytoplankton, with a (known) significant diurnal as well as seasonal variation. This fact makes chlorophyll-a fluorescence measurements not directly applicable for validation purposes, unless calibrated wrt. the more accurate chlorophyll-a HPLC analysis. NIVA has made a comparison between fluorescence and HPLC chlorophyll-a measurements in 2004 and found them well correlated - about 80% of the variance was explained. Highest confidence should accordingly be given to the validation performed using the HPLC chlorophyll-a measurements although also some in vivo chlorophyll-a fluorescence data are included in the analysis of the MarCoast chlorophyll-a products for 2007. The comparison with the in vivo chlorophyll-a fluorescence has been based on the night-time measurements (i.e. not during the time of the MERIS coverage of the area). Although the time difference wrt. the MERIS satellite overpass is larger for these night-time fluorescence measurements, they seems to be better representing the actual chlorophyll-*a* distribution.

Given that the satellite overpasses the study area at about 11 AM local time, the explanation is thought to be residing in the sinusoidal diurnal cycle of the phytoplankton fluorescence (Figure 8): the fluorescence signal is minimum at the local noon, and its maximum falls on

[2] From FerryBox Results and Reports, Revision 2.0, CD-ROM Publication, April 2006.

the local midnight. The ratio between *chl* and *chl* fluorescence varies on the same diurnal scale, due to the varying photosynthetic state of the phytoplankton cells. Thus the correlation between *chl* concentration measured by the satellite and *chl* fluorescence measured by the Ferrybox system varies accordingly.

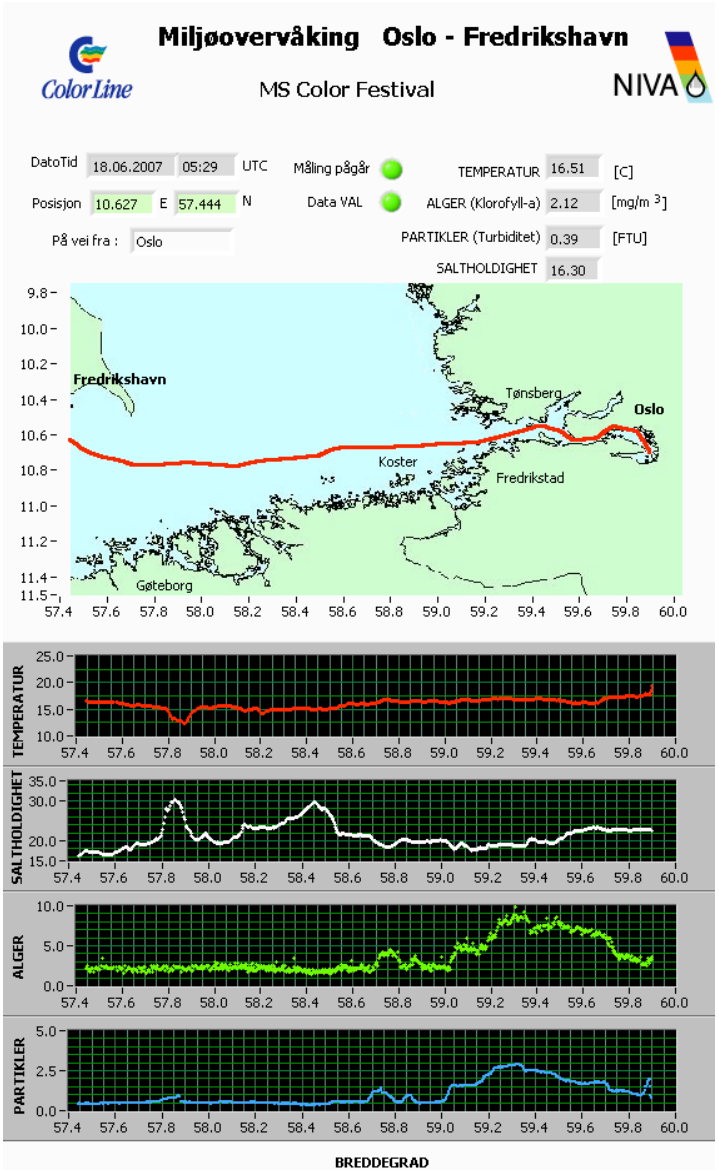

Figure 7: An example of the near real-time out put of the NIVA Ferrybox system between Oslo and Fredikshavn. Source: http://www.niva.no.

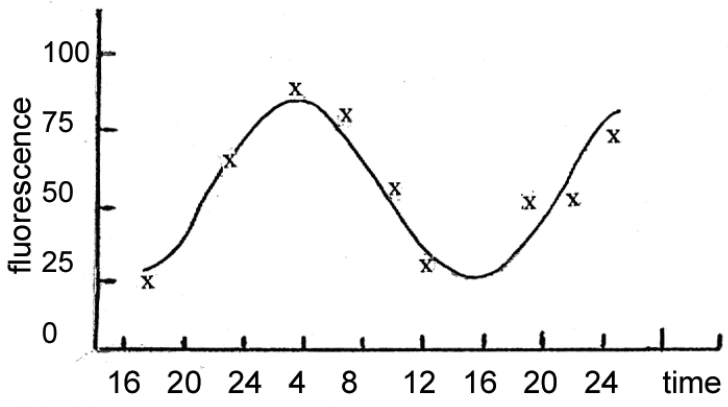

Figure 8: Diurnal cycle of the phytoplankton fluorescence (0 m, relative units) recorded in the Riga Bay, May 1975; t- local time (hrs).

3.2.2 Total Suspended Matter (TSM)

The total suspended matter (TSM) products derived from the MERIS data has also been validated with some measurements from the Ferrybox system. A Polymetron turbidity sensor is used in the Ferrybox system, measuring the scattered light in the red range of the spectrum. In the EU Ferrybox project comparison of these types of observations in 2003 vs. laboratory samples of turbidity have been performed, resolving a relation with about 74% explained variance [Sørensen, 2006]. The study concludes with an absolute turbidity accuracy of ± 0.4 FTU on a yearly basis. These data was based on the Seapoint sensor while in 2007 the data are improved using a self-cleaning sensor (Polymetron). Further it is assumed that the relation between turbidity and TSM is constant in the region, which has been justified for measurements in the Skagerrak and North Sea waters.

3.3 COMPARISON OF SATELLITE AND GROUND TRUTH DATA

For validation of remote sensing NRT products (algal_1, algal_2 and lmchl) and of Ferrybox NRT product (fluorescence) the water samples and HPLC based laboratory analysis of phytoplankton chlorophyll-a concentrations were used. Samples of water are taken by a Ferrybox system, installed onboard of a ferry cruising between Norway and Denmark, are collected automatically or on request at selected positions along each transect. The water samples are stored onboard, collected and processed later at the NIVA laboratory with the HPLC method for evaluating amount of chlorophyll in phytoplankton cells.

For validation of remote sensing NRT products (total_susp and lmsm) turbidity measurements taken by a Ferrybox system were used. The turbidity is measured in the flow of seawater each minute by an optical sensor installed onboard the ferry. Turbidity values are converted into concentrations of total suspended matter using a linear regression algorithm developed at NIVA [see e.g. Sørensen, 2006].

These independent validation data were provided in a database with an entry for each measurement and with the following information: date, time, latitude, longitude, HPLC, fluorescence, turbidity, TSM.

The validation was carried out in two experiment and gave rather interesting results important both for assessment of the remote sensing data accuracy and better understanding of the validation technique itself. The difference in the analysis was simply related to the allowed time span between the satellite and the *in situ* observations, which initially was allowed to be 24 hrs. and for the results presented in this report limited to 2 hours – in accordance with satellite EO data validation protocols. This time constraint reduced the number of available high quality match-up data from 57 to 20 observations for entire 2007, limiting both the seasonal period of observations, the range of available concentrations as well as the number of observations along each transect.

In the assessments matching remote sensing and *in situ* HPLC measurements were selected in the following order:

- For each in situ HPLC point a satellite image acquired not earlier/later than 2 hours was selected (this is longer than the general constraints used in validation protocols);
- The closest pixel was selected from the MERIS image and values of ***algal_1***, **algal_2**, **lmchl** were extracted and stored in the database of matchups together with value of HPLC chlorophyll;
- The database of matchups was analyzed and matchups fulfilling the following criteria were excluded:
    - Pixel is closer that 5 km to the border of a cloud (due to higher uncertainties atmospheric correction procedure);

- Image is acquired earlier than February or later than November (due to the Sun zenith angle);
- Pixel is close to land (due to edge effect, less quality of the atmospheric correction);
- Pixel is close to the Kattegat waters (different hydro-optical model);

- Scatter plots were created for the remaining points and statistical regression parameters were calculated.

Matching remote sensing and in situ TSM measurements were selected in the same order.

Due to the large diurnal, seasonal and state variations in the in vivo fluorescence measurements of Chlorophyll-a a direct comparison of fluorescence and remote sensing products was not carried out. Instead transect plots were created for comparison of Ferrybox fluorescence, HPLC and remote sensing chlorophyll concentrations.

3.4 VALIDATION RESULTS

3.4.1 Chlorophyll-a products – HPLC samples

Totally 20 individual HPLC Chlorophyll-a samples have been available for the MarCoast validation study, with a majority of the measurements in the range of low concentrations (Chlorophyll-a <2 μg/l, see Figure 9). These HPLC samples where obtained from 9 transects across Skagerrak (one individual stations per transect) during the months of March, June and August 2007. The statistical comparison of Chlorophyll-a values measured *in situ* with ***algal_1***, ***algal_2*** and ***lmchl*** products (respectively in Figure 10-12) show in general good agreements. The statistical correlation between Chlorophyll-a concentrations (summarized in Table 3) indicates that ***lmchl*** product has a higher level of correlation with the *in situ* measurements, followed by respectively ***algal_2*** and ***algal_1***, as expected since the ***algal_2*** products has been developed for the optical conditions of the North Sea and Skagerrak regions. This indicates that use of our regionally tuned and developed algorithm, developed for specifically the Skagerrak waters in this case, can improve the capability and accuracy of retrieval of the Chlorophyll-*a* distribution, compared to use of the standard and global products provided by the space agencies delivering satellite EO data.

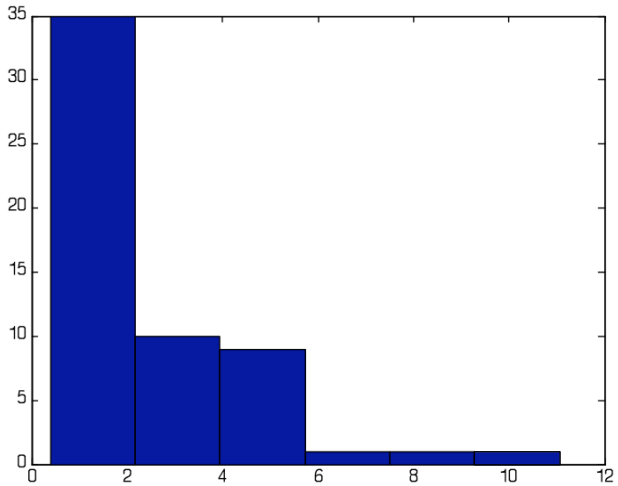

Figure 9: Frequency distribution of HPLC chlorophyll values [µg/l] measured in situ available for this validation study. 20 of these 57 values meet the requirements of being within 2 hours of the MERIS satellite overpass and are used for the high quality validation of the satellite cased chl-a products.

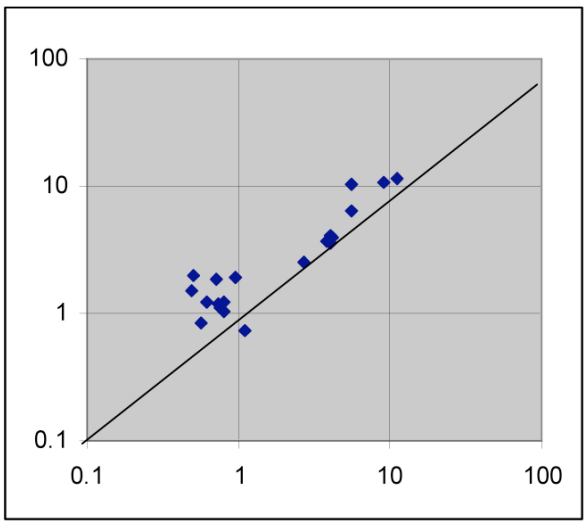

*Figure 10: Comparison of **algal_1** product (Y-axis, µg/l) and HPLC chlorophyll. y = 1.06x, R^2 = 0.90, p = 95%, RMSE = 1.24.*

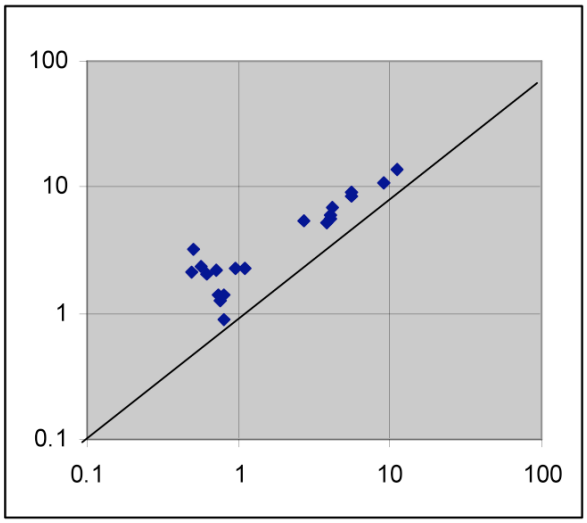

*Figure 11: Comparison of **algal_2** product (Y-axis, µg/l) and HPLC chlorophyll. y = 1.16x, R^2 = 0.95, p = 95%, RMSE = 1.96*

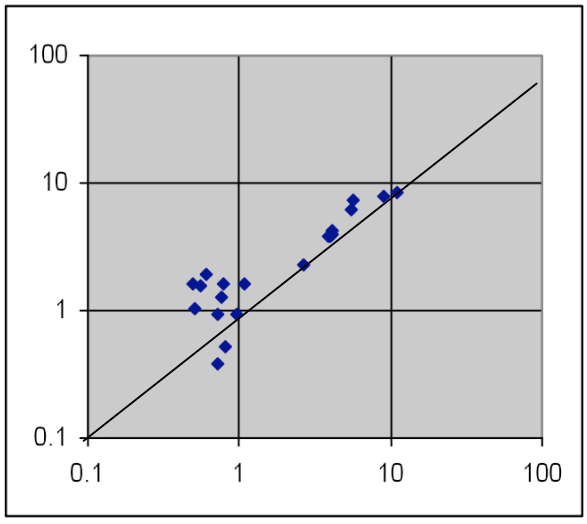

*Figure 12: Comparison of **lmchl** product (Y-axis, µg/l) and HPLC chlorophyll. y = 0.81x, R^2 = 0.92, p = 95%, RMSE = 0.94.*

Product/ Parameter	Samples (n)	Explained variance (R^2)	p	RMSE	slope (a)[3]	intercept (b)*	sat-min	sat-max	In situ-min	In situ-max
algal_1	20	0.90	95%	1.24	1.06	0.47	0.72	11.4	0.49	11.1
algal_2	20	0.95	95%	1.96	1.16	1.28	0.90	13.6	0.49	11.1
lmchl	20	0.92	95%	0.94	0.81	0.74	0.38	8.5	0.49	11.1
FLH	20	0.67	95%	3.72E-08	1.00E-04	0.79	3.18E-05	5.73E-04	0.49	11.1
Night fluorescence	16	0.83	95%	2.7						
Day fluorescence	24	0.38	95%	8.85						
total_susp	20	0.84	95%	0.35	1.28	-0.31	0.47	3.19	0.73	2.49
lmsm	20	0.46	95%	0.47	0.63	0.64	0.67	2.31	0.73	2.49

Table 3: Summary of validation statistics, comparing the MERIS and Ferrybox based MarCoast service products, following approved EO data validation protocols. The lower parts of the table are discussed in the following Sections.

3.4.2 Chlorophyll-a products – Fluorescence ferry box data

Along the Ferrybox transect fluorescence (Figure 13) and HPLC Chlorophyll-a concentrations were measured and remote sensing Chlorophyll-a measurements were extracted for the *in situ* station locations. As expected, comparison of *in situ* Chlorophyll-a values measured with HPLC and fluorescence methods shows that the night-time (Figure 14) fluorescence product has higher quality than day-time fluorescence measurements (Figure 15).

An example of such transects data are presented for 25th March 2007 for comparison of respectively ***algal_1*** and ***algal_2*** products with night-time fluorescence transects (Figure 16-17) shows that both standard ESA algorithms are generally overestimating Chlorophyll-a concentrations. Along the same transect a much better coincidence of Chlorophyll-a values from ***lmchl*** products both with HPLC values and night-time fluorescence transect is observed at Figure 18. Comparison of the same remote sensing data products and HPLC products with day-time fluorescence values shows significantly higher errors (Figure 19, shown as an example). Similar diurnal variations have been observed in other data available in MarCoast and are known from the scientific literature (see Section 3.2.1).

In conclusion the day-time fluorometric measurements can not be used for validation of satellite optical EO products, however using the night time measurements, which will be more correct with respect to resolve the Chlorophyll-a concentrations, these will be too fare off in time to be useful for comparisons of the satellite data due to the hydrodynamic and algae community changes over the time difference between the night and the AM to mid-day satellite overpass.

[3] y=ax+b, where y=satellite and x= in situ

Figure 13: Location of the Ferrybox transect on 25.03.2007. Purple line indicates the entire transect. Red line indicates the part of the ship transect which was used for comparison with remote sensing data.

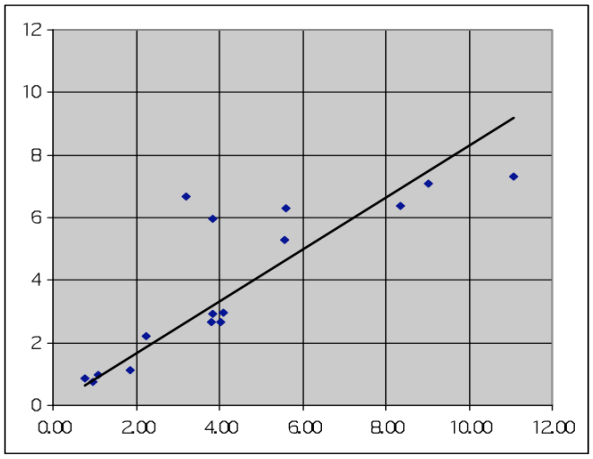

*Figure 14: Comparison of **in situ HPLC chlorophyll** (X-axis, µg/l) and **Ferrybox night fluorescence** product (Y-axis, µg/l). y = 0.83x, R^2 = 0.83, p = 95%, RMSE = 2.7.*

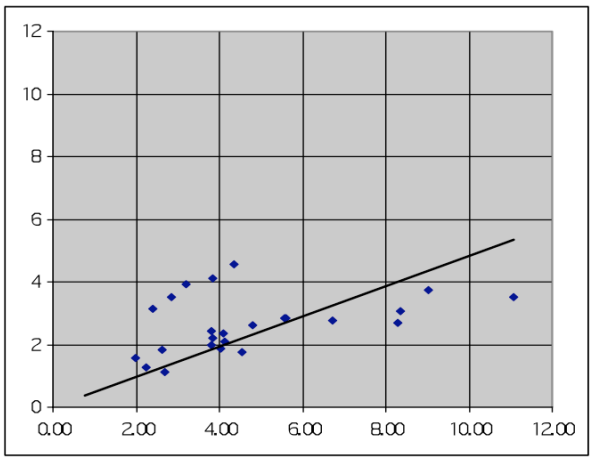

*Figure 15: Comparison of **in situ HPLC chlorophyll** (X-axis, µg/l) and **Ferrybox day fluorescence** product (Y-axis, mg/l). y = 0.48x, R^2 = 0.46, p = 95%, RMSE = 8.85.*

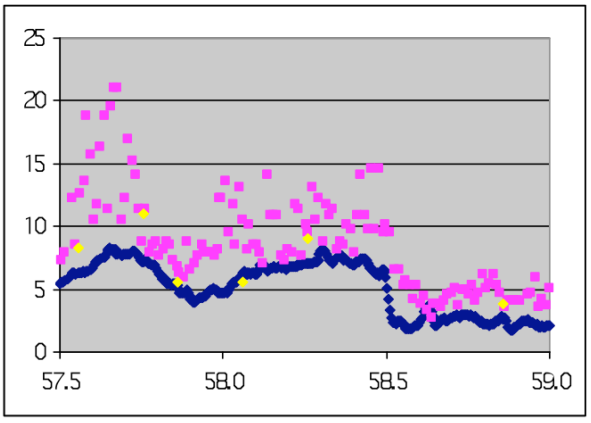

*Figure 16: Comparison of Ferrybox **night fluorescence chlorophyll** products (blue dots) with **algal_1** product (pink dots) and with **HPLC chlorophyll** measurements (yellow dots) taken along the transect.*

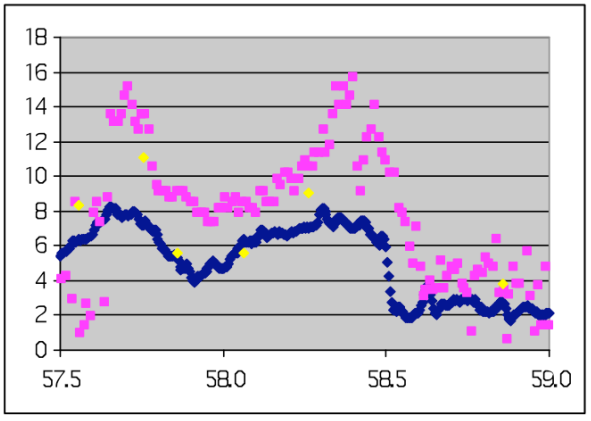

*Figure 17: Comparison of Ferrybox **night fluorescence chlorophyll** products (blue dots) with **algal_2** product (pink dots) and with **HPLC chlorophyll** measurements (**yellow dots**) taken along the transect.*

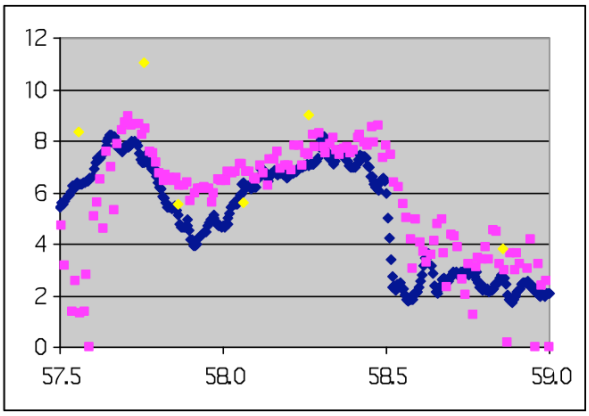

*Figure 18: Comparison of Ferrybox **night fluorescence chlorophyll** products (blue dots) with **lmchl** product (pink dots) and with **HPLC chlorophyll** measurements (**yellow dots**) taken along the transect.*

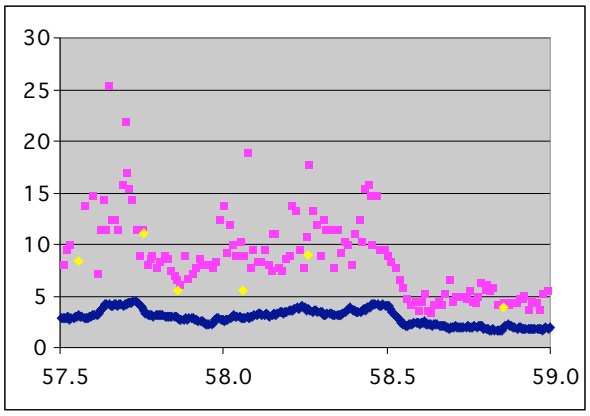

*Figure 19: Comparison of Ferrybox **day fluorescence chlorophyll** products (blue dots) with **algal_1** product (pink dots) and with **HPLC chlorophyll** measurements (yellow dots) taken along the ship transect. Similar results are obtained for the other chlorophyll-a parameters and hence not shown.*

3.4.3 Total Suspended Matter

Totally 20 Ferrybox *in situ* measurements of the water turbidity have been used in the validation of the satellite EO TSM data products. Extracted Ferrybox turbidity (sensor) measurements were taken at the same time and position as water sampling used for HPLC analysis. The rationale for doing this is that high quality of satellite data is already secured through the chlorophyll analysis.

Sørensen [2006] have found a constant relation for Skagerrak waters in the range 0-15 mg/m^3 between the laboratory measured turbidity and the Total Suspended Matter (TSM) as measured in situ. NIVA has used this relation to convert the turbidity measurements to TSM for comparison with the MERIS derived TSM measurements. The turbidity was converted to TSM by the equation; TSM=0.9553*turb + 0.1402, [ref Ferrybox project report].

The comparison of **TSM** values estimated from Ferrybox turbidity measurements and MERIS products shows acceptable to quite poor agreements with 84 and 46 % explained variance for respectively the **total_susp** and **lmsm** (Table 3 and Figure 20-21). This cannot be attributed only to a poor quality of remote sensing data. The linear regression approach which is used for calculating TSM values from turbidity does not take into account other factors influencing turbidity (e.g. absorption by algae cells and suspended minerals) and may give errors of the same order of magnitude. The Ferrybox data can be used for validation of satellite TSM products. However, the conversion factors between *in situ* turbidity and in situ TSM should

be determined and tuned for the area of interest. This has already been done for Skagerrak in the Ferrybox project report and is the basis for this study.

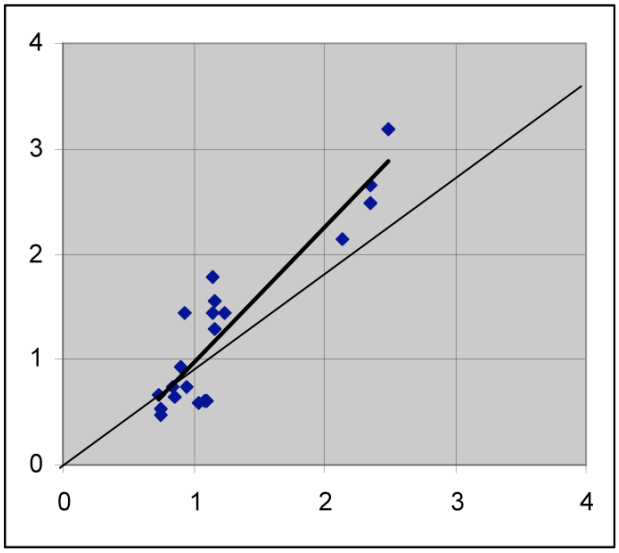

*Figure 20: Comparison of **total_susp** (x-axis, mg/l) and **TSM** (calculated from ferrybox turbidity). y = 1.28x -0.31, R^2 = 0.84, p = 95%, RMSE = 0.35.*

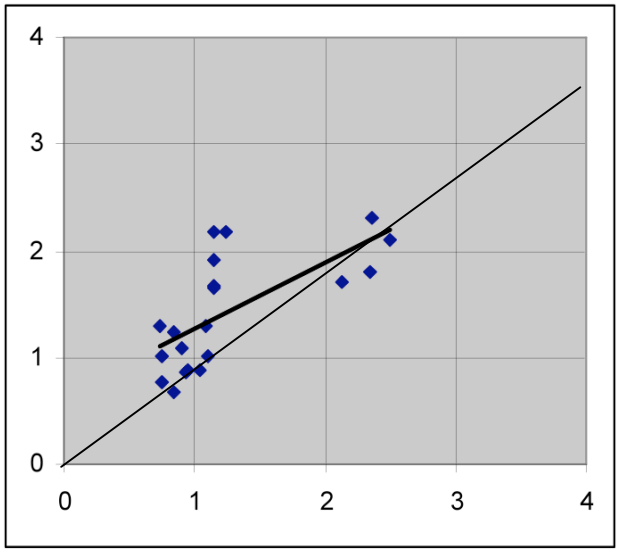

*Figure 21 Comparison of **lmtsm** (x-axis, mg/l) and **TSM** (calculated from ferrybox turbidity). y = 0.63x + 0.64, R^2 = 0.46, p = 95%, RMSE = 0.47.*

3.4.4 The Fluorescence Line Height product

Fluorescence line height (FLH) method is applied for estimating magnitude of chlorophyll fluorescence in red part spectra, where the phytoplankton chlorophyll fluorescence at 681 nm is measured and 667 nm and 709 nm are the two shouldering bands. This is used as an indicator for the biological activity of the phytoplankton. By constructing a baseline using bands on either side of the fluorescence band, we can estimate the deviation from the amount of radiance expected for pure water that results from chlorophyll fluorescence. This increase in radiance (centered at 683 nm for chlorophyll) has been noted for decades in measurements of the light field in the ocean.

20 individual HPLC Chlorophyll-a samples taken in May, Aug., Oct. and Nov. are used to validate FLH product (Figure 22). However, its correlation coefficient is r^2 = 0.67, the performance is particularly poor in the low range Chlorophyll-a concentrations (Table 3). Fluorescence quantum yield may be a reason for that, which varied as a function of environment and incident irradiance, and it may vary significantly (up to one magnitude) both spatial and temporally. It is need caution to interpret the FLH product.

Two transects data are presented for on 25th (night- Figure 23 and day-time Figure 24) and 17th March (night- Figure 26 and day-time Figure 27) 2007 for comparison between FerryBox fluorescence chlorophyll products and the FLH values extracted from MERIS data, which is an average value with 3*3 box. The results show good agreement on 25th March 2007, and the correlation coefficient value of day-time fluorescence product is higher, which time difference between satellite and in situ measurements are smaller. But regression equations are different. For the other days, the agreement is not as good as expected. But the trends of satellite and in situ measurements are similar. In case 2 water, the FLH algorithm is complicated by the overlap of fluorescence and elastic reflectance, especially the effects of high concentration of mineral. The high values circled in Figure 27 are with high TSM values, which are derived from the satellite data.

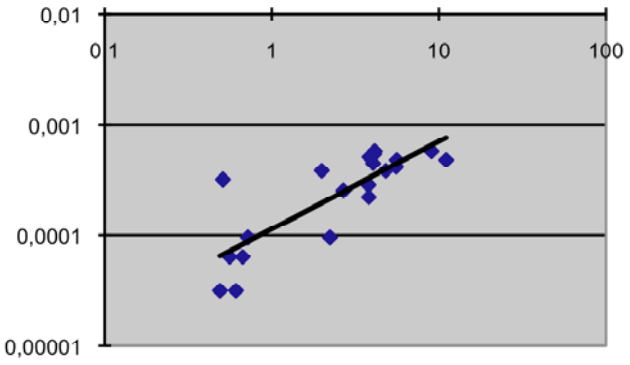

Figure 22: Comparison of in situ HPLC chlorophyll (X-axis, µg/l) and FLH product (Y-axis, 1/sr) in logarithmic axis.). $y = 0.0001*x^{0.7902}$, $R^2 = 0.67$, p = 95%, RMSE = 3.72E-04.

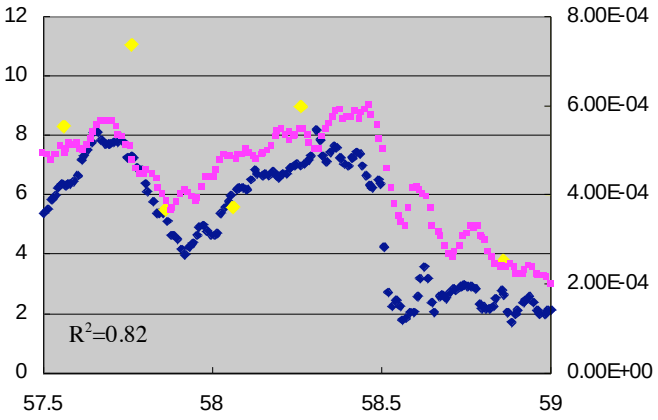

Figure 23: Comparison of Ferrybox night-time fluorescence chlorophyll products (blue dots) with FLH product (pink dots) and with HPLC chlorophyll measurements (yellow dots) taken along the transect acquired on March 25, 2007.

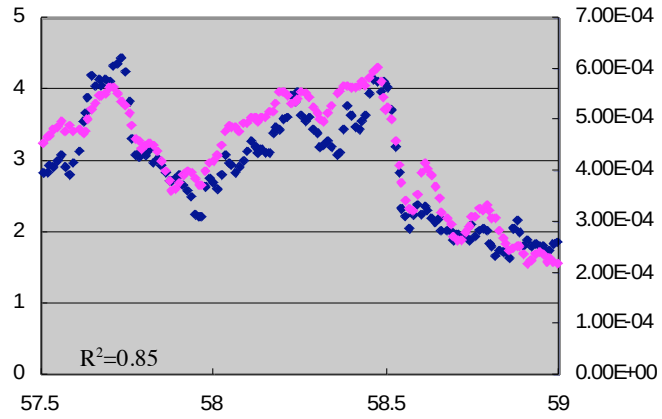

Figure 24: Comparison of Ferrybox day-time fluorescence chlorophyll products (blue dots) with FLH product (pink dots) taken along the transect acquired on March 25, 2007.

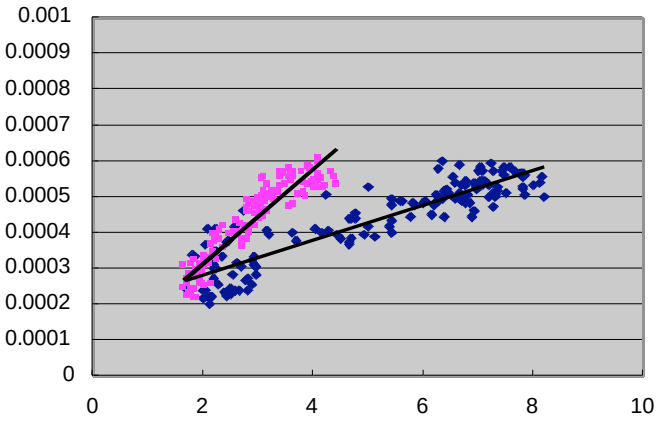

Figure 25: Comparison of Ferrybox night- (blue dots) and day-time (pink dots) fluorescence chlorophyll products (X-axis, 1/sr) and FLH values (Y-axis, µg/l) for the data acquired on March 25.

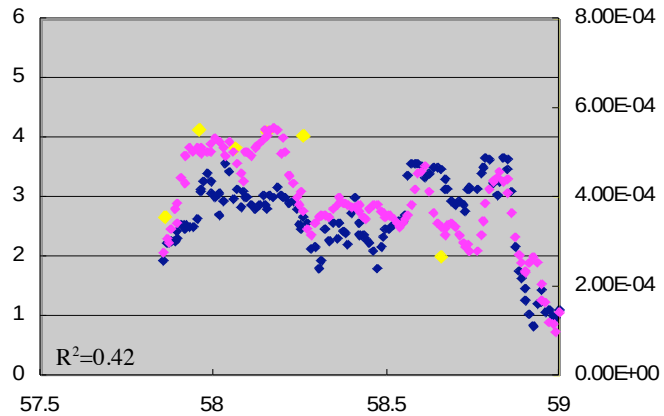

Figure 26: Comparison of Ferrybox night-time fluorescence chlorophyll products (blue dots) with FLH product (pink dots) taken along the transect acquired on March 17, 2007.

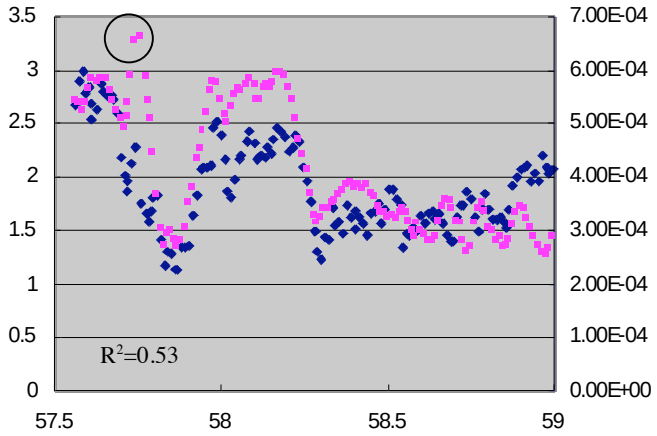

Figure 27: Comparison of Ferrybox day-time fluorescence chlorophyll products (blue dots) with FLH product (pink dots) taken along the transect acquired on March 17, 2007.

3.5 ASSESSMENT OF THE VALIDATION RESULTS

Validation of remote sensing data processed by standard and a regionally tuned algorithms showed that the products delivered to the end users are of high quality and can be used for daily monitoring of both the phytoplankton and sediment concentrations of the Skagerrak area.

At the same time it was shown that following MERIS validation protocols is highly recommended for achieving more adequate validation results of the analysis. According to these protocols one have to select only those *in situ* measurements taken within ±2 hours interval of the satellite overpass. The consequence is a dramatic decrease of the number of *in situ* observations that are valid and of high quality for use in the validation analysis. Typically only 20 such high quality match-up observations were available from the NIVA Ferrybox system and the MERIS data during 2007, despite the field data system in "continuous" daily operations the overall annual numbers of high quality match-up observations are very limited. Accordingly we need much more *in situ* measurements in the future for thorough validation of the system performance, including inter annual sampling.

4 Suggestions for improvement of service and products

4.1 APPLICATION OF ADVANCED ATMOSPHERIC CORRECTION

A new code has been developed at GKSS for atmospheric correction (AC) of MERIS data. The code is embedded into the ESA Beam/Visat software and can be launched as a command line tool. The code has been evaluated and used in several projects over different water bodies - Lake Ladoga, the Kara Sea, Bay of Biscay. This ne AC code performs significantly better in several respects. First, no negative spectral reflectance values are generated for the water leaving radiance. Second, remote sensing reflectance spectra calculated in different definitely Case-II waters are in very good accordance with theoretical assumptions. Experiments where we processed the MERIS images after the GKSS AC with the BOREALi code showed that RMSE (root mean square error between measured and retrieved spectra) is very low even for highly turbid waters of Ob and Yenisey rivers. That

also gave us the opportunity to train a neural network for retrieval of the water quality parameters (chl-a, TSM and DOC) from MERIS data using BOREALi.

We suggest implementing additional line of MERIS data processing in order to improve the atmospheric correction of the spectral measurements. This line will include download of L1 MERIS data from Rolling archive; processing with the Beam/Visat based GKSS atmospheric correction and, finally, processing with artificial neural networks.

The major improvement resides in the higher quality of the retrieved water quality products. Using the developed neural network implies that the computing time will be decreased dramatically. However, the size of the downloaded data files will be approximately doubled, which requires changes to the operational service production line.

Obviously such approach will also require validation and comparison with in situ measurements of *chl* and *tsm*. In principle, the shipborne data from the MarCoast 2007 project may be re-utilized and eventually supplemented with additional Ferrybox data from 2008.

4.2 CALCULATION OF CHLOROPHYLL-A CONCENTRATIONS FROM THE FLH PRODUCT

So far the FLH product is equal only to the difference between the base-line and 7th MERIS band and is measured in sr^{-1}. It is known that FLH is related to the phytoplankton fluorescence and, eventually, to phytoplankton chlorophyll and biomass. However, all algorithms for calculating chl from FLH require regional tuning. Data on HPLC *chl*, fluorescence chlorophyll and FLH product can be used for tune an algorithm for retrieval of chl concentration from the FLH product. This algorithm can be introduced into the processing chain for the North Sea and for other areas.

5 Conclusions

During 2008 the software routines for the processing and hardware storage capacities have been improved for the MarCoast WQ & HAB Service, (http://HAB.nersc.no) in order to make the service performance more robust and that the routine processing can be more efficient and with less human interaction in case of irregularities. A level-by-level processing solution has been developed, with a structure that makes it re-usable for new geographical areas and for introduction of new products and processing methods.

Validation for satellite EO data products with *in situ* observations is a challenge and specific validation protocols have been developed in cooperation between the oceanographic research community and the space agencies for dedicated validation studies of various satellite based EO products. The unique spatial and temporal coverage of the satellite data, integrating the surface signal over e.g. a one or more km^2, and the *in situ* sampling providing detailed, but some times instrument and method specific, accurate measurement of one or more parameter based on sampling of often a “few litres” of water at the surface or at some specific depth. The different nature of these types of environmental observations implies that a one-to-one comparison cannot be expected in such validation investigations. Also the natural spatial and temporal variability of the oceans, and in particular in the coastal waters, is within these spatial and temporal scales of variation and similar in magnitude as expected to be found between satellite and *in situ* observations. Following accepted validation protocols we have limited the temporal difference of observations in situ and from satellites to 2 hours. The fact that the observed parameters often are not exactly the same and needs to be converted to the same units also introduces errors in the validation exercise. Additionally also *in situ* measurements of the “same parameter” is pending on the measurement protocol or instrumentation used, which may cause additional errors. However, under these conditions it is still viable to do a validation comparison between satellite and *in situ* measurements of water quality and algae bloom events as part of the MarCoast service portfolio.

Two of three service product categories have been validated with independent field data. The Ferrybox data available for this MarCoast validation have provided independent measurements of both Chlorophyll-a (using two methods – *in vivo* fluorescence measurements and water samples for HPLC laboratory analysis) and of measurements of the turbidity, which has been converted to total suspended matter (TSM) for the central Skagerrak waters. Unfortunately no data for validation of the yellow substance has been available.

The three different Chlorophyll-a products based on the MERIS RR data have been validated against HPLC measurements of the phytoplankton content of water samples. All the satellite ***algal_1***, ***algal_2*** and ***lmchl*** products compares well with the HPLC measurements and the "regionally tuned" algorithms performs slightly, but significant, better (r^2=0.95 for ***algal_2*** and 0.92 for ***lmchl***) than the "global" ***algal_1*** algorithm (r^2=0.90).

The comparison of total suspended material (TSM) values estimated from Ferrybox turbidity measurements and MERIS products show a less accurate agreement with only 84 and 46 % explained variance for respectively the **total_susp** and **lmsm**. Most likely this is due to the linear regression approach which is used for calculating TSM values from turbidity measurements does not take into account other factors influencing the turbidity (e.g. absorption by algae cells and suspended minerals) and may give errors of the same order of magnitude. The Ferrybox data can be used for validation of satellite TSM products. However, the conversion factors between *in situ* turbidity and in situ TSM should be determined and tuned for the area of interest.

The validation efforts could have been significantly more extensive in order to obtain a more extensive validation of the EO data products, however as long as the service is fit for the users needs and expectations, including being aware of their inaccuracies, the validation activities are sufficient for the service assessment.

For monitoring of the Norwegian coastal waters it is essential to have information from the up-stream part of the general ocean/coastal circulation system, staring in the southern North sea and evolving up along the south an west coast of Norway. Used in combination with the regular field observations along the Norwegian coast this MarCoast service is complementary and sufficient and delivering information products of both type and quality in accordance with the users needs.

For improvement of the service data products a parallel line for advanced processing of MERIS data will include improved atmospheric correction and artificial neural networks. The improvement is thought to increase accuracy of the retrieved products – firstly the derived water leaving spectral radiances and secondly the derived water quality parameter products. The use of the neural network processing approach will also decrease the computing time.

Another product, fluorescence line height chlorophyll (flh_chl), will be added to the list of products. It will contain values of chlorophyll calculated from values of FLH products. The algorithms should be developed and tuned on the basis of *in situ* and corresponding remote sensing data of chl and FLH, using data from 2007 and possible also 2008.