| | |
|---|---|
| **Document title** | Proposal for including EO and SOOP data to the HELCOM eutrophication assessment database |
| **Code** | 3-2 |
| **Category** | DEC |
| **Agenda Item** | 3 – Setting up assessment data flow |
| **Submission date** | 29.8.2014 |
| **Submitted by** | Finland, with input from the EU-RSC data -project |
| **Reference** | EUTRO-OPER 1-2014 |

## Background

One of the conclusions of 'Eutrophication status of the Baltic Sea (BSEP 143)' was, that the confidence of chlorophyll-*a* and Secchi depth indicators would be substantially increased by including also remotely sensed and automated observations to the update of these indicators. EUTRO-OPER 1-2014 agreed to investigate further the possibilities of updating the chlorophyll-*a* indicator using also these data-types, and including them into the assessment data flow. Finland was welcomed to take lead of the work.

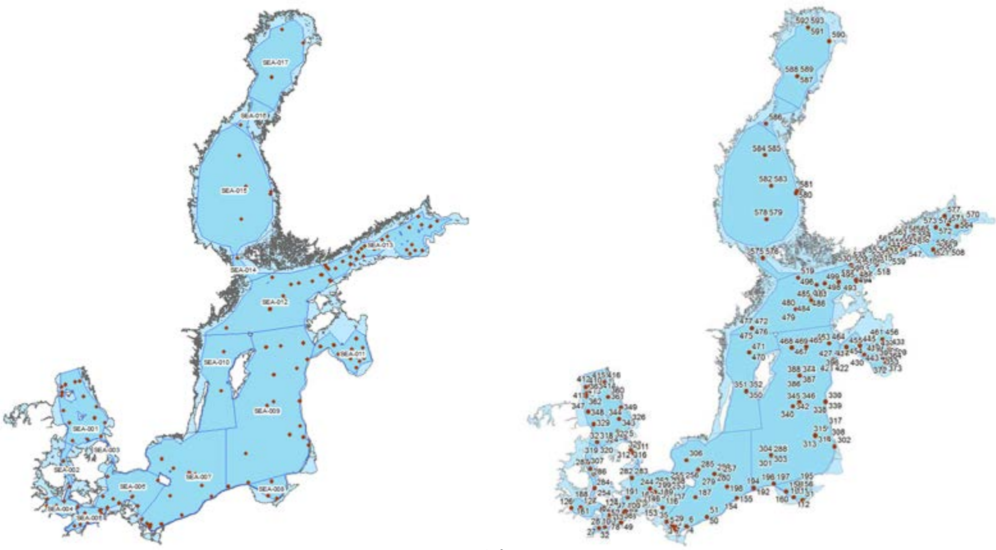
## Action required

The Meeting is invited to take note of the results and proposals, and agree on next steps, with the aim of including EO- and Ferrybox-data into the assessment data flow.

This report provides considerations for inclusion of EO (Earth Observation, i.e. remote sensing) and ship-of-opportunity (SOOP) data to the HELCOM Eutrophication assessment database. The inclusion of EO and Alg@line flow-through data can complement the monitoring station measurements in the assessment of the state of the Baltic Sea, particularly in areas and seasons out of reach of traditional methods. The results part of this report gives examples of EO data comparisons against monitoring station data for a test period of years 2007-2011. Appendix A provides a more detailed description of calculations and methods of comparison.

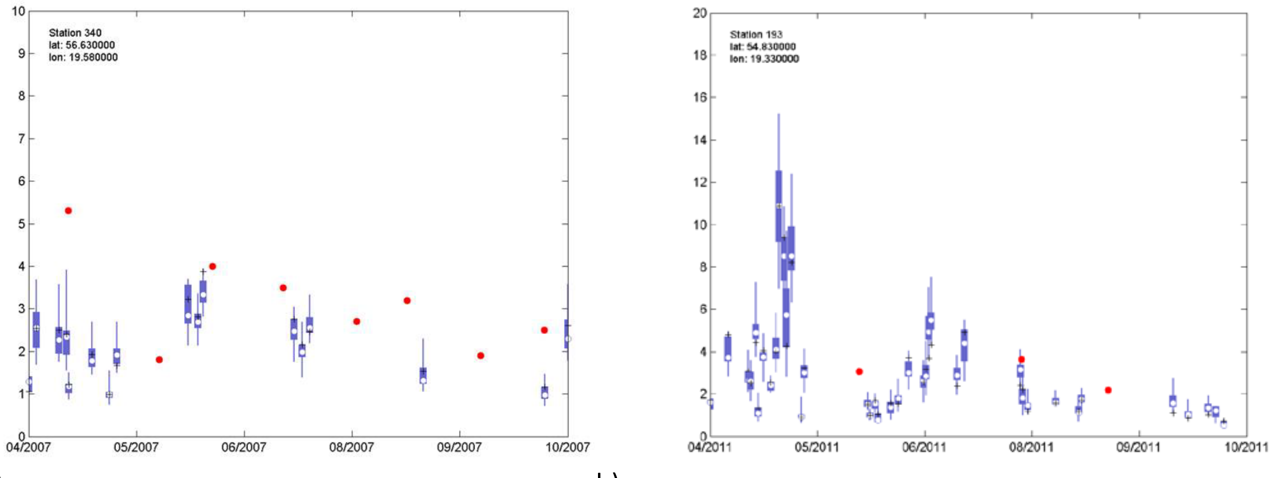
In comparison with the monitoring station measurements, the volume of EO and flow-through instrument observations is large. For example, during year 2011, 37 station measurements were made on HELCOM area SEA-012, Northern Baltic Proper. In contrast, 29,6 million EO non-cloudy observations were recorded during the growing season on 2011. Thus, it is not feasible to add all EO observations to the assessment database, but rather to include regional and sub-regional statistics for given period.

**Results of comparisons between EO and monitoring station data**



a) b)

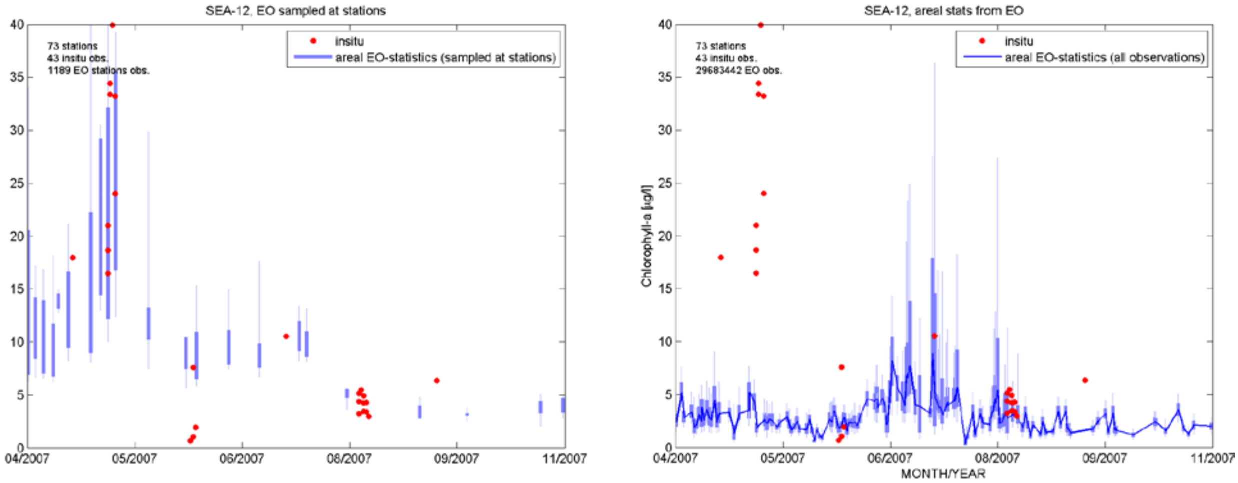
Figure 1. a) HELCOM regions and b) monitoring station locations identified by numbers on the Baltic Sea. Monitoring station data is utilized in comparisons against EO data (2007-2011). Monitoring station data provided by Vivi Fleming.



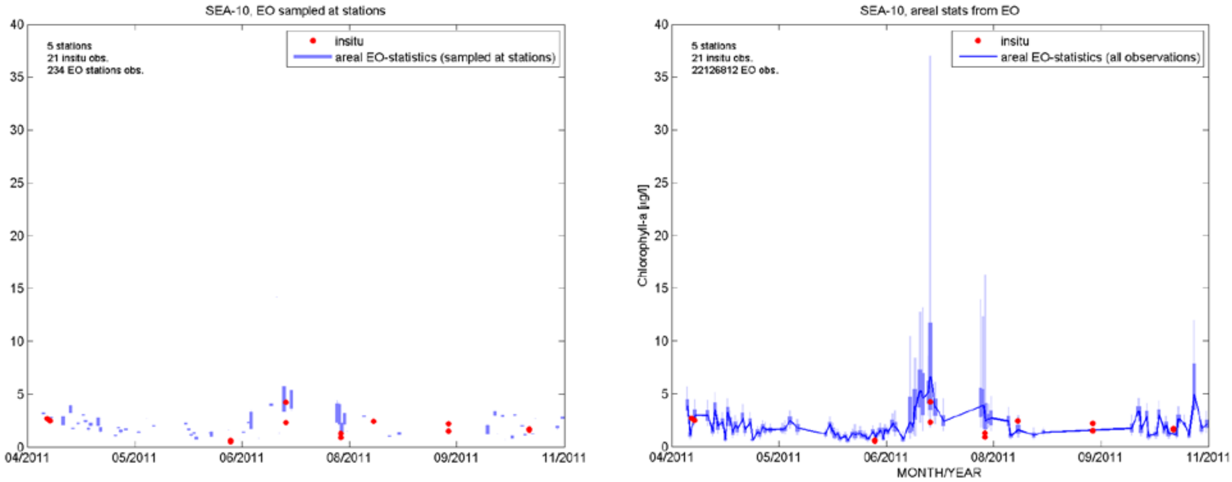

a) b)

Figure 2. Example of time series of chlorophyll-*a* [µg/l] observations using EO (blue bars) and monitoring station measurements (red dots). EO data is collected using 3x3 pixels around a) station 340, year 2007 and b) station 193, year 2011 (see Figure 1 for the location of the stations).

a)




b)




c)

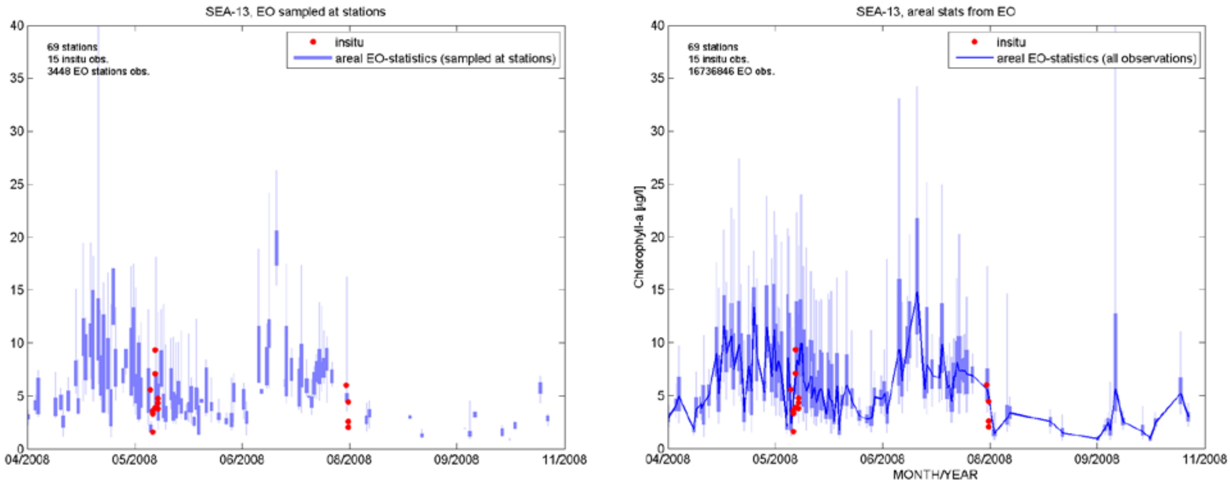



Figure 3. Examples of time series of chlorophyll-a [µg/l] observations using EO (blue) and monitoring station measurements (red dots). EO data represents observations from assessment areas a) SEA-12, year 2007, b) SEA-10, year 2011, c) SEA-13, year 2008 and d) SEA-15. See Figure 1 a for assessment area IDs.  EO data on the figures on the left hand

size represent the locations of the monitoring station. The EO data on the right hand side figures is collected from the whole assessment area. Monitoring station data represents all station measurements that locate on the assessment area and have measurements during the presented example year.

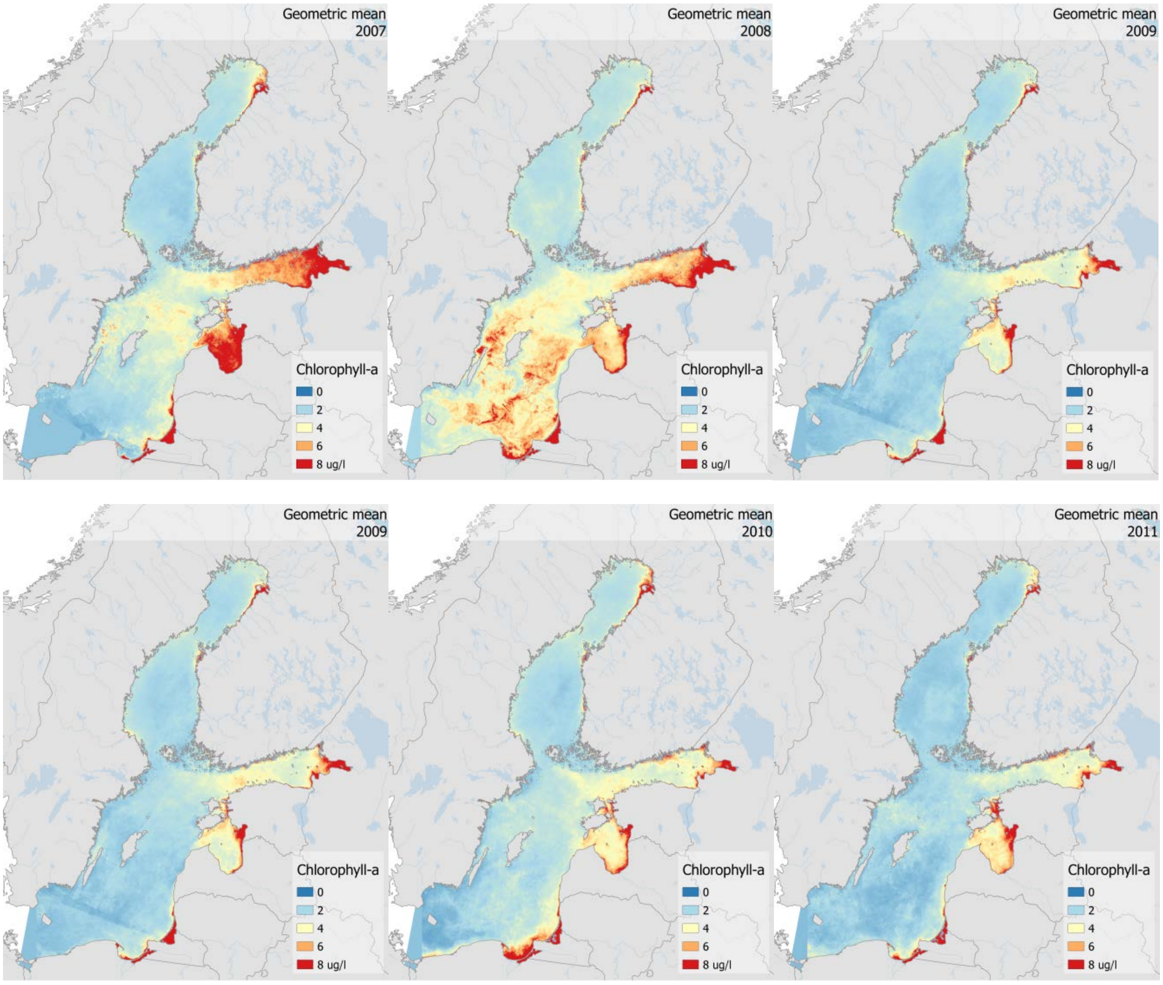



Figure 4. Maps of annual geometric mean (period: 1.6.-31.9.) for years 2007 -2011.

Table 1. Median of monitoring station measurements, EO observations around stations (3x3 pixels) and areal observations of EO, representing the whole assessment area. Annual period: 1.6.-31.9. HELCOM assessment areas: SEA-007-SEA-017.

| Name | Bornholm Basin | Gdansk Basin | Eastern Gotland Basin | Western Gotland Basin | Gulf of Riga | Northern Baltic Proper | Gulf of Finland | Åland Sea | Bothnian Sea | The Quark | Bothnian Bay |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | SEA-007 | SEA-008 | SEA-009 | SEA-010 | SEA-011 | SEA-012 | SEA-013 | SEA-014 | SEA-015 | SEA-016 | SEA-017 |
| **2007** | | | | | | | | | | | |
| station | 2,60 | 2,55 | 1,73 | 3,35 | 2,45 | 4,24 | 1,17 | 2,90 | 2,10 | | 3,37 |
| EO st | 1,44 | 2,97 | 2,10 | 2,91 | 3,60 | 8,70 | 5,21 | 2,33 | 2,08 | 2,08 | 2,49 |
| EOareal | 1,74 | 2,84 | 2,86 | 2,84 | 6,96 | 2,79 | 5,80 | 2,52 | 2,01 | 2,53 | 2,86 |
| **2008** | | | | | | | | | | | |
| station | 3,13 | 4,20 | 4,37 | 2,75 | 2,82 | | 3,52 | 1,57 | 1,82 | | 1,60 |
| EO st | 4,10 | 3,87 | 5,71 | 4,72 | 4,13 | 5,03 | 4,87 | 2,76 | 2,61 | 2,58 | 2,64 |
| EOareal | 3,64 | 6,50 | 4,11 | 3,99 | 5,11 | 3,87 | 5,26 | 2,71 | 2,49 | 2,38 | 2,50 |
| **2009** | | | | | | | | | | | |
| station | 2,60 | 2,90 | 3,51 | 2,75 | 2,66 | 1,01 | 3,07 | 2,48 | 3,28 | | 1,51 |
| EO st | 1,78 | 1,91 | 2,62 | 2,26 | 3,33 | 3,55 | 3,75 | 2,49 | 2,30 | 2,31 | 2,53 |
| EOareal | 1,36 | 3,76 | 1,82 | 1,71 | 3,53 | 2,45 | 3,41 | 2,24 | 2,09 | 2,52 | 2,46 |
| **2010** | | | | | | | | | | | |
| station | 2,30 | 2,76 | 5,04 | 3,40 | 2,82 | 3,09 | 5,24 | 4,62 | 2,16 | | 2,76 |
| EO st | 1,59 | 2,54 | 4,25 | 2,56 | 3,19 | 4,55 | 3,61 | 2,30 | 2,04 | 2,68 | 2,61 |
| EOareal | 1,70 | 6,59 | 2,48 | 2,32 | 5,70 | 2,74 | 3,85 | 2,69 | 2,38 | 2,89 | 2,63 |
| **2011** | | | | | | | | | | | |
| station | 2,78 | 2,40 | 3,71 | 1,50 | 2,74 | 0,95 | 1,02 | 3,58 | 3,22 | 2,34 | 2,49 |
| EO st | 1,43 | 1,97 | 2,65 | 1,78 | 2,30 | 4,87 | 3,09 | 2,23 | 2,05 | 1,83 | 2,10 |
| EOareal | 1,67 | 2,68 | 1,84 | 2,02 | 4,35 | 2,66 | 3,53 | 2,27 | 1,92 | 2,04 | 2,08 |

**Benefits of utilizing EO (and Alg@line) data in assessment**

Chlorophyll-*a* observations with EO data show similar concentration levels as monitoring station measurements in examples presented in Figures 2 and 3. However, on most stations, the number of EO observations is considerably higher than station monitoring measurements as can be seen on time series in Figure 2 and 3. For example, on assessment area SEA-13 (Gulf of Finland), 15 measurements were made on stations, 16.7 million via EO (3448 of these represent the station locations) during year 2008. Thus, the dynamics of algae growing season are often captured well in EO data time series. Another important benefit of utilizing EO data is that also spring bloom can be observed in much more detail. This is also shown in the examples of time series in Figures 2 and 3.

Areal EO shows often different time series and statistics than those collected from stations. This is due to the spatial coverage of EO data. The monitoring station data does not include all stations during all years, thus the station data can occasionally represent only part of the assessment area. EO data is observed from the whole area and, thus, typically shows more spatial and temporal variation and representativeness. Another explanation for differences during the assessment period are the surface floating algae blooms that are more often observed with EO data than with monitoring station data. In addition, on the surface floating blooms, EO data represents the surface layer only, while monitoring stations provide profile also beneath the surface bloom. Thus, the surface floating cyanobacteria bloom observations are likely to increase the statistics calculated using EO interpretations of chlorophyll-*a*.

The Alg@line flow-through data represents the mixed layer, as it is based on measurements made on 5 m depth. An example of this is presented in Figure 5.

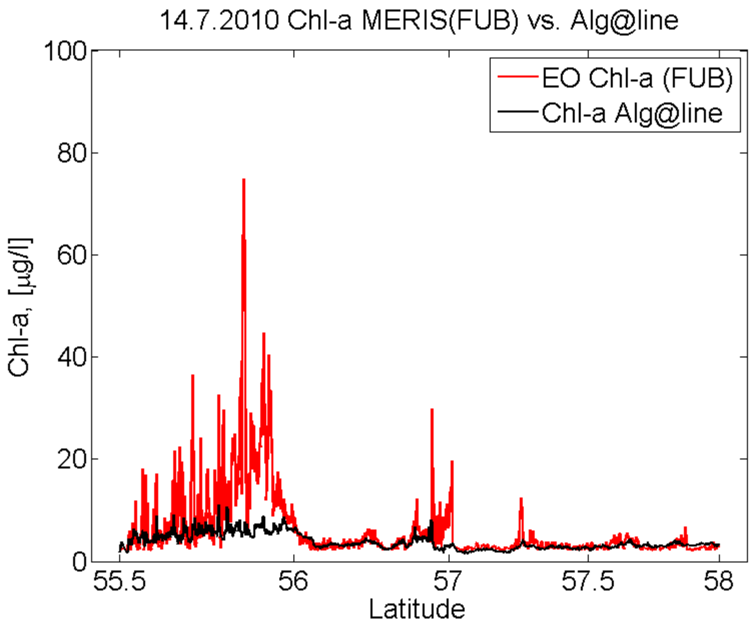



Figure 5. EO (red line) and Alg@line flow-through (black line) observations on a surface bloom (latitudes between 55.5 and 56) and on situation where algae is mixed in the water layer (latitudes between 57 and 58).

**Recommendation**

We propose to add EO chlorophyll-*a* data to HELCOM assessment database. As the annual volume of individual observations (i.e. pixels) is easily 65 million for assessment area, inclusion of statistics for each HELCOM assessment area is more preferable than inclusion of all observations (pixels). Thus, we propose at this stage to include statistics per defined period (annual, monthly, etc) for each assessment area

- arithmetic and geometric mean
- standard deviation
- percentiles (5,25, 50, 75, 95)
- min, max
- N of observations that were used to derive statistics

In addition, for annual periods, it is suggested to include also the mode of the histogram in statistics. EO data allows for derivation of time series and histograms for assessment area. This type of data is may not be storable in the database. Nevertheless, this data can be stored otherwise for assistance of assessment work.

In addition to the whole assessment area, we see that for intercomparisons between the monitoring station information and EO data, it is advisable to store also data around most commonly used monitoring stations. This requires the identification of monitoring station locations that are of interest. For these locations, a median value using 3x3 pixels from monitoring station location is most preferable.

As the number of EO observations on assessment area and monitoring station location varies due to cloudiness and availability of EO data, we also suggest that regular grid data is stored using HELCOM grid. This report does not give examples for this approach, but is to be presented later. The use of gridded data may be used for accounting and balancing possible cloudy periods on EO data.

The information collected with Alg@line flow-through instruments can be utilized using similar procedure to derive similar statistics (and time series) for the assessment areas for defined time periods. This will complement the caps produced by clouds in EO data.

**Appendix A**

Daily overpasses of satellite instruments have resulted in time series of observations from the period of 2002-2011 with the most optimal instrument MERIS on-board Envisat satellite. MERIS instrument collected observations using 300mx300m pixel size (spatial resolution). After the loss of Envisat satellite (spring 2012), the current EO-monitoring of chlorophyll-*a* (by SYKE) has been performed using MODIS instrument that observes with 1km spatial resolution. For the future monitoring and assessment, satellite instruments Sentinel-3a OLCI can provide the best functionality for the estimation of parameters such as chlorophyll-a, transparency and turbidity. This is related to both their improved spatial resolution and optimal spectral characteristics (particularly OLCI). The estimated time for OLCI to be operational is during 2016.

**Methods to describe the data processing and matching of EO and monitoring station data**

Examples of EO data comparisons with monitoring station data are given for the test period of 2007-2011. The image pixels that are covered by clouds or cloud shadows have been removed from each EO-image through a combined automatic and manual screening process, leaving a final product with only the pixels which have an undisturbed visibility to the satellite instrument.

*Statistics for the HELCOM-area*

The areal statistics for each HELCOM-area were computed from the clear pixels for each day for which an EO-image is available within the annual assessment period (1.6.-30.9.). The statistics comprise arithmetic and geometric means of EO-observation of all clear pixels, median and mode values, percentiles of the distribution (5..95 % in 5 % increments, and 2/98%).

*Sampling the EO-observations to match the monitoring station measurements*

To complement the areal statistics, sampling of EO-data at each monitoring station was computed. For each day within the annual assessment period, the EO-observations that match to each monitoring station were extracted, which produces a dataset of daily EO-observations for each measurement station.

To reduce the effect of the small noise present in EO-observations and to smooth out the spatial variability of chlorophyll-*a* nearby the measurement station a sampling algorithm was used. The windowed sampling algorithm finds the closest EO-observation pixel to match each monitoring station and extracts the EO-observations for the 3 x 3 pixel window centred at the monitoring station (spatial coverage 900 m x 900 m). The median value from the window was used to represent the EO-observation for the monitoring station. The 5/95% and 25/75% percentiles of the values from the 3 x 3 pixel window were used to estimate the confidence limits for the observation.

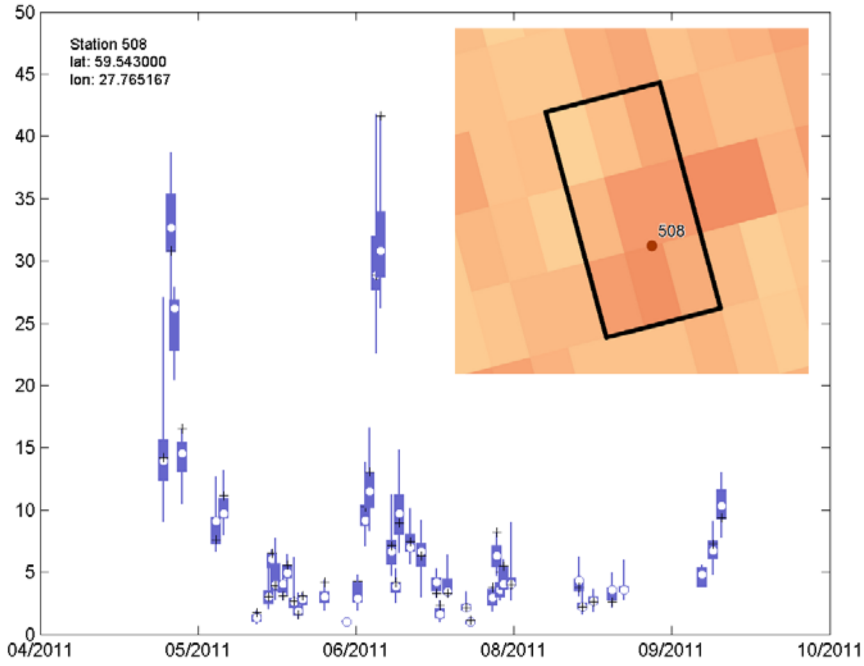



*Figure A1. An example of the 3 x 3 pixel sampling window used to sample the EO-observations for monitoring station no. 508, and a graph of the produced samplings that represent the observations at that monitoring station. Each pixel has a spatial coverage of 300 m x 300 m (image is distorted due to the used map projection).*

*Sampled statistics of the EO-observations for the HELCOM-areas*

The sampled EO-observations for each monitoring station were used to compute daily statistics for each HELCOM-area. For each day within the assessment period, the algorithm collecs the monitoring stations that fall inside the particular HELCOM-area, and uses the sampled EO-observations to compute a representative statistics for the whole area (arithmetic and geometric mean, median and percentiles with 5% increments). The distribution of the sampled values is visualized by the 5/95% percentiles and 25/75% percentiles.