Guidelines for monitoring of chlorophyll *a* via EO

1. Background

1.1 Introduction

Increase in phytoplankton biomass is a direct consequence of advancing eutrophication. For monitoring purposes, phytoplankton biomass is estimated by chlorophyll a (Chl *a*) concentration.

The Chl *a* concentration provides information about the trophic state of a water body. Although this parameter is not a direct index for phytoplankton biomass because of a highly variable ratio of cellular carbon to Chl *a* in phytoplankton (Geider, 1987) it is easily-determinable providing a wealth of data and being thus a key variable in trophic evaluation of waters.

The scope of this guideline is the determination of Chl *a* concentration by using EO (Earth Observation), i.e., by employing satellite instrumentation.

1.2 Purpose and aims

Monitoring of Chl *a* provides information that is used for assessing direct effects of eutrophication. The aim is to provide spatiotemporal information for the determination of an aquatic trophic status and long-term trends in it, and to ensure that the data acquired is compatible with the HELCOM core indicator 'Chlorophyll a'. The indicator description, including its monitoring requirements, is given in the HELCOM core indicator web site [helcom.fi/baltic-sea-trends/indicators/chlorophyll-a]. Chl *a* is a multiparametric indicator and is based on a combination of station data, and in the open sea areas, Ferrybox data (the ship-of-opportunity (SOOP)) and EO data.

2. Monitoring methods

2.1 Monitoring features

Chl *a* is the main photosynthetic pigment in eukaryotic phytoplankton. It exhibits two dominant peaks in the absorption spectrum of phytoplankton, at blue (440 nm) and red (665 nm) wavebands. Phytoplankton absorption together with backscattering by phytoplankton cells shape the emergent flux (Kirk, 1994). Determination of Chl *a* via EO is based on analyzing and modelling this flux.

EO data can be used for the mapping of Chl *a* within the frames of available and properly calibrated satellite instruments. Chl *a* determination methods are available for ENVISAT MERIS instrument (2003 – 2011) and in forthcoming assessments for Sentinel-3-series OLCI instruments (see Annex A for more information).

2.2 Time and area

The EO monitored area covers the entire Baltic Sea area, both the open sea and coastal areas. EO-based determination of Chl *a* brings information with 300 m ground resolution throughout the growth season.

EO-based determination of Chl *a* is dependent on the sufficiency of insolation and, in the northern Baltic Sea, the absence of ice; in the northern Baltic Sea this qualifies months April – October, offering the possibility of assessing mean values for the spring / autumn season, and for the entire growth season.

Clouds prevent the proper interpretation of the EO-mapping based on the transmission of visible light. For the Baltic Sea area, approximately 70% of all observations are cloudy.

In optically extreme areas, being subject to a considerable load of CDOM (Colored dissolved organic matter) and / or suspended matter (e.g. river estuaries especially in the northernmost part of the Baltic Sea), the method is not reliable for determining Chl *a* concentration.

2.3 Monitoring procedure

2.3.1 Monitoring strategy

Chl *a* patterns vary substantially both in space and time, including inter-annual variation, obscuring the determination of temporal trends. EO can provide environmental information of the open sea sub-basins in good spatial and temporal coverage. Annually, the number of valid individual observations (pixels) from open area sub-basins is typically between 20 and 30 million depending on the sub-basin size and cloudiness within the assessment period (1.6. – 30.9.).

2.3.2 Sampling method

The representative depth of penetration of light (the layer in which the satellite signal is formed) depends on the clarity of the water. Therefore, Chl *a* estimated by EO is not based on any specific depth interval:

- In principle, an EO signal stems from the sunlit water layer (from the surface down to Secchi depth). In practice, it is primarily influenced by optically active substances (seston, including phytoplankton, and CDOM) causing light absorption and scattering from the surface down to about 0.5*Secchi depth. If the surface layer is uniformly mixed then the EO products describe the whole surface mixed layer (varying typically between 8 – 20 m in the open Baltic Sea).
- Whenever cyanobacteria form surface accumulations the contribution of deeper water to the EO signal decreases drastically with increasing depth. The modelled Chl *a* concentration does not represent the water column below the scum (e.g. Kutser T., 2004; Reinart and Kutser, 2006).
- In very turbid or CDOM-rich estuaries the depth of penetration of light is small (< 1.5 m). These areas are not relevant parts of the open sea assessment areas.

2.4 Data analysis and quality control

The input satellite data used for Chl *a* determination should utilize the latest image re-processing (calibration) of the responsible organization providing the observations (e.g. ESA, EUMETSAT or NASA). These are recommended and qualified by the validation team of each EO instrument (e.g. for OLCI, the Sentinel-3 Validation Team). Cloudy areas and areas shallower than the penetration of light in the water should be excluded and quality assured well to avoid non-valid observations. The EO data delivery format and description reported to ICES should follow the Guidelines for EO based information (HELCOM 2015 and Annex B).

The EO data analysis is based on the determination of phytoplankton absorption together with backscattering by phytoplankton cells, from which the Chl *a* is derived. The inversion model (Schroeder et al., 2007a, 2007b) estimates Chl *a* concentrations with the best overall accuracy in the Baltic Sea (e.g. Kratzer et al., 2008; Kratzer and Vinterhav, 2010; Beltrán-Abaunza et al., 2014; Attila et al., 2018). Annex A describe background information on EO observations and cross-comparison against station sampling. It also describes briefly the data analysis and handling for the MERIS instrument that was suggested in HELCOM (2015). The formula of descriptions in Annex A should be used as a basis for forthcoming EO data reporting. The data analysis and quality control followed in the EO Chl *a* reported previously is described in detail in Attila et al. (2018).The modelling method and its quality control have been cross-compared against monitoring station samples (e.g. HELCOM, 2015; Kauppila et al., 2016). Annex A gives information on how to reach station-wise time series cross-comparisons between EO and station sampling Chl *a* over ICES station sites for the MERIS timeline.

3. Data reporting and storage

EO Chl *a* data reporting to HELCOM is done within HELCOM 20 km grids containing daily statistics (5, 25, 50, 75, 95 percentiles) of observations according to the guidelines as described in (HELCOM 2015; Annex B). The 20km data is kept at the eutrophication assessment database hosted by ICES (HELCOM 2015, HELCOM core indicator 'Chlorophyll a'). Currently, EO Chl *a* data has been provided by Finland. This present dataset is validated, stored and reported to ICES by Finnish Environment Institute (SYKE) (Annex A and HELCOM 2015).

4. Contacts and references

4.1 Contact persons

Jenni Attila, Finnish Environment Institute

Mika Raateoja, Finnish Environment Institute

4.2 References

Attila, J., Kauppila, P., Alasalmi, H., Kallio, K., Keto, V., Bruun, E., Koponen S. (2018). Applicability of Earth Observation chlorophyll-a data in assessment of water status via MERIS – with implications for the use of OLCI sensors. Remote Sensing of Environment, 212, 273-287. https://doi.org/10.1016/j.rse.2018.02.043

Babin M. (2008). Phytoplankton fluorescence: theory, current literature and in situ measurement. In: Babin M, Roesler CS, Cullen JJ (eds) Real-time coastal observing systems for marine ecosystem dynamics and harmful algal blooms: Theory, instrumentation and modeling. Unesco, Paris, p. 237-280.

Beltrán-Abaunza, J.M., Kratzer, S., & Brockmann, C. (2014). Evaluation of MERIS products from Baltic Sea coastal waters rich in CDOM. Ocean Science, 10, 377–396. DOI: 10.5194/os-10-377–2014.

Geider R. (1987). Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. New Phytologist 106: 1-34.

HELCOM (2015) Eutrophication Assessment Manual, Annex 3A. http://www.helcom.fi/helcom-at-work/projects/eutro-oper/. Last revision 2015.

Kauppila P., Eremina T., Ershova A., Maximov A., Lips I., Lips U., Alasalmi H., Anttila S., Attila J., Bruun J.-E., Kaitala S., Kallio K., Keto V., Kuosa H. and Pitkänen H., Distribution and trends of phytoplankton chlorophyll a and surface blooms, In: Raateoja M. and Setälä O., (Eds.), Eutrophication in the Gulf of Finland (Reports of the Finnish Environment Institute, No. 27), 2016, ISBN 978-952-11-4577-3 or (PDF) 978-952-11-4578-0.

Kirk J. (1994). Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Second edition. 509 p.

Kratzer S., Brockmann C., & Moore G. (2008). Using MERIS full resolution data to monitor coastal waters – a case study from Himmerfjärden, a fjord-like bay in the northwestern Baltic Sea. Remote Sensing of Environment, 112, 2284–2300.

Kratzer S., & Vinterhav C. (2010). Improvement of MERIS level 2 products in Baltic Sea coastal areas by applying the Improved Contrast between Ocean and Land processor (ICOL) – data analysis and validation. Oceanologia, 52(2), 211–236.

Kutser, T. (2004). Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. Limnol. Oceanogr., 49(6), 2179–2189.

Reinart A. & Kutser T. (2006). Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. Remote Sensing of Environment, 102, 74–85.

Schroeder, T., Schaale, M., & Fischer, J. (2007a). Retrieval of atmospheric and oceanic properties from MERIS measurements: A new Case-2 water processor for BEAM. International Journal of Remote Sensing, 28(24), 5627–5632.

Schroeder, T., Behnert, I., Schaale, M., Fischer, J., & Doerffer, R. (2007b). Atmospheric correction algorithm for MERIS above case-2 waters. International Journal of Remote Sensing, 28, 1469–1486.

Annex A

Background on EO observations, data handling and cross-comparison against station sampling

Background

One of the conclusions of 'Eutrophication status of the Baltic Sea (BSEP 143)' was, that the confidence of Chl *a* indicators would be substantially increased by including also Earth Observations. EUTRO-OPER-project investigated further the possibilities of updating the Chl *a* indicator using also EO-based Chl *a* and to include it into the assessment data flow. As a result, HELCOM accepted the use of EO data for the HELCOM core indicator 'Chlorophyll a'. The indicator description, including its monitoring requirements, is given in the HELCOM core indicator web site: [http://helcom.fi/baltic-sea-trends/indicators/chlorophyll-a].

The EUTRO-OPER- project (HELCOM, 2015) proposed solutions for the format of describing how EO Chl *a* data should be included into the assessment database, and how the information should be described and validated before used in updating the indicator (updated version in Annex B). The EO instrument used for the report was MERIS (MEdium Resolution Imaging Spectrometer) by ESA (European Space Agency). The MERIS-based Chl *a* dataset provided to the assessment database covers years 2003 – 2011. OLCI (Ocean and Land Color Instrument), a follow-up instrument by ESA onboard Sentinel-3 satellite series, can be used to derive Chl *a* from May 2016 onward. The OLCI instrument will be onboard four consecutive satellites ensured by ESA with continuing observations guaranteed until 2030. Consistent and complete OLCI observation datasets covering the years from 2016 onward has been released by EUMETSAT on Feb 2018. The technical descriptions of the EO data handling in this Annex and in HELCOM 2015 are according to the EO data procedure followed in the previous reporting by Finland, but can be updated once other member countries wish to report EO data to HELCOM. The data handling of OLCI observations and conversion to Chl *a* concentrations should follow the procedures described for the MERIS instrument, but are to be updated in technical details during 2019.

EO observation principle

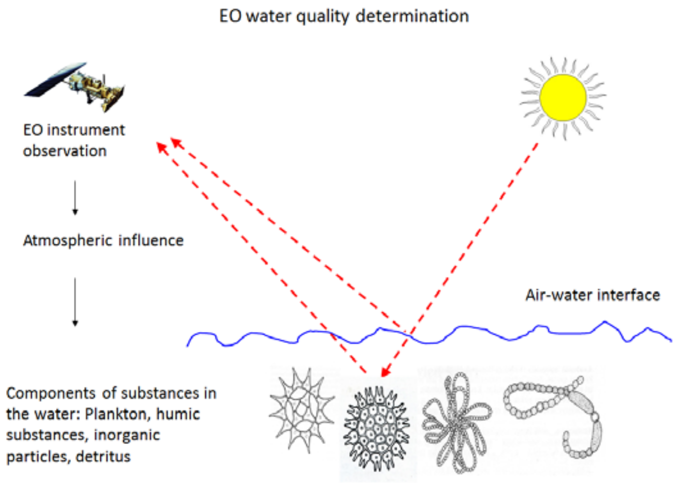


Figure 1. The measurement principle of a passive optical satellite instrument and the substances in water that affect the observed radiances.

The measurement principle of passive optical satellite instruments is based on observing sunlight, reflected by water to the direction of the satellite (Figure 1). The measured light is affected by the atmosphere and the optical properties of the water body. Therefore, the observed radiance is a mixture of surface properties and the optical properties of the upper layer of the water (from the surface to the depth of approximately up to one half of Secchi depth). In shallow areas – where the transparency of the water is higher than the depth of the water – also the seabed and its composition affect the reflected light (Kirk, 1994).

Clouds prevent the observation of ground targets, and approximately 70% of the observations are cloudy over the Baltic Sea. EO algorithms used for determining the water constituents such as the concentration of Chl *a* and suspended particles utilize the known properties of absorption and scattering at the different wavelength regions observed by the EO instruments.
The absorption and scattering by substances depend highly on their concentration in the water body. Thus, concentrations of different substances, together with their specific optical properties, determine the variations in the spectral absorption. The total absorption detected from the water column is a sum of its components according to

$$a_{tot}(\lambda) = a_w(\lambda) + a_{CDOM}(\lambda) + a_{ph}(\lambda) + a_{nap}(\lambda)$$

where $a_w(\lambda)$, $a_{CDOM}(\lambda)$, $a_{ph}(\lambda)$ and $a_{nap}(\lambda)$ are the absorption coefficients of pure water, CDOM, phytoplankton and non-algal particles (detritus and inorganic particles) (Kirk K., 1994; Kallio K., 2012), respectively.
Since the specific absorption features of different constituents, such as phytoplankton and CDOM, occur at certain wavelengths their concentrations can be determined from the measured absorption spectrum (Kirk K., 1994).

EO Chl *a* determination accuracy in comparison to station sampling

The correspondence between station sampling and EO Chl *a* observations has been analyzed e.g. for the coastal and open waters adjacent to Sweden (Kratzer et al., 2008; Kratzer and Vinterhav, 2010; Beltrán-Abaunza et al., 2014), for the Gulf of Finland (Kauppila et al., 2016), for the coastal waters of Finland (Attila et al., 2018) and for the HELCOM assessment areas and ICES stations (HELCOM, 2015). Similarly to the results presented here for ICES stations, these studies indicated that with respect to the EO processors available for MERIS the inversion model developed at the Free University of Berlin, the 'FUB model' (Schroeder et al., 2007a, 2007b), provided the best estimates of Chl *a* concentrations in the Baltic Sea. The comparison analysis between EO and ICES was made for the period of 2007 – 2011 (HELCOM, 2015) e.g. by matching EO data to the station sampling at the ICES station location and Alg@line water sampling sites as a time series (using 3x3 pixel median values of EO data around the station location). The EO and MS histograms produced from the match-up datasets showed good correspondence and skewed distribution over the same day and location of both datasets (HELCOM 2015; Attila et al., 2018). When the EO images were taken on the same day as the station observations, the absolute bias that describes differences between ICES and EO Chl *a* results is less than 5 µg l^{-1} in 76.3 % of the dataset and less than 2 µg l^{-1} for 54% of the match-ups.

The differences between EO and station sampling Chl *a* concentrations for the assessment period were identified to be mainly due to very different magnitudes of the number of observations. Depending on the size of the assessment area, individual EO observations of Chl *a* typically amount up to several million within one year (e.g. in 2011, 30 million non-cloudy EO Chl *a* observations (individual pixels) were recorded during the growing season on HELCOM area SEA-012, Northern Baltic Proper, whereas station sampling amounted to 37 Chl *a* recordings). HELCOM (2015) found the HELCOM 20K grid units with daily EO Chl *a* statistics (percentiles) as the most suitable calculation units for delivering spatially and temporally balanced EO Chl *a* for the assessment. The comparisons made over the HELCOM assessment areas with each of the monitoring method show good reciprocal correspondence (HELCOM 2015). The geometric mean maps summarize well the EO Chl *a* distribution during the assessment period (HELCOM 2015; Kauppila et al., 2016). Analysis about the use of EO data for the assessment of larger assessment areas has been published (HELCOM 2015; Kauppila et al., 2016; Attila et al., 2018). Individual large differences between station sampling and EO Chl *a* data may occur during intensive cyanobacteria scum and spring phytoplankton bloom, both of which are dynamic periods with high temporal and spatial variation (e.g. Kutser T. 2004; Attila et al., 2018). According to analysis of Chl *a* concentrations derived from EO vs. monitoring stations, the mode and the geometric mean of the EO Chl *a* data were found suitable statistics for utilization in status assessments (HELCOM 2015; Attila et al., 2018). The EO Chl *a* observations are not normally distributed, therefore the use of the EO arithmetic mean in status assessments is not recommended for water areas with intensive momentary phytoplankton blooms.

Description of EO data handling

This part describes EO data processing details followed in previous reporting (by Finland) and serves as a base of formula for future reporting.

Reported statistics for the HELCOM 20 km grid division

The statistics for each HELCOM-20 km grid were computed from non-cloudy pixels for each day for which an EO-image was available. The statistics comprise arithmetic and geometric means of EO-observation of all valid pixels, median and mode values, percentiles of the distribution (5% to 95 % in 5 % increments, and 2%/98%).

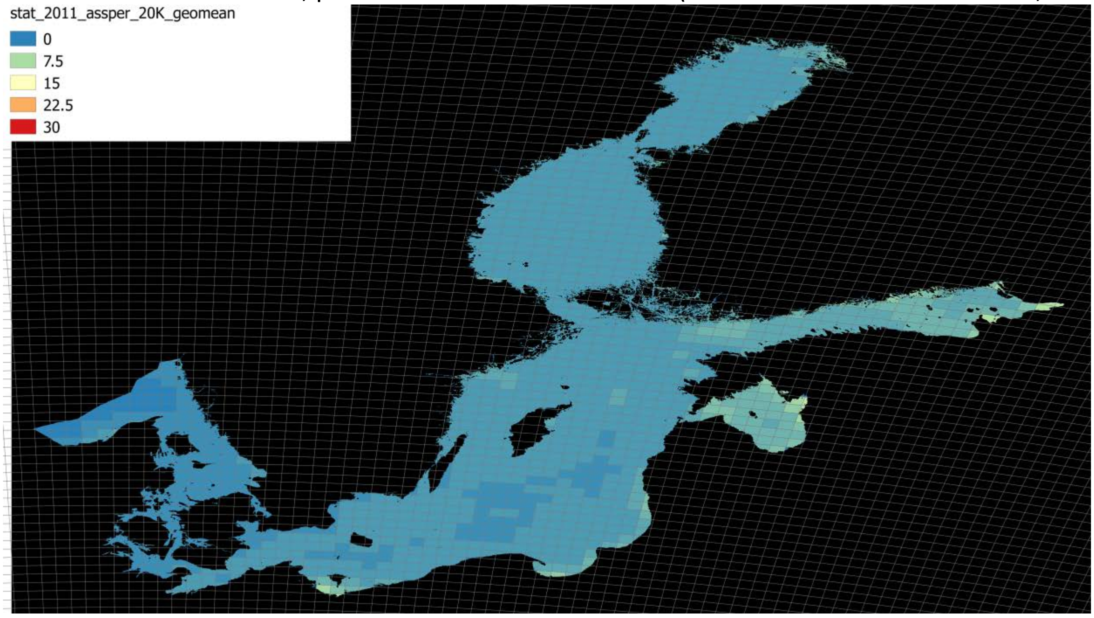


Figure 2. HELCOM 20 km grid division used for EO Chl *a* statistics.

MERIS instrument

In the dataset reported by Finland, MERIS L1B data (3rd reprocessing) was corrected with AMORGOS 3.0 (Accurate MERIS Ortho-Rectified Geo-location Operational Software). Chl *a* concentrations were derived with the FUB processor — specifically, the MERIS Case-2 Water Properties Processor, (Schroeder et al., 2007a, 2007b). The FUB processor is based on a combination of neural networks (NNs): one for the atmospheric correction and three separate NNs that derive the water-quality components (chl-a, TSM, and CDOM) directly from top-of-atmosphere radiance values. A forward radiative transfer model is used to simulate light propagation through the ocean and the atmosphere (Schroeder et al., 2007b). The image pixels that were covered by clouds or cloud shadows were removed from each EO-image automatically (IDEPIX processor). The FUB built-in land mask was complemented by a national shoreline vector data. The depth of the seabed was identified by means of available datasets: accurate depth raster data inherited from the national nautical charts of the Finnish Maritime Administration for most of Finnish coastal areas, and for the other areas the lower-accuracy Baltic Sea Bathymetry Database (BSHC, 2013). Pixels with water depth identified less than 5 m were excluded.

Table 1. Technical details of reported MERIS Chl *a* (as in Annex B).

Processing responsible: country, organization, institute, PI	Finland, Finnish Environment Institute, Jenni Attila
Satellite instrument(s)	MERIS
Products derived from satellite data	Chl *a*
Details of the collection sensor	ESA website: https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat/instruments/meris
Resolution of original data	300 m
Algorithm used for deriving product	FUB (Schroeder et al., 2007a, 2007b)
Atmospheric correction scheme and cloud masking	FUB (Schroeder et al., 2007a, 2007b), IDEPIX
Level of spatial aggregation used: either HELCOM assessment area or HELCOM 20 km grid	HELCOM 20 km grid, (originating from general EEA grids,https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2)

Level of temporal aggregation used: daily or annual assessment period	daily
Uncertainties on product estimates	see HELCOM 2015
Date and time of the start and end of the sampling (UTC)	Period 1.4 – 31.10. during years 2003-2011
Position estimate (latitude and longitude degrees and minutes or decimal degrees can be used. Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs.)	According to the specifications of HELCOM 20 km grid division.

OLCI instrument

A consistent and complete OLCI dataset covering the years starting from 2016 has been released by EUMETSAT on February 2018 and preparations have been started to report EO Chl *a* based on this instrument for the next Chl *a* assessment. The data handling of OLCI observations and conversion to Chl *a* concentrations will follow the procedures described for the MERIS instrument. Some technical details will be updated during 2019. According to recent evaluations, the best approach for Chl *a* determination is to use the C2RCC inversion model (Doerffer and Schiller, 2007). The results for OLCI + C2RCC show accuracies that are similar to MERIS + FUB processor. A more comprehensive analysis is ongoing and will be finished in 2019.

Station-wise validation of the reported MERIS instrument Chl *a* dataset

Cross-comparison results between EO and station site observed Chl *a* reported to ICES are presented in the TARKKA service at www.syke.fi/tarkka/en

Below is a view of the TARKKA web application with ICES stations and HELCOM assessment area division shown over the basemap. In the TARKKA interface, ICES-stations and HELCOM-area division can be activated by clicking the 'HELCOM ICES stations' on (and all other datasets off) from the layer selection panel on the left:

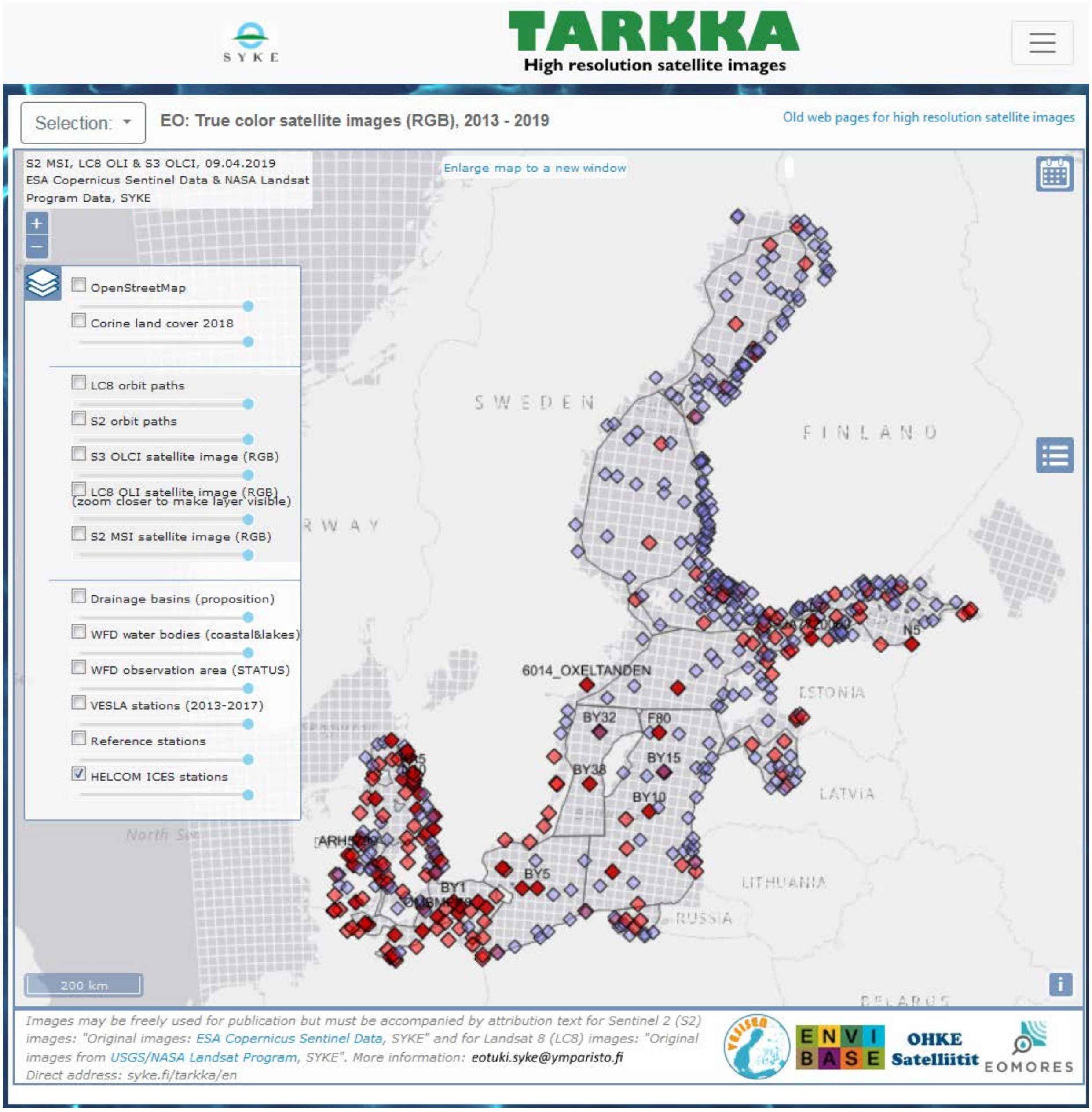


Figure 3. TARKKA web application view to comparisons between EO and ICES station sampling Chl *a*.

A particular station can be selected by clicking the left mouse button over the station marker. The stations most frequently monitored are marked with red (50 most frequently observed stations), the stations with less frequent measurements are presented with lighter-shaded red color, and the sparsely monitored stations are marked with blue. If the station sites are overlapping in the view, the color is darker-shaded.

When clicking at the station marker, a view to the station time series is opened:

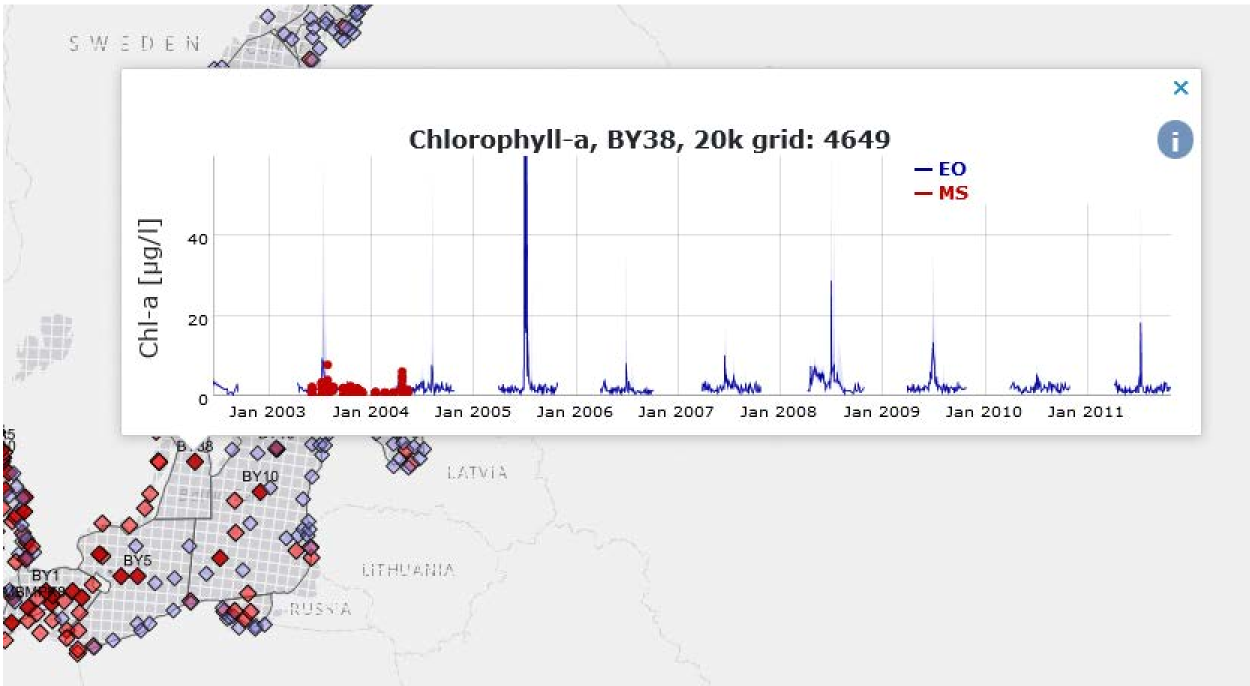


Figure 4. Example view to station BY38 for the MERIS timeline 2003 – 2011.

The time series plots are dynamic. For example, one can zoom to a particular year or concentration range by holding the left mouse key and moving the mouse either in x-axis dimension (timeline, e.g. one of the years) or in y-axis direction (concentration range).

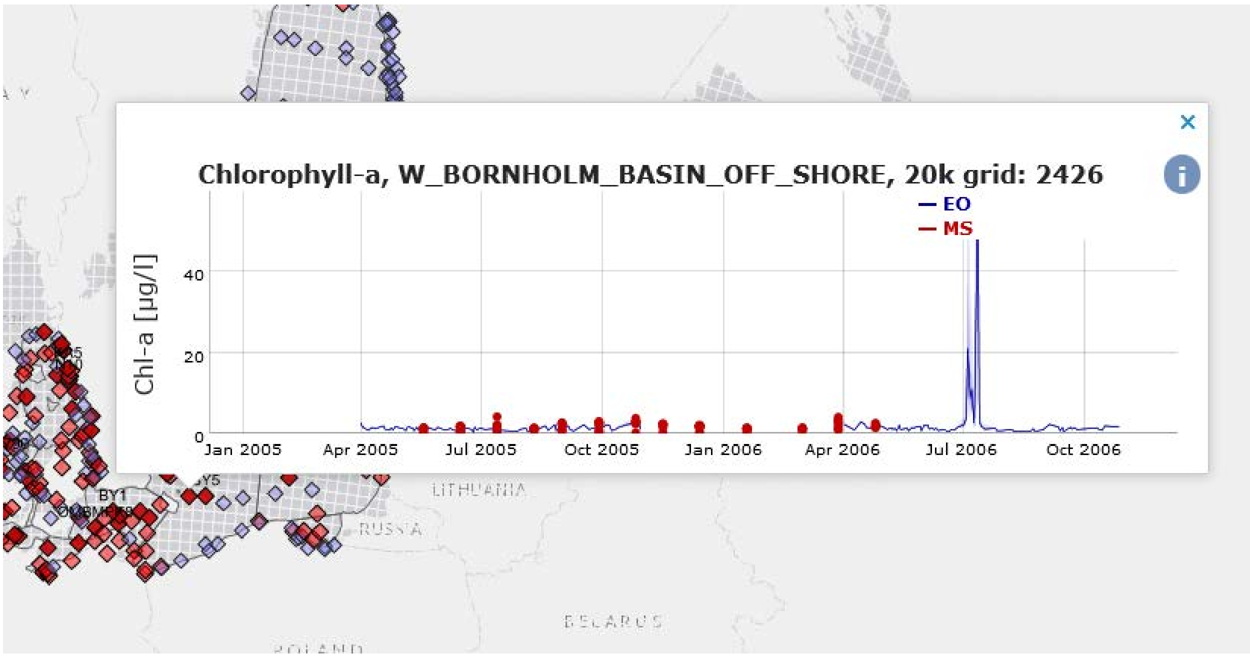


Figure 5. Example view zoomed to years 2005 and 2006.

The default settings can be returned by double clicking the left mouse button over the graph. All station depths are included at the moment. We are currently working on rejecting all other than surface samples in case the measurement depth is clearly defined for the particular site measurement in the ICES dataset.

References

Attila, J., Kauppila, P., Alasalmi, H., Kallio, K., Keto, V., Bruun, E. (2018). Applicability of Earth Observation chlorophyll-a data in assessment of water status via MERIS – with implications for the use of OLCI sensors. Remote Sensing of Environment, 212, 273-287. https://doi.org/10.1016/j.rse.2018.02.043

Beltrán-Abaunza, J.M., Kratzer, S., & Brockmann, C. (2014). Evaluation of MERIS products from Baltic Sea coastal waters rich in CDOM. Ocean Science, 10, 377–396. DOI: 10.5194/os-10-377–2014.

Doerffer, R. & Schiller, H. (2007). The MERIS Case 2 algorithm. International Journal of Remote Sensing, 28 (3–4), 517–535. doi:10.1080/01431160600821127.

HELCOM (2015). Eutrophication Assessment Manual, Annex 3A. http://www.helcom.fi/helcom-at-work/projects/eutro-oper/. Last revision 2015.

Schroeder, T., Schaale, M., & Fischer, J. (2007a). Retrieval of atmospheric and oceanic properties from MERIS measurements: A new Case-2 water processor for BEAM. International Journal of Remote Sensing, 28(24), 5627–5632.

Schroeder, T., Behnert, I., Schaale, M., Fischer, J., & Doerffer, R. (2007b). Atmospheric correction algorithm for MERIS above case-2 waters. International Journal of Remote Sensing, 28, 1469–1486.

Kallio K., (2012). Water quality estimation by optical remote sensing in boreal lakes. PhD Thesis. Monographs of the Boreal Environment Research, 39, 1-54. ISBN 978-952-10-7405-9, ISBN 978-952-10-7406-6 (PDF).

Kauppila, P., Eremina, T., Ershova, A., Maximov, A., Lips, I., Lips, U., Alasalmi, H., Anttila, S., Attila, J., Bruun, J.-E., Kaitala, S., Kallio, K., Keto, V., Kuosa, H., & Pitkänen, H. (2016). Distribution and trends of phytoplankton chlorophyll a and surface blooms. In M. Raateoja & O. Setälä (eds), Eutrophication in the Gulf of Finland (Reports of the Finnish Environment Institute, No. 27). ISBN 978-952-11-4577-3 or (PDF) 978-952-11-4578-0.

Kirk J. (1994). Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Second edition. 509 p.
Kratzer, S., & Vinterhav, C. (2010). Improvement of MERIS level 2 products in Baltic Sea coastal areas by applying the Improved Contrast between Ocean and Land processor (ICOL) – data analysis and validation. Oceanologia, 52(2), 211–236.

Kratzer S., Brockmann C., & Moore G. (2008). Using MERIS full resolution data to monitor coastal waters – a case study from Himmerfjärden, a fjord-like bay in the northwestern Baltic Sea. Remote Sensing of Environment, 112, 2284–2300.

Kratzer S., & Vinterhav C. (2010). Improvement of MERIS level 2 products in Baltic Sea coastal areas by applying the Improved Contrast between Ocean and Land processor (ICOL) – data analysis and validation. Oceanologia, 52(2), 211–236.

Annex B

Guidelines for describing EO-based information

This Annex is updated version of the guidelines for describing EO-based information given in HELCOM (2015). In this guideline, EO-based information is considered to be any information derived from satellite images for the use of updating indicators. The observations are in validated form, and may be aggregated temporally and spatially to a specific level.

The data must include a distinct time and position information. In the case of aggregated information, these may represent average values.

Receiving data

The Data Centres require the following information to be supplied by the data supplier together with the data. When receiving data, the Data Centres shall strive to meet the following guidelines.

Data standard

All satellite-derived products must be clearly specified and described. If product codes are to be used, then the source data dictionary consistency must be specified (e.g. CF Metadata Convention) . Product units must be clearly stated, and the algorithms used in the computations should be stated. The data should be fully checked for quality and pre-edited or flagged for erroneous values. An explicit statement should be made of the checks and edits applied to the data. A brief description, or a reference to the data collection and processing methods (e.g. reference to a specific technique or specific project protocols) must be included and should contain information regarding:

- Methods and procedures applied to the analysis of original raw data
- Methods / protocols and dataset(s) used for validation, or refer to their original source
- Internal or external quality assurance procedures (e.g. NASA, ESA protocols, QA4EO guidance[1])

A brief description of the data processing procedures must be included and should contain information regarding:

- editing/quality control methods
- how are trace values (values below the detection limit) identified
- how are missing values handled (null vs. zero, or “blanks”)
- what is the precision of the methods (e.g. number of significant figures)
- what units are used
- describe what quality flags are used if any
- supply a validation document

If a report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

Format description

EO data and related metadata will be provided primarily via open and standard interfaces (INSIPIRE compatible format).

[1] http://qa4eo.org

Data format, in case individual observation data is provided, should be documented for example NetCDF-4 or INSPIRE compliant format. If in doubt about the suitability of any particular format, advice from the Data Centre should be sought. Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is used. The contents of the data and ancillary information should adhere to the convention for CF (Climate and Forecast) metadata (http://cfconventions.org) or equivalent (e.g. Copernicus Marine Service).

Collection and processing details

Pertinent information to be included in the data transfer to the Data Centre includes:

- Processing responsible: country, organization, institute, PI
- Satellite instrument(s)
- Products derived from satellite data
- Details of the collection sensor
- Resolution of original data
- Algorithm and data processing used for deriving product
- Atmospheric correction scheme and cloud masking
- Level of temporal and spatial aggregation used
    - spatial: either HELCOM assessment area or HELCOM 20 km grid
    - temporal: daily or annual assessment period
- Uncertainties on product estimates
- Date and time of the start and end of the sampling (UTC)
- Position estimate (latitude and longitude degrees and minutes or decimal degrees can be used. Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs.)

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use. For additional information on quality control procedures, metadata requirements for particular parameters and collection instrumentation, see CF Convention (http://cfconventions.org).

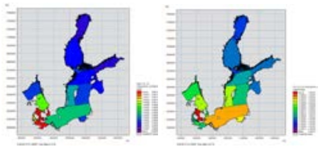
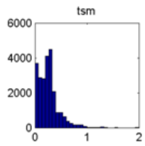
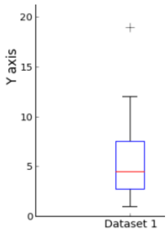
Validation details

Validation is prerequisite to ensure the distribution of quantitative data products and their subsequent application by the user community. Information on the validation process of the provided data should be able to prove the reliability and consistency of satellite-derived products. Pertinent information includes:

Well documented validation protocol used as an example see e.g. for ocean colour (Mélin and Franz 2014 and MarCoast validation protocols). Detailed characteristics of the validation data, i.e. match-up data sets (in case of direct comparison between satellite product and contemporaneous and co-located in-situ measurements of the same quantity). Use of existing database (e.g. AERONET, Zibordi et al. 2006) and ICES for the chlorophyll-a. Uncertainties associated with field observations in case these are given (e.g. ICES). The data used for validation, its temporal and spatial coverage must be described and the validation procedure must be described. The validation must concern the Baltic Sea region. Validation metrics/statistics should be given or referred to accepted publication (e.g. Table 1, number of match-ups, scatter and systematic difference or bias between the distributions, rms).

Table 1. Table of statistical measures used to describe EO validation. Notations: n = number of observations, $\bar{X}$ = mean of variable X, σX = standard deviation of variable X, X = independent (*in situ*) data,

Y = dependent (EO) data, E = Y - X = Error. References L09 = Lehmann et al. 2009, A07 = Allen et al. 2007. Table continued on next page. Adopted from EU/FP7-project CoBiOS deliverable 5.3&5.7.

Statistical measure	**Formula**	**Ref.**
Descriptive statistics		
Maps of dependent and independent data or and/or time series plots		
Frequency distributions or boxplots	 or 	
Scale	Linear	
Geometric mean (as a tribute to log-normal distributions)	$\left(\prod_{i=1}^{n} a_i\right)^{1/n} = \sqrt[n]{a_1 a_2 \cdots a_n}$	
Outlier detection		
Statistical outlier detection (alternatively by expert judgement)	Outlier range: < P25 – 3 * (P75 – P25) or > P75 + 3 * (P75 – P25), where P25 and P75 are the 25th and 75th percentile respectively	
Regression and correlation		
Regr. and corr. results	A, b, r, r^2, n, p (single sided)	
Error statistics		
Mean absolute error	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert Y_i - X_i \rvert$	L09
Bias	$Bias = \frac{1}{n}\sum_{i=1}^{n} (Y_i - X_i) = \bar{Y} - \bar{X}$	
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum (y_i - x_i)^2}{n}}$	L09
Ratio of standard deviations	$\frac{\sigma_y}{\sigma_x}$	A07[2]
Percentage model bias *i.e.* model – data)/data	$Pbias = \frac{\sum_{i=1}^{n} (Y_i - X_i)}{\sum_{i=1}^{n} X_i}$	A07

Statistical measure	**Formula**	**Ref.**
Median error	50^{th} percentile of the error distribution.	
Model efficiency (Nash Sutcliffe Model Efficiency)	$ME = 1 - \frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$	A07
Skewness of error distribution	$s_0 = \frac{\sqrt{n(n-1)}}{n-2} \frac{\frac{1}{n}\sum_{i=1}^{n}(E_i - \bar{E})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - \bar{E})^2}\right)^3}$	
Cost function (Normalized bias)	$CF = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert Y_i - X_i \rvert}{\sigma_X}$	A07
Target diagram		
1.1 Standardized Bias	$Bias^* \frac{1}{n}\sum(y_i - x_i) / \sigma_x$	
1.2 Standardized unbiased RMSE	$RMSE^* = \sqrt{\frac{\sum((Y - \bar{Y}) - (X - \bar{X}))^2}{n}} / \sigma_x$	
2.1 Standardized Bias (Median)	$Bias^* \frac{1}{n}\sum(y_i - x_i) / Median_x$	
2.2 Standardized unbiased RMSE (Median)	$RMSE^* = \sqrt{\frac{\sum((Y - \bar{Y}) - (X - \bar{X}))^2}{n}} / Median_x$	

References

Allen J. I., Holt J. T., Blackford J., Proctor R. (2007a): Error quantification of a high-resolution coupled hydrodynamic ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM. Journal of Marine Systems 68, 381–404.

Allen J. I., Somerfield P. J., Gilbert F. J. (2007b): Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. Journal of Marine Systems 64, 3–14.

Lehmann M. K., Fennel K., He R. (2009): Statistical validation of a 3-D bio-physical model of the western North Atlantic. Biogeosciences, 6, 1961–1974.

Mélin F., and B.A. Franz (2014). Assessment of satellite ocean colour radiometry and derived geophysical products. In G. Zibordi, G.J. Donlon, and A.C. Parr (eds.) Optical Radiometry for Ocean Climate Measurements. Chap. 6.1 Vol. 47 Experimental Methods in the Physical Sciences. Elsevier Inc.

Zibordi G. Et al. (2006). A network for standardized ocean colour validation measurements. EOS Trans. AGU 87: 293-297.