General Comments
This article provides an interesting insight into the regional dynamics of phytoplankton Chlorophyll at a regional scale and makes good use of a multi-sensor time series to provide new information on phenology. The justification of the dataset and algorithm used appears valid and the data processing is clearly explained. The linking of the chlorophyll dynamics to sea-surface temperature would benefit from a cross correlation analysis between the two time series to allow a numerical support to the statements made. A couple of minor corrections are required to the text (mostly for ease of reading) and a couple of the plots (such as missing legends or the addition of regression lines).

We appreciate that the reviewer liked the article and thank her/him for the suggestions that certainly improved the quality of the paper. The current version of the manuscript takes account of all the comments from the Reviewer. Moreover, the comments from both Reviewers stimulated a thorough and careful re-reading of the entire manuscript. We have more strongly stressed the focus of the paper, which is worth reminding develops in the operational context of CMEMS. Main limitations of this work are now mentioned and discussed in the manuscript.

The entire manuscript went through a general re-editing:

- *The abstract and Introduction have been nearly totally rewritten.*
- *Section Data (now Data and methods) now includes a new section with the description of the statistical analysis (formerly present at the beginning of the results).*
- *Similarly, the structure of section Results (now Results and discussion) has been improved by including a forth subsection (Algorithm regional calibration) to better ease the reading.*
- *Conclusions should now read less cumbersomely as more general statements did replace lots of the many summarizing and purely technical and repetitive sentences.*

All figures have been edited.

- *Figure 1 (formerly showing the matchup spatial distribution only) now includes the space-temporal matchup distribution along with the in situ CHL frequency distribution.*
- *Figures 2 and 3 about the scatterplot of satellite versus in-situ CHL now include a qualitative colour legend and regression lines for each plot.*
- *Figure 4 now includes a panel to show more clearly the variability and robustness of the regression lines obtained within the bootstrapping exercise.*
- *In Figure 5 the zero is now part of one bin of the histogram, while formerly was the boundary between two bins.*
- *Former figure 6 on the full time series of CHL has been removed because the discussion was and still is only focused on the climatological values (current figure 6) and the supplementary material already contained detailed figures of each CHL yearly time series. On the other hand former figure 8 on the full time series of SST have been moved to the supplementary material. Both figures 6 and 8 were not adding any specific information while potentially distracting the reader.*
- *A brand new figure 7 has been added to show that the adopted algorithms can also be used to monitor the space-time variability of the cyanobacteria in the Central Baltic, consistent with previous findings (Kahru and Elmgren, 2014).*
- *In figure 8 only the years with corresponding SST data are shown.*

The specific issues raised by the Reviewer are addressed below (in italic).

Specific Comments
page 2289: line 15: The authors state that there are “many more” invalid data in the MLP product. It might be of use to give a mean % valid pixels difference for s sample of representative images so that readers can quantitatively consider the difference.

On average, for all the time series, MLP had around 15 % more flagged pixels that OC4 and OC5. This has been specified in the manuscript.

pg2290 lines 5-20: There is no information provided on the methodology of the in situ chlorophyll a measurements. Were the data used from the database measured using HPLC, extracted fluorometric methods, both?

Relevant info has been added to the section, including the non-fully traceability of the methods used by all data providers to data repositories: data are included into the repository regardless whether measurement protocols are specified.

pg2291 line 20: The authors state that they discard outliers, where outliers are defined as possessing an error greater than a set criteria (ratio). I understand the need to remove outliers, especially in log normally distributed data, but are the points discarded evenly distributed across the range of the dataset? The distribution of the 'very poor' data is of interest in itself. Also, the authors state that 3.5% of the values are discarded using the ratio threshold. One assumes that the approximately normal distribution of satellite and in situ measurements also leads to a somewhat normal distribution of errors. Would it be possible to state that points where errors were outside the range 2nd-98th percentile of all errors were removed as outliers? If not then could the authors justify the error ratio chosen to dictate outliers?

As explicitly mentioned in the main text, "The discarded data were evenly distributed over the entire range of CHL variability and without any specific temporal or spatial patterns". Moreover, the criterion implemented in this work discards (3.5%) as many outliers as if the other criterion (2nd < X < 98th) was chosen (4% by definition). However, our method has the advantage of ensuring the analysis to be conducted over exactly the same datasets, thus allowing proper statistical inter-comparison among algorithms.

pg 2295 line 12: The authors state that the horizontal-averaged CHL for OC4v6corr were computed for images with a minimum of 1000 valid pixels. I assume that this is referring to the mean CHL across the entire basin? If so then it might be worth stating the total number of pixels that make up the region of study. This would then allow a statement along the lines of "Horizontal-averaged (whole Baltic) CHL for OC4v6corr were computed only for images with at least 1000 valid pixels (x% of basin observed)." Obviously observing the whole basin in a single day is near extremely unlikely but it would be good to know if 1000 pixels corresponds to 10%, 20%, 30% etc when considering the following phenological discussion.

Relevant info has now been added to the main text to incorporate the Reviewer comment. In this respect, the whole Baltic has 21424 pixels. Gulf of Bothnia has 5750, Skagerrak and Kattegat have 2625 and the Central Baltic has 13049. 1000 pixels correspond to 5 %, 17 %, 38 % and 7 % of their respective surfaces. It is a statistical property that the expected error of the horizontal average depends on the absolute number of pixels, not on the relative. For this reason, we decided to establish the minimum 1000 pixels equally for all occasions when horizontal averages are computed.

pg 2297: lines 2-8: The authors state that there is a relationship between SST anomaly and CHL anomaly time series. It would be worth performing a brief cross correlation analysis of the two time series such that the link could be quantified beyond "high-amplitude temperature anomalies induce similar growth and decay in CHL.". This may also provide information of the lag of a system response to forcing factors, which would be of interest to those cited in the discussion of lines 9-29.

The coefficient of determination R2 of such cross correlation has been added. Temporally, the range has been restricted between the days 150 and 250 (May 30th and September 7th) to focus onto the summer cyanobacterial bloom that is supposed to have a tighter link to temperature. As it is now clearly mentioned, the scope of this work is to assess the best performing algorithm among the many available for operational use. The lagged correlation analysis would surely be of interest to better highlight the link between SST and CHL, and in this respect we added some text and reference on both the complexity of the Baltic ecosystem dynamics and on the fact that "In the specific context of this cross-correlation analysis, we are implicitly

assuming that both SST and CHL respond to the calm weather conditions with the same time lag". Moreover, according to Kahru et al. (1993), cyanobacteria can heat the water column from their absorbed light, but of course cyanobacteria also respond to temperature variations, so the final observed cause-effect relationship might be not obvious.

Technical Comments

pg2283 Title of article: The authors should hyphenate 'multi-sensor' as both terms apply to the following noun.

Corrected.

pg 2284 line 2: "Fifteen-year" should be "A fifteen-year"

Corrected.

line 8: Authors should expand the sentence "Statistics showed low linearity". Is this for all the provinces, individually and combined? Is this for all algorithms?

We changed "Statistics showed low linearity" with "In general, statistics showed low linearity" to better refer to the general outcome of the whole matchup analysis.

pg 2285 lines 2-6: A very long sentence. Perhaps best to break after Øresund? Change "Great Belt and Øresund, thus leaving the Skagerrak..." to "Great Belt and Øresund. This leaves the Skagerrak..."

Introduction significantly changed and this sentence has been removed.

line 18: It would be more clear if "Thus statistics should be consequently calculated." were changed to "Therefore the statistical assessment of algorithm performance should be performed on the area as a whole."

Introduction significantly changed and the concept expressed by this sentence has been rephrased and moved later into the introduction to better match the new version.

pg2286 line 4: "to cite a few" seems superfluous.

Removed.

line 27: perhaps include chl in the units for "1 mg m-3" otherwise the value could technically refer to concentrations of some other optical component.

Within the new version of the Introduction this bit was totally rephrased.

pg2287 line 4: "Case 2" to "Case II" for consistency with cited literature.

Corrected.

pg 2288 line 4: change "remains up to date" to "is currently"

Removed.

pg 2292 lines 3-4: change "that it can be derived" to "that can be derived"

Corrected.

line 9: change "BIAS respect to the..." to "BIAS with respect to the..."

Corrected.

line 27-28: It might be more clear to state that “In each region, OC4v6 overestimates CHL by > 40% on average.” as “In all cases, OC4v6 overestimates CHL more than 40 %.” could be interpreted to mean that for OC4v6 overestimates CHL by > 40% for all data points.

Corrected.

pg 2295 line 19: Change “reflect this phenomenon (Fig. 3a). Causes could be” to “reflect this phenomenon (Fig. 3a), possibly due to”

This entire sentence has been removed.

line 22: Change “are excessively risen” to “are excessively raised”

This entire sentence has been removed.

line 25: Possibly change “eventual coccolithophore” to something like “occasional coccolithophore” or “annual coccolithophore” to reflect the actual frequency of coccolithophore blooms in this region.

This sentence has been moved to the conclusions.

line 28: Might be clearer to change “However, few spikes in the time series...” to “Additionally, a few spikes inthe whole-Baltic time series...”.

This sentence has been removed.

pg 2298 line 5: Change “Fifteen years-long merged multi sensor daily CHL data” to “A fifteen-year merged-multi-sensor-daily dataset of CHL”

Corrected.

line 14: Change “the dynamics was similar as in the Central Baltic” to “the dynamics were similar to the Central Baltic”

This sentence has been rephrased.

lines 16-20: These would be supported by metrics from a cross correlation analysis.

Done.

line 24: It is not just a higher R2 that would be preferable but a smaller RMS and data collection over a full range of regional conditions to avoid the situation seen at high CHL in the Skagerrak and Kattegat region.

This sentence has been rephrased.

pg 2306 Figure 2: This figure has a colour scale for density of points but there is no legend to show what the colour scale range. Also, it may be of help to the reader to show the linear regression on the plots in addition to the 1:1 line (as is done in figure 3) so that one can get a better understanding of the regression parameters at a glance.

A qualitative colour legend indicating this low-to-high density scale has been added to both Fig. 2 and Fig. 3. Here, what really matters is the density of the matchup population within the plot, considering that a quantitative estimation of the point distribution would not add any particular crucial information.

Regression lines have been added in all scatterplots of the Matchups and Validation sections.

pg2307 Figure 3: The caption states that the line of equal value is dashed and black, but in the plot it is solid and black. Either the text or the plot require correction. This plot also requires the addition of a colour scale legend.

Done.

Remote sensing of chlorophyll in the Baltic Sea at basin scale from 1997 to 2012 using merged multi-sensor data

Jaime Pitarch[(1),*], *Gianluca Volpe*[(1)], *Simone Colella*[(1)], *Hajo Krasemann*[(2)], *Rosalia Santoleri*[(1)]

(1) Institute for Climate and Atmospheric Sciences, Italian National Research Council. Via del Fosso del Cavaliere 100, I-00133 Rome, Italy.

(2) Helmholtz-Zentrum Geesthacht, Centre for Materials and Coastal Research GmbH. Max-Planck-Strasse 1, D-21502 Geesthacht, Germany.

*Corresponding author: jaime.pitarchportero@artov.isac.cnr.it, phone +39 06 4993 4313

Keywords: Baltic Sea; chlorophyll; remote sensing; ocean colour; multisensor; algorithms; in-situ data; calibration; validation; time series; phytoplankton phenology

Abstract

A fifteen-year (1997-2012) time series of chlorophyll-a (CHL) in the Baltic Sea, based on merged multi-sensor satellite data were analysed. Several available CHL algorithms were sea-truthed against the largest in-situ publicly available CHL dataset ever used for calibration and validation over the Baltic region. To account for the known biogeochemical heterogeneity of the Baltic, matchups were calculated for three separate areas: (1) the Skagerrak and Kattegat, (2) the Central Baltic, including the Baltic Proper and the gulfs of Riga and Finland, and (3) the Gulf of Bothnia. Similarly, within the operational context of the Copernicus Marine Environment Monitoring Service (CMEMS) the three areas were also considered as a whole in the analysis. In general, statistics showed low linearity. However, a bootstrapping-like assessment did provide the means for removing the bias from the satellite observations, which were then used to compute basin average time series. Resulting climatologies confirmed the three regions to display completely different CHL seasonal dynamics. The Gulf of Bothnia displays a single CHL peak during spring, whereas in the Skagerrak and Kattegat the dynamics is less regular and made of highs and lows during winter towards a small bloom in spring and a minimum in summer. In the Central Baltic, CHL follows a dynamics of a mild spring bloom followed by a much stronger bloom in summer. Surface temperature data are able to explain a variable (with years) fraction of the intensity of the summer bloom, in the Central Baltic.

1. Introduction

Global to regional monitoring of the surface ocean is believed to be an essential element for the sustainability of the ocean resources. In Europe, the Ocean Colour (OC) Thematic Assembly Centre (TAC) is the entity devoted to produce and provide ocean colour remote sensing data; and this is performed in the context of the Copernicus Marine Environment Monitoring Service (CMEMS). OC data are currently provided at both global and regional scales. These two scales refer to both the geographical limits and the algorithms used to process the data. The OCTAC is thus meant to provide an added value by not only zooming the data from the global domain to the single regional European seas, but also and especially for the application of tailored *ad hoc* regional algorithms for chlorophyll (CHL) retrieval. The present work aims at assessing the performance of existing CHL algorithms for operational applications over the Baltic Sea. CHL is routinely measured over the world oceans with two main kinds of algorithms: i) those using the blue-to-green reflectance ratio (e.g., empirical) and ii) the semi-analytical, e.g., those using the inherent optical properties to infer the chlorophyll concentration. The former builds on the common experience that water colour spans from blue to green as CHL increases, in open ocean (Case I waters). The latter are mathematically more complex and thus based on a larger number of assumptions; nevertheless, they are believed to be more suited for optically complex waters (known as Case II waters) where the colour of the ocean is determined by several non-covarying constituents, such as CHL, Coloured Dissolved Organic Matter (CDOM) and non-algal particles. Both types of algorithms are very sensitive to the in-situ observations used to calibrate them, thus providing the motivation of the regionalization approach adopted within CMEMS. Those based on neural network constitute a third kind of algorithms for CHL retrieval whose limitations are, in theory, the same as the first two: strong dependency on the training datasets that limit their overall applicability. Here, all three kinds of algorithms are tested.

55 The Baltic Sea is a semi-enclosed basin bordering the North Sea in correspondence of the Danish
56 archipelago. Skagerrak and Kattegat are generally not associated with the Baltic Sea. However, the
57 Baltic domain that is defined within CMEMS extends the eastern limit to the meridian 9.24 °E, thus
58 including most of the Skagerrak and the entire Kattegat basins. The Baltic is characterized by
59 significant CDOM concentration due to high river runoff. It is known that high CDOM
60 concentration reduces the water-leaving radiance making the seawater darker (Berthon and
61 Zibordi, 2010), and this constitutes one of the main challenges for ocean colour algorithms to work
62 properly over the Baltic Sea (Mélin and Vantrepotte, 2015). Despite the fact that the Baltic Sea is
63 widely recognized as a challenging test bed for remote sensing, literature on calibration and
64 validation of CHL algorithms is not abundant. Standard algorithms are those provided by the space
65 agencies for global and operational applications. The application of these algorithms to the in-situ
66 Remote Sensing Reflectance (R_{rs}) collected in 707 stations off Poland between 1993 and 2001
67 revealed uncertainties exceeding 100% when the output was compared against collocated CHL
68 measurements (Darecki and Stramski, 2004). Even less encouraging results were obtained when
69 four standard CHL algorithms were applied to Sea-Viewing Wide Field-of-View Sensor (SeaWiFS)
70 images between 2000 and 2001 (HELCOM, 2004). Matchup with 75 CHL profiles across all the
71 Baltic Sea, with predominance of Swedish coastal waters, gave virtually null correlation, with
72 satellite CHL underestimating the in-situ CHL by 180% to 500%, in contradiction with Darecki and
73 Stramski (2004). More recently, the Case II Regional, Boreal, and Eutrophic MERIS processors were
74 applied to images between 2006 and 2009 (Attila et al., 2013). Matchup with 312 stations in the
75 Gulf of Finland and in the central Baltic Sea showed large CHL overestimation. However, when the
76 standard bio-optical relationships of these processors were tuned with the local in-situ CHL, the
77 bias did reduce significantly (Fig. 6 in Attila et al., 2013). The heterogeneity of results combined
78 with the limited spatial and temporal representativeness of the in-situ observations used in the

79 mentioned data comparisons prompts for further investigation. This work aims at filling this gap by
80 using the largest and publicly available in-situ dataset ever used over the Baltic Sea for validation
81 activities.

82 There is an extensive literature on the biogeochemistry of the Baltic Sea and its relation to physics.
83 River outflows release large amounts of organic matter, which sinks to the bottom and lowers the
84 oxygen concentration leading to large amounts of phosphate to be released by the sediment and
85 upwelled through complex mixing processes (Reissmann et al., 2009). In spring, a nutrient-
86 enriched *hypolimnion* and warmer temperatures trigger diatom and dinoflagellate blooms,
87 depleting nitrogen. In summer, nitrogen-fixing cyanobacteria take advantage of the relatively
88 phosphate-rich, calm and warm surface waters, producing another bloom (Reissmann et al.,
89 2009). The Central Baltic Sea is characterized by summer blooms of cyanobacteria that are known
90 to have buoyancy regulation ability (e.g., N. Spumigena and Aphanizomenon sp., Ibelings et al.,
91 1991) and that, under calm conditions, can accumulate at the sea surface (Ploug, 2008).
92 Cyanobacteria blooms are commonly observed in the central Baltic Proper but not in the
93 Skagerrak and Kattegat nor in the Gulf of Bothnia (Wasmund and Uhlig, 2003). Skagerrak and
94 Kattegat are subject to much higher influence from the North Sea, so that the phytoplankton
95 dynamics, here, is expected to be different than that of the Baltic Sea. Thus there is a strong need
96 for the calibration and validation of the proposed algorithms to take account of the complex
97 morphology and biogeochemistry of the basin. Algorithms are then tested in four geographical
98 areas: (1) Skagerrak and Kattegat, (2) the Baltic Proper and the gulfs of Riga and Finland, here
99 referred to as “Central Baltic”, (3) the Gulf of Bothnia, and (4) the entire area (1 to 3).

100 Ocean colour has cloud cover as additional problem, which is particularly high over northern
101 Europe. To increase the spatial coverage of daily products, the International Ocean-Colour

Coordinating Group (IOCCG) recommended the merging of ocean colour data from multiple missions (IOCCG, 2007). At European level, the Climate Change Initiative (CCI) program (www.esa-oceancolour-cci.org) and the Globcolour (www.globcolour.info) project followed this recommendation and reprocessed archived data from various medium-resolution sensors. Here, the CCI-derived R_{rs} are used as input to the CHL algorithms for the comparison exercise (see section 2.1 for their description). One of the limitations of this approach is given by the fact that the CCI does not include any bands in the near-infrared, which are known to be better suited than the blue-green for Case II waters (Odermatt et al., 2012). On the other hand, merged data spans for longer time periods (1997-2012) than any of the individual sensors alone and provide higher coverage on a daily basis.

Applications of remote sensing in the Baltic Sea have been mainly focused on a few main topics: cyanobacteria blooms (Reinart and Kutser, 2006), light penetration (Pierson et al., 2008) and management of various coastal areas (Kratzer et al., 2008). A good overview of such different applications can be found in Siegel and Gerth (2008). Long-term multisensor satellite data were recently used to develop an indicator of surface cyanobacteria accumulation over defined Baltic regions for trend analysis (Kahru et al., 2007;Kahru and Elmgren, 2014). However, long-term phytoplankton dynamics over the entire Baltic region is still lacking, despite the fact that this is required by the European Water Framework Directive for coastal and inland waters and by the Marine Strategy Framework Directive for open ocean waters. In this article, we aim to partially fill this gap by focusing on long-term remote sensing of CHL at basin scale.

2. Data and methods

2.1 Satellite CHL data

Table **1** summarizes the four satellite CHL products evaluated in this article with their respective references.

Acronym	Input	CHL Algorithm	Reference
GLC	Rrs from single sensors	GSM	(Maritorena and Siegel, 2005)
OC4v6	ESA-CCI Rrs	OC4v6	(Werdell, 2010)
OC5	ESA-CCI Rrs	OC5	(Gohin et al., 2002)
MLP	ESA-CCI Rrs	MLP	(D'Alimonte et al., 2011)

Table 1: summary of the algorithms used in the validation analysis with the acronym used in this work along with the required input for each of them. GLC stands for GlobColour, OC4v6 for Ocean Colour four bands algorithm (version 6), OC5 for Ocean Colour five bands, and MLP for Multi-Layer Perceptron.

The GlobColour dataset (GLC hereafter) was developed in the framework of the European Space Agency Data User Element program to support global carbon cycle research. Daily GlobColour data were downloaded from the project web site (www.globcolour.info). Products are obtained by merging MERIS, MODIS, SeaWiFS and VIIRS data. Validation at global scale was carried out by Maritorena et al. (2010). Downloaded data are 2nd reprocessing Level 3 binned images (L3b), having a resolution of 1/24 of degree at the equator (i.e., around 4.63 km) and consisting of the accumulated data of all merged level 2 products, corresponding to periods of one day. Merged data are generated by the GSM model (Maritorena and Siegel, 2005), which also produces the CHL parameter, delivered as product named CHL1. CHL1 parameter is meant to provide best performances over case I waters and thus it is not recommended for use over optically complex waters, but no alternative is given for the Baltic Sea (further details in the Product User Guide, GlobColour, 2015).

142 The ESA Ocean Colour CCI program has the goal to provide stable, long-term, multisensor satellite
143 products. The dataset consists of the merged SeaWiFS, MODIS, and MERIS data, by shifting MODIS
144 and MERIS R_{rs} to the SeaWiFS wavebands, before merging (ESA-OC-CCI, 2014). Data are mapped at
145 4 km resolution and are available through the OC-CCI (www.oceancolour.org) and the CMEMS
146 portals (marine.copernicus.eu). Standard CHL products are global-ocean daily mean sea surface
147 CHL. ESA-CCI retrieves CHL through the application of the OC4v6 algorithm (O'Reilly et al.,
148 2000;Werdell, 2010) to the merged R_{rs}. The dataset available from CMEMS also includes an
149 additional CHL product by applying the OC5 algorithm (Gohin et al., 2002), developed as an
150 adaptation of the OC4 to French Atlantic coastal waters (further details in the Product User
151 Manual, CMEMS, 2015). Calibrated R_{rs} are also available for the application of custom algorithms.
152 We used these R_{rs} to test a Baltic Sea-specific CHL algorithm, available for the SeaWiFS bands,
153 developed by D'Alimonte et al. (2011). This algorithm is based on a Multi-layer perceptron (MLP)
154 and was trained with in-situ R_{rs} and CHL. MLP was only validated with in-situ R_{rs} and CHL
155 (D'Alimonte et al., 2012), thus not taking into account all the known issues linked to the
156 atmospheric correction over the basin.

157 An image pre-analysis revealed ~15 % more flagged (invalid) pixels for MLP than for OC4v6 and
158 OC5, despite all are derived from the same CCI reflectances. The cause is the frequent occurrence
159 of negative $R_{rs}(412)$ most likely due to aerosol optical thickness overestimation in the blue
160 together with high CDOM. In contrast, OC4v6 does not use $R_{rs}(412)$, the most sensible band to the
161 atmospheric correction procedure, thus allowing for problematic pixels (those with $R_{rs}(412)<0$) to
162 be retrieved as well. Similarly, OC5 is insensitive to negative $R_{rs}(412)$ values, thus allowing CHL to
163 be retrieved also under the extreme conditions of atmospheric correction failure.

2.2 In-situ CHL data

We downloaded publicly available in-situ CHL data, contained in the repositories at Seadatanet (www.seadatanet.org), the Baltic Marine Environment Protection Commission (www.helcom.fi) and the NOAA World Ocean Database (www.nodc.noaa.gov/OC5/WOD/pr_wod.html). CHL is computed from bottle samples using standard laboratory techniques. The technique used to collect and measure CHL spans from fluorimetry to spectrophotometry and HPLC. The amount of information provided depends upon each environmental agency or research institution that collected and uploaded the data. For their part, data repositories have additional quality control criteria based on outlier estimation. All data collected in the Baltic region during the period covered by the satellite observations (1997-2012) were merged and duplicates were eliminated.

Since the remote sensing signal can be fairly considered as a weighted average within the first optical depth, in-situ observations must be treated accordingly. In-situ CHL consisted either of a single sub-surface reading or CHL profiles derived from a few depth readings. In this latter case, a proper vertical average of a CHL profile is needed for comparison to remote-sensing data. The vertical weighting function depends on the inherent optical properties (IOPs) that cannot be inferred solely from CHL in case II waters. In rigor, coincident IOP measurements are needed to perform the vertical averaging, but such measurements are scarce and not publicly available. In case I waters, vertical averaging can be performed with the method by Morel and Berthon (1989) with input CHL profile data. The remaining applicable options to our in-situ data were either to select only the sub-surface CHL value or to average the profiles with the method by Morel and Berthon (1989), despite the theoretical inconsistencies. Calculations (not shown) revealed that satellite-in-situ correlations did improve (even if only slightly) if available profiles were vertically averaged (and this is the approach used in this work) instead of taking only the uppermost

reading. To avoid bottom contribution to the water-leaving radiance, only stations with a bottom depth of at least 10 m were selected.

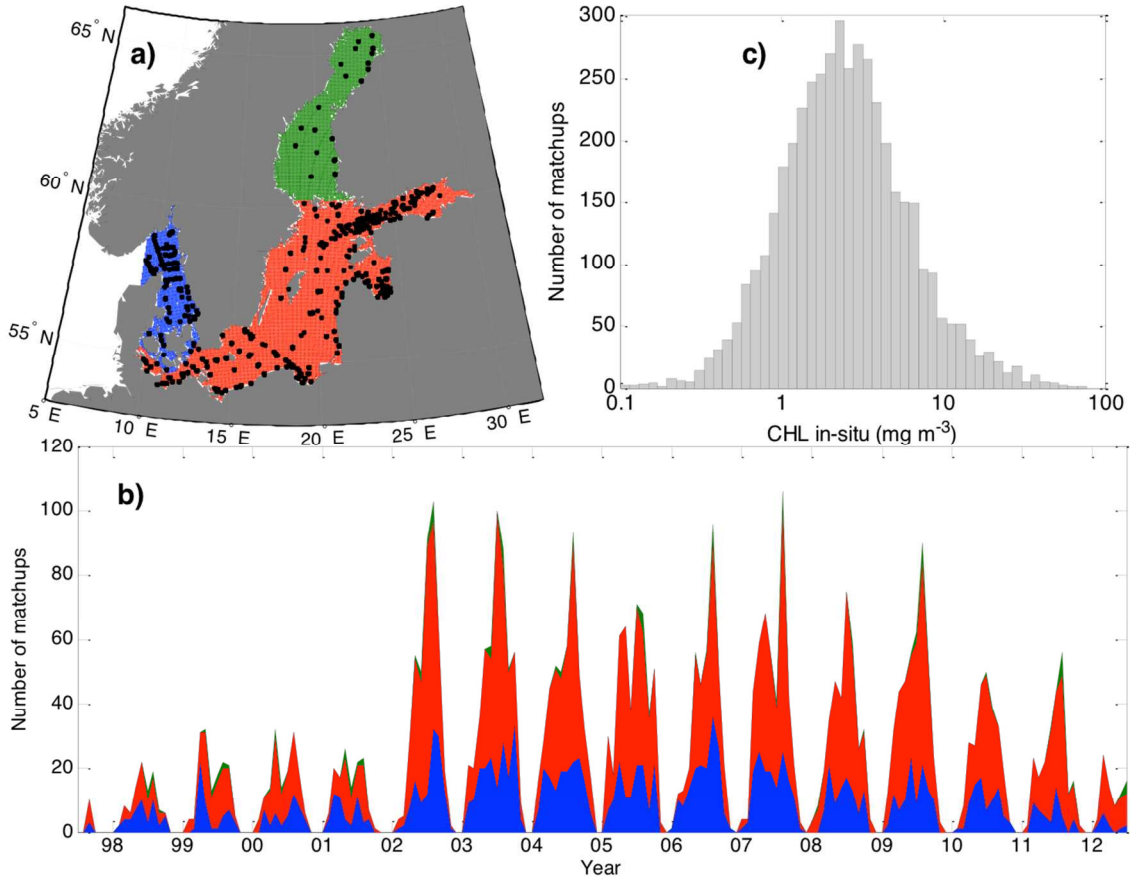


Fig. 1: a) Spatial distribution of the 4492 in-situ stations used in the matchup analysis (see section 3.1) along with the partition of the area of study. Skagerrak and Kattegat is highlighted in blue with 1456 matchup points. Central Baltic is highlighted in red with 2922 matchup points, and the Gulf of Bothnia is green with 114 stations. Temporal station distribution is also shown using the same colour code (b). The frequency distribution of the entire in-situ CHL is shown in panel c.

Similarly, to ensure representativeness of the data in the case of CHL stratification, only stations with the uppermost reading shallower than 2 m were retained for the analysis. The spatial location of matchup stations is shown in Fig. 1a. Although covering the entire Baltic region, data are not uniformly distributed, as the dataset is built from different sources, in which individual institutions

and agencies are interested in specific zones. The number of matchups increases significantly when both MODIS-Aqua and MERIS started to be operational in 2002 (Fig. 1b), thus providing further evidence of the utility of merging different sensors for oceanographic research. The CHL in-situ dataset used in the following of this work is log-normally distributed around the mean value of ~ 2.46 mg m^{-3} and spanning from 0.1 mg m^{-3} to 77 mg m^{-3} (Fig. 1c). Fleming and Kaitala (2006) reported CHL values 7 to 12 mg m^{-3} in the northern Baltic Proper during the spring bloom. Our gathered in-situ matchup dataset during April in the northern Baltic Proper (35 samples) shows CHL to range from 1.39 to 14.7 mg m^{-3}, consistent with these previous findings.

2.3 Statistical evaluation

Satellite CHL was extracted from single pixels without further spatial windowing. To calculate the mean bias and the RMS we applied decimal logarithm-transformation to the CHL data, and returned to percentage linear scale:

$$\mathrm{BIAS} = \left[10^{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)} - 1 \right] \cdot 100 \quad (1)$$

$$\mathrm{RMS} = \left[10^{\frac{1}{N}\sqrt{\sum_{i=1}^{N}(y_i - x_i)^2}} - 1 \right] \cdot 100 \quad (2)$$

where x_i and y_i are the $\log_{10}$-transformed in-situ and satellite CHL, respectively. N is the number of matchups. The best linear fits were found using the log-transformed CHL. The corresponding coefficient of determination (R^2), slope (m) and intercept (n) are also presented. The whole area was divided into regions with expected bio-optical differences (see Fig. 1a). The number of observations available from the Gulf of Bothnia is very limited, so the statistical information that can be derived from the regressions must be interpreted with caution. Nevertheless, results are

219 presented for completeness. The p-value of the regressions was zero for all regions except for the
220 Gulf of Bothnia, where it was $p<10^{-3}$.

221 Outliers were defined as data in which any of the four algorithms gave CHL outside the range
222 within one twentieth and twenty times the in-situ CHL. In applying this criterion, roughly 3.5 % of
223 the data were discarded and led N to become 1873. Most of these discarded matchups were
224 rejected because of the GLC underestimation, together with the high scattering (Fig. 2a). The
225 discarded data were evenly distributed over the entire range of CHL variability and without any
226 specific temporal or spatial patterns. For comparison issues among algorithms, only matchups
227 with coincident valid pixels for all four satellite products within the same day were considered, but
228 once the best performing algorithm was identified, all available matchup stations for this
229 algorithm were used to provide its full record of statistics (N = 4492).

230 # 3. Results and discussion

231 ## 3.1 Matchups

232 In general, satellite and in-situ data show modest agreement in the Baltic. This can be intuitively
233 associated with both the non-full traceability of the methods used to assemble the in-situ dataset
234 and the satellite algorithms. MLP and GLC provide poor R^2 and negative BIAS with respect to the
235 in-situ data. Results of OC4v6 (R^2=0.43) are consistent with findings by Darecki and Stramski
236 (2004). The positive bias of 44 % here (Fig. 2b) is smaller than 119 %, as found by Darecki and
237 Stramski (2004), but still confirms the OC4v6 to overestimate CHL in the Baltic Sea. OC4v6 matches
238 better the in-situ data for high CHL, whereas tends to saturate for low values. OC5 has similar
239 linearity (R^2=0.44) but significantly improves in terms of bias (-14 %) with respect to OC4v6.
240 Besides the similar R^2, we appreciated graphical similarities between the scatter plots of OC4v6

241 and OC5. Guided by this hint, we performed a linear regression in log form between OC4v6 and
242 OC5 satellite derived CHL (not shown). Regression analysis revealed a very high linear dependence
243 (R^2=0.97), although the relationship is more complex in theory (Gohin et al., 2002), and this will
244 have implications for the rest of this work (see below).

245 Geographical partition of the matchup dataset highlighted significant differences in the statistical
246 behaviour of algorithms. For instance, the performance of MLP strongly degrades in Skagerrak and
247 Kattegat (Fig. 2h) with respect to the central Baltic Sea (Fig. 2l). MLP was calibrated with data only
248 inside the Baltic Sea, and not in the Skagerrak and Kattegat (D'Alimonte et al., 2012, Fig. 2d). It
249 appears then that such algorithm design is highly dependent on the calibration data. GLC performs
250 always worst in all regions, and the scatter plots seem like undefined clouds, which is best
251 highlighted by the large RMS errors. OC4v6 displays similar statistics at both sides of the Danish
252 Strait, although the slope of the regression line decreases towards Skagerrak and Kattegat. In each
253 region, OC4v6 overestimates CHL by more than 40 %. The behaviour of OC5 is always in
254 accordance with OC4v6, with a shifted BIAS, given the very high correlation between the two. Due
255 to the much simpler applicability of OC4v6 and its wider diffusion in the community, the following

analysis will be based on the OC4v6.

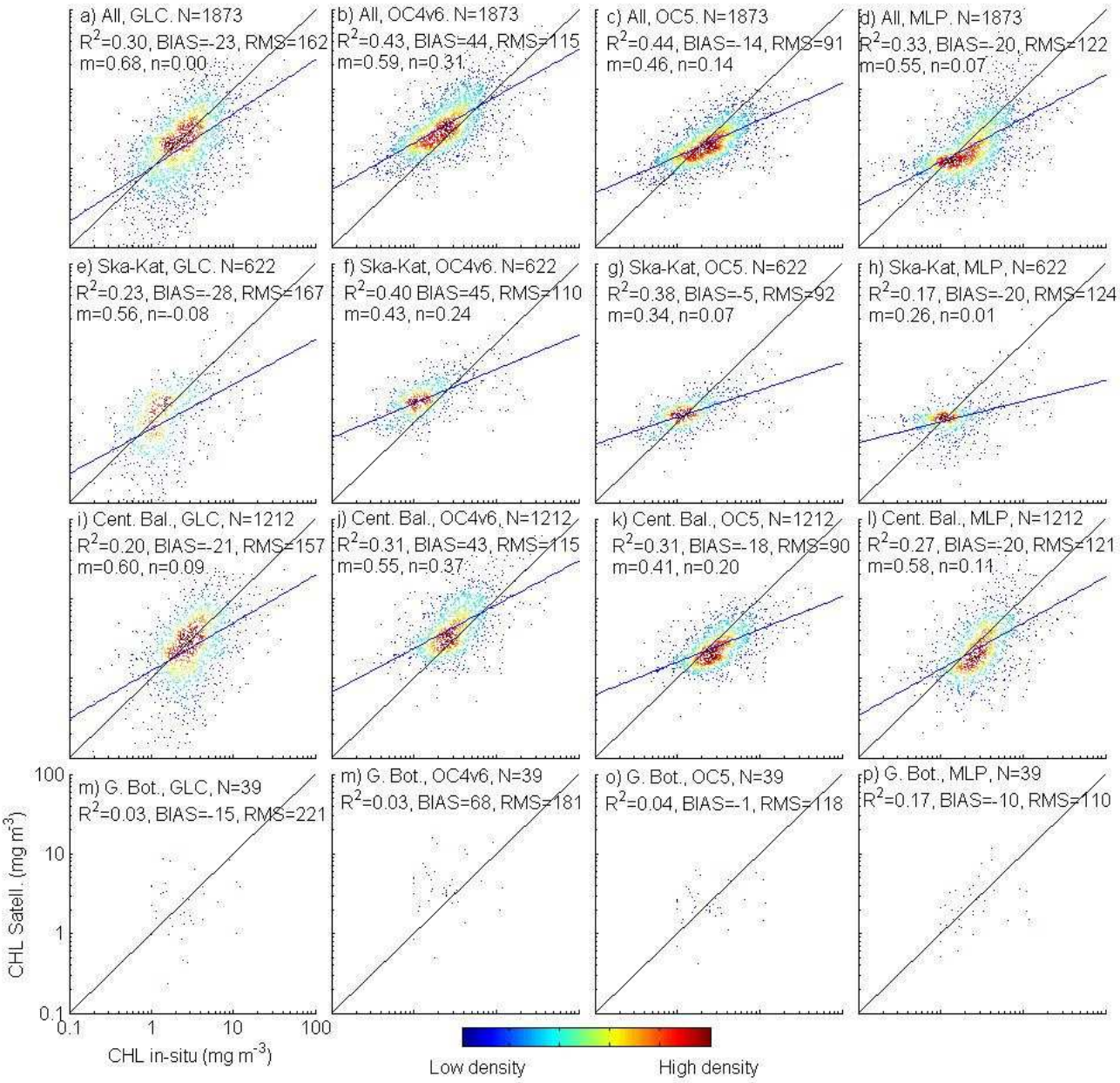


Fig. 2 Density scatter plots of in-situ versus satellite-retrieved CHL for all algorithms providing meaningful values. The line of best fit (blue) and that of equal value (black) are superimposed together with relevant statistics.

The matchup analysis is here repeated with the same conditions, including the definition and removal of the outliers, but now for OC4v6 alone. Only 22 matchups were discarded (< 0.5 % of the data), of which 17 due to overestimation (i.e., higher than twenty times the in-situ

counterpart). As mentioned, when the coincidence with the other algorithms is removed, the number of matchups increases to 4492, distributed as 1456 in Skagerrak and Kattegat, 2922 in the Central Baltic and 114 in the Gulf of Bothnia. Fig. 3 shows the corresponding density scatter plots and statistics. In general, the interpretation from Fig. 2 still holds, with the bigger size of the matchup dataset providing increased confidence level of the derived statistics. However, since the additional data were previously discarded (not used to produce Fig. 2), it is not surprising that the latter statistics did degrade (R^2 = 0.43, BIAS = 72%, RMSE = 151%, m = 0.57, n = 0.41, N = 2619). The orders of magnitude of the uncertainties found here (Fig. 3) are in line with those available from the literature (Darecki and Stramski, 2004) even considering the wider space and time distribution of the data (both in-situ and satellite) used here.

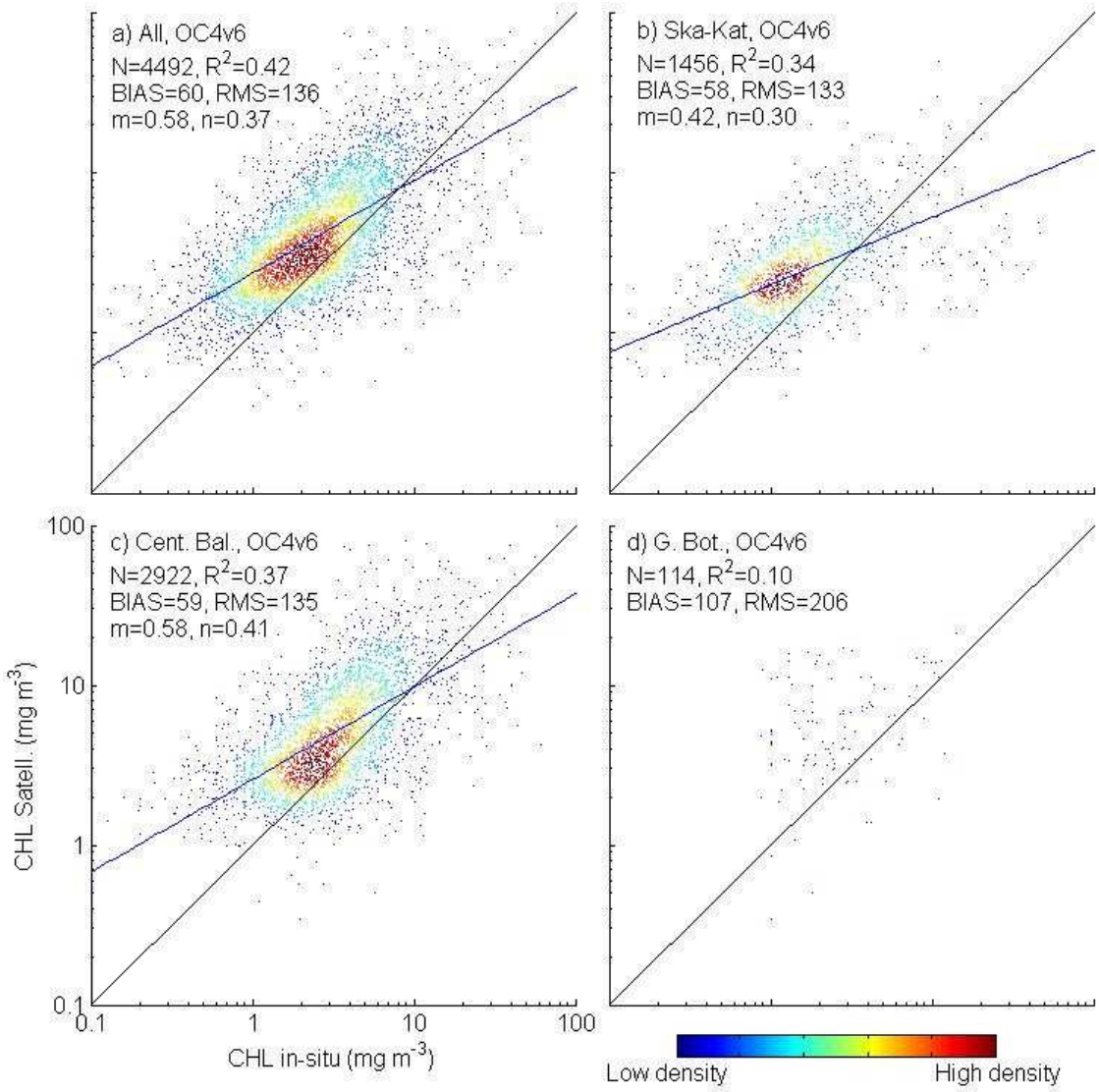
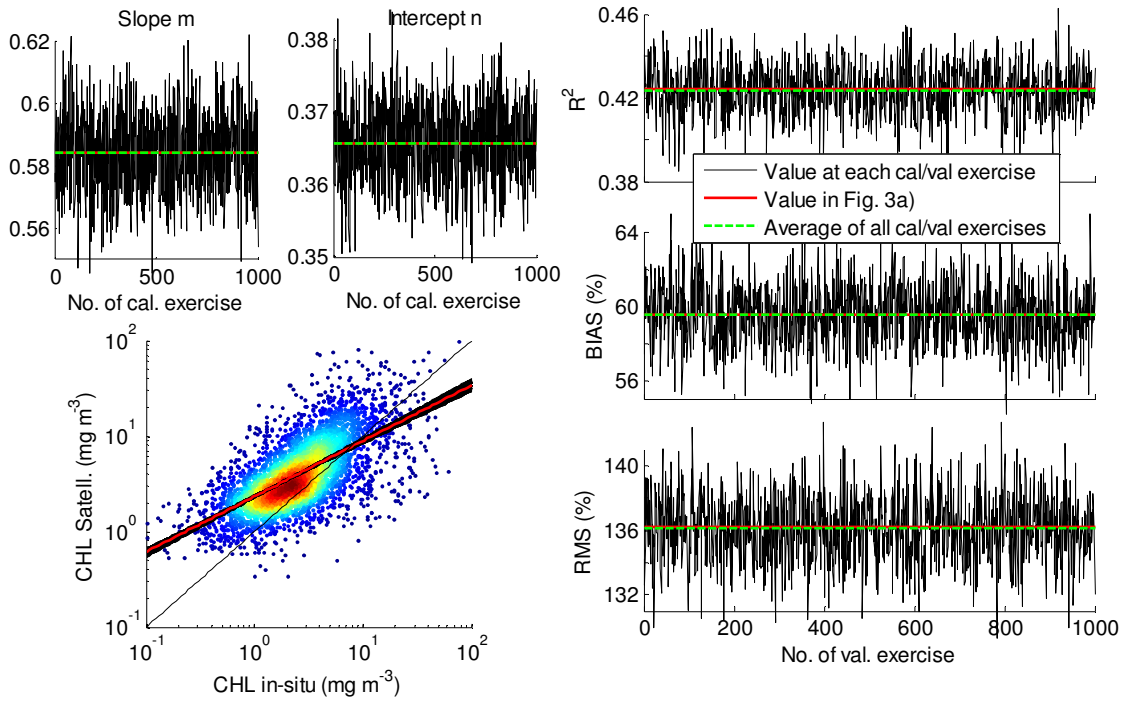


Fig. 3 Density scatter plots of in-situ versus satellite-retrieved CHL for OC4v6 algorithm. The best linear regression (blue) and the line of equal value (black) are superimposed along with relevant statistics.

3.2 Validation

When the regression coefficients are used to re-calibrate existing algorithms, the validity and robustness of the matchup statistics needs to be validated against independent data. Starting from the matchups for OC4v6 alone (Fig. 3a), we performed a sensitivity study to test the dataset homogeneity by a bootstrapping-like assessment (Efron, 1979) as used in recent validation

283 exercises (Brewin et al., 2013). The whole dataset (N = 4492) was partitioned a thousand times
284 into two randomly chosen halves: calibration (N_{cal} = 2246) and validation (N_{val} = 2246). Each
285 calibration dataset is used to compute the linear regression coefficients (m,n) which are then
286 applied to the corresponding complementary validation half to compute the associated statistics.
287 The obtained series of coefficients and statistics are shown in Fig. 4. Results are remarkably
288 robust: the averages of the regressions found (m=0.5843, n=0.3657, green dashed line in Fig. 4)
289 are almost equal to those when the whole dataset is used (m=0.5845 and n=0.3656, red line in Fig.
290 4). Moreover, the dispersion is very small with the coefficient of variation being 2.07 % when
291 computed over the slopes and 1.38 % over the intercepts. As for the validation statistics, their
292 mean values (graphically shown in green in Fig. 4) R^2 = 0.4236, BIAS = 59.55 %, RMS = 136.13 % are
293 very similar to those obtained for the whole dataset (Fig. 3a, R^2 = 0.4241, BIAS = 59.53 %, RMS =
294 136.19 %).



295

296 Fig. 4 Left up, in black: best linear fits (slope m and intercept n) of 1000 randomly chosen
297 calibration datasets (N_{cal} = 2246, X axes) of $\log_{10}(CHL_{in\text{-}situ})$ versus $\log_{10}(CHL_{OC4v6})$. Left down:

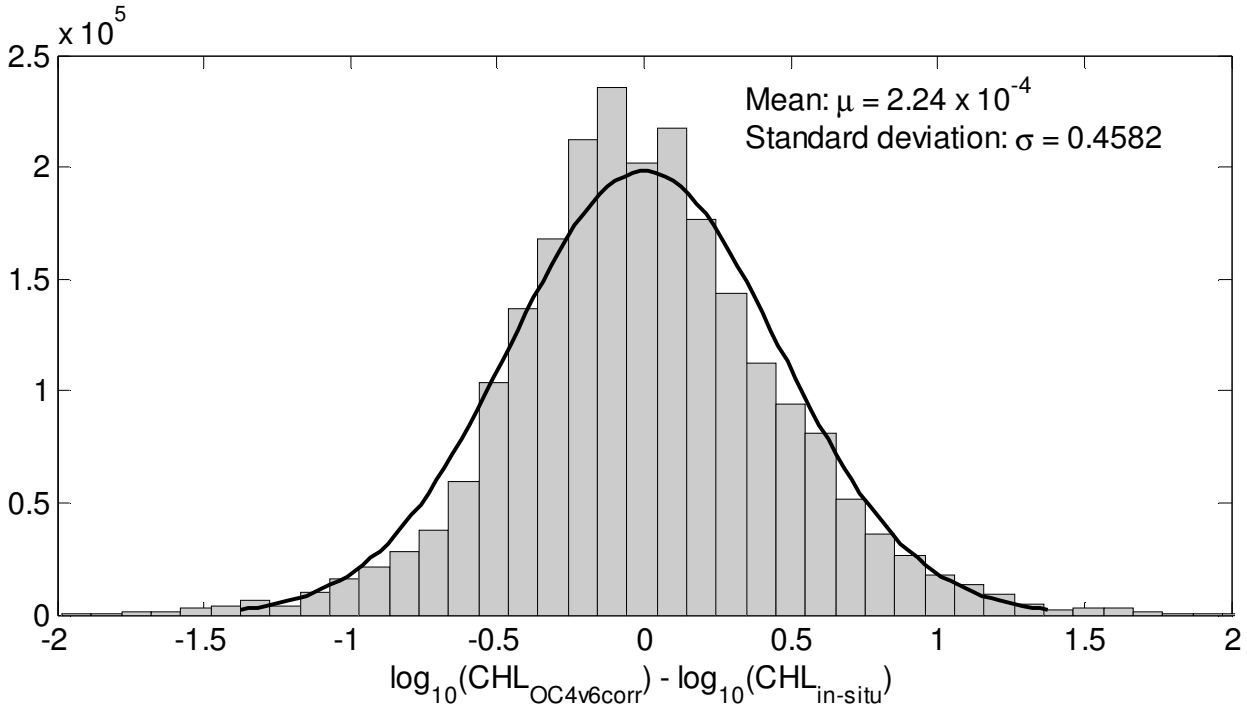
application of all 1000 (m,n) pairs to the OC4v6 vs. in-situ scatter cloud. In red, slope and intercept for the whole dataset, as shown in Fig. 3a. In green, average of the 1000 calibration results. Right, in black: statistics when applying each m and n pairs from the left side to the complementary validation datasets (N_{val} = 2246, X axes). These are: coefficient of determination, BIAS (eq (1) and RMS (eq. (2)). In red: same statistics found for the whole dataset, as shown in Fig. 3a. In green, average of the 1000 validation results.

3.3 Algorithm regional calibration

Efficient and useful algorithm regionalization needs appropriate bio-optical in-situ data. Unfortunately, the Baltic lacks of such publicly available in-situ dataset that therefore prevents a canonical regionalization. This, together with the high confidence level associated with the described statistics and in view of obtaining an unbiased proxy for CHL, with the available data, prompt for using the computed coefficients (m and n in Fig. 4) for recalibrating the OC4v6, as follows:

$$\log_{10}(\mathrm{CHL}_{\mathrm{OC4v6corr}}) = \frac{\log_{10}(\mathrm{CHL}_{\mathrm{OC4v6}}) - n}{m} \quad (3)$$

Errors between eq. (3) and the complementary in-situ validation matchups were calculated. Each of the 1000 chosen combinations generated a vector of errors with length N_{val} = 2246. Their accumulation led to a total of 2246000 error estimates, whose distribution is shown in Fig. 5, together with the fitted Gaussian curve. The recalibration via eq. (3) removed the bias, resulting in a zero-centred error distribution. It is worth reminding that within the calibration and validation exercises the two datasets are independent. The standard deviation (σ = 0.4582) includes all errors not taken into account by the system, i.e. atmospheric noise, errors in the in-situ measurements and, most of all, the limited suitability of algorithms as the OC4v6.



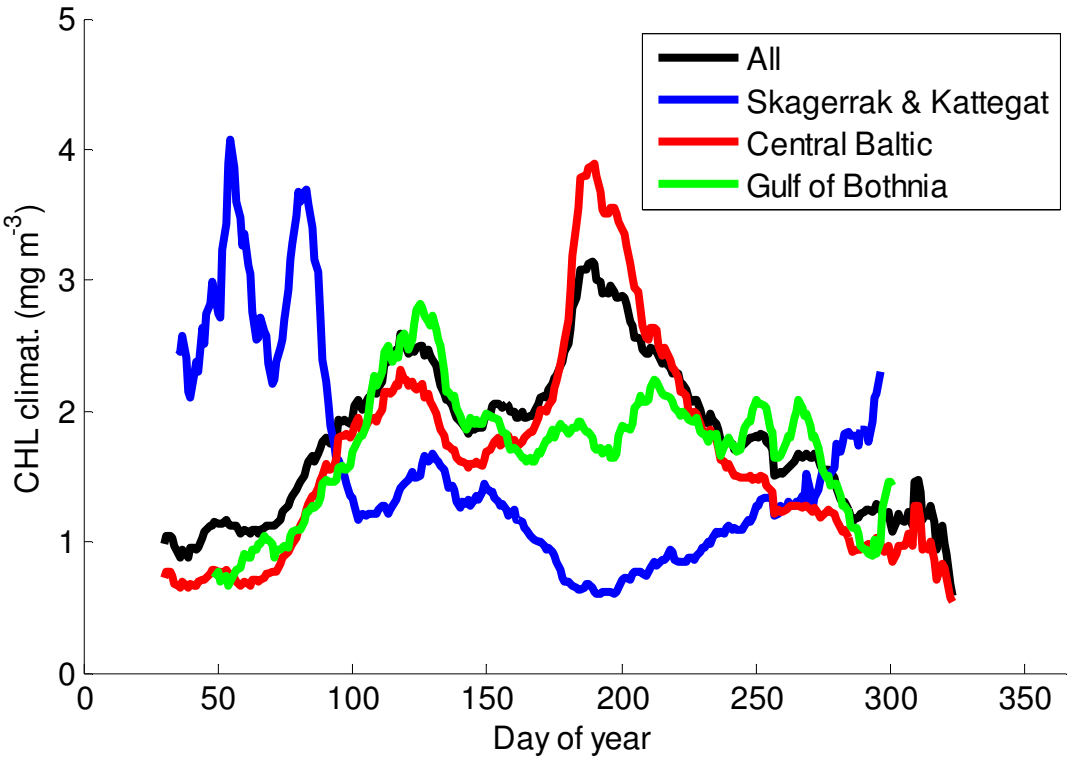
320

321 Fig. 5 Histogram of the absolute error between $OC4v6_{corr}$ and in-situ CHL, both in logarithmic form.
322 Associated mean and standard deviation are also shown and used to compute relevant fitted
323 Gaussian distribution (black line).

324 The symmetric and zero-centred error distribution (Fig. 5) obtained with the application of eq. (3)
325 within the bootstrapping-like assessment warrants a high level of confidence when basin averages
326 are calculated; all the errors at the level of individual pixels are expected to cancel out when a
327 horizontal (pixel-wise) average is performed over a large region. Although the former statement
328 implies that the statistical properties of the matchup dataset can be extrapolated to the whole
329 Baltic area, the good spatial and temporal coverage of the former (see Fig. 1) is a good asset to
330 support this argument. From this point, we defined the algorithm $OC4v6_{corr}$ through eq. (3), with
331 the coefficients (m = 0.5884, n = 0.3751) of Fig. 3a. This enabled the bias to be removed.
332 Nevertheless, RMS was altered, rising up to 187 %, in agreement with $\sigma = 0.4582$ in Fig. 5 through
333 eq. (2). The mathematical explanation of the latter relationship is that the RMS and the standard
334 deviation of a zero-mean distribution are equal.

Among all regions in which the Baltic Area has been divided, Fig. 3 highlights different best linear fits. Given the coefficients of variation 2.07 % and 1.38 % for the slope and intercept respectively found in the bootstrapping assessment, the coefficients in Fig. 3 are significantly different. If OC4v6 is linearly adjusted with eq. (3), the coefficients must be different for each region, in particular, equal to those found in Fig. 3. Therefore, for Skagerrak and Kattegat, they were set to 0.4212 and 0.3027, respectively for m and n. Due to the lack of enough data, the stations in the Gulf of Bothnia were aggregated to those of the Central Baltic. Resulting statistics for these two regions were almost equal to those of the Central Baltic alone: R^2 = 0.35, BIAS = 60.45 %, RMS = 138.64 %, m = 0.5632, and n = 0.4206. These linear coefficients were applied to recalibrate OC4v6 for the Central Baltic and the Gulf of Bothnia. Even if the same algorithm was used results are presented separately for the two basins.

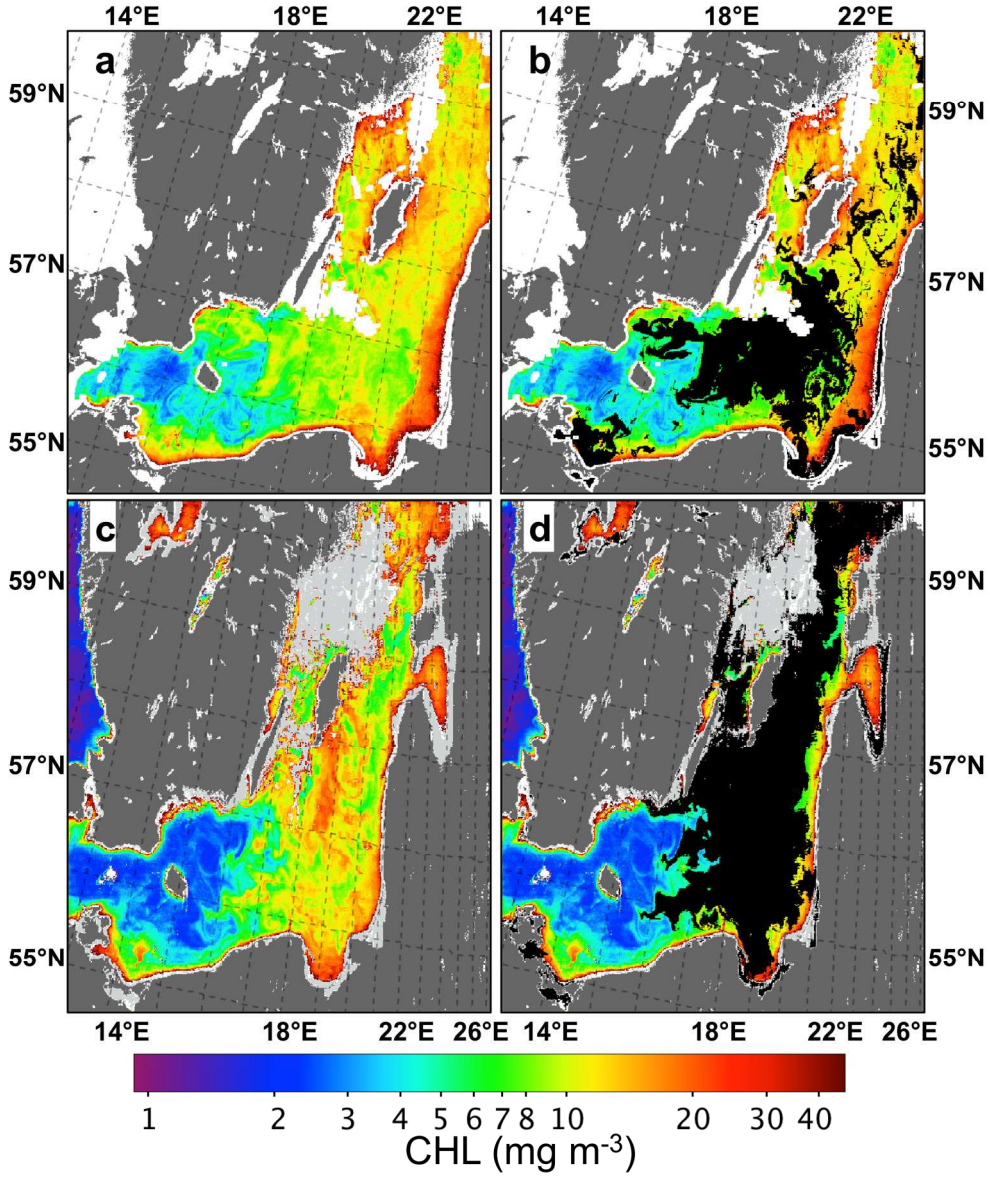
3.4 Satellite derived basin averages



348 Fig. 6 CHL daily Climatology. For any given day of the year, the average was computed only if data
349 for a minimum of six years were available. Plots of individual time series with their associated
350 standard deviation bars can be found in the supplementary material. To improve the plot
351 readability, all time series were smoothed with a one-week moving average.

352 Horizontally-averaged CHL for $OC4v6_{corr}$ were computed only for images with a minimum number
353 of 1000 valid pixels. The entire Baltic has 21424 pixels, with the Gulf of Bothnia contributing with
354 5750 pixels, Skagerrak and Kattegat with 2625 pixels and the Central Baltic with 13049 pixels. One
355 thousand pixels correspond to 5 %, 17 %, 38 % and 7 % of their respective surfaces. CHL dynamics
356 strongly varies among regions at both seasonal (Fig. 6) and interannual time scales (supplementary
357 material). In Skagerrak and Kattegat, the dynamics consists of intermittent growth periods in late
358 winter (up to ~ 4 mg m^{-3}) and a small bloom in spring, reaching a minimum in summer (~ 0.5 mg m^{-3}), consistent with other works (Edelvang et al., 2005) . In the Gulf of Bothnia, the overall range of
360 CHL variability is limited to ~ 2 mg m^{-3} (0.7 – 2.8 mg m^{-3}) with minima in winter and a series of
361 bloom-like pulses from spring to fall. The spring bloom is the most intense and lasts longer than
362 the others (Carstensen et al., 2015). Given the prolonged winter darkness, the length of this data
363 time series is shorter than those from the other regions. Moreover, in winter the Gulf of Bothnia is
364 normally ice-covered and some ice remains in the northern part until May, thus not the entire
365 domain contributed to the displayed CHL. A non-trivial point is that this time series has to be
366 interpreted with caution due to lack of a significant number of data for specific calibration in this
367 area. In the Central Baltic, the dynamics is completely different. Two distinct CHL maxima are
368 appreciable (Reissmann *et al*., 2009): the first one peaks at the end of April, reaching ~ 2.5 mg m^{-3},
369 which is at the lower end of the variability previously observed by Schneider et al. (2006); the
370 intensity of the second peak, in mid-July, (~ 4.6 mg m^{-3}) is consistent with previous observations in
371 the area (Schneider et al., 2006), and from which it steadily decreases and reaches a minimum in

372 winter. The dynamics of the entire domain (black line in Fig. 6) is clearly dominated by the Central
373 Baltic due to its major weight in terms of area coverage. The summer bloom that occurs in the
374 Central Baltic is known to be due to cyanobacteria taking advantage of the milder weather
375 conditions and of the increased water temperature. As cyanobacteria can form surface scum, it is
376 worth questioning whether such data would be masked during the operational image processing.
377 A previously documented mild cyanobacteria bloom on the 11th of July, 2010 was visible from
378 space via qualitative RGB image, and for which surface accumulation was not reported (SMHI,
379 2010). To assess whether the standard processing is able to provide reliable observations also in
380 these conditions, MODIS-Aqua Level-1A was downloaded and processed to L2 using the same
381 settings used to produce the CCI input data. Fig. 7a shows the Central Baltic blooming also in the
382 areas identified as cyanobacteria by the SeaDAS Level-2 flag TURBIDW (Fig. 7b) used to
383 discriminate the accumulation of cyanobacteria (Kahru and Elmgren, 2014). During summer 2005,
384 the Baltic experienced the second largest cyanobacteria bloom (Kahru and Elmgren, 2014) that
385 covered 25% of the entire domain (183000 km^2). As for the 2010 bloom and apart from the small
386 area classified as too bright in the north Baltic Proper (in light grey in Fig. 7c and 7d), the standard
387 processing demonstrated its ability to provide valid data also under these conditions. Therefore,
388 the data used here appear suitable for the study of phytoplankton dynamics throughout the year,
389 even during cyanobacteria bloom events, during which only a negligible percentage of pixels is
390 affected by atmospheric correction failures (Kahru and Elmgren, 2014).



391

392 Fig. 7: MODIS Level-1A of the 11th of July, 2010 (a and b) and 2005 (c and d) were downloaded
393 from the OBPG website (Ocean Biology Processing Group, oceancolor.gsfc.nasa.gov) and
394 processed to Level-2 using the standard settings within SeaDAS version 7.3 (seadas.gsfc.nasa.gov).
395 Kahru and Elmgren (2014) recently identified the presence of cyanobacteria accumulating on the
396 sea surface using the SeaDAS Level-2 flag TURBIDW ("Turbid water") when the flag MAXAERITER
397 ("Maximum Aerosol Iterations") is turned off within the Level-1 to Level-2 processing. Here, CHL
398 images without (panels a and c) and with (panels b and d) the application of the TURBIDW flag is
399 shown; pixels affected by TURBIDW are coloured black. As mentioned by Kahru and Elmgren

400 (2014), the MAXAERITER flag is, by default, turned on within the NASA standard processing (e.g.,
401 the same used here); light grey area (panels c and d) in the north-western Baltic Proper is
402 perceived by the operational processing as too bright (i.e., masked as MAXAERITER) and as such
403 not processed.

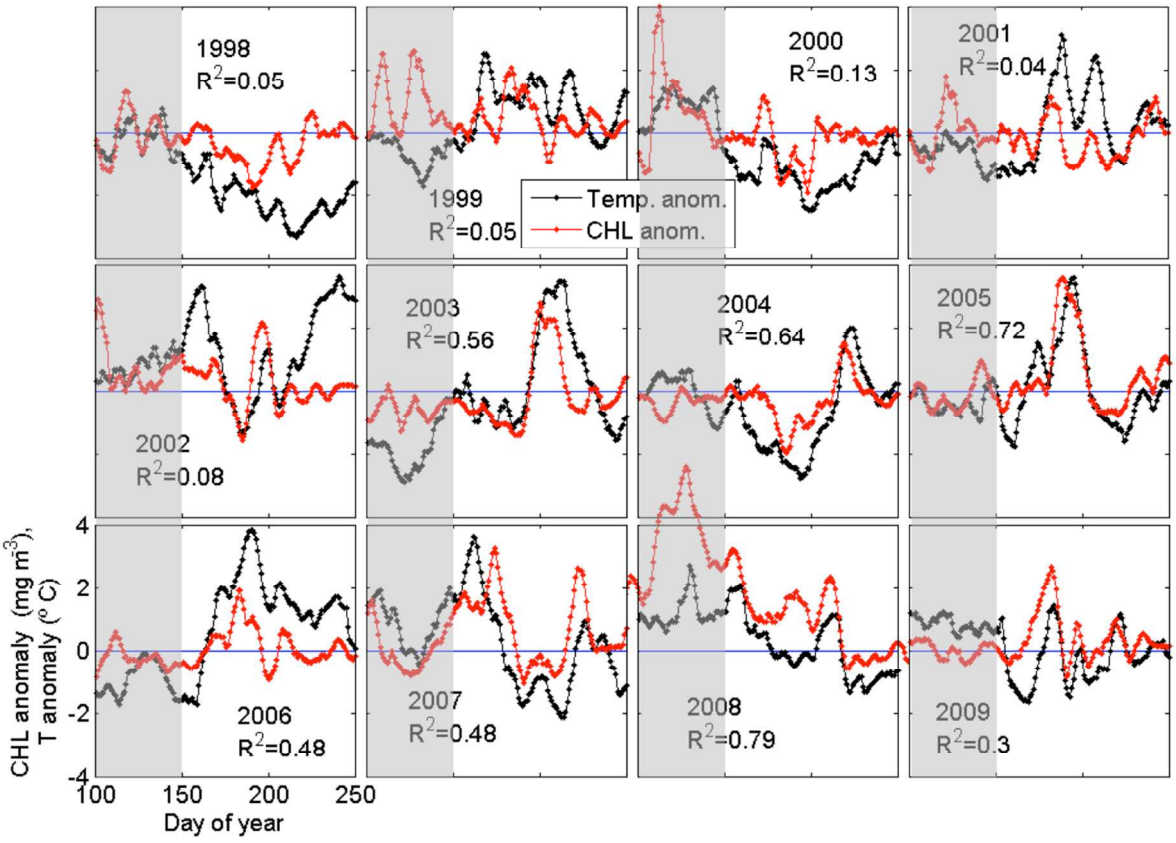
404 Fig. 6 shows that the strongest signal in the Central Baltic is given by the summer bloom.
405 Cyanobacteria-like species are known to bloom under warm and calm weather conditions (Ploug,
406 2008). High sea surface temperature (SST) are known to enhance the growth of cyanobacteria,
407 both directly through higher growth rates, and indirectly by increasing the stability of the water
408 column to allow cyanobacteria to take advantage of their buoyancy regulation ability (Ibelings et
409 al., 1991). Analogously, cyanobacteria were demonstrated to provide positive feedbacks to the
410 surface temperaure by absorbing the incoming radiation (Kahru et al., 1993). It is then reasonable
411 to investigate whether CHL and SST covary over the Central Baltic during summer. In the specific
412 context of this cross-correlation analysis, we are implicitly assuming that both SST and CHL
413 respond to the calm weather conditions with the same time lag. For this matter, daily-averages
414 SST data (1998-2009) over the Baltic Sea were downloaded from the CMEMS website. The SST
415 dataset is the merged product from the sensors AVHRRs (series 7, 9, 11, 14, 16, 17, 18), Envisat
416 ATSR1 and ATSR2, and the AATSR (see CMEMS (2015) for details and Supplementary Material for
417 their basin-average time series). Both CHL and SST data time series were deseasonalized by
418 computing the anomalies with respect to their climatologies, which were used as input for the
419 cross correlation analysis. Fig. 8 shows the two time series anomalies along with correlation values
420 computed over the summer period (between the Julian days 150 and 250) for all years for which
421 SST was available. Prior to the correlation analysis, the CHL anomaly time series was further
422 smoothed with a one-week moving average. Here, the basic underlying assumption is that warm
423 waters, as proxy of calm weather conditions, can explain the dynamics of cyanobacteria. Thus

when cyanobacteria do represent a high fraction (in terms of their space and time presence) of the CHL signal, the correlation is expected to be high, and vice versa.

Fig. 8 shows quite a surprising relationship between both quantities with high-amplitude SST correlating with those of CHL. This related behaviour is somewhat unexpected, because we are comparing here not absolute CHL and temperature, but their differences with respect to their climatological values. Generally, during the second half of the time series, from 2003 on, the correlation appears to be tighter. The causes of the dynamics shown are undoubtedly complex involving considerations on the circulation and the peculiar biogeochemistry of the basin (Reissmann et al., 2009). Nevertheless, this article is focused on the remote sensing aspect and the intensity of the cyanobacteria bloom appears to depend on the timing of the summer temperature peak: although 2004 had a high SST peak, such peak happened late in the season (August 10th), which appeared not favourable for cyanobacteria growth. On the contrary, years 2002, 2003, 2005 and 2006 had SST peaks of similar or lower intensity, but much earlier in the season. Instead, 2001 displayed two marked positive SST anomalies that were only mildly followed by CHL anomalies. Despite the CHL and SST anomalies are poorly correlated during 1998 (Fig. 8), they are both negative suggesting that in that year the cyanobacteria bloom, generally dominating the summer signal in the Central Baltic, was only partially contributing to the overall dynamics. This is clearly documented in Kahru and Elmgren (2014), who found the Fraction of Cyanobacteria Accumulations of only 6%, in 1998; FCA being the ratio of the number of pixels classified as cyanobacteria to the number of cloud-free sea-surface views during the period July to August.

On the other hand, the year 2008 was completely anomalous with respect to both the climatology value and timing of the summer bloom, with a maximum at the beginning of May. This massive and early bloom has already been documented (Majaneva et al., 2012;Larsson et al., 2014), with

447 the dominant species being *Prymnesium polylepis*. Responsible abiotic factors were exceptionally
448 calm and sunny weather during October 2007, resulting in high light availability and low
449 turbulence above the thermocline (Majaneva et al., 2012;Larsson et al., 2014). These conditions
450 enabled *P. polylepis* to build up a considerable biomass. The following winter was the mildest since
451 more than a century, which allowed *P. polylepis* to persist throughout the winter. Improving
452 weather and plenty of nutrients allowed further growth until a maximum in spring.



453

454 Fig. 8 Time series of the CHL and SST anomalies with respect to their climatologies, over the
455 Central Baltic. The reference value 0 is also displayed. Shaded areas indicate the part of the time
456 series not used for the computation of the cross-correlation coefficient, which is indicated on each
457 year. Full size plots of individual years can be found in the supplementary material.

4. Conclusions

A fifteen-year merged-multi-sensor-daily dataset of satellite-derived CHL contains very valuable information for ecological studies if information is properly processed. Matchup analysis was undertaken with the largest in-situ database ever used for calibration and validation purposes over the Baltic region. Standard algorithms demonstrated easy to apply but, in the Baltic Sea, required further adjustments before an unbiased estimation of the basin-average CHL was obtained. Our derived time series take advantage of the independence of the error added by other water constituents and additional sources. The error distribution of the CHL estimates is such that, when averaging over a large number of observations, tends to zero, thus demonstrating that more accurate observations can be achieved when averaging over large areas.

The $OC4v6_{corr}$-derived climatology in Skagerrak and Kattegat revealed strong productivity in winter and a rather inactive summer. However, it should be noted that the blue-green CHL algorithms are not optimal for the coccolithophore detection (Gordon et al., 2001), commonly observed in this area. In the Gulf of Bothnia, CHL exhibits a single bloom during spring and experiences lower variability than the Skagerrak and Kattegat regions or the Central Baltic. In the latter region, the productivity in late fall, winter and early spring is severely inhibited. A first growth period with a maximum at the end of April is detected, followed by a stronger summer bloom peaking at the second week of July. The summer bloom in the Central Baltic constitutes the most intense signal found in this work, and attributed to cyanobacteria-like species. CHL and SST anomaly time series were cross-correlated to assess the cyanobacteria contribution to the overall CHL dynamics during the summer period of the Central Baltic. For example, the exceptionally warm winter 2007/2008 triggered an intense spring bloom in 2008 that also altered the normal dynamics throughout the year.

The Baltic region is widely recognized as a challenging test bed for ocean colour remote sensing. The interfering CDOM at blue wavelengths suggests that better CHL algorithms should use red and NIR bands, like the fluorescence line height or the maximum chlorophyll index algorithms (Odermatt et al., 2012, Fig. 1). Most of the Baltic CHL values range between ~ 1 and 10 mg m^{-3} and are at the lower part of the retrievable concentrations, via these algorithms (Odermatt et al., 2012, Fig. 1). These algorithms are only applicable to the archived MERIS data (2002-2012). The Ocean and Land Colour Instrument, on-board the Sentinel-3 will provide continuity with MERIS and algorithms will be adapted. The addition of the 400 nm band will expectedly aid in the separation of the CDOM contribution, given that proper atmospheric correction is achieved.

5. Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme through HORIZON 2020 under grant agreement no. 210129802 (Copernicus Marine environment monitoring service). Seadatanet, HELCOM and NOAA along with all single contributors are thanked for the in-situ data and CMEMS, ESA-CCI and GlobColour for the satellite data. Vega Forneris is thanked for technical support and Vittorio Brando for suggestions on the manuscript. Two anonymous reviewers are thanked for their comments and suggestions.

6. References

Attila, J., Koponen, S., Kallio, K., Lindfors, A., Kaitala, S., and Ylöstalo, P.: MERIS Case II water processor comparison on coastal sites of the northern Baltic Sea, Remote Sensing of Environment, 128, 138-149, http://dx.doi.org/10.1016/j.rse.2012.07.009, 2013.

501 Berthon, J.-F., and Zibordi, G.: Optically black waters in the northern Baltic Sea, Geophysical
502 Research Letters, 37, n/a-n/a, 10.1029/2010GL043227, 2010.

503 Brewin, R. J. W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E.,
504 Doerffer, R., Fomferra, N., Franz, B., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H.,
505 Lee, Z., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton,
506 J., Werdell, J., and White Iii, G. N.: The Ocean Colour Climate Change Initiative: III. A round-robin
507 comparison on in-water bio-optical algorithms, Remote Sensing of Environment, In press,
508 http://dx.doi.org/10.1016/j.rse.2013.09.016, 2013.

509 Carstensen, J., Klais, R., and Cloern, J. E.: Phytoplankton blooms in estuarine and coastal waters:
510 Seasonal patterns and key species, Estuarine, Coastal and Shelf Science, 162, 98-109,
511 http://dx.doi.org/10.1016/j.ecss.2015.05.005, 2015.

512 Product User Manual for Baltic Sea Physical Reanalysis Products:
513 http://marine.copernicus.eu/documents/PUM/CMEMS-OC-PUM-009-ALL.pdf, 2015.

514 D'Alimonte, D., Zibordi, G., Berthon, J. F., Canuti, E., and Kajiyama, T.: Bio-optical algorithms for
515 European seas: Performance and applicability of neural-net inversion schemes, Joint research
516 Centre, IspraJRC66326, 2011.

517 D'Alimonte, D., Zibordi, G., Berthon, J.-F., Canuti, E., and Kajiyama, T.: Performance and
518 applicability of bio-optical algorithms in different European seas, Remote Sensing of Environment,
519 124, 402-412, http://dx.doi.org/10.1016/j.rse.2012.05.022, 2012.

520 Darecki, M., and Stramski, D.: An evaluation of MODIS and SeaWiFS bio-optical algorithms in the
521 Baltic Sea, Remote Sensing of Environment, 89, 326-350,
522 http://dx.doi.org/10.1016/j.rse.2003.10.012, 2004.

523 Edelvang, K., Kaas, H., Erichsen, A. C., Alvarez-Berastegui, D., Bundgaard, K., and Jørgensen, P. V.:
524 Numerical modelling of phytoplankton biomass in coastal waters, Journal of Marine Systems, 57,
525 13-29, http://dx.doi.org/10.1016/j.jmarsys.2004.10.003, 2005.

526 Efron, B.: Bootstrap methods: another look at the jackknife, The annals of Statistics, 7, 1-26, 1979.

527 Product User Guide: http://www.esa-oceancolour-cci.org/?q=webfm_send/318, 2014.

528 Fleming, V., and Kaitala, S.: Phytoplankton Spring Bloom Intensity Index for the Baltic Sea
529 Estimated for the years 1992 to 2004, Hydrobiologia, 554, 57-65, 10.1007/s10750-005-1006-7,
530 2006.

531 Product User Guide: http://www.globcolour.info/CDR_Docs/GlobCOLOUR_PUG.pdf, 2015.

532 Gohin, F., Druon, J. N., and Lampert, L.: A five channel chlorophyll concentration algorithm applied
533 to SeaWiFS data processed by SeaDAS in coastal waters, International Journal of Remote Sensing,
534 23, 1639-1661, 10.1080/01431160110071879, 2002.

535 Gordon, H. R., Boynton, G. C., Balch, W. M., Groom, S. B., Harbour, D. S., and Smyth, T. J.: Retrieval
536 of coccolithophore calcite concentration from SeaWiFS Imagery, Geophysical Research Letters, 28,
537 1587-1590, 10.1029/2000GL012025, 2001.

538 HELCOM: Thematic Report on Validation of Algorithms for Chlorophyll a Retrieval from Satellite
539 Data in the Baltic Sea Area, Helsinki Commission-HELCOM, Ispra94, 2004.

540 Ibelings, B. W., Mur, L. R., and Walsby, A. E.: Diurnal changes in buoyancy and vertical distribution
541 in populations of Microcystisin two shallow lakes, Journal of Plankton Research, 13, 419-436,
542 10.1093/plankt/13.2.419, 1991.

543 IOCCG: Ocean-colour data merging, IOCCG, Dartmouth, Canada6, 2007.

544 Kahru, M., Savchuk, O. P., and Elmgren, R.: Satellite measurements of cyanobacterial bloom
545 frequency in the Baltic Sea: interannual and spatial variability, Marine Ecology Progress Series,
546 343, 15-23, 10.3354/meps06943, 2007.

547 Kahru, M., and Elmgren, R.: Multidecadal time series of satellite-detected accumulations of
548 cyanobacteria in the Baltic Sea, Biogeosciences, 11, 3619-3633, 10.5194/bg-11-3619-2014, 2014.

549 Kratzer, S., Brockmann, C., and Moore, G.: Using MERIS full resolution data to monitor coastal
550 waters — A case study from Himmerfjärden, a fjord-like bay in the northwestern Baltic Sea,
551 Remote Sensing of Environment, 112, 2284-2300, http://dx.doi.org/10.1016/j.rse.2007.10.006,
552 2008.

553 Larsson, K., Hajdu, S., Kilpi, M., Larsson, R., Leito, A., and Lyngs, P.: Effects of an extensive
554 Prymnesium polylepis bloom on breeding eiders in the Baltic Sea, Journal of Sea Research, 88, 21-
555 28, http://dx.doi.org/10.1016/j.seares.2013.12.017, 2014.

556 Majaneva, M., Rintala, J.-M., Hajdu, S., Hällfors, S., Hällfors, G., Skjevik, A.-T., Gromisz, S.,
557 Kownacka, J., Busch, S., and Blomster, J.: The extensive bloom of alternate-stage Prymnesium

558 polylepis (Haptophyta) in the Baltic Sea during autumn–spring 2007–2008, European Journal of
559 Phycology, 47, 310-320, 10.1080/09670262.2012.713997, 2012.

560 Maritorena, S., and Siegel, D. A.: Consistent merging of satellite ocean color data sets using a bio-
561 optical model, Remote Sensing of Environment, 94, 429-440,
562 http://dx.doi.org/10.1016/j.rse.2004.08.014, 2005.

563 Maritorena, S., d'Andon, O. H. F., Mangin, A., and Siegel, D. A.: Merged satellite ocean color data
564 products using a bio-optical model: Characteristics, benefits and issues, Remote Sensing of
565 Environment, 114, 1791-1804, http://dx.doi.org/10.1016/j.rse.2010.04.002, 2010.

566 Mélin, F., and Vantrepotte, V.: How optically diverse is the coastal ocean?, Remote Sensing of
567 Environment, 160, 235-251, http://dx.doi.org/10.1016/j.rse.2015.01.023, 2015.

568 Morel, A., and Berthon, J.-F.: Surface pigments, algal biomass profiles, and potential production of
569 the euphotic layer: Relationships reinvestigated in view of remote-sensing applications, Limnology
570 and Oceanography, 34, 1545-1562, 1989.

571 Odermatt, D., Gitelson, A., Brando, V. E., and Schaepman, M.: Review of constituent retrieval in
572 optically deep and complex waters from satellite imagery, Remote Sensing of Environment, 118,
573 116-126, http://dx.doi.org/10.1016/j.rse.2011.11.013, 2012.

574 Pierson, D. C., Kratzer, S., Strömbeck, N., and Håkansson, B.: Relationship between the attenuation
575 of downwelling irradiance at 490 nm with the attenuation of PAR (400 nm–700 nm) in the Baltic
576 Sea, Remote Sensing of Environment, 112, 668-680, http://dx.doi.org/10.1016/j.rse.2007.06.009,
577 2008.

578 Ploug, H.: Cyanobacterial surface blooms formed by Aphanizomenon sp. and Nodularia spumigena
579 in the Baltic Sea: Small-scale fluxes, pH, and oxygen microenvironments, Limnology and
580 Oceanography, 53, 914-921, 10.4319/lo.2008.53.3.0914, 2008.

581 Reinart, A., and Kutser, T.: Comparison of different satellite sensors in detecting cyanobacterial
582 bloom events in the Baltic Sea, Remote Sensing of Environment, 102, 74-85,
583 http://dx.doi.org/10.1016/j.rse.2006.02.013, 2006.

584 Reissmann, J. H., Burchard, H., Feistel, R., Hagen, E., Lass, H. U., Mohrholz, V., Nausch, G., Umlauf,
585 L., and Wieczorek, G.: Vertical mixing in the Baltic Sea and consequences for eutrophication – A
586 review, Progress in Oceanography, 82, 47-80, http://dx.doi.org/10.1016/j.pocean.2007.10.004,
587 2009.

588 Schneider, B., Kaitala, S., and Maunula, P.: Identification and quantification of plankton bloom
589 events in the Baltic Sea by continuous pCO2 and chlorophyll a measurements on a cargo ship,
590 Journal of Marine Systems, 59, 238-248, http://dx.doi.org/10.1016/j.jmarsys.2005.11.003, 2006.

591 Siegel, H., and Gerth, M.: Optical Remote Sensing Applications in the Baltic Sea, in: Remote
592 Sensing of the European Seas, edited by: Barale, V., and Gade, M., Springer Netherlands,
593 Dordrecht, 91-102, 2008.

594 A mild algal bloom in 2010: http://www.smhi.se/en/news-archive/a-mild-algal-bloom-in-2010-
595 1.12999, 2010.

596 Wasmund, N., and Uhlig, S.: Phytoplankton trends in the Baltic Sea, ICES Journal of Marine
597 Science: Journal du Conseil, 60, 177-186, 10.1016/s1054-3139(02)00280-1, 2003.

Ocean color chlorophyll (OC) v6: http://oceancolor.gsfc.nasa.gov/REPROCESSING/R2009/ocv6/, 2010.

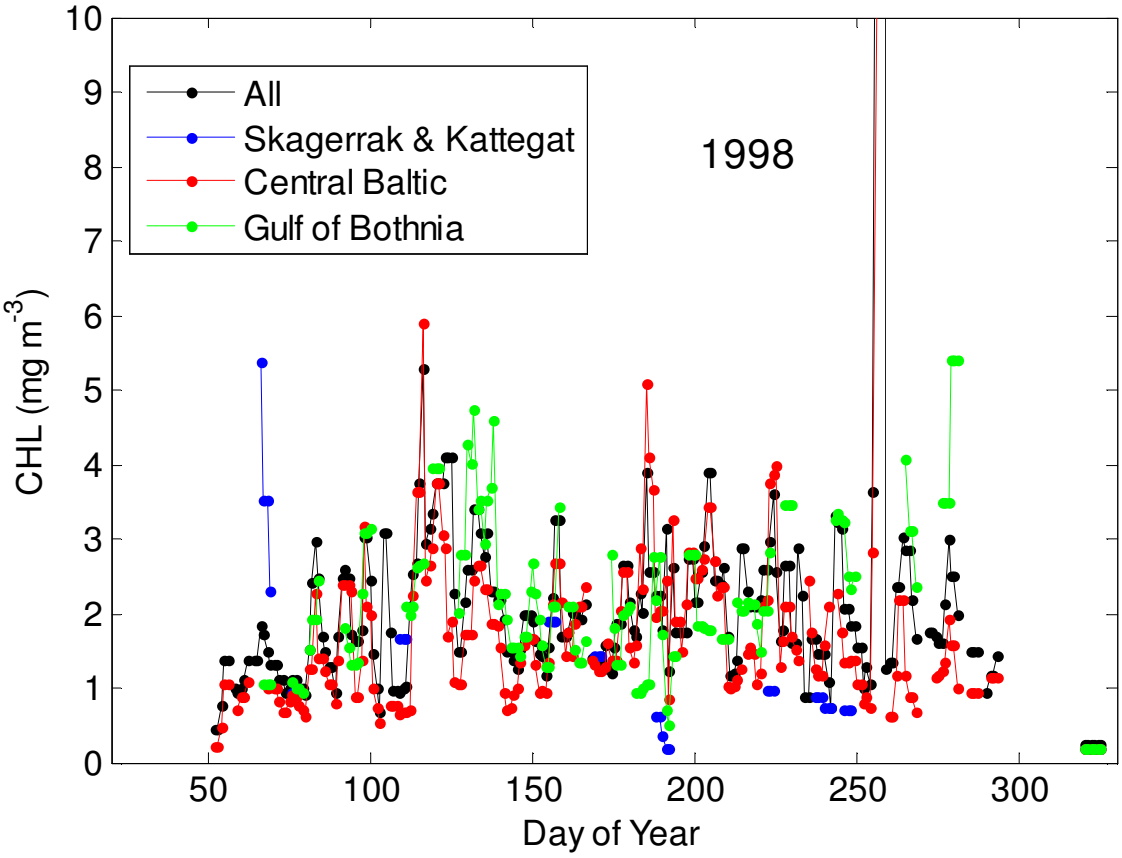


Fig. S1 CHL basin averages in 1998.

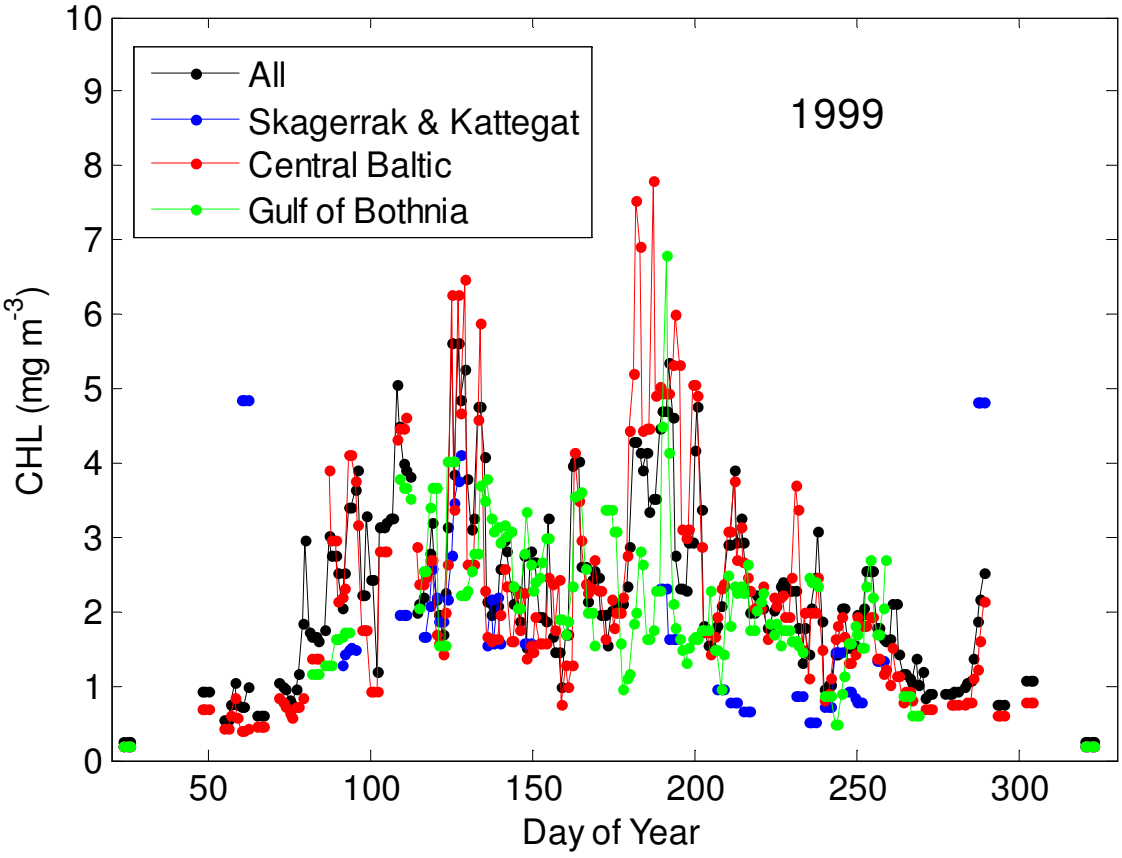


Fig. S2 CHL basin averages in 1999.

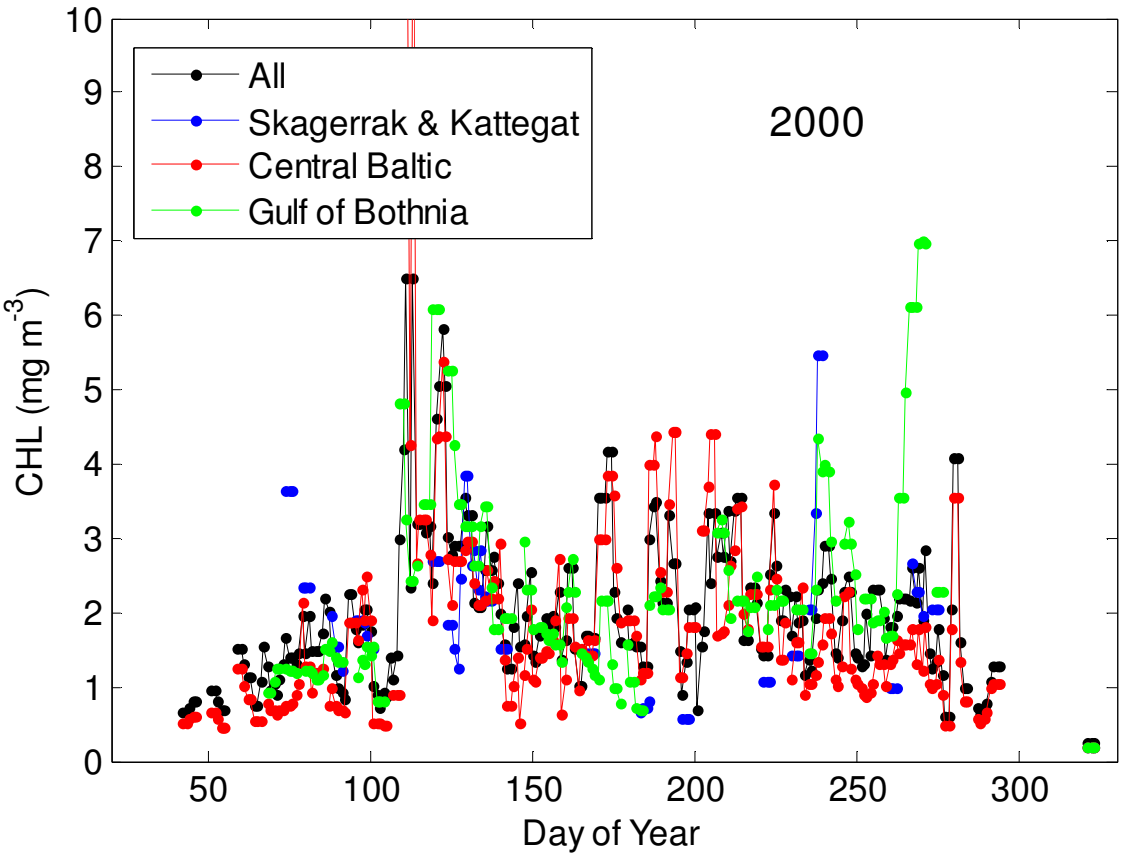


Fig. S3 CHL basin averages in 2000.

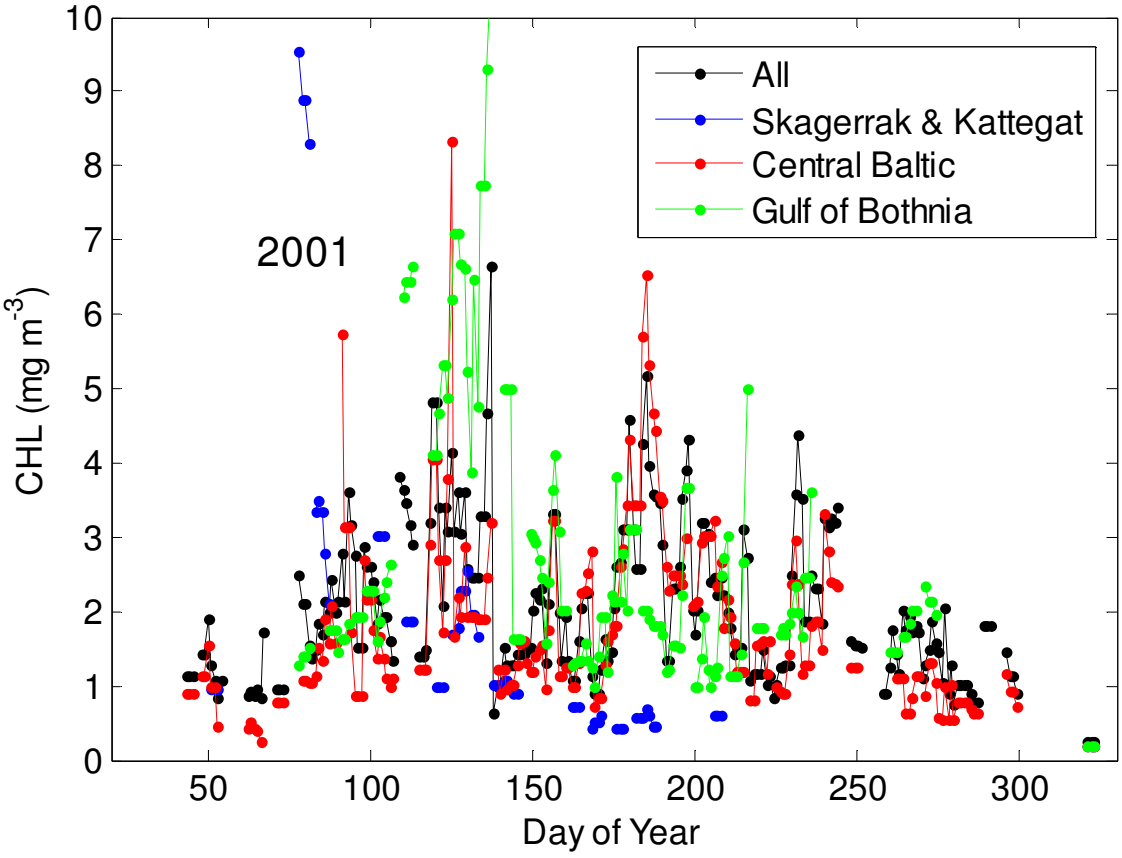


Fig. S4 CHL basin averages in 2001.

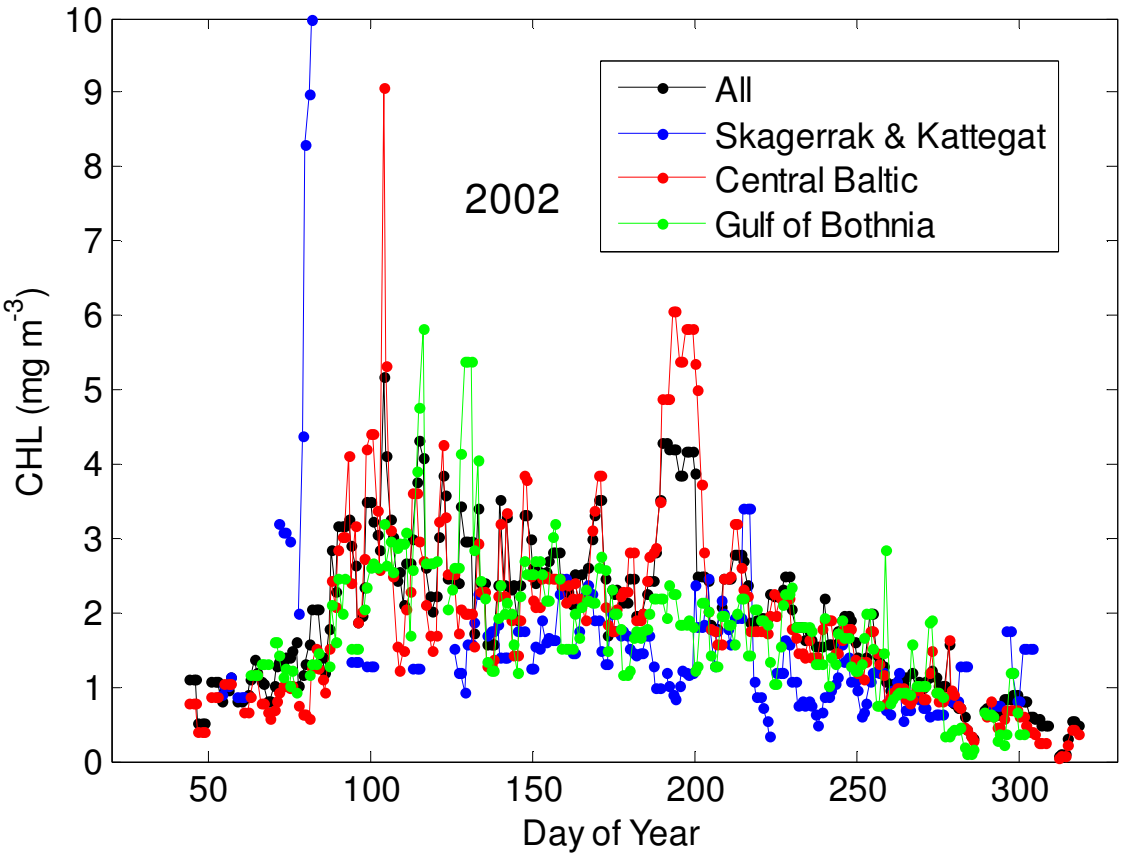


Fig. S5 CHL basin averages in 2002.

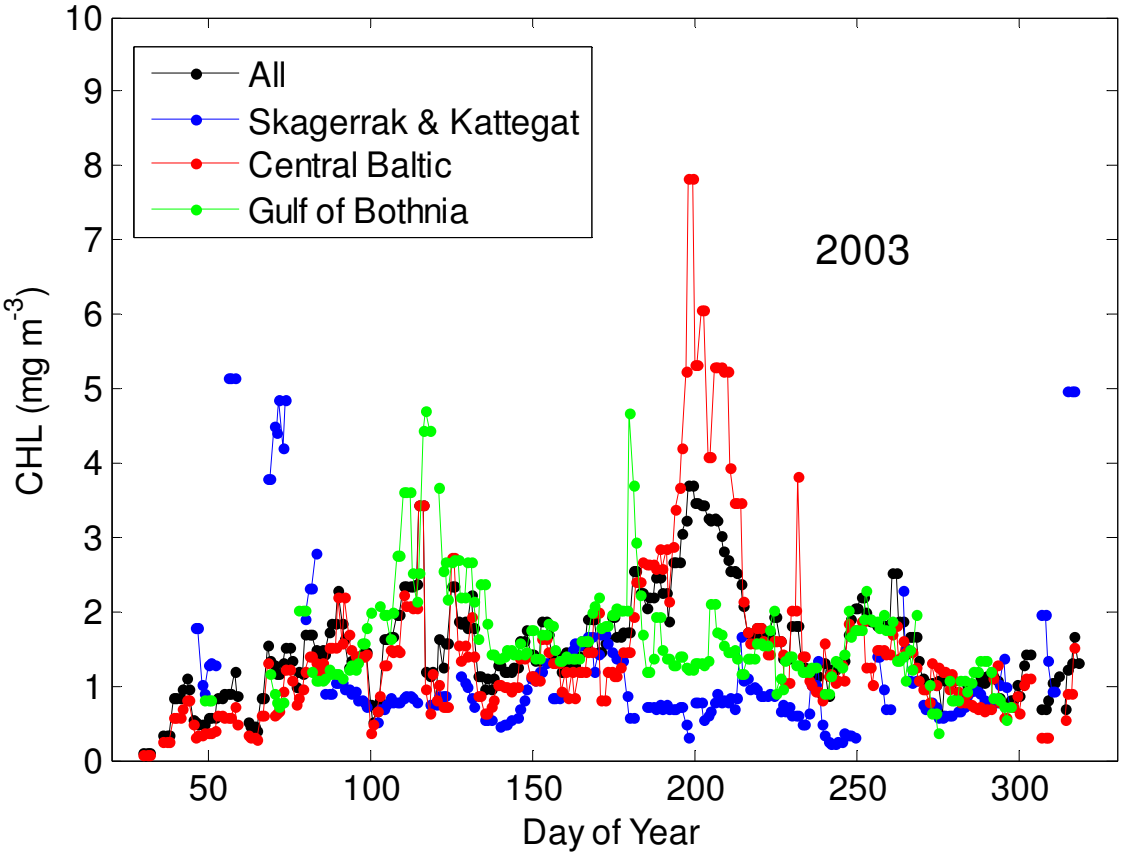


Fig. S6 CHL basin averages in 2003.

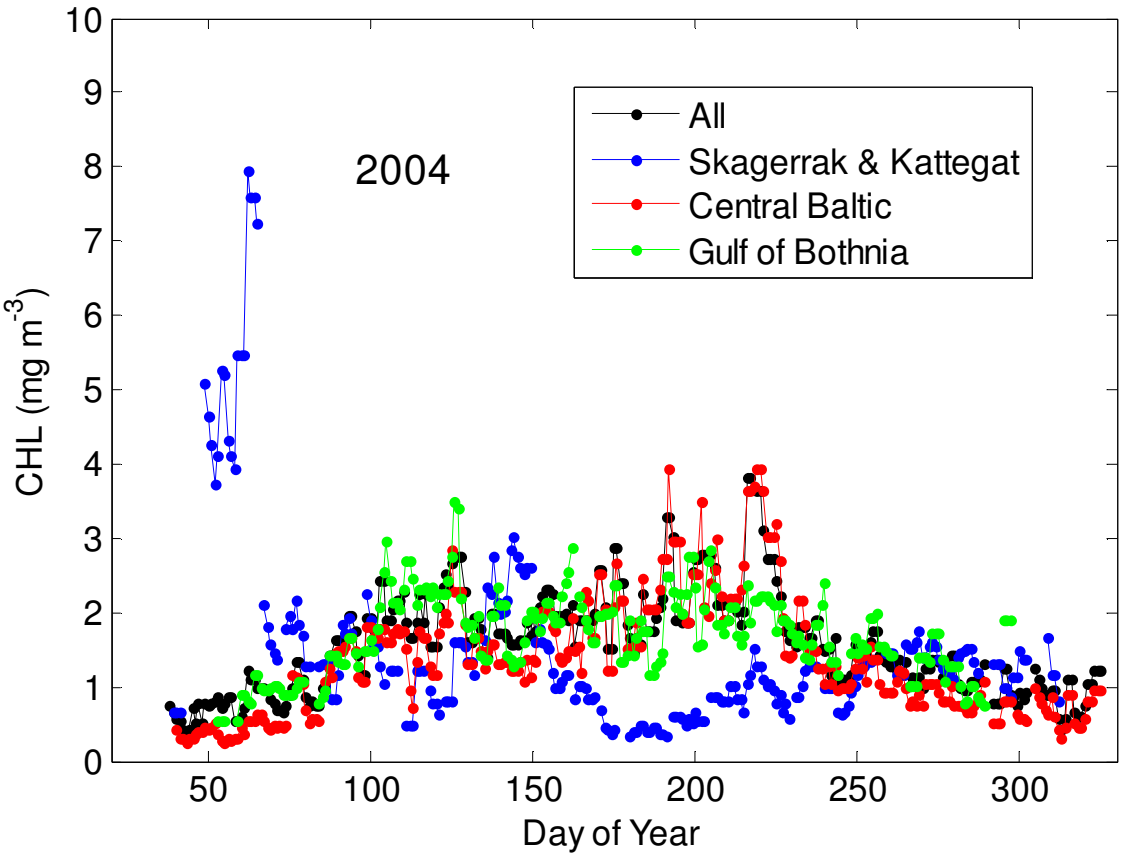


Fig. S7 CHL basin averages in 2004.

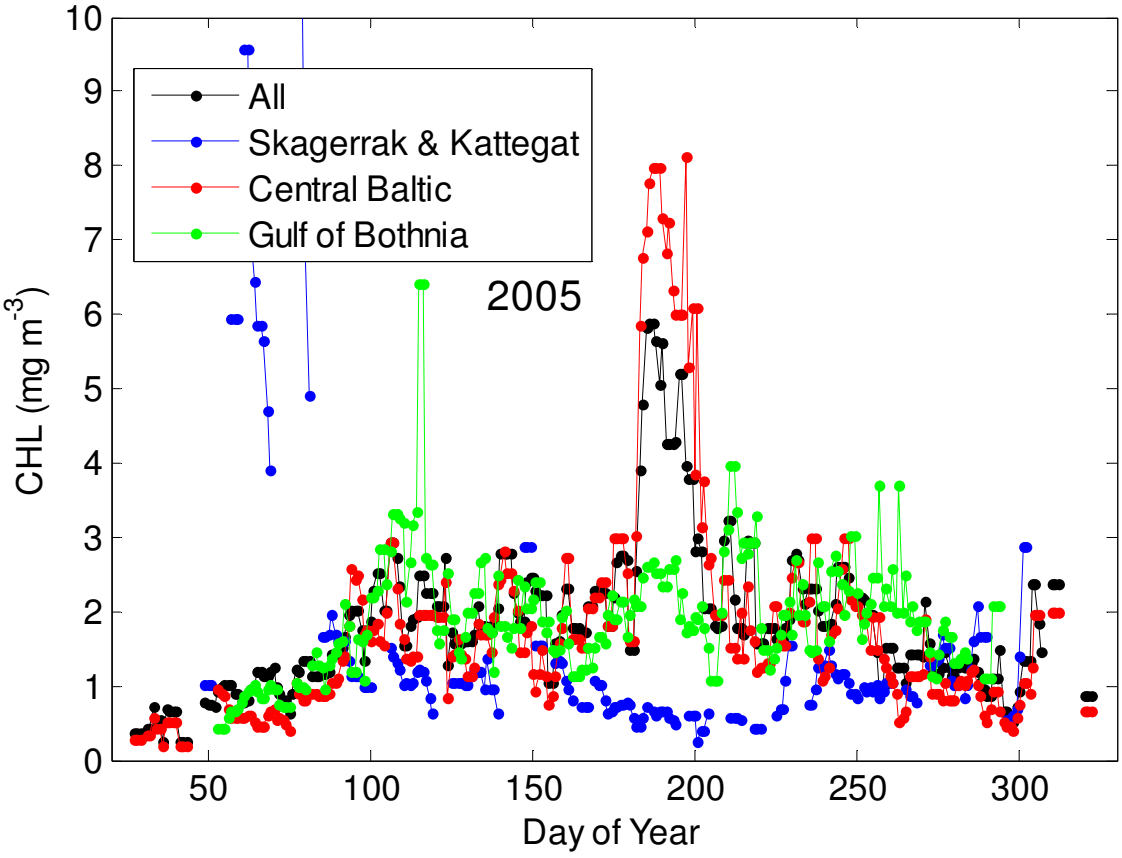


Fig. S8 CHL basin averages in 2005.

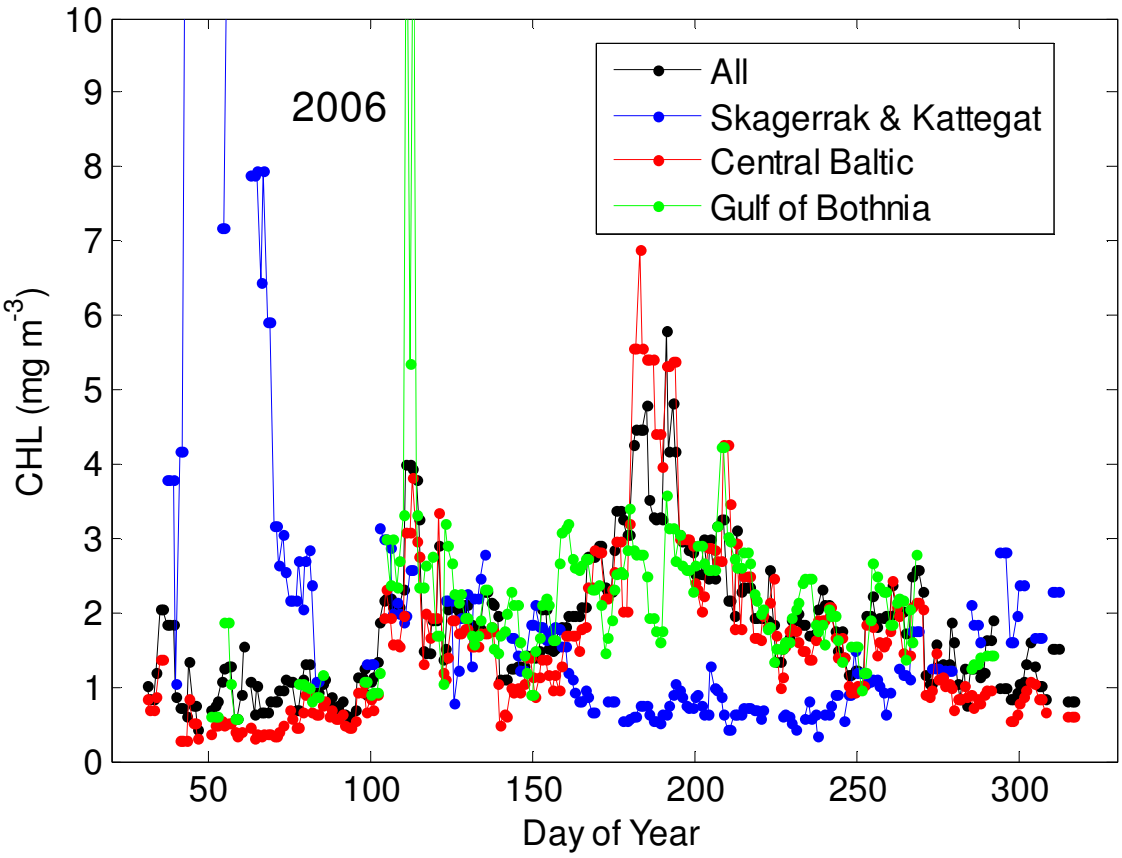


Fig. S9 CHL basin averages in 2006.

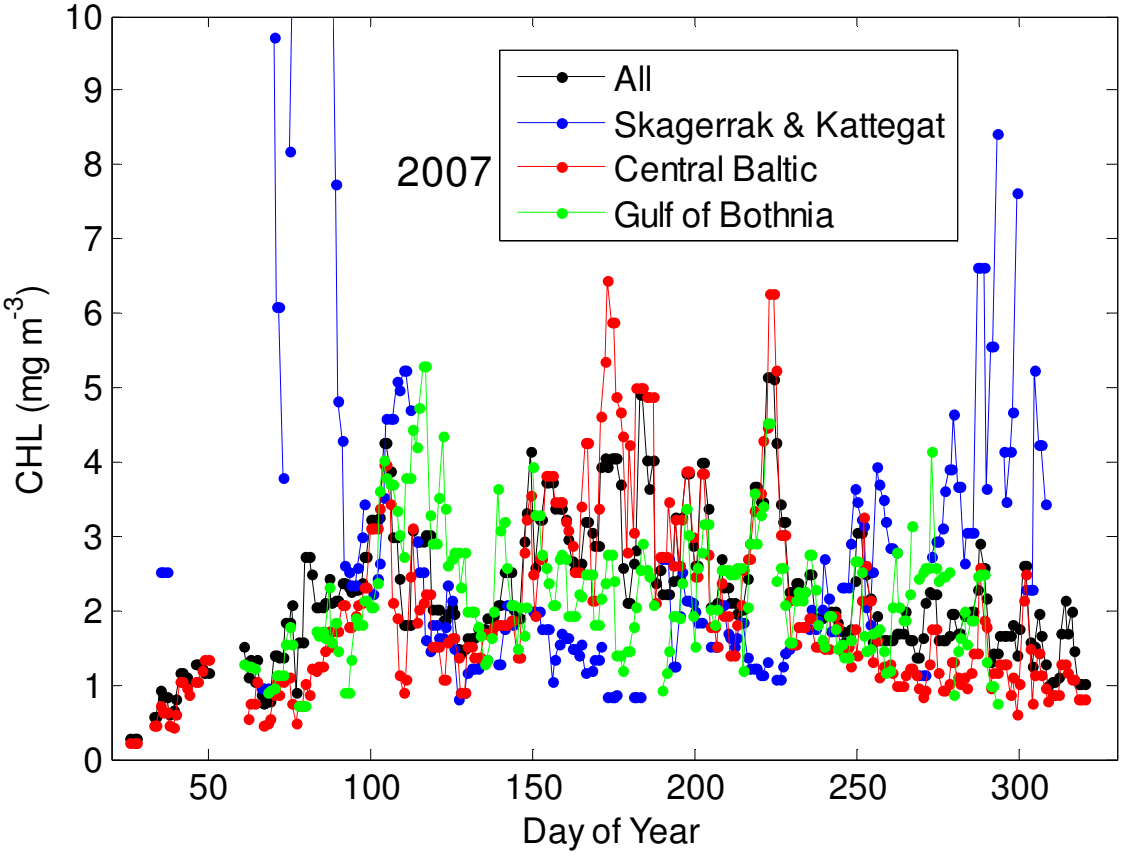


Fig. S10 CHL basin averages in 2007.

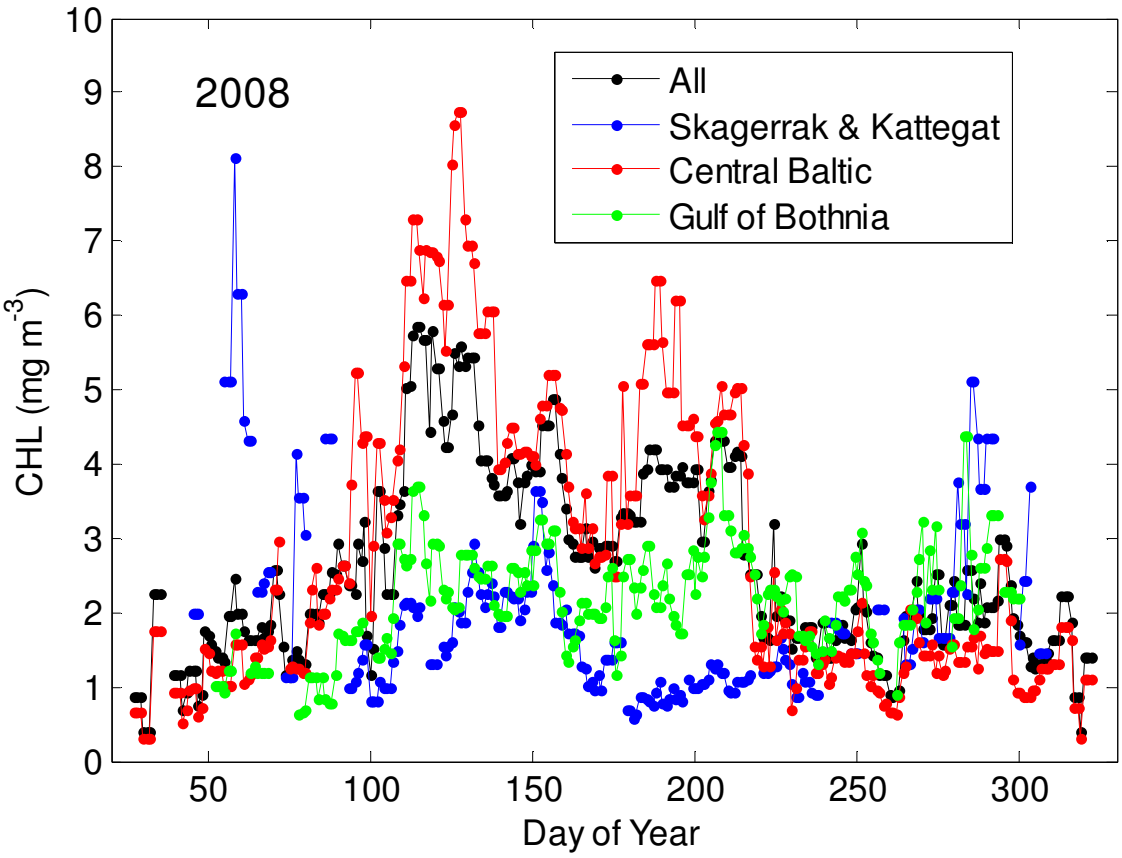


Fig. S11 CHL basin averages in 2008.

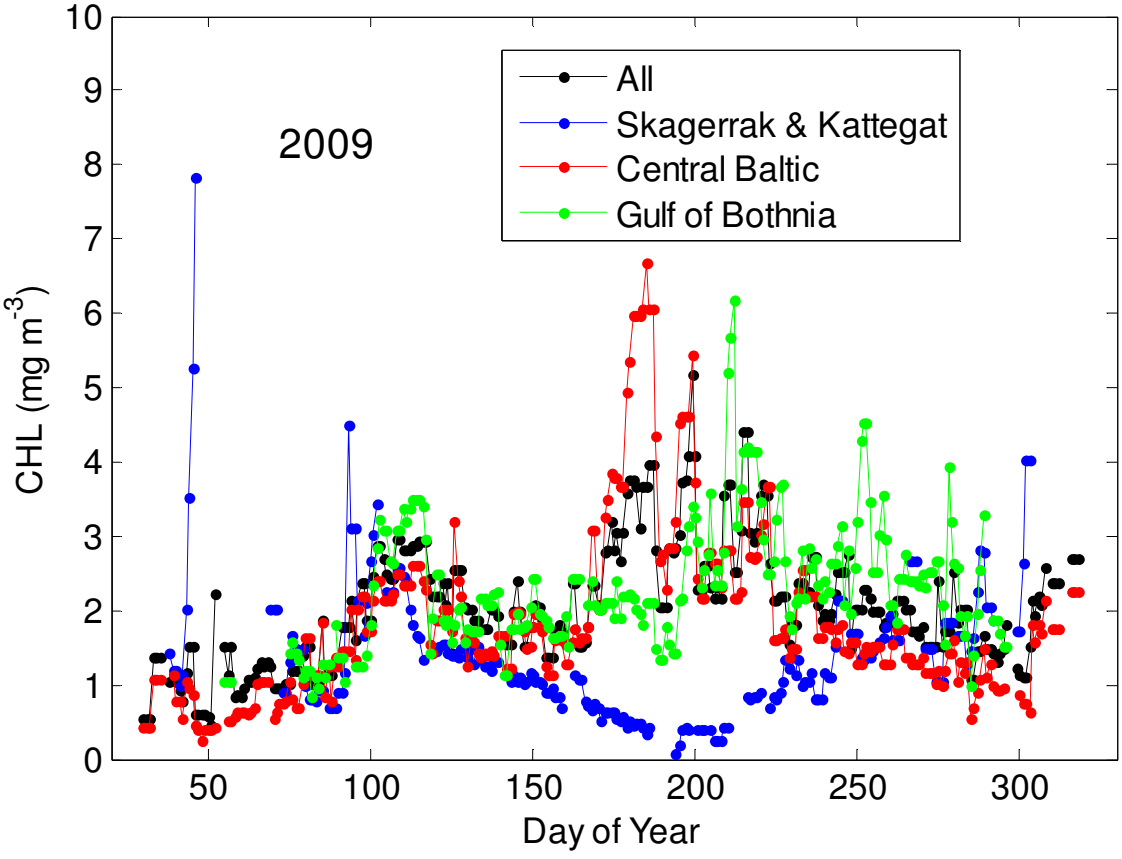


Fig. S12 CHL basin averages in 2009.

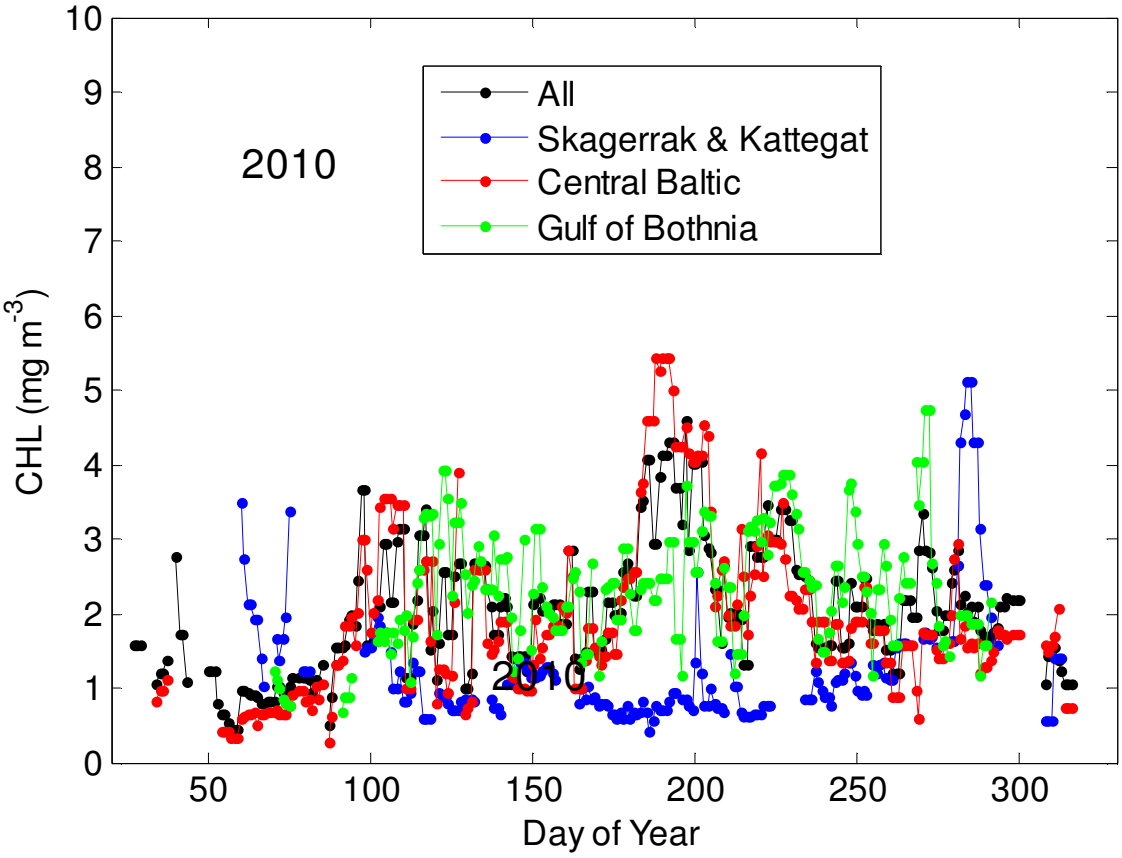


Fig. S13 CHL basin averages in 2010.

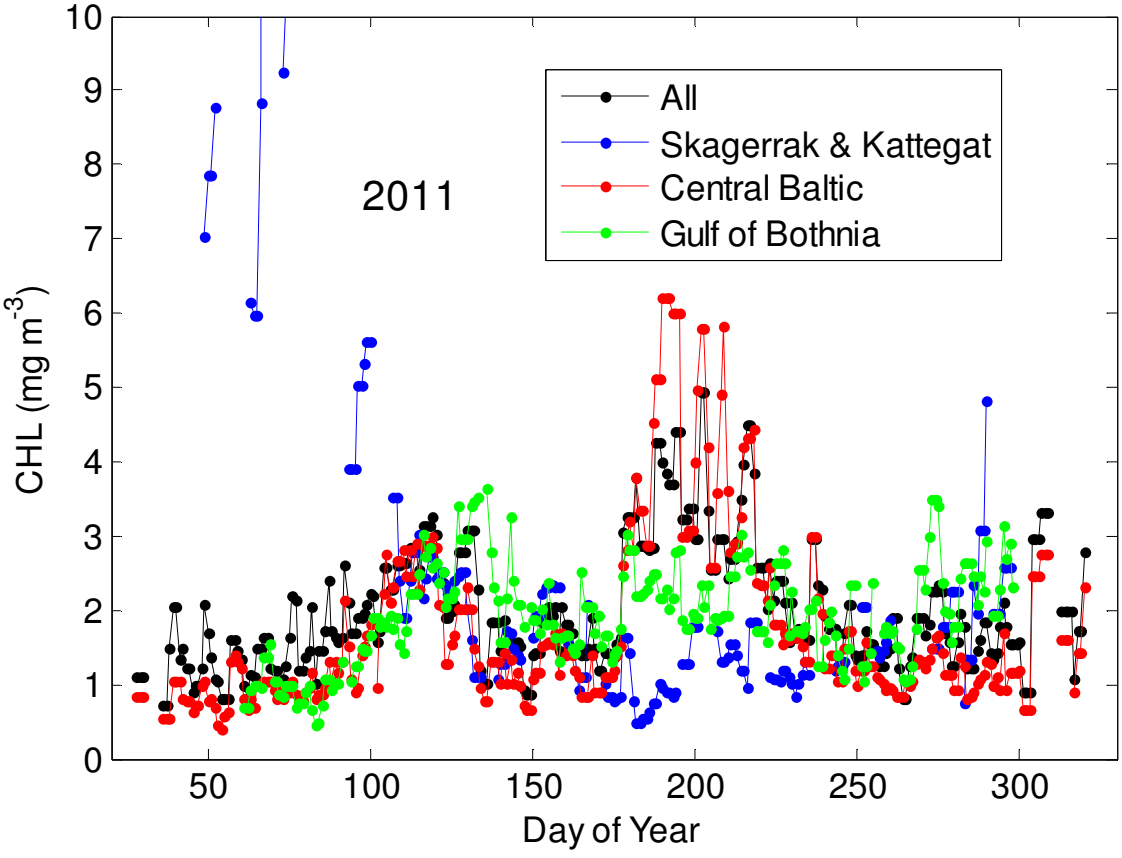


Fig. S14 CHL basin averages in 2011.

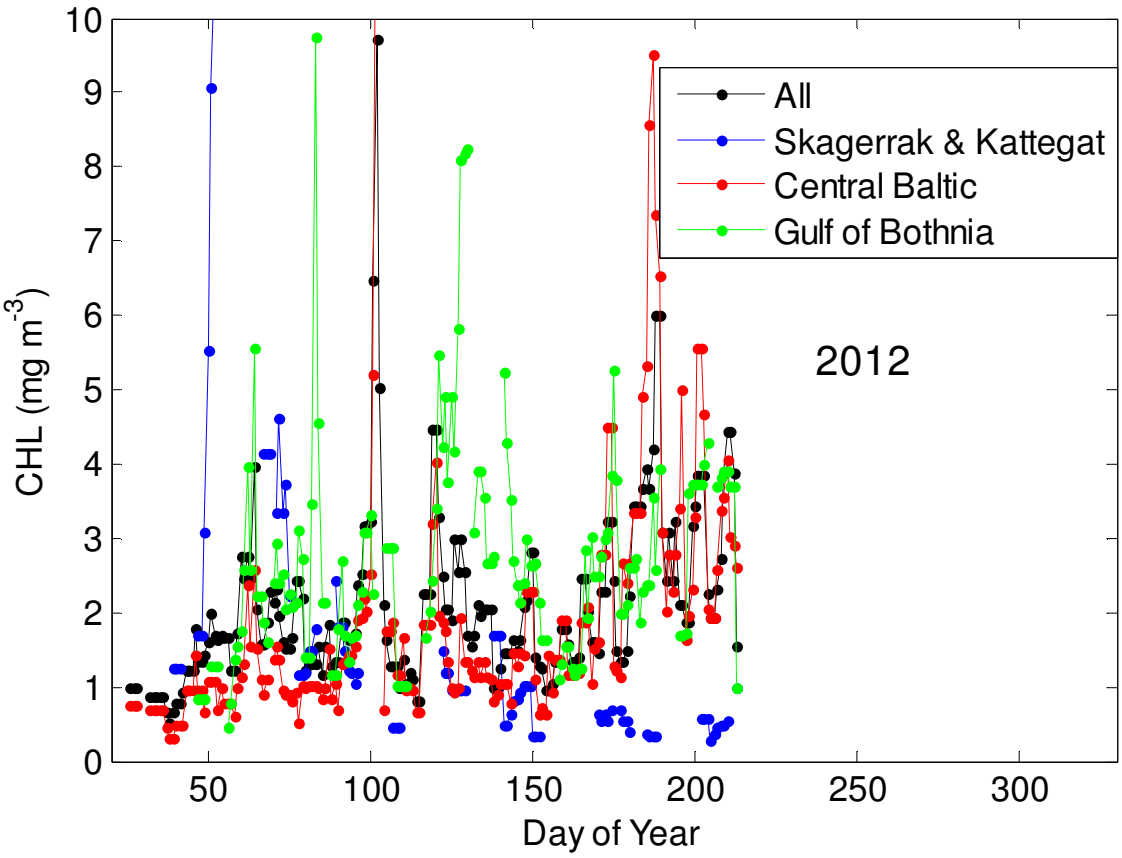


Fig. S15 CHL basin averages in 2012.

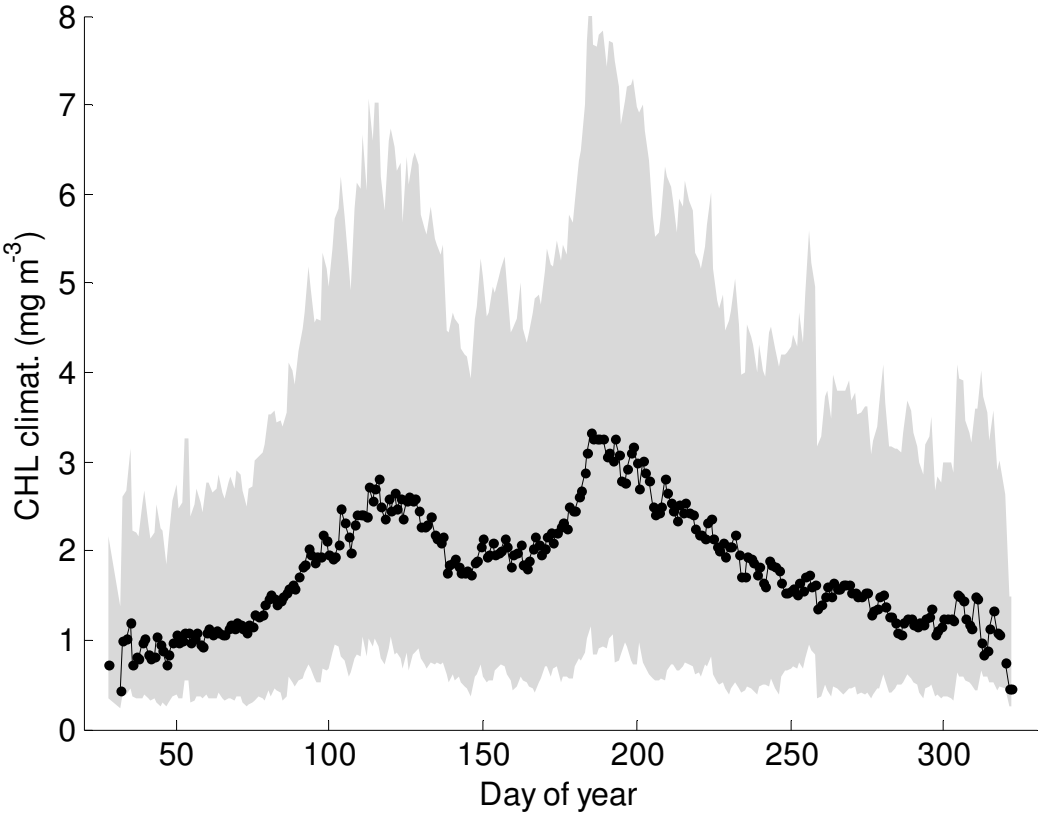


Fig. S16 Climatologic CHL of the whole domain, inside the plus-minus standard deviation band.

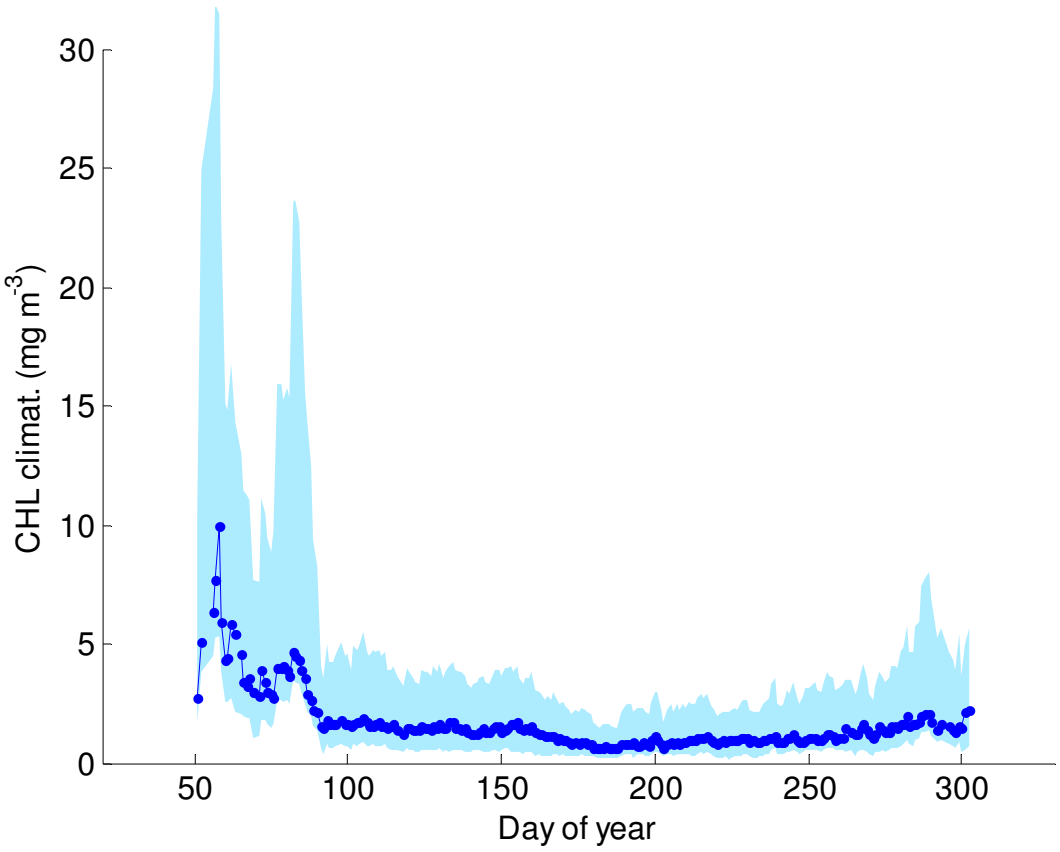


Fig. S17 Climatologic CHL of Skagerrak and Kattegat, inside the plus-minus standard deviation band.

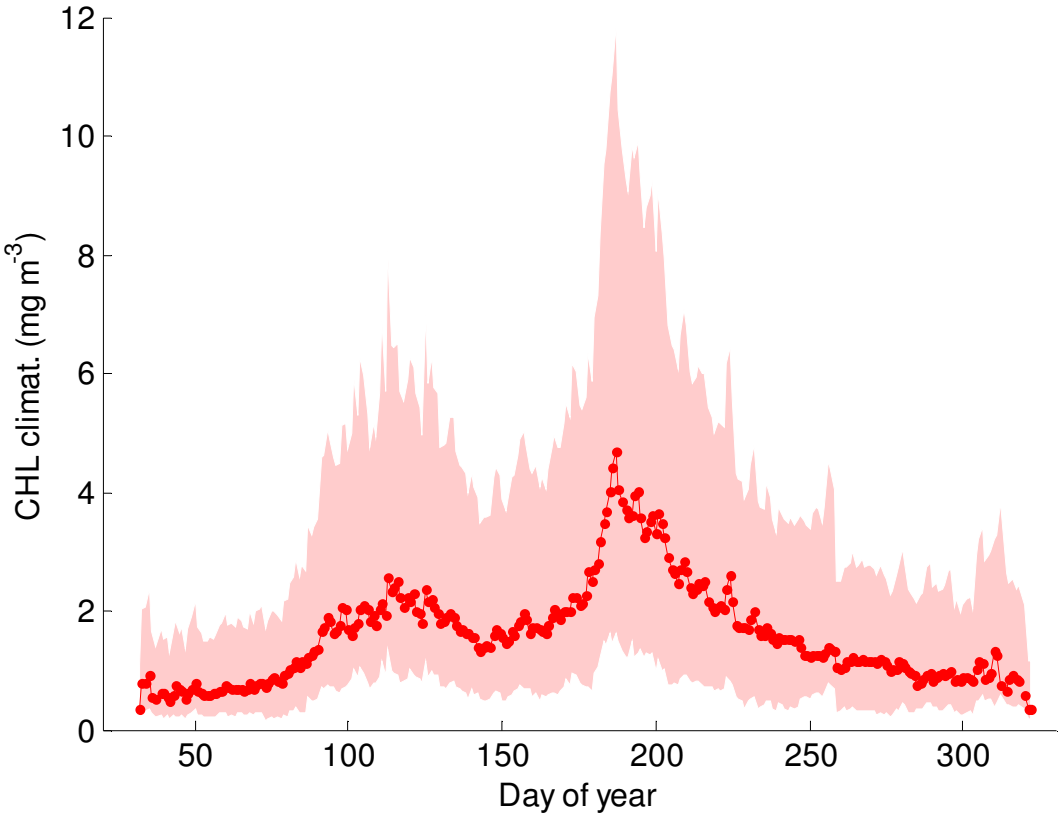


Fig. S18 Climatologic CHL of the Central Baltic, inside the plus-minus standard deviation band.

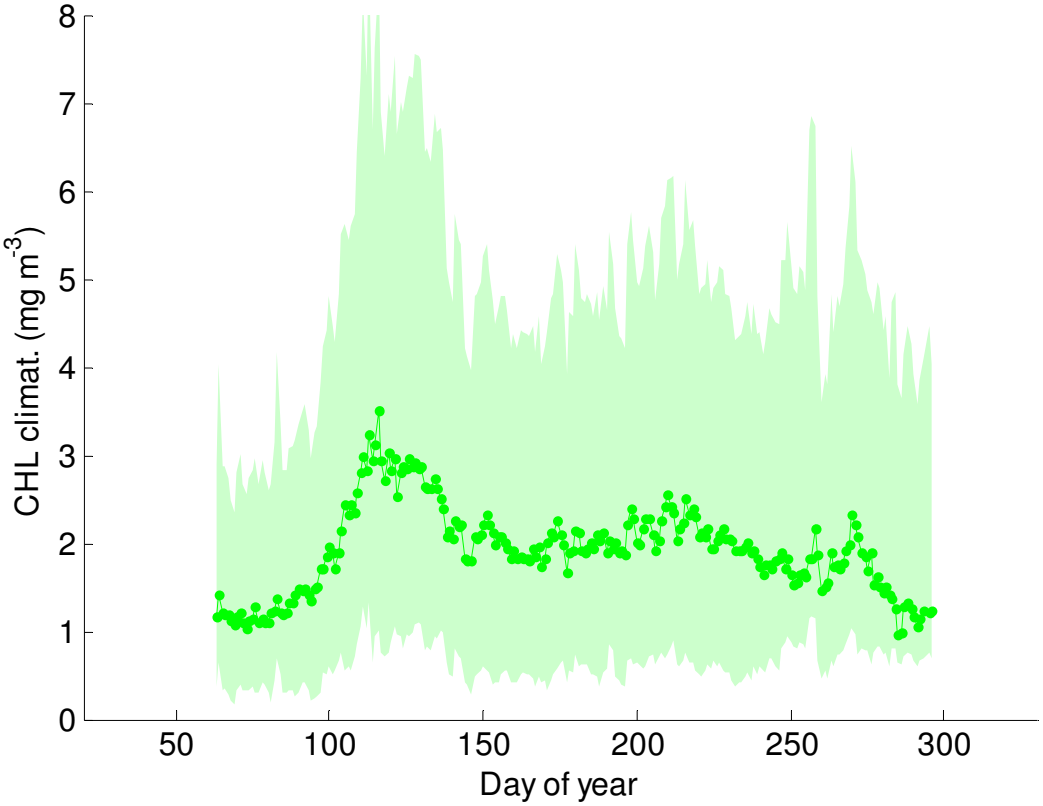


Fig. S19 Climatologic CHL of the Gulf of Bothnia, inside the plus-minus standard deviation band.

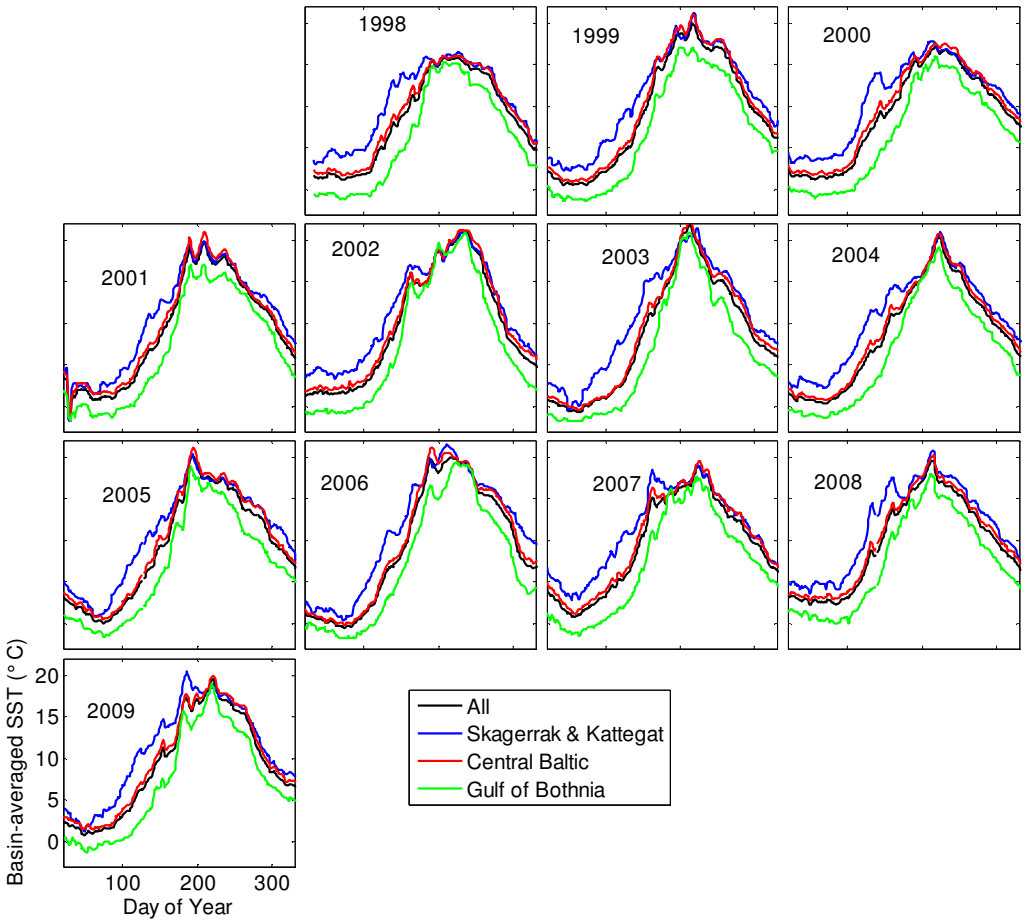


Fig. S19 Basin averages of the satellite-derived sea surface temperatures.

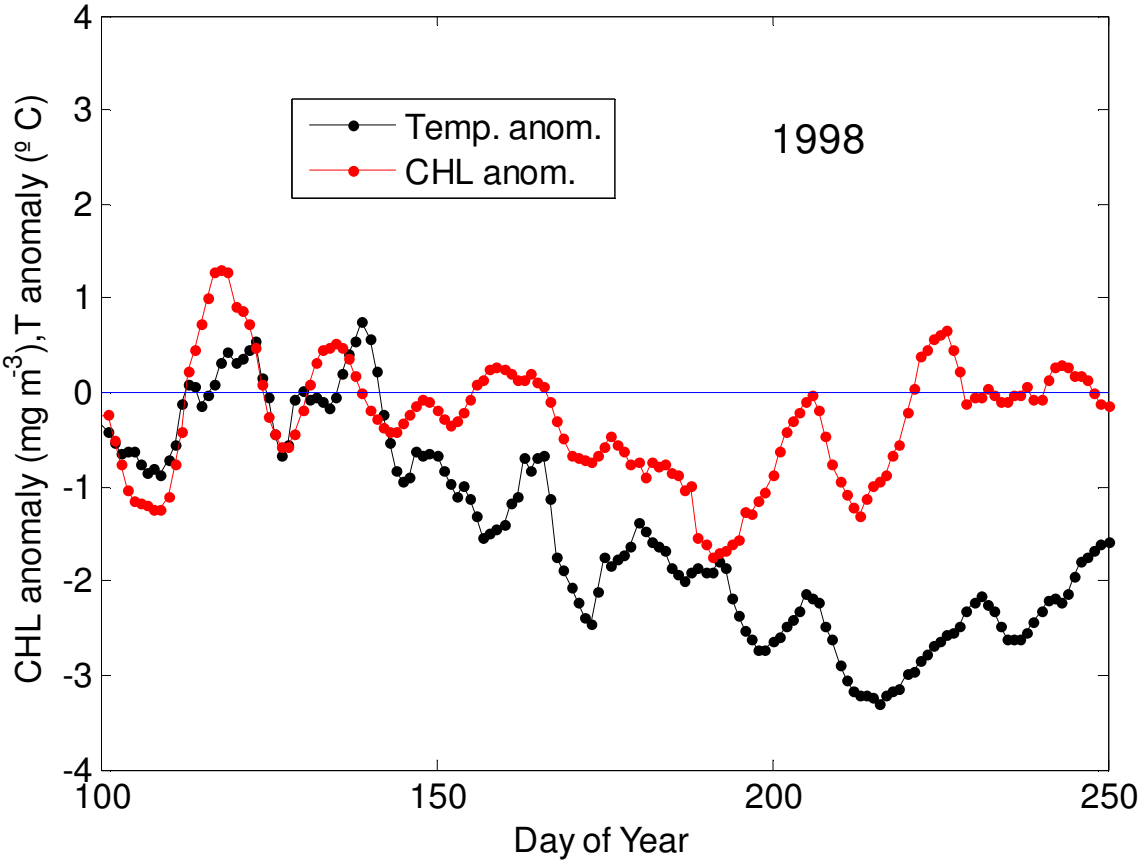


Fig. S20 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 1998.

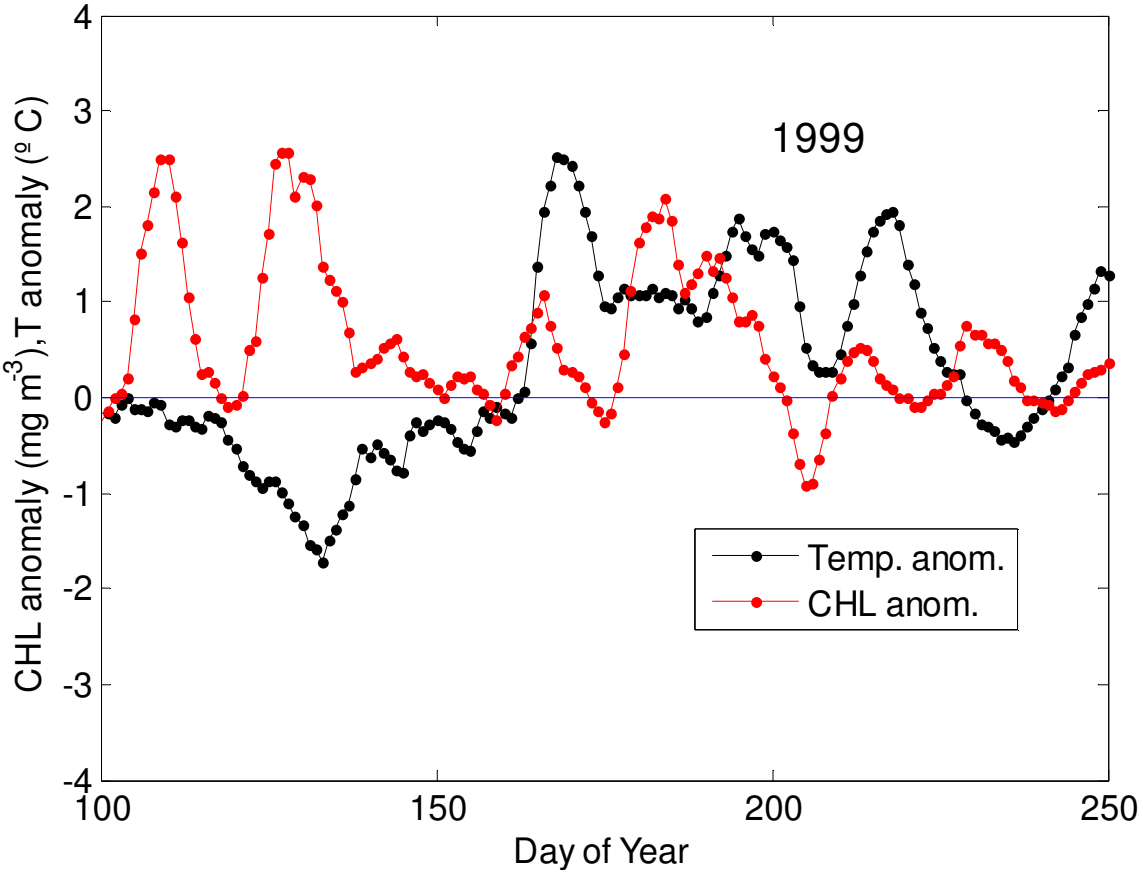


Fig. S21 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 1999.

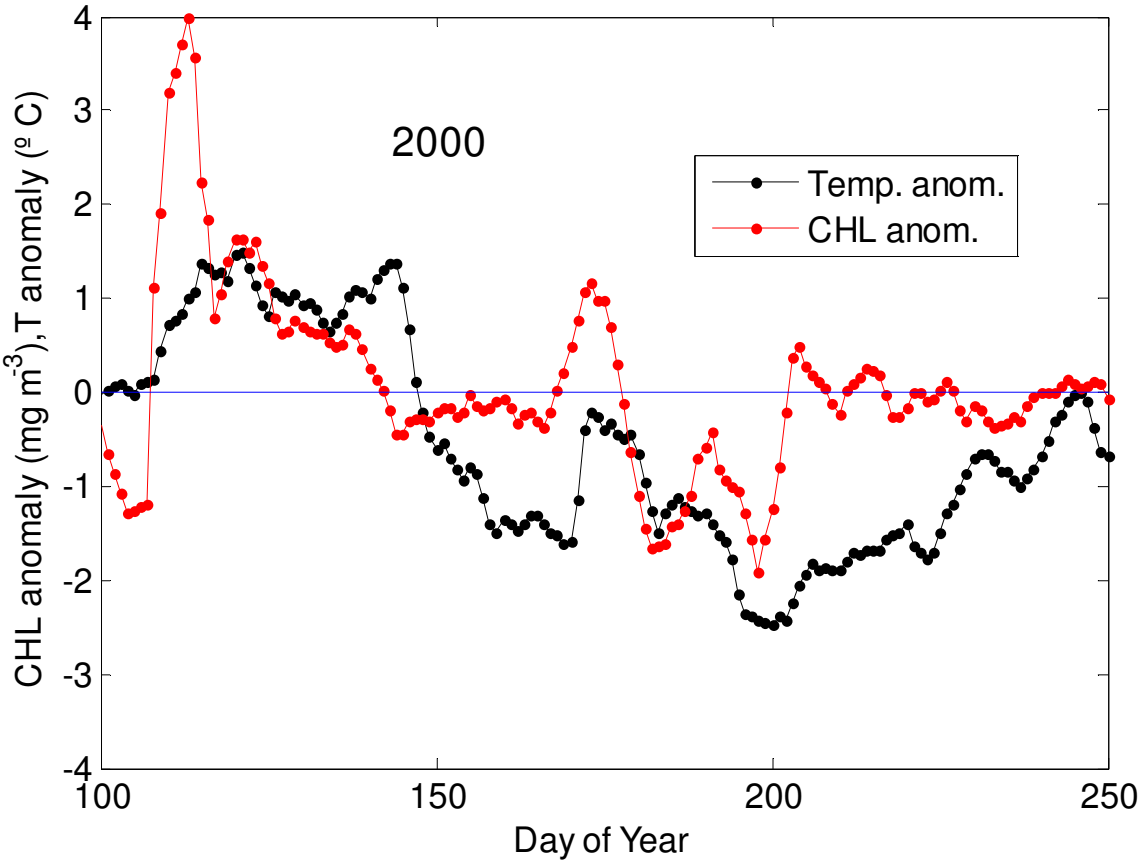


Fig. S22 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2000.

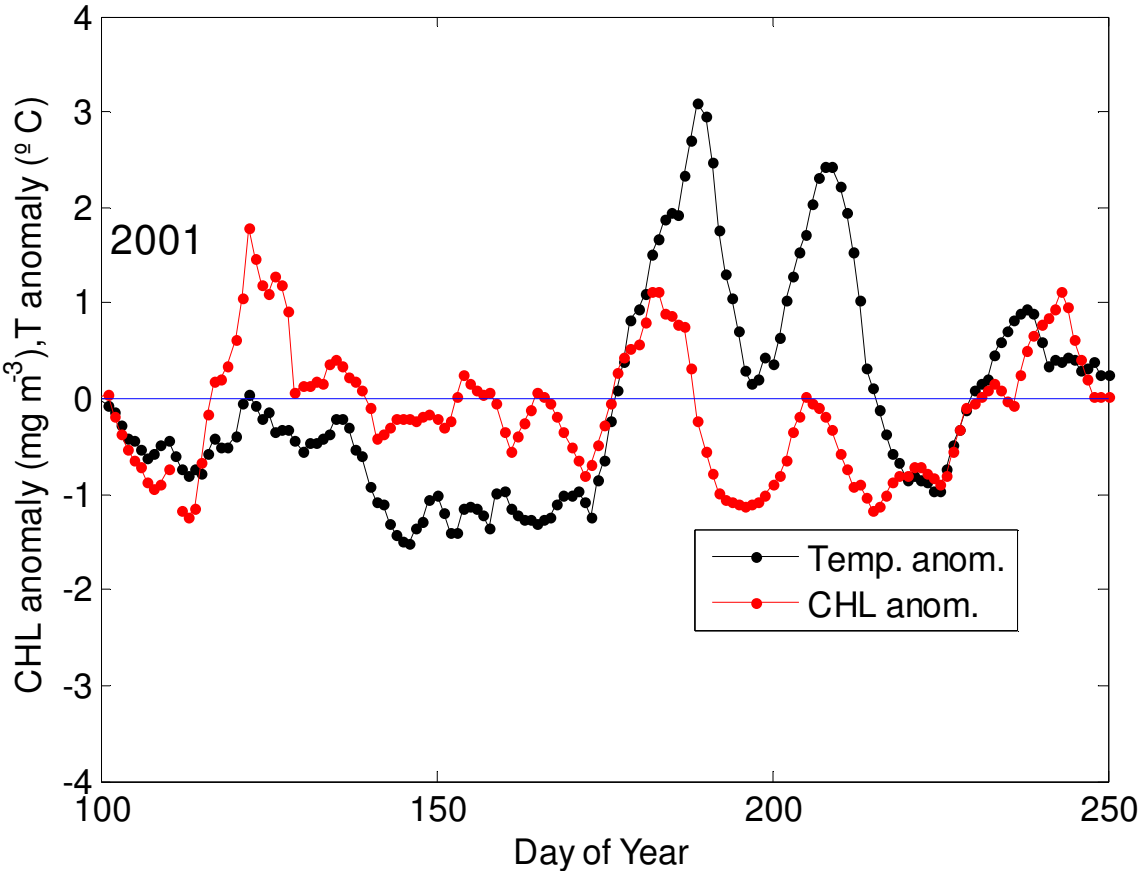


Fig. S23 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2001.

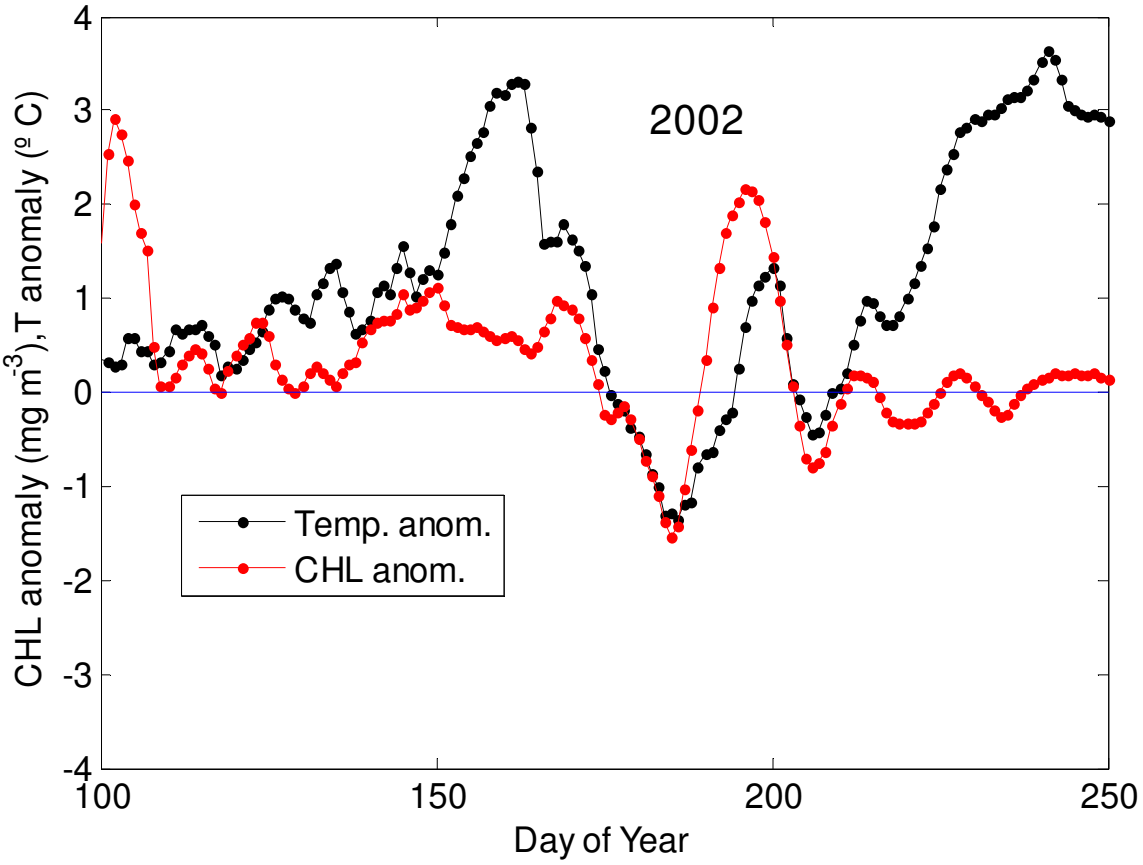


Fig. S24 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2002.

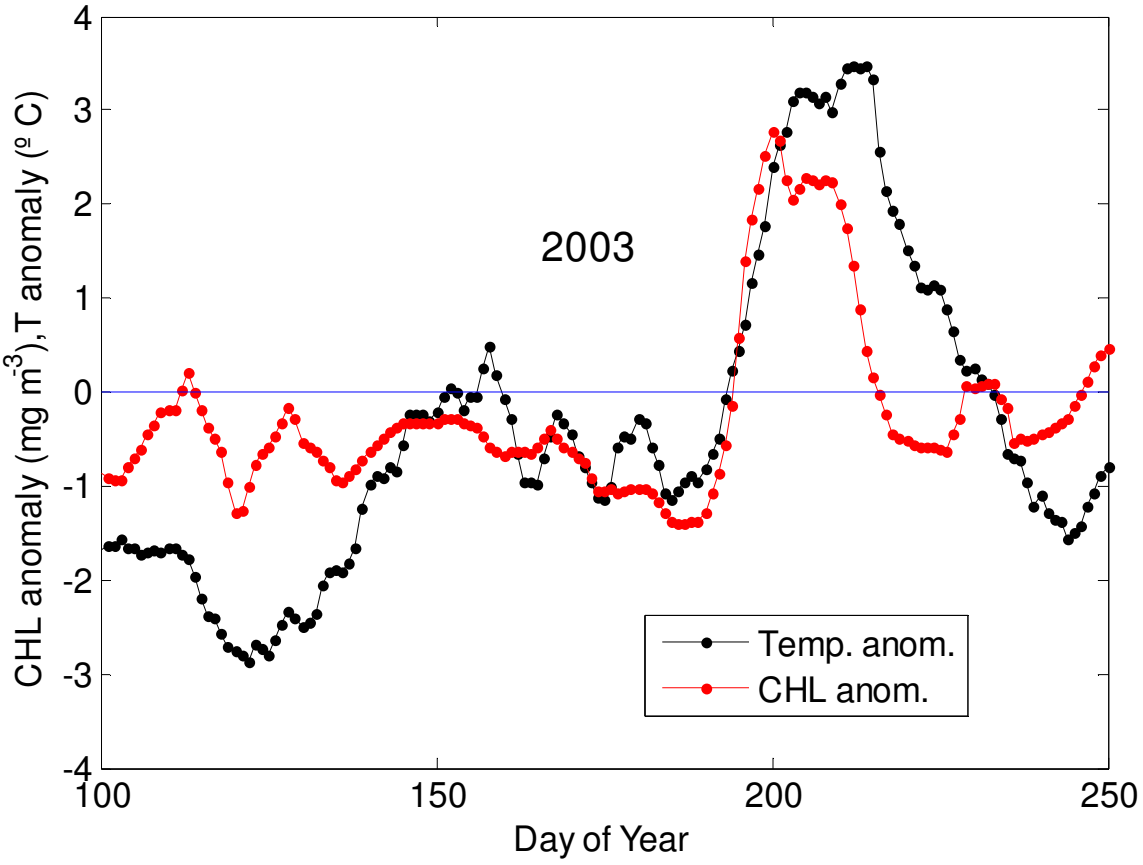


Fig. S25 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2003.

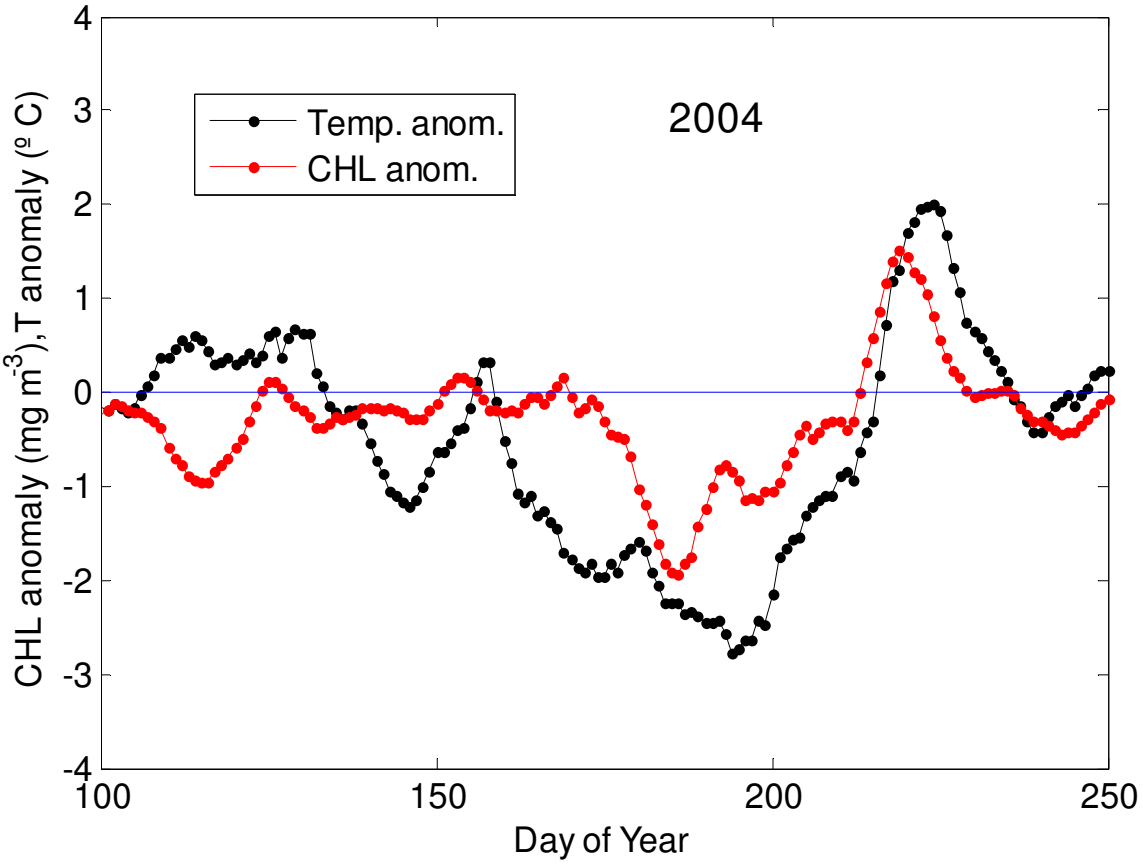


Fig. S26 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2004.

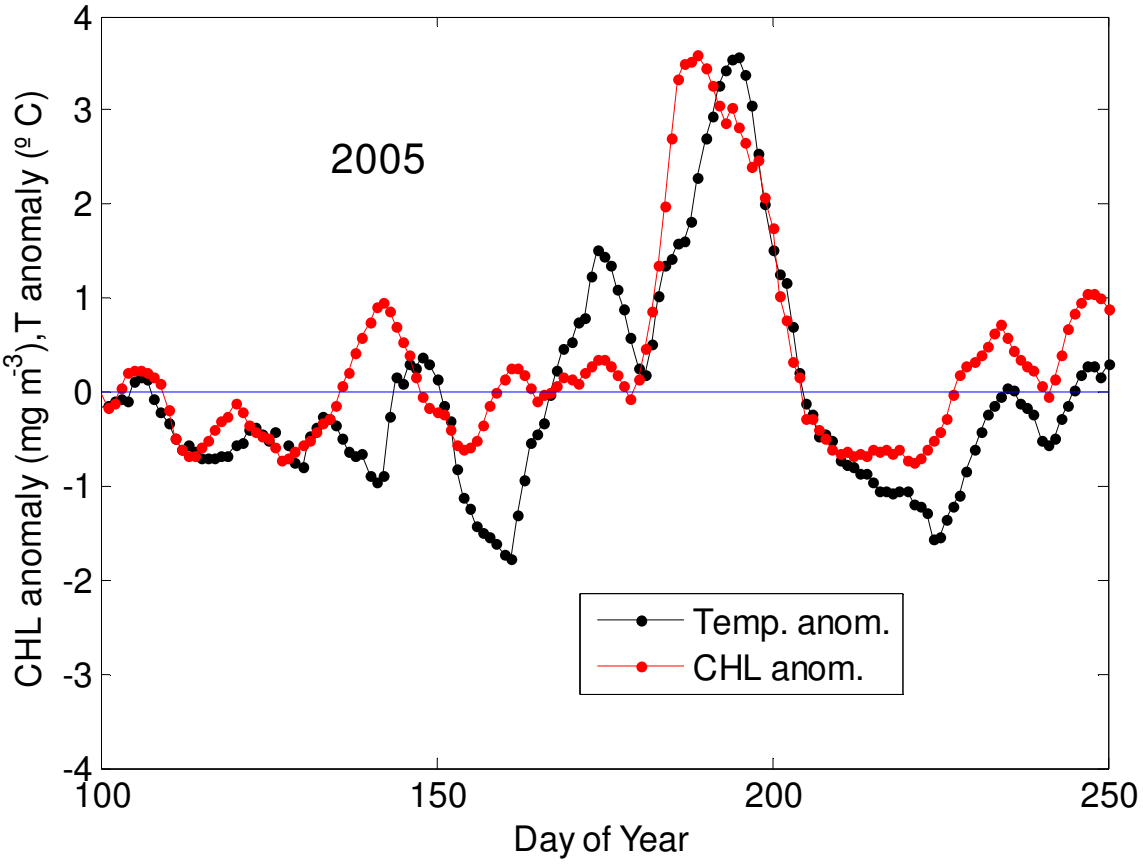


Fig. S27 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2005.

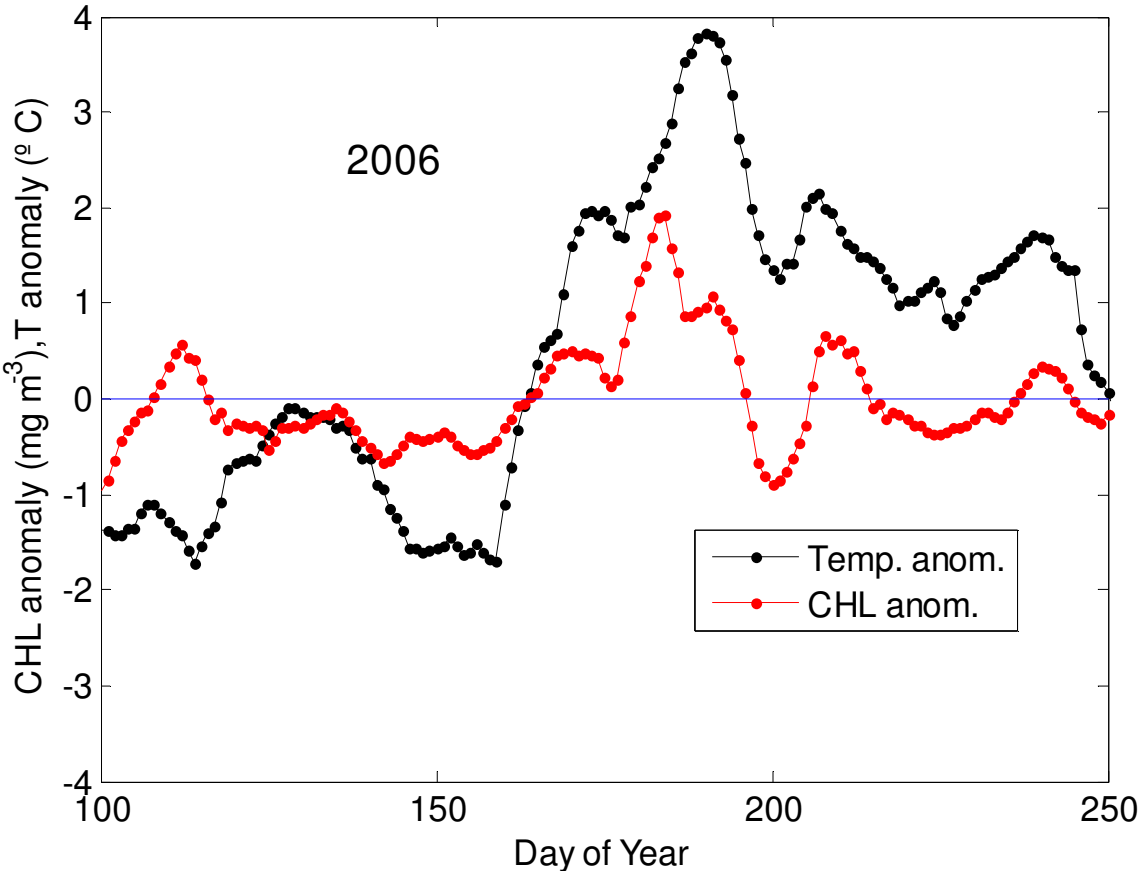


Fig. S28 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2006.

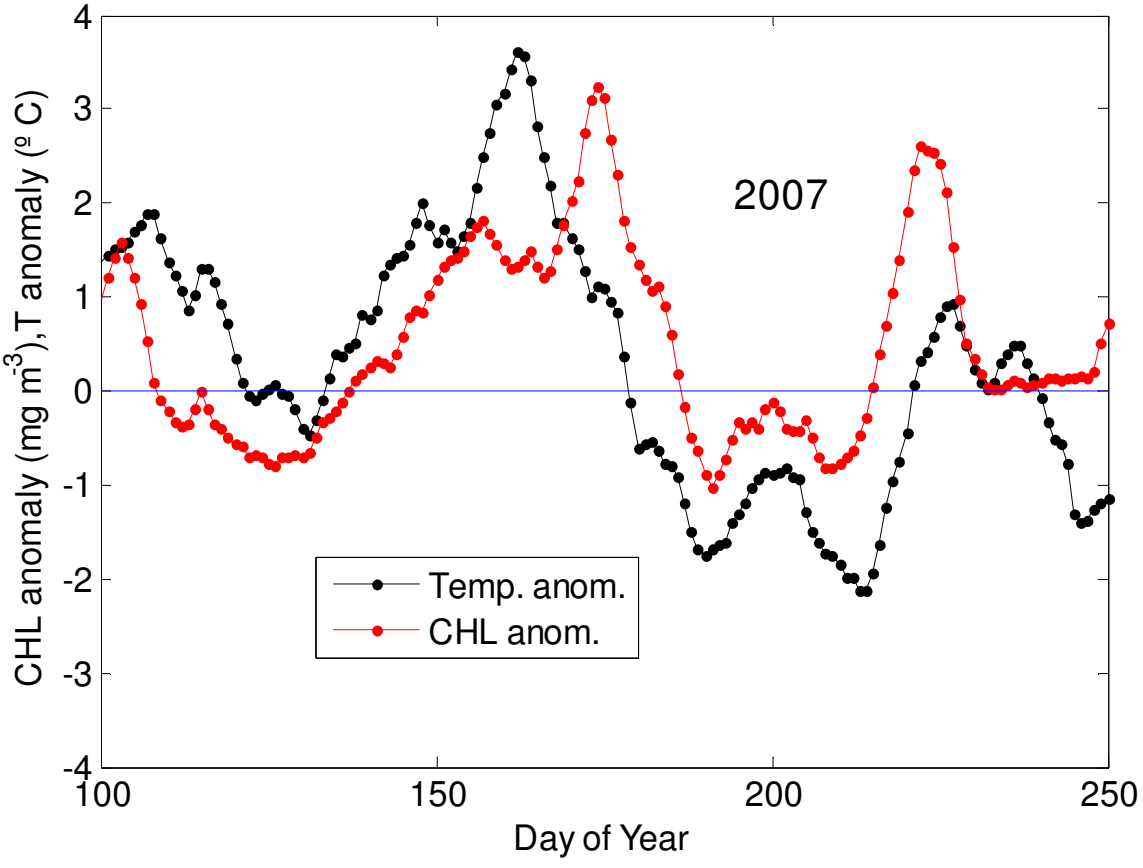


Fig. S29 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2007.

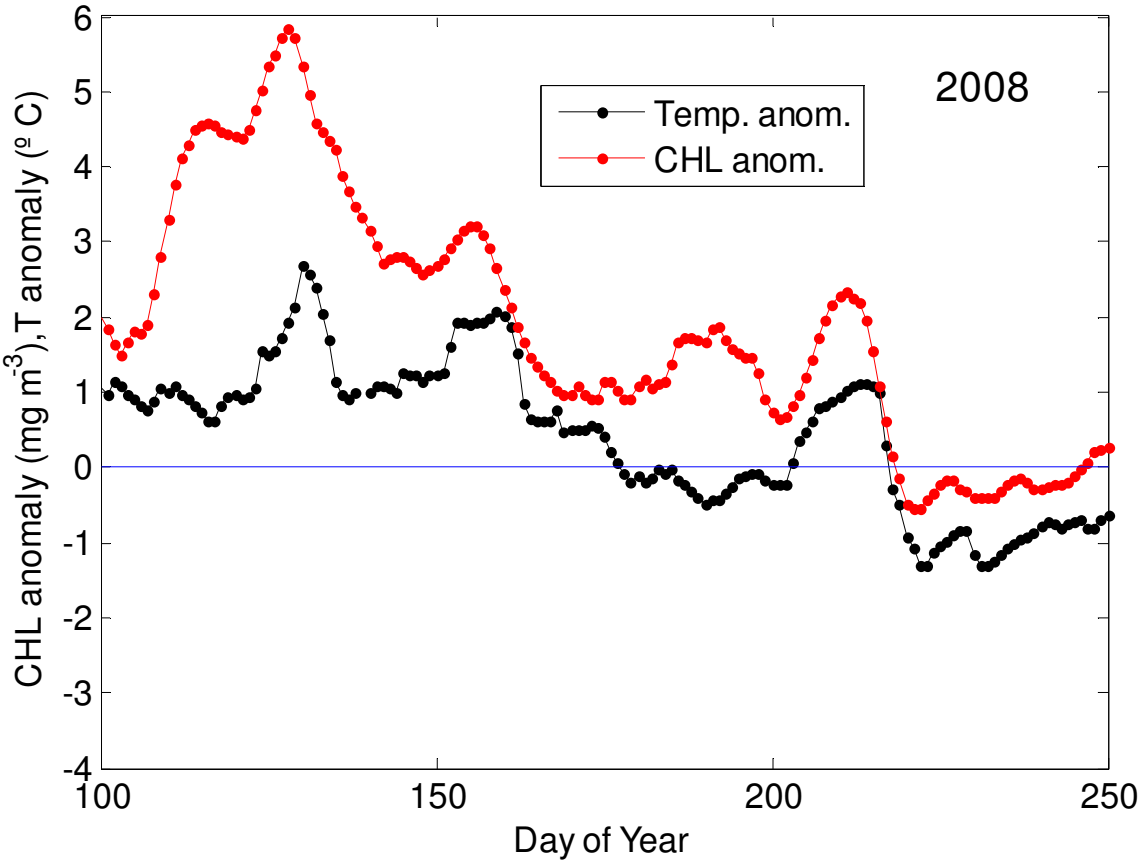


Fig. S30 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2008.

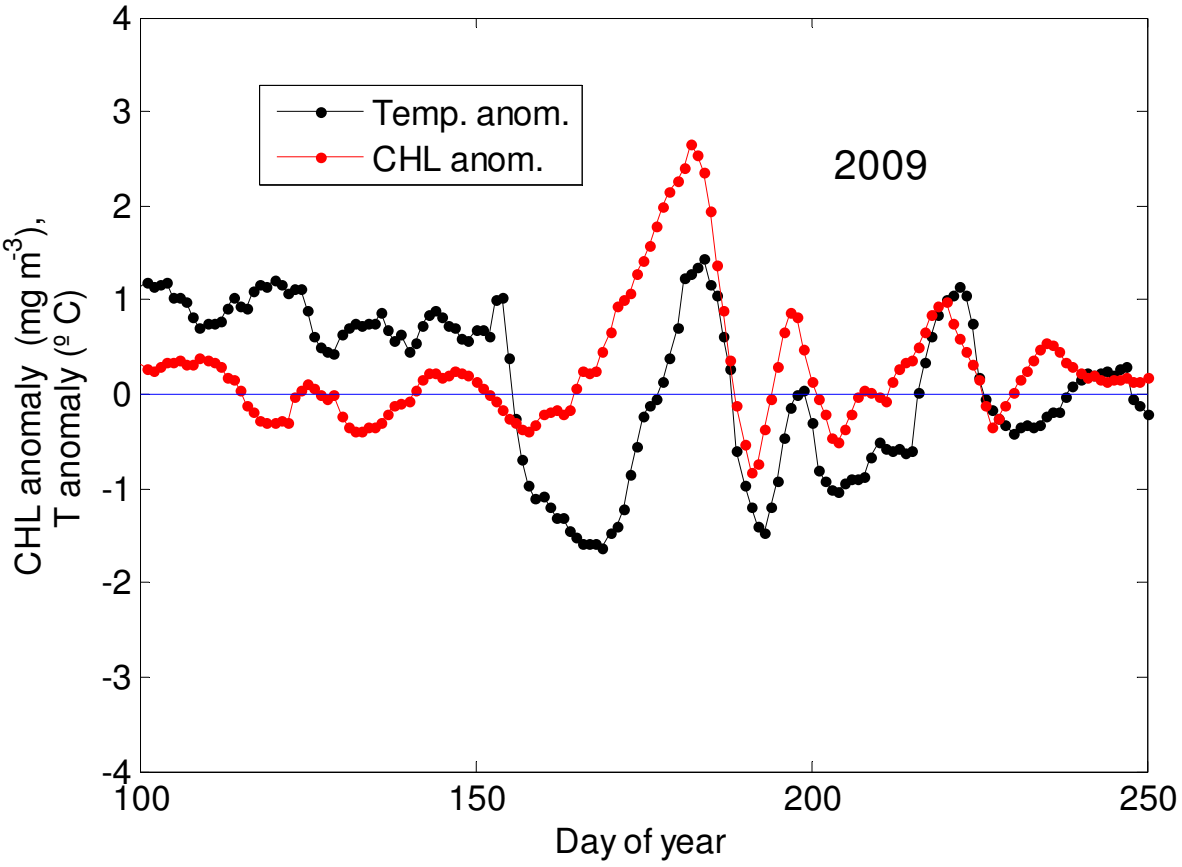


Fig. S31 Anomalies of CHL and SST over climatologies at the Central Baltic. Year 2009.