

Abstract

Studies of coastal seas in Europe have brought forth the high variability in the CO₂ system. This high variability, generated by the complex mechanisms driving the CO₂ fluxes makes their accurate estimation an arduous task. This is more pronounced in the Baltic Sea, where the mechanisms driving the fluxes have not been as highly detailed as in the open oceans. In addition, the joint availability of in-situ measurements of CO₂ and of sea-surface satellite data is limited in the area. In this paper, a combination of two existing methods (Self-Organizing-Maps and Multiple Linear regression) is used to estimate ocean surface $p\text{CO}_2$ in the Baltic Sea from remotely sensed surface temperature, chlorophyll, coloured dissolved organic matter, net primary production and mixed layer depth. The outputs of this research have an horizontal resolution of 4 km, and cover the period from 1998 to 2011. The reconstructed $p\text{CO}_2$ values over the validation data set have a correlation of 0.93 with the in-situ measurements, and a root mean square error is of 38 μatm . The removal of any of the satellite parameters degraded this reconstruction of the CO₂ flux, and we chose therefore to complete any missing data through statistical imputation. The CO₂ maps produced by this method also provide a confidence level of the reconstruction at each grid point. The results obtained are encouraging given the sparsity of available data and we expect to be able to produce even more accurate reconstructions in the coming years, in view of the predicted acquisitions of new data.

1 Introduction

The ocean plays an important role as a major carbon reservoir for carbon dioxide (CO₂) emitted to the atmosphere from fossil fuel burning, cement production, biomass burning, deforestation and other land use change. The ocean is at present acting to slow the rate of climate change by absorbing about 30 % of human emissions of CO₂ emitted to the atmosphere since the industrial revolution (Stocker et al., 2013). The exchange

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



of CO₂ between coastal environments and the atmosphere is significant for the global carbon budget (e.g. Borges et al., 2005; Chen and Borges, 2009; Laruelle et al., 2010). The estimation of the overall sink of CO₂ in continental shelf sea is $-0.22 \text{ Pg C yr}^{-1}$ (Laruelle et al., 2010), corresponding to 16% of the open oceanic sink (Takahashi et al., 2009), while continental shelves only represent 7% of the oceanic surface area and less than 0.5% of the ocean volume. These estimates are prone to a large uncertainty related to sparse data coverage in time and space. The monitoring of the oceanic partial pressure of CO₂ ($p\text{CO}_2$) at monthly and seasonal time scale is essential to estimate the regional and global air-sea CO₂ flux and improve this uncertainty. Due to technical as well as financial restrictions, in situ measurements of marine $p\text{CO}_2$ are sparse in spatial and temporal distribution. However, over the last decade, technical improvements and cooperation with the shipping industry have allowed for the installation of several autonomous underway systems on board commercial vessels routinely crossing the ocean basin. Those instruments perform quasi-continuous measurements, offering temporal and spatial coverage which allows for regional analysis of the highly variable spatial and temporal distribution of $p\text{CO}_2$ (e.g. Lefèvre et al., 2004; Lüger et al., 2004; Corbière et al., 2007; Schneider et al., 2003). In spite of the increased number of measurements in the Baltic Sea, the assessment of the carbon fluxes in the Baltic Sea remains particularly challenging due to the non-linearity of the emission and absorption system. Neural network techniques can be generally described as empirical statistical tools that resolve, to a certain degree, the nonlinear and often discontinuous relationships among proxy parameters without any a priori assumptions. In the past decade a handful of authors have reported the application of an neuronal network technique to basin-scale $p\text{CO}_2$ sea analysis (Lefèvre et al., 2005; Jamet et al., 2007; Friedrich and Oschlies, 2009; Telszewski et al., 2009), concentrating mainly on the North Atlantic Ocean. Most recently, (Telszewski et al., 2009) successfully applied a self-organizing-map (SOM) based neuronal network technique to reconstruct $p\text{CO}_2$ sea distribution in the North Atlantic (10.5° to 75.5° N, 9.5° E to 75.5° W) for three years (2004 to 2006) by examining non-linear/discontinuous relationship between $p\text{CO}_2$ sea

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



and ocean parameters of sea surface temperature (SST), mixed layer depth (MLD), and chlorophyll *a* concentration (CHL). One of the main benefits of this approach over the more traditional techniques, such as multiple linear regression (MLR), is that there are numerous empirical relationships established (e.g., 2220 in Telszewski et al., 2009) between examined parameters, allowing for more accurate representation of the highly variable system of interconnected water properties. In this paper, we departed from a similar, but rougher SOM classification of the available explicative oceanic parameters in the Baltic sea. Making the assumption that the explicative parameters of each of the obtained classes could locally be considered to be linearly related to the $p\text{CO}_2$, we performed a MLR of the $p\text{CO}_2$ from these explicative parameters in each of the retrieved classes. The purpose of the method developed is to use the classes and MLR parameters calculated for each class, to estimate the $p\text{CO}_2$ in the Baltic Sea and to improve the estimation of air-sea CO₂ fluxes in the future. The manuscript has been structured in four parts. In the first part we present a synopsis of the problem studied, including the existing studies performed on the reconstruction of the $p\text{CO}_2$ in other maritime regions. We followed that with a presentation of the available data and a brief description of the methodologies used. In the third part of the article we presented our results, namely the topological maps obtained as well as the reconstructions performed with them. We concluded the article with a discussion on the results obtained and future possible improvements.

2 Materials and methods

2.1 Study area

The Baltic Sea is a semi-enclosed sea with limited exchange with the North Atlantic through the North Sea–Skagerrak system. Previous investigations from the Baltic Proper showed large variability of $p\text{CO}_2$ in time and in space. The amplitude of the annual cycle of $p\text{CO}_2$ varies significantly depending on the region (Schneider and Kaitala,

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



system deployed on a cargo ship. The Leibniz Institute for Baltic Sea Research, Warnemünde, Germany (IOW Institut für Ostseeforschung Warnemünde) has made continuous measurements of $p\text{CO}_2$ at a depth of 5 m on board the cargo vessel Finnpartner. This ship crosses between Lübeck and Helsinki at a two-day interval, alternately crossing the eastern and western Gotland Sea (Schneider and Kaitala, 2006; Schneider et al., 2009). Data from Finnpartner were taken between July 2003 to December 2005.

- Swedish Meteorological and Hydrological Institute (SMHI) data set: pH (measured with the method from Grasshoff et al., 1999) and alkalinity (TA) (measured by potentiometric titration described in Grasshoff et al., 1999) are measured continuously at a monthly or semi-monthly resolution in the Baltic Sea at different stations. All data used are from 5 m depth below the surface. The uncertainty for the pH is ± 0.03 pH units and for the TA is $\pm 5\%$ (Wesslander et al., 2009). $p\text{CO}_2$ is estimated from the pH, TA, salinity and temperature measurements and used the standard CO₂SYS program (Lewis and Wallace, 1998) with the same equilibrium constant from Weiss (1974) and Merbach et al. (1973) as refitted by Dickson and Millero (1987) like in Wesslander et al. (2009).

To apply a neural network method to reconstruct the $p\text{CO}_2$ from 1998 to 2011, we used several satellite data set.

2.3 Remote sensing data

The satellite data used for this study have different sources. We used a monthly time resolution and a spatial resolution based from the lower spatial resolution of our products. The lower spatial resolution is the CDOM (Coloured Dissolved Organic Matter) product.

We used five parameters from different sources:

SST Sea Surface Temperature: for the SST several products are used, we combine two types of products for 2007 and 2011. For 2005–2011: we used the

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



data from Federal Maritime and Hydrographic Agency (BSH) processing the data from AVHRR-NOAA for 2005–2011 and the data from GRHSST (Group for High Resolution Sea Surface Temperature) product for Baltic Sea 2007–2011. The spatial resolution is 0.03° at daily scale (http://podaac.jpl.nasa.gov/dataset/DMI-L4UHfnd-NSEABALTIC-DMI_OI). From 1998–2004, the data is a re-analysis of the NOAA/NASA Advanced Very High Resolution Radiometer (AVHRR) data stream conducted by the University of Miami's Rosenstiel School of Marine and Atmospheric Science (RSMAS) and the NOAA National Oceanographic Data Center (NODC). It consists of 4 km monthly SST (in °C) extracted from version 5.2 of the AVHRR Pathfinder project (Casey et al., 2010, <http://www.nodc.noaa.gov/SatelliteData/pathfinder4km/>).

Chl Chlorophyll *a*: The dataset consists of monthly averages from the following sensors: SeaWiFS (September 1998–December 2002) 4 km monthly and MODIS-Aqua (July 2002–June 2011) 4 km monthly (Casey et al., 2010). A lognormal distribution was assumed for the Chl.

CDOM Coloured Dissolved Organic Matter come from MODIS data, at 4 km monthly average. The CDOM index quantifies the deviation in the relationship between CDOM and chl concentration, where 1.0 represents the mean relationship for Morel and Gentili (2009) case 1 waters, and values above or below 1.0 indicate excess or deficit in CDOM relative to that mean relationship, respectively. The algorithm and its application is fully described in Morel and Gentili (2009)

NPP Primary Production come from two products. The first one come from EMIS: This model is depth-integrated but allows for depth-dependent variability in the diffuse attenuation coefficient, which is calculated from a multiple-component semi-analytical inversion algorithm (Lee et al., 2005). The primary production calculation is based on the formulation obtained through dimensional analysis by Platt and Sathyendranath (1993). The assignment of the photosynthetic parameters is achieved by the combined use of a temperature-dependent relationship for the

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



maximum growth rate (Eppley, 1972) and a variable formulation to retrieve the C:Chl following the empirical relation of Cloern et al. (1995). The dataset in EMIS consists of monthly average values from October 1997 to September 2008. The second for 2009–2011 is from the Vertically Generalized Production Model (VGPM) (Behrenfeld and Boss, 2006) as the standard algorithm. The VGPM is a “chlorophyll-based” model that estimate net primary production from chlorophyll using a temperature-dependent description of chlorophyll-specific photosynthetic efficiency. For the VGPM, net primary production is a function of chlorophyll, available light, and the photosynthetic efficiency. Standard product are based on MODIS chlorophyll and temperature data, SeaWiFS PAR, and estimates of euphotic zone depth from a model developed by Morel and Berthon (1989) and based on chlorophyll concentration. A correction between this two product has be done for the maximum value.

MLD Mixed Layer Depth: There is also two sources for the MLD. At monthly averages from 1998 to 2007, come from one3D hydrodynamic model currently used at the JRC/IES is the public domain GETM model (General Estuarine Transport Model – <http://www.getm.eu>), which has its roots within developments at the JRC/IES (Burchard and Bolding, 2002). GETM simulates the most important hydrodynamic and thermodynamic processes in coastal and marine waters and includes flexible vertical and horizontal coordinate systems. Different turbulence schemes are incorporated from the GOTM (General Ocean Turbulence – <http://www.gotm.net>). Between 2008 and 2011, we used the data from Carbon-based Production Model at monthly scale (Behrenfeld et al., 2005).

Some corrections were applied for each parameter of the data to render the different products coherent between themselves.

In the Baltic Sea the satellite data have a lot of gaps, due to the high proportion coastal waters, where the satellite products are less reliable, and the frequent large-scale cloud coverage. To improve the number of our data we used a monthly time scale

of the following year, and to therefore be able to situate the process in relation to its seasonality.

In total, 1445 data of $p\text{CO}_2$ were used for this study. Each parameter (SST, Chl, CDOM, NPP and MLD) was located around each data of $p\text{CO}_2$. During winter (October to March), more data were missing (Table 1, column 1), in particular for Chl, CDOM and NPP which is the period when it's more difficult to measure or estimate these parameters. Between April and September, the number of data missing for SST, Chl, CDOM and MLD are relatively low compared to the total amount of data. Missing data represents less than 3%. To increase the number of data available, we completed the data with a training of the topological map. Further detailed is given in Sect. 2.5.

2.5 Methodology

The relationship between $p\text{CO}_2$ and the environmental parameters available is highly non-linear. As mentioned in the introduction, we chose to combine two statistical approaches: *self-organizing maps* (Kohonen, 1990; Dreyfus, 2005) and *linear regression*. The SOM are a subfamily of the neuronal networks algorithms, used to perform multidimensional classification. One of the defining characteristics of the SOMs is their ability of their classes to locally represent a gaussian distribution centred around its typical profile of the environmental parameters. We used this hypothesis to classify the environmental parameters dataset, and then estimate for each class the parameters of a linear regression. In the following section we will present an overview of the two statistical algorithms and their application to our data sets.

2.5.1 Self organizing maps

Self-organizing topological maps (SOM) is a clustering method based on neural networks. They provide a clustering of a learning data set into a reduced number of subsets, called classes, with common statistical characteristics.

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing algorithm for sea surface CO₂

G. Parard et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



The generation of a SOM requires the creation of a training database that contains homogenous vectors. After a training phase, we obtain a Self-Organizing Map. Each class is represented by its referent vector r_i which is an approximation of the mean value of the elements belonging to it and it index that positions it on the map in relation to the other classes (Fig. 4). The topological aspect of the maps can be justified by considering the map as an undirected graph on a two-dimensional lattice whose vertices are the N classes. This graph structure permits the definition of a discrete distance $d(C_i, C_j)$ between two classes C_i and C_j , defined as the length of the shortest path between C_i and C_j on the map. The nature of the SOM training algorithm forces a topological ordering upon the map and, therefore, any neighbouring classes C_i and C_j on the map ($d(C_i, C_j) = 1$) have referent vectors r_i and r_j that are close in the Euclidean sense in the data space.

Let us consider a vector x that is of the same dimensions and nature as the data used to generate the topological map; we can find the index of the class to which it is classified by choosing: $\text{index} = \arg \max_i (\|x - r_i\|)$, therefore assigning it to the class whose referent is closest to it in the Euclidean sense (Fig. 5). A classified vector x will be represented by its class index, C_{index} . In the case we are trying to classify a vector that has some missing values, the comparison is performed between the existing values of x and the corresponding values of each r_i .

As a version of the Expectation–Maximization algorithm the SOM algorithm perform an iterative training. During the early phases of this training, the referent vectors of each class are strongly affected by the changes imparted on their neighbours' referent vectors in order to capture the shape of the data cloud. Depending on the training parameters of the SOM, in the latter phases of the training, the effect of the neighbouring vectors on the determination of the referent vector can be considered null. In these cases, each referent vector approximates, locally, the mean value of a multidimensional Gaussian random distribution that generated the training data assigned to that class (Dreyfus, 2005).

2.5.2 Multiple linear regression

A multiple linear regression is a modelling method that expresses the value of one response variable y (in our study $p\text{CO}_2$) as a linear function of other explicative variables $X = X_1, X_2, \dots, X_i$ (in our study SST, Chl, CDOM, NPP, MLD, time_{sin} , time_{cos}). The purpose of performing a multiple linear regression generally is either to interpret the relationship of the variable y with each of the other predictive variables X_j , or to predict, from a dataset of vectors containing the values of X , the corresponding value of y . In this paper, we used both aspects of MLR.

However, in order to perform our MLRs, we had to take into account their limitations and the nature of our problem. More specifically, in order to perform a MLR we are obliged to assume that the relationship between the predictor variables and the response variable is *linear*. However, this is not the case in our data sets. The $p\text{CO}_2$ does not follow a linear relationship with the variables presented when considering the entirety of the problem presented. However as noted above in Sect. 2.5.1, if we consider the classes created by the SOM, they are very localised regions of the combined explicative and response data space that can be considered to approximate, locally, the mean value of a multidimensional Gaussian random distribution. We therefore make the assumption that, if performed in the reduced neighbourhood of a SOM class, the relationship between $p\text{CO}_2$ and the explicative variables is linear.

3 Application and results

3.1 Statistical imputation

As described in Sect. 2.4, the data available for the application were presenting many missing values. In order to complete them we chose to use an imputation method similar to those described by Schafer and Graham (2002) and Malek et al. (2008). The main idea of these methods is to use the classifying abilities of the SOMs in order to

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



regroup the data in typical situations and replace the missing values of the explicative data with the corresponding values of the referent vector of the class it belongs to.

We first selected the database containing, as mentioned before: SST, Chl, CDOM, NPP, MLD, time_{sin} , time_{cos} . We then sorted its vectors according to the number of missing values in them and noting the placement of missing values for each vector. We chose all complete data vectors and the first 5 % of the sorted vectors with missing data and trained a SOM. We proceeded by replacing the missing values of these initial 5 % with the corresponding values of the referent vector of the class they each belonged to. We performed the same process by iteratively increased by 5 % the amount of the vectors with missing values included in the training of a new SOM, and by replacing all missing values, even those completed in previous iterations, by the corresponding values of the referent vector of the class that each vector with missing values belonged to. A more detailed presentation of this method is the subject of another forthcoming article, but a schematic representation version of the imputation method used can be seen in Fig. 6.

After this imputation of the missing data through the iterative training give a good representation of the data as presented in Fig. 7. The repartition of $p\text{CO}_2$ (Fig. 7a) is well representative of the variability of the data with a large range of value. Some very high in particular local events like a coastal upwelling but most of the data are range between $180 \mu\text{atm}$ (value observed in summer) and $550 \mu\text{atm}$ (observed in winter). The SST (Fig. 7b) is well representative of the variability in the Baltic Sea with a maximum of observation between July and September in all the basin around 18°C (Siegel and Gerth, 2012). The NPP variability is quite homogenous except the pic at $10 \text{ mg C m}^{-2} \text{ d}^{-1}$ this is due to the first model which have is maximum of NPP at $10 \text{ mg C m}^{-2} \text{ d}^{-1}$. So the correction on the satellite data of NPP take into account this maximum.

The variability of chlorophyll gives a part of data with a low value and an other part with a higher value than 6 mg m^{-3} , which can be explain by the fact that the Baltic Sea is a narrow sea so the coast are quite important and the two bloom take place during

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



spring/ summer, during this period the value of chlorophyll can be very high, and the reconstruction give a mean value for this characteristic. A pic at 10 mg m^{-3} is observed on the chlorophyll data, this pic is not due to the reconstruction but to the maximum value in the satellite data file. The low level of MLD occurs in summer and in the model the minimum is 10 m depth, which appear to be around the minimum value observed on the Fig. 7f. Absorption by CDOM decreased with increased distance from the riverine source and reached a relatively stable absorption background in the open sea. Most of our data are more in open sea conditions, the value are quite low.

Once the data was completed we could combine the two datasets again and train the SOM used for the reconstruction of $p\text{CO}_2$.

3.2 $p\text{CO}_2$ estimation

3.2.1 Topological map

In this study we classified the explicative variables (SST, Chl, CDOM, NPP, MLD, time_{sin} , time_{cos}) into classes that share similar characteristics. We separated our data set in two parts: 90 % of the completed data (1300 vectors) were used for the training phase, with the remaining 10 % split into 5 % (72 vectors) for testing and 5 % for the validation of our method. We iteratively tested classifications with varying number of classes, and selected the parameters of our SOM based on the performances on the test dataset.

At the end of our optimization, we selected a SOM consisting of 77 classes. The number of observations captured by each class ranges from 0 to 38 (Fig. 8). The order of magnitude of the number of observations is constant throughout the SOM, and we can consider that it has been well deployed to represent the data space of the explicatory parameters. The presence of classes that did not capture any elements can be justified as preventive: they preserve the topological aspect of the SOM by preventing classes that are not similar enough from becoming neighbours.

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



In order to estimate the average concentration of $p\text{CO}_2$ in each class, its measurements associated with vectors consisting of SST, CDOM, NPP, MLD and CHL components were presented to the already trained SOM as input data (Fig. 9). The average value computed for the vectors belonging to each class corresponds to the average value of $p\text{CO}_2$ for that class.

In the final map, the distribution of $p\text{CO}_2$ is strongly dependent of the SST distribution with low values of $p\text{CO}_2$ correlating with high value of SST (Fig. 9). This is in agreement with the seasonal cycle of $p\text{CO}_2$, which is characterized by a large amplitude, ranging from high value in winter ($\approx 500 \mu\text{atm}$) and low value in summer ($\approx 150 \mu\text{atm}$) Wesslander (2011). From Schneider and Kaitala (2006), the high winter value of $p\text{CO}_2$ are a consequence of mixing with deeper water layer enriched in CO_2 which is in agreement with the distribution of the MLD (Fig. 9h) with the higher value during winter and autumn correlate with the high value of $p\text{CO}_2$. Wesslander (2011) explained also that it can be related to the mineralization, which exceed production in winter. Biological production starts in spring when sunlight and nutrients are sufficient. The chlorophyll begin to increase in March/April due to the spring phytoplankton bloom, which have for effect to reduce the $p\text{CO}_2$ during this period. The more intensive decrease take place in April and May which is consistent with the higher value of NPP (Fig. 9). The studies in the central Baltic Sea shows two summer minima, the first is in April/May and a second July/August which is resulting from a second production period. The higher variability are observed during this period with standard deviation between $39 \mu\text{atm}$ and $50 \mu\text{atm}$ from different region (Wesslander, 2011; Schneider and Kaitala, 2006).

3.2.2 Linear regression in the neurons

As mentioned in Sect. 2.5.2, in order to perform a MLR we are obliged to assume that the relationship between the predictor variables and the response variable is *linear*, which we could only take as a valid hypothesis when performing the MLR in the reduced neighbourhood of a SOM class, where the relationship between $p\text{CO}_2$ and the explicative variables can be assumed to be linear.

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



this identifiable point, it is quite easy to organize a system of flag. This flag can give an information about the quality in function of the position of the neurons. The difference obtained for $p\text{CO}_2$ in each neuron range from 0 to 56 μatm (Table 2), but 58 % of the values observed are under 30 μatm of difference. The difference for each parameters can be quite high for some parameter like SST with a maximum value of 1.9 °C, but most of the value are lower than 1 °C and CDOM range between 0 and 5.15 (Table 2). The other parameters have quite low variability like MLD range between 0 and 9.7 m. The average is between two or three time lower than the maximum value observed, which give low value for all the satellite parameters.

The validation data set give a quite good correlation ($R = 0.93$) with the reconstruction method (Fig. 12), the root mean square (RMS) is 36.7 μatm . 12 % have a value higher than 20 μatm and 45 % between 20 μatm and 30 μatm (Fig. 13). The characteristic in time, SST, MLD, CDOM, Chl and NPP, do not explain the difference observed in the reconstruction.

A reconstruction has been done with the satellite data from 1997 to 2011. The seasonal cycle is well reproduced and in agreement with other studies. The maximum is observed during winter with 437 μatm in average and 274 μatm in summer. This values are comparable to the average estimated in central Baltic with 500 μatm in summer and 150 μatm in winter (Wesslander, 2011). In April, the $p\text{CO}_2$ decrease due to the biology and increase slowly in September (Fig. 14).

A simple flag was constructed to monitor the reconstruction quality. The flag is equal to 1 for classes where the average difference is lower than 20 μatm , equal at 2 for a difference between 20 μatm and 30 μatm and equal at 3 for higher average differences. In the example shown here the values are high (3) so the confidence on the reconstruction is low but some point have flag of 1 or 2 (Fig. 14d–f) and the reconstruction is more reliable. On the geographic map (Fig. 14d–f) the values of 4 correspond at the presence of ice which is estimated with the satellite product of the National Snow and Ice and Ice Data Center based on NOAA Level 3 products (Njoku, 2007).

$p\text{CO}_2$ measurements, the colocalized satellite data were frequently incomplete. This lead us to having to complete our database using a novel imputation method based on SOMs.

In comparison, existing studies performed over the North Atlantic and North Pacific, based on at a minimum 10000 data points (which take into account all the data from SOCAT) to at maximum 800 000 data (e.g. Friedrich and Oschlies, 2009; Landschützer et al., 2013; Hales et al., 2012). Friedrich and Oschlies (2009) had a RMSE error of $19 \mu\text{atm}$. A similar study over the totality of the Atlantic Ocean give a RMSE of $17 \mu\text{atm}$ for the independent time series (Landschützer et al., 2013). Hales et al. (2012) had a RMS deviation of $20 \mu\text{atm}$ with a correlation coefficient of 0.81. The RMSE obtained in our study were higher than the previous study in Atlantic Ocean, but, taking into account the highly reduced amount of data available, the results presented are promising.

The organisation of the values different coefficients of the MLR over each class ascertained that all the parameters from satellite data are important to reconstructing the $p\text{CO}_2$ in the Baltic Sea, even if it is for some particular cases. Therefore the improvement of the satellite data availability could also improve the performance of our reconstruction.

The methodological approach presented could potentially be further developed when reconstructing spatial fields of $p\text{CO}_2$ by taking the information of the classes attributed to the neighbouring grid points of geographic study area, when selecting the class whose linear regression coefficients to use in the reconstruction of the $p\text{CO}_2$. This aspect is however dependent on the acquisition of additional in-situ measurements.

There exist many programs in place for the acquisition of new data. The data at Östergarnsholm site are still forthcoming, the year 2012 did not give too much data but 2013 and 2014 need to be validated. The SMHI station also continue in time could be also add to our data. The VOS transect are not yet available for 2012 to 2014 but this measurement will also in continue, and some data will soon be available. There also some data from ferry boat from Gothenburg (Gothenburg–Kemi–Oulu–Lübeck–Gothenburg). The Gothenburg transect is weekly <http://www.hzg.de/>

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



imperia/md/content/ferryboxusergroup/presentations/fb-ws2011_karlson.pdf. The first test was made in 2010 and 2011 so some data could be soon available. New measurement of $p\text{CO}_2$ began in 2012 at the Utö Atmospheric and Marine Research Station <http://en.ilmatieteenlaitos.fi/GHG-measurement-sites#Uto>.

5 Given the amount of new data soon to be available we remain optimistic that the comprehension and statistical modeling of the $p\text{CO}_2$ in the Baltic Sea will continue to improve in the coming years.

Acknowledgements. The authors would like to thank Tiit Kutser and Melissa Chierici for their valuable help with data collection. We would like to thank Sylvie Thiria for her advising over the SOM classifications and Marion Leduc-Leballeur for her help with the method to fill the gap on satellite data. We would like to thank Swedish National Space Board (nr 120/11:3) for financing the project.

References

15 Behrenfeld, M. J. and Boss, E.: Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass, *J. Mar. Res.*, 64, 431–451, 2006. 12262

Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M.: Carbon-based ocean productivity and phytoplankton physiology from space, *Global Biogeochem. Cy.*, 19, GB1006, doi:10.1029/2004GB002299, 2005. 12262

Bergstrom, S.: River runoff to the Baltic Sea: 1950–1990, *Ambio*, 23, 280–287, 1994. 12259

20 Borges, A. V., Delille, B., and Frankignoulle, M.: Budgeting sinks and sources of CO_2 in the coastal ocean: diversity of ecosystems counts, *Geophys. Res. Lett.*, 32, L14601, doi:10.1029/2005GL023053, 2005. 12257

Burchard, H. and Bolding, K.: GETM: A General Estuarine Transport Model; Scientific Documentation, European Commission, Joint Research Centre, Institute for Environment and Sustainability, 2002. 12262

25 Casey, K. S., Brandon, T. B., Cornillon, P., and Evans, R.: The past, present, and future of the AVHRR Pathfinder SST program, in: *Oceanography from Space*, Springer, 273–287, 2010. 12261

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Chen, T. A. and Borges, A. V.: Reconciling opposing views on carbon cycling in the coastal ocean: continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO₂, *Deep-Sea Res. Pt. II*, 56, 578–590, 2009. 12257
- 5 Cloern, J. E., Grenz, C., and Vidergar-Lucas, L.: An empirical model of the phytoplankton chlorophyll: carbon ratio-the conversion factor between productivity and growth rate, *Limnol. Oceanogr.*, 40, 1313–1321, 1995. 12262
- Corbière, A., Metz, N., Reverdin, G., Brunet, C., and Takahashi, T.: Interannual and decadal variability of the oceanic carbon sink in the North Atlantic subpolar gyre, *Tellus B*, 59, 168–178, 2007. 12257
- 10 Dickson, A. and Millero, F.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, *Deep-Sea Res. Pt. I*, 34, 1733–1743, 1987. 12260
- Dreyfus, G.: *Neural Networks: Methodology and Applications*, Springer, Berlin New York, 2005. 12264, 12265
- 15 Dutt, A. and Rokhlin, V.: Fast Fourier transforms for nonequispaced data, II, *Appl. Comput. Harm. A.*, 2, 85–100, 1995. 12263
- Eppley, R. W.: Temperature and phytoplankton growth in the sea, *Fish. B.-NOAA*, 70, 1063–1085, 1972. 12262
- Friedrich, T. and Oschlies, A.: Neural network-based estimates of North Atlantic surface pCO₂ from satellite data: a methodological study, *J. Geophys. Res.-Oceans*, 114, C03020, doi:10.1029/2007JC004646, 2009. 12257, 12273
- 20 Grasshoff, K., Kremling, K., and Ehrhardt, M.: *Methods of Seawater Analysis*, 3rd edn., VCH Publishers, 1999. 12260
- Greengard, L. and Lee, J.-Y.: Accelerating the nonuniform fast Fourier transform, *SIAM Rev.*, 46, 443–454, 2004. 12263
- 25 Hales, B., Strutton, P. G., Saraceno, M., Letelier, R., Takahashi, T., Feely, R., Sabine, C., and Chavez, F.: Satellite-based prediction of pCO₂ in coastal waters of the eastern North Pacific, *Prog. Oceanogr.*, 103, 1–15, 2012. 12273
- Hjalmarsson, S., Wesslander, K., Anderson, L. G., Omstedt, A., Perttilä, M., and Mintrop, L.: Distribution, long-term development and mass balance calculation of total alkalinity in the Baltic Sea, *Cont. Shelf. Res.*, 28, 593–601, 2008. 12259
- 30 Jamet, C., Moulin, C., and Lefèvre, N.: Estimation of the oceanic pCO₂ in the North Atlantic from VOS lines in-situ measurements: parameters needed to generate seasonally mean maps, *Ann. Geophys.*, 25, 2247–2257, doi:10.5194/angeo-25-2247-2007, 2007. 12257

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



- Kohonen, T.: The self-organizing map, *Proc. IEEE*, 78, 1464–1480, 1990. 12264
- Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J.: A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, *Biogeosciences*, 10, 7793–7815, doi:10.5194/bg-10-7793-2013, 2013. 12273
- Laruelle, G. G., Dürr, H. H., Slomp, C. P., and Borges, A. V.: Evaluation of sinks and sources of CO₂ in the global coastal ocean using a spatially-explicit typology of estuaries and continental shelves, *Geophys. Res. Lett.*, 37, L15607, doi:10.1029/2010GL043691, 2010. 12257
- Lee, Z.-P. P., Darecki, M., Carder, K. L., Davis, C. O., Stramski, D., and Rhea, W. J.: Diffuse attenuation coefficient of downwelling irradiance: an evaluation of remote sensing methods, *J. Geophys. Res.-Oceans*, 110, C02017, 2005. 12261
- Lefèvre, N., Watson, A. J., Olsen, A., Ríos, A. F., Pérez, F. F., and Johannessen, T.: A decrease in the sink for atmospheric CO₂ in the North Atlantic, *Geophys. Res. Lett.*, 31, 2004. 12257
- Lefèvre, N., Watson, A. J., and Watson, A. R.: A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data, *Tellus B*, 57, 375–384, 2005. 12257
- Lewis, E. and Wallace, D. W. R.: Program Developed for CO₂ System Calculations, Tech. rep., ORNL/CDIAC, US Department of Energy, Oak Ridge, Tennessee, 1998. 12260
- Lüger, H., Wallace, D. W., Körtzinger, A., and Nojiri, Y.: The pCO₂ variability in the midlatitude North Atlantic Ocean during a full annual cycle, *Global Biogeochem. Cy.*, 18, 2004. 12257
- Malek, M. A., Harun, S., Shamsuddin, S. M., and Mohamad, I.: Imputation of time series data via Kohonen self organizing maps in the presence of missing data, *Engin. Technol.*, 41, 501–506, 2008. 12266
- Merbach, C., Culbertson, C. H., and Hawley, J. E.: Measurements of the apparent dissociation constants of carbonic acid in seawater of atmospheric pressure, *Limnol. Oceanogr.*, 18, 897–907, 1973. 12260
- Morel, A. and Berthon, J.-F.: Surface pigments, algal biomass profiles, and potential production of the euphotic layer: relationships reinvestigated in view of remote-sensing applications, *Limnol. Oceanogr.*, 34, 1545–1562, 1989. 12262
- Morel, A. and Gentili, B.: A simple band ratio technique to quantify the colored dissolved and detrital organic material from ocean color remotely sensed data, *Remote Sens. Environ.*, 113, 998–1011, 2009. 12261

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Njoku, E.: AMSR-E/Aqua L2B Surface Soil Moisture, Ancillary Params, & QC EASE-Grids v002, Información digital, Actualizado diariamente, EEUU: National Snow and Ice Data Center, Boulder, Colorado, 2007. 12271

Norman, M., Raj Parampil, S., Rutgersson, A., and Sahlée, E.: Influence of coastal upwelling on the air-sea gas exchange of CO₂ in a Baltic Sea Basin, *Tellus B*, 65, doi:10.3402/tellusb.v65i0.21831, 2013. 12259

Omstedt, A., Elken, J., Lehmann, A., and Piechura, J.: Knowledge of the Baltic Sea physics gained during the BALTEX and related programmes, *Prog. Oceanogr.*, 63, 1–28, 2004. 12259

Omstedt, A., Gustafsson, E., and Wesslander, K.: Modelling the uptake and release of carbon dioxide in the Baltic Sea surface water, *Cont. Shelf. Res.*, 29, 870–885, 2009. 12259

Platt, T. and Sathyendranath, S.: Estimators of primary production for interpretation of remotely sensed data on ocean color, *J. Geophys. Res.-Oceans*, 98, 14561–14576, 1993. 12261

Rutgersson, A., Norman, M., Schneider, B., Pettersson, H., and Sahlée, E.: The annual cycle of carbon dioxide and parameters influencing the air–sea carbon exchange in the Baltic Proper, *J. Marine Syst.*, 74, 381–394, 2008. 12259

Schafer, J. L. and Graham, J. W.: Missing data: our view of the state of the art., *Psychol. Methods*, 7, 147–77, 2002. 12266

Schneider, B. and Kaitala, S.: Identification and quantification of plankton bloom events in the Baltic Sea by continuous *p*CO₂ and chlorophyll *a* measurements on a cargo ship, *J. Marine Syst.*, 59, 238–248, 2006. 12258, 12260, 12269

Schneider, B., Nausch, G., Nagel, K., and Wasmund, N.: The surface water CO₂ budget for the Baltic Proper: a new way to determine nitrogen fixation, *J. Marine Syst.*, 42, 53–64, 2003. 12257

Schneider, B., Kaitala, S., Raateoja, M., and Sadkowiak, B.: A nitrogen fixation estimate for the Baltic Sea based on continuous *p*CO₂ measurements on a cargo ship and total nitrogen data, *Cont. Shelf. Res.*, 29, 1535–1540, 2009. 12260

Schomberg, H. and Timmer, J.: The gridding method for image reconstruction by Fourier transformation, *IEEE T. Med. Imaging*, 14, 596–607, doi:10.1109/42.414625, 1995. 12263

Siegel, H. and Gerth, M.: Baltic Sea environment fact sheet Sea Surface Temperature in the Baltic Sea in 2011, HELCOM Baltic Sea Environment Fact Sheets, available at: <http://www.helcom.fi/baltic-sea-trends/environment-fact-sheets/>, 2012. 12267

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Stocker, T. F., Qin, D., Plattner, G.-K. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M.: Climate Change 2013. The Physical Science Basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change-Abstract for decision-makers, Tech. rep., Groupe d'experts intergouvernemental sur l'évolution du climat/Intergovernmental Panel on Climate Change-IPCC, C/O World Meteorological Organization, 7bis Avenue de la Paix, CP 2300 CH-1211 Geneva 2 (Switzerland), 2013. 12256

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A. O., Tilbrook, B., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R. R., and De Baar, H. J. W.: Climatological mean and decadal change in surface ocean $p\text{CO}_2$, and net sea–air CO₂ flux over the global oceans, Deep-Sea Res. Pt. II, 56, 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009. 12257

Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D. W. R., and Wanninkhof, R.: Estimating the monthly $p\text{CO}_2$ distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405–1421, doi:10.5194/bg-6-1405-2009, 2009. 12257, 12258

Weiss, R. F.: CO₂ in water and seawater: the solubility of a non ideal gas, Mar. Chem., 2, 203–215, 1974. 12260

Wesslander, K.: The carbon dioxide system in the Baltic Sea surface waters, Ph.D. thesis, University of Gothenburg, 2011. 12259, 12269, 12271

Wesslander, K., Omstedt, A., and Schneider, B.: Inter-annual variation of the air-sea CO₂ balance in the southern Baltic Sea and the Kattegat, in: EGU General Assembly Conference Abstracts, vol. 11, p. 8629, 2009. 12260

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

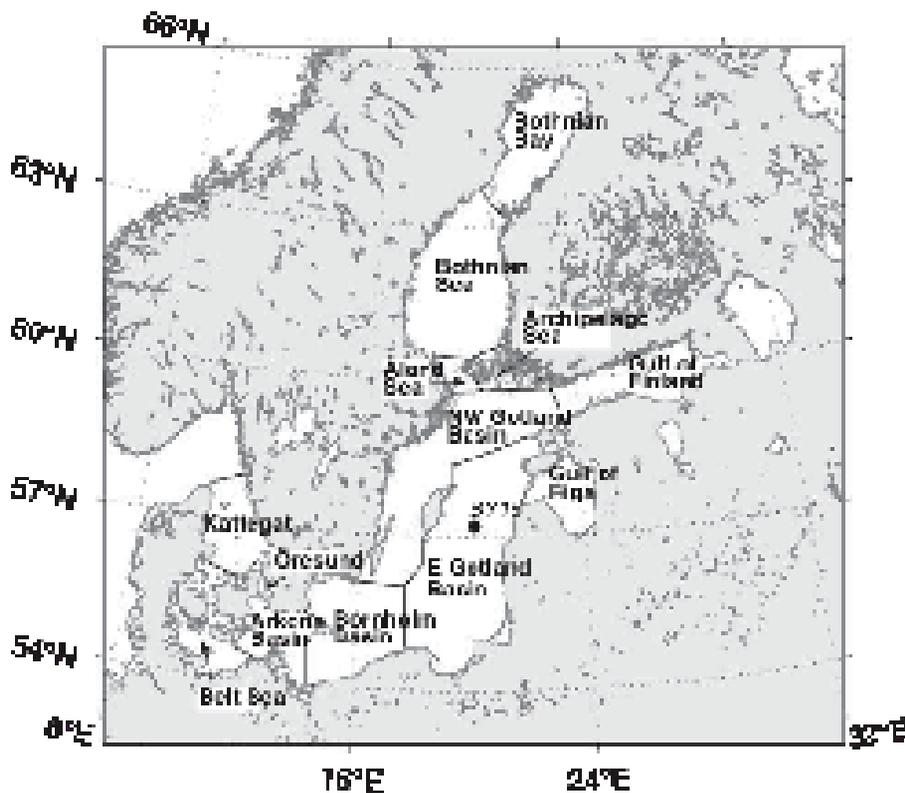
[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Figure 1. Map of the Baltic Sea. The major basins and sea regions are named.

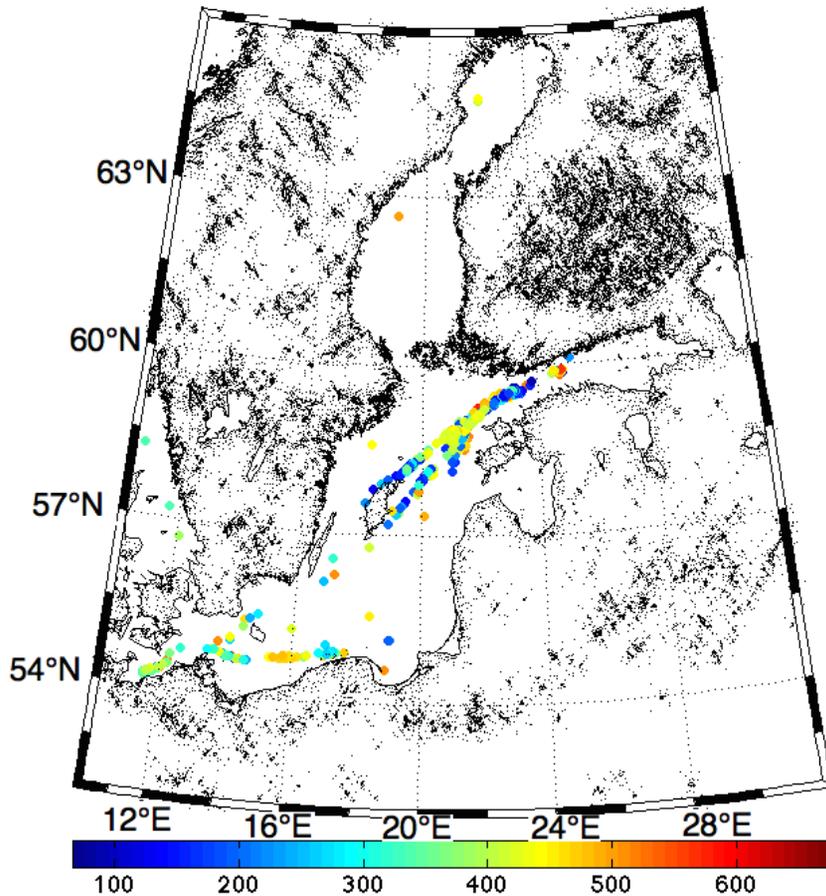


Figure 2. Monthly data available from 1998 to 2011 in Baltic Sea. The colorbar described the $p\text{CO}_2$ value in μatm .

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO_2

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



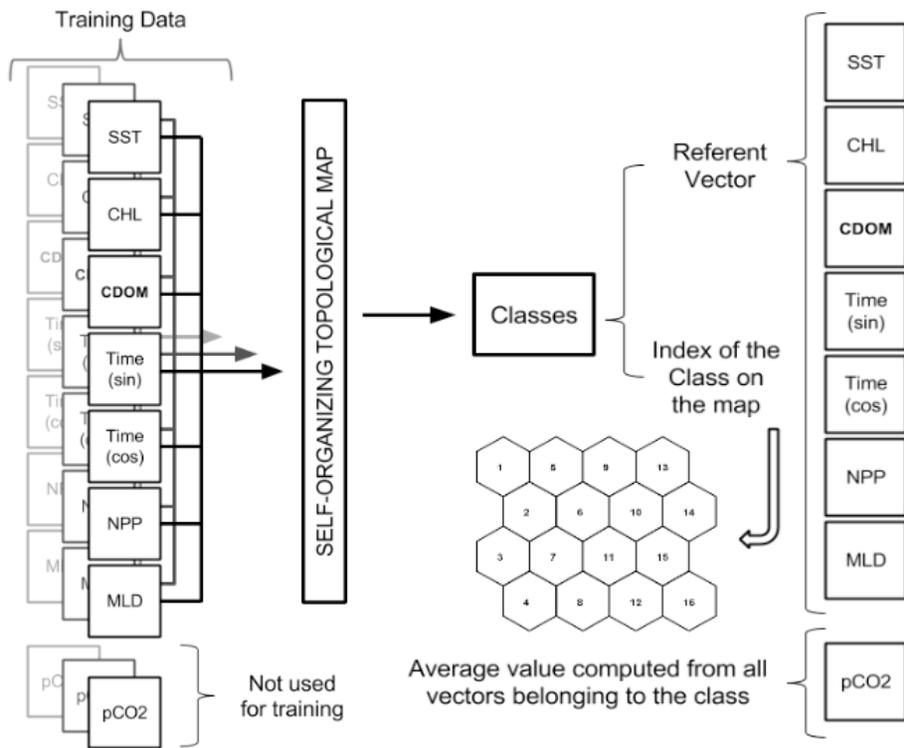


Figure 4. The different elements of the training of the self-organizing map.

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

[Back](#) [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



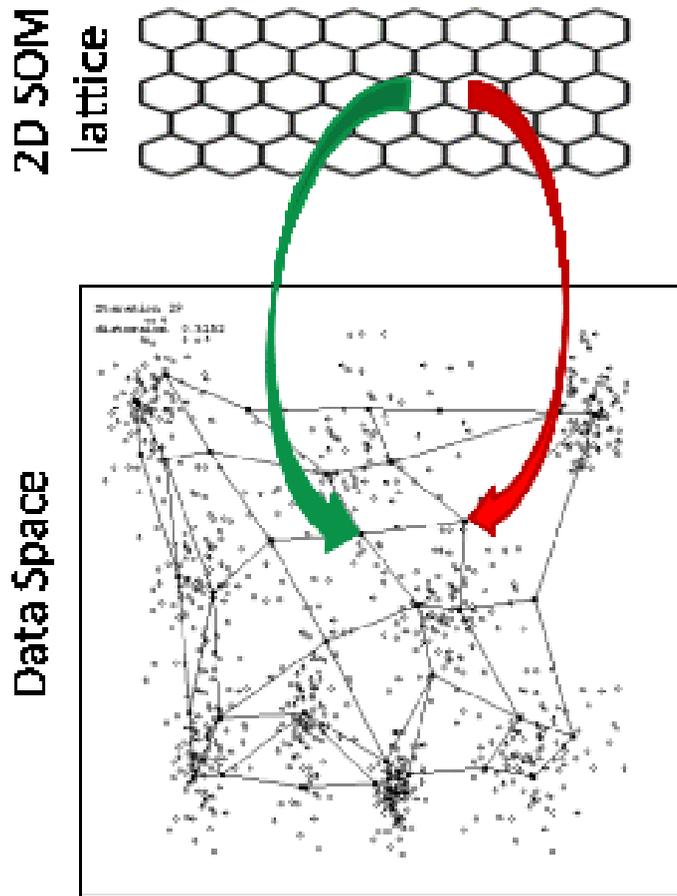


Figure 5. Adjacent classes on the SOM lattice correspond to adjacent areas in the multidimensional data space.

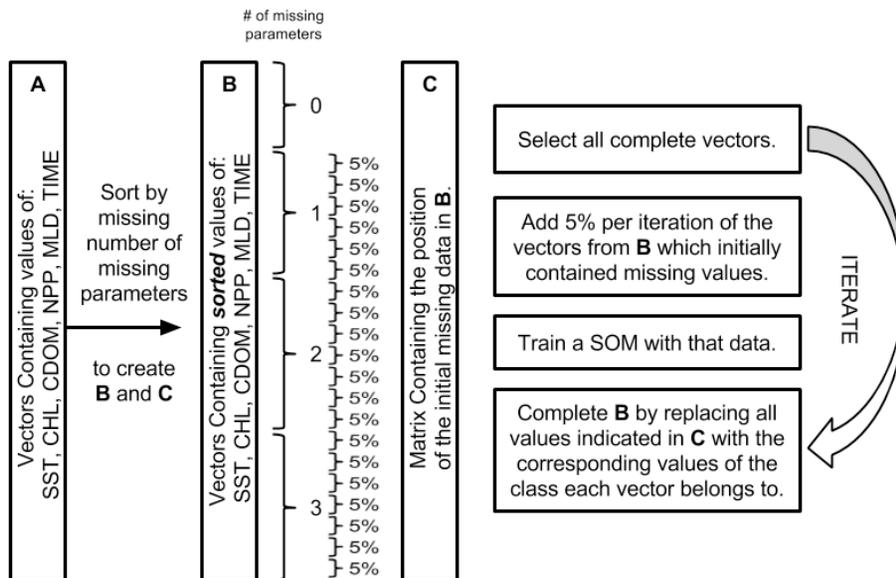


Figure 6. A schematic representation of the imputation method used.

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

[Back](#) [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Remote sensing
algorithm for sea
surface CO₂

G. Parard et al.

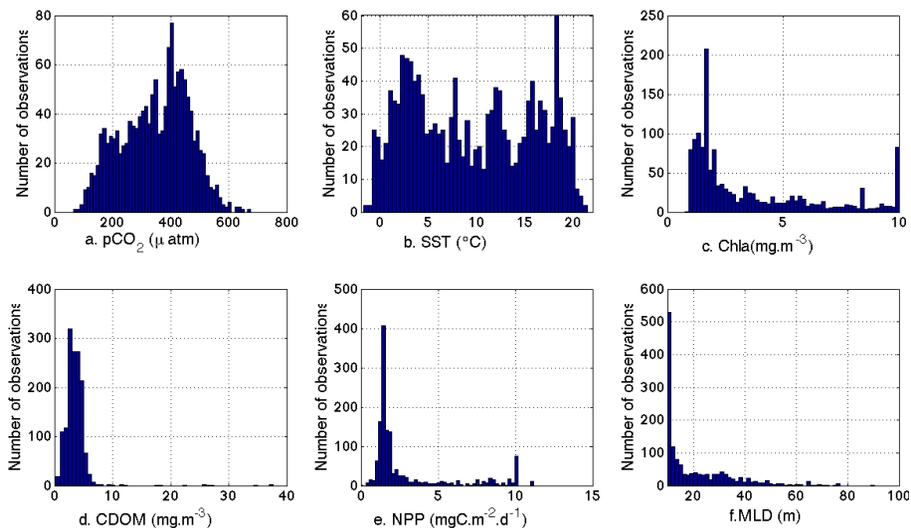


Figure 7. $p\text{CO}_2$ and satellite data (SST, Chl, NPP, CDOM and MLD) available for the SOM after reconstruction.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

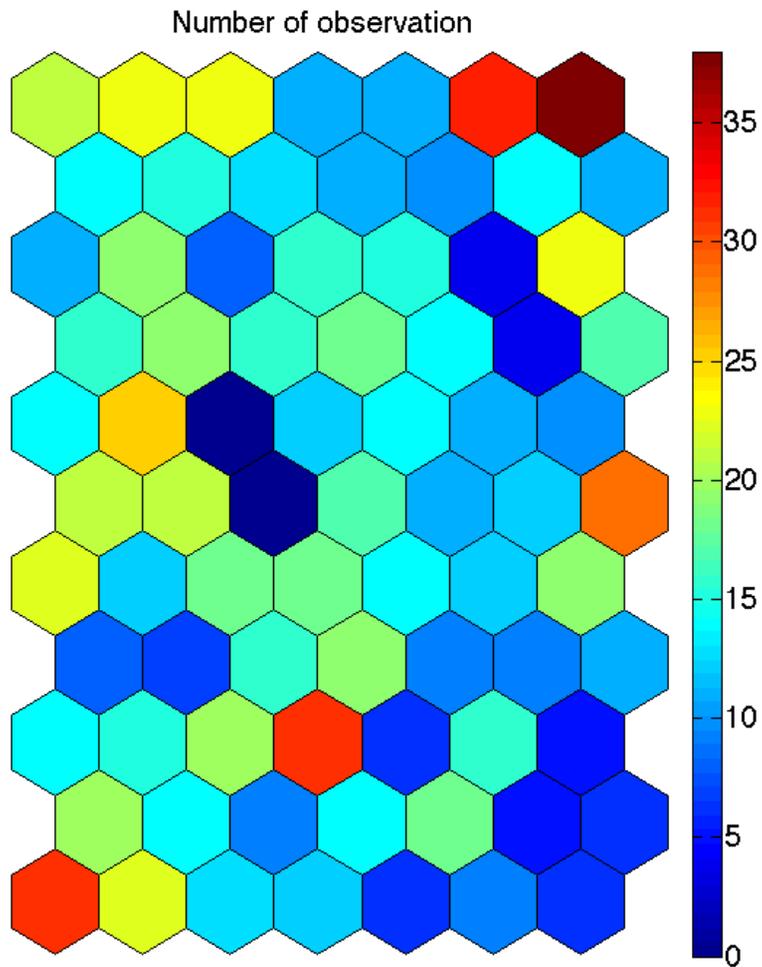


Figure 8. Number of observations attributed to each class. The colorbar corresponds at the number of data used in each neuron.

12288

BGD

11, 12255–12294, 2014

Remote sensing algorithm for sea surface CO₂

G. Parard et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing algorithm for sea surface CO₂

G. Parard et al.

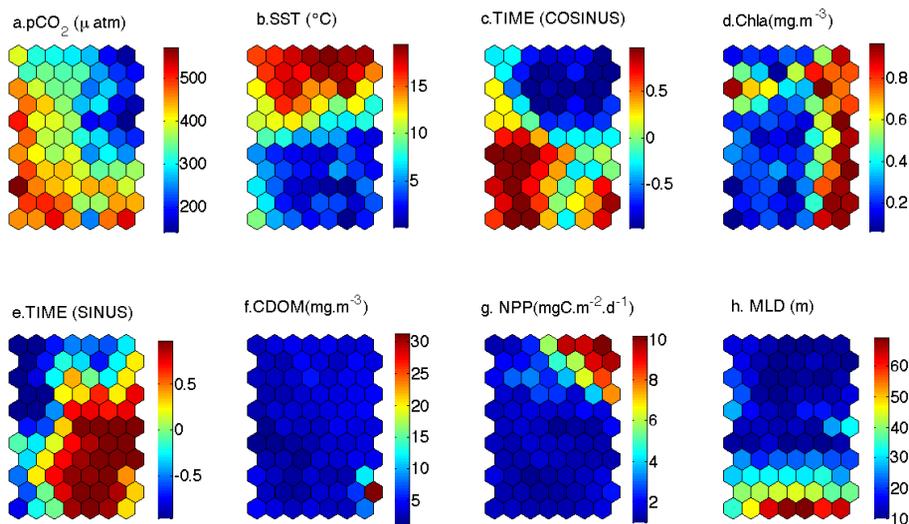


Figure 9. Distribution of each parameter in the neural map. **(a)** $p\text{CO}_2$ in μatm **(b)** SST in $^{\circ}\text{C}$ **(c and e)** time respectively cosine and sine, **(d)** Chl a in $\text{mg}\cdot\text{m}^{-3}$, **(g)** NPP in $\text{mg}\cdot\text{C}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ and **(h)** MLD in m.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing
algorithm for sea
surface CO₂

G. Parard et al.

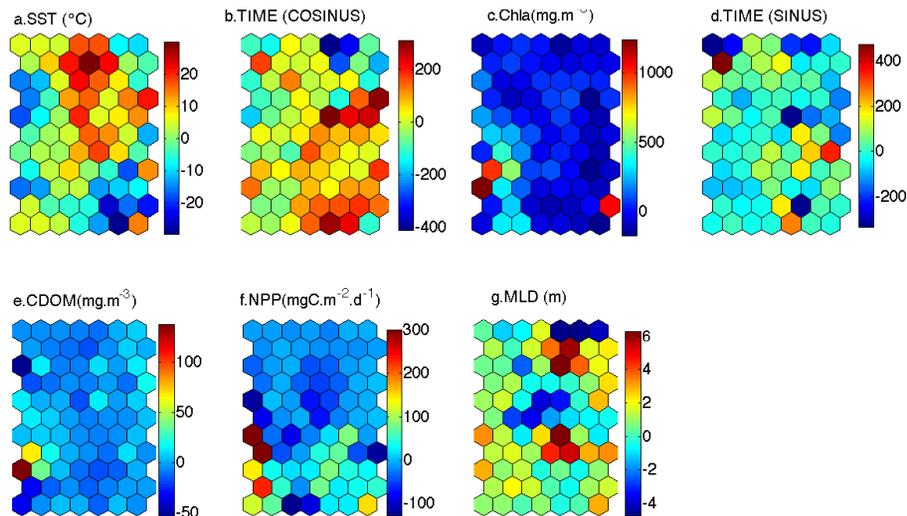


Figure 10. Coefficient from linear regression for each parameters. **(a)** SST in °C **(b and d)** time respectively cosine and sine, **(c)** Chl *a* in mg m^{-3} , **(f)** NPP in mg C m^{-2} and **(g)** MLD in m.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing
algorithm for sea
surface CO₂

G. Parard et al.

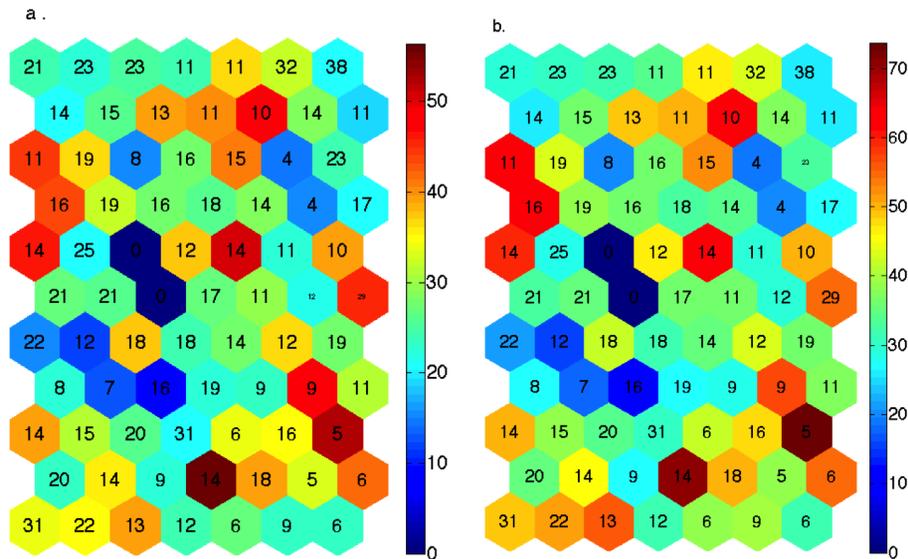


Figure 11. (a) The colorbar represents the average difference in each neuron for $p\text{CO}_2$. (b) The colorbar represents the standard deviation (std) in each neuron for $p\text{CO}_2$. For (a) and (b) the number inside the neurons correspond at the number of data.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Remote sensing
algorithm for sea
surface CO₂

G. Parard et al.

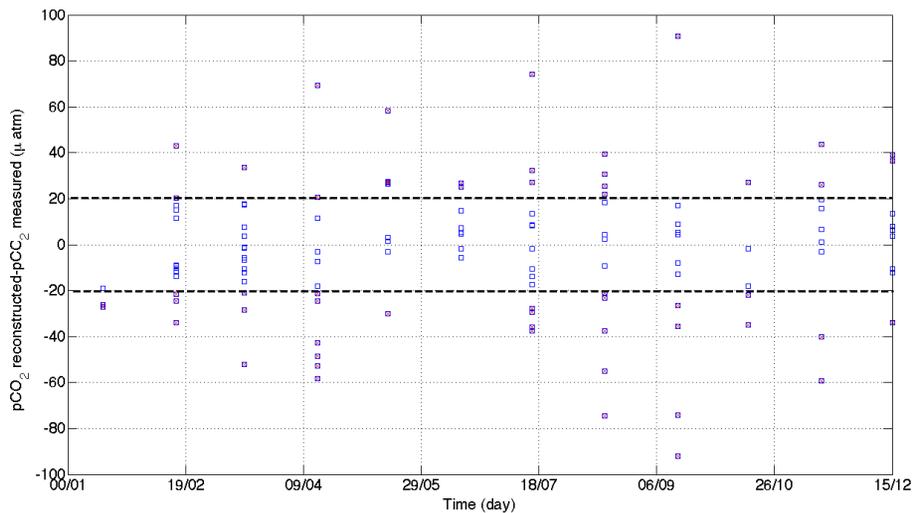


Figure 13. Difference between $p\text{CO}_2$ reconstructed and measured for the validation data set.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Remote sensing
algorithm for sea
surface CO₂

G. Parard et al.

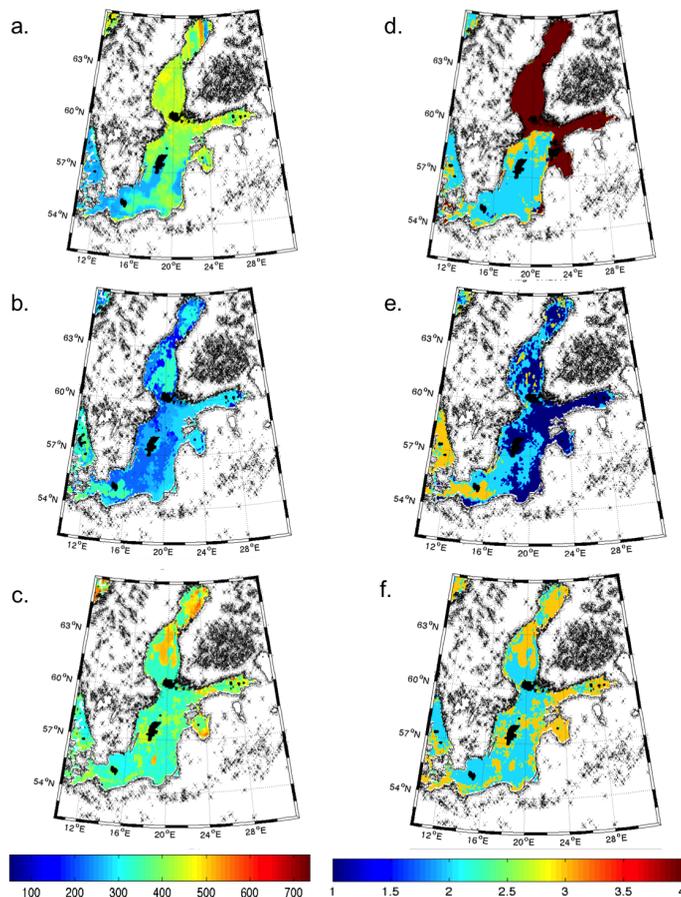


Figure 14. (a–c) Reconstruction of the $p\text{CO}_2$ map and (d–f) the flag for each map in (a and d). March 2010 (b and e). July 2010 (c and f) September 2010. The flag value correspond to: 1: high confidence, 2: medium confidence, 3: poor confidence.