Global high resolution monthly pCO_2 climatology for the coastal ocean derived from neural network interpolation

Running head: Global coastal pCO_2 maps

Goulven G. Laruelle[1], Peter Landschützer[2], Nicolas Gruber[3], Jean-Louis Tison[1], Bruno Delille[4], Pierre Regnier[1]

1. Department Geoscience, Environment & Society (DGES), Université Libre de Bruxelles, Belgium
2. Max Planck Institute for Meteorology, Hamburg, Germany
3. Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland
4. Unité d'Oceanographie Chimique, Astrophysics, Geophysics and Oceanography department, University of Liège, Belgium

Corresponding author: Goulven G. Laruelle

Revised version of manuscript bg-2017-64 (Minor revisions)

Abstract

In spite of the recent strong increase in the number of measurements of the partial pressure of CO_2 in the surface ocean (pCO_2), the air-sea CO_2 balance of the continental shelf seas remains poorly quantified. This is a consequence of these regions remaining strongly under-sampled both in time and space, and of surface pCO_2 exhibiting much higher temporal and spatial variability in these regions compared to the open ocean. Here, we use a modified version of a two-step artificial neural network method (SOM-FFN, Landschützer et al., 2013) to interpolate the pCO_2 data along the continental margins with a spatial resolution of 0.25 degrees and with monthly resolution from 1998 until 2015. The most important modifications compared to the original SOM-FFN method are (i) the much higher spatial resolution, and (ii) the inclusion of sea-ice and wind speed as predictors of pCO_2. The SOM-FFN is first trained with pCO_2 measurements extracted from the SOCATv4.0 data base. Then, the validity of our interpolation, both in space and time, is assessed by comparing the generated pCO_2 field with independent data extracted from the LDVEO2015 data base. The new coastal pCO_2 product confirms a previously suggested general meridional trend of the annual mean pCO_2 in all the continental shelves with high values in the tropics and dropping to values beneath those of the atmosphere at higher latitudes. The monthly resolution of our data product permits us to reveal significant differences in the seasonality of pCO_2 across the ocean basins. The shelves of the western and northern Pacific, as well as the shelves in the temperate North Atlantic display particularly pronounced seasonal variations in pCO_2, while the shelves in the southeastern Atlantic and in the South Pacific reveal a much smaller seasonality. The calculation of temperature normalized pCO_2 for several latitudes in different oceanic basins confirms that the seasonality in shelf pCO_2 cannot solely be explained by

temperature-induced changes in solubility, but are also the result of seasonal changes in circulation, mixing, and biological productivity. Our results also reveal that the amplitudes of both thermal and non-thermal seasonal variations in pCO_2 are significantly larger at high latitudes. Finally, thanks to this product having been extended to cover open ocean areas as well, it can be readily merged with existing global open ocean products to produce a true global perspective of the spatial and temporal variability of surface ocean pCO_2.

1. Introduction

The quantitative contribution of the coastal ocean to the global oceanic uptake of atmospheric CO_2 is still being debated (Borges et al., 2005; Chen and Borges, 2009; Cai, 2011; Wanninkhof et al., 2013; Gruber, 2015), but several recent studies have suggested that the flux density, or uptake per unit area, is greater over continental shelf seas than over the open ocean (Chen et al., 2013; Laruelle et al., 2014). Laruelle et al. (2014) used more than $3 \cdot 10^6$ pCO_2 measurements from the SOCATv2 database (Pfeil et al., 2014 Bakker et al., 2016) to demonstrate very strong disparities in air-seawater CO_2 exchange at the regional scale as well as pronounced seasonal variations, especially at temperate latitudes. Furthermore, it was suggested that, despite the presence of a seasonally varying sea-ice cover, Arctic continental shelves are a regional hotspot of CO_2 uptake (Bates et al., 2006; Laruelle et al., 2014; Yasunaka et al., 2016). Yet, even with this much larger dataset compared to previous studies, large regions of the global coastal ocean remained either void of data or very poorly monitored in space and time, including the seasonal cycle. These data gaps not only limit our ability to reduce uncertainties in flux estimates and to unravel whether they differ from the adjacent open ocean, but also hamper the identification and quantification of the many processes controlling the source-sink nature of the coastal ocean (Bauer et al., 2013). Laruelle et al., (2014) attempted to overcome this limitation by combining various upscaling methods depending on data density in different regions, e.g., resorted to using annual means, wherever the seasonal coverage was deemed to be insufficient. But they could not overcome the limitation that the data alone are insufficient to assess whether there are any trends in coastal fluxes. This is a serious gap when considering that the influence of human activity on coastal system is increasing rapidly (Doney, 2010; Cai, 2011; Regnier et al., 2013; Gruber, 2015).

In the open ocean, novel statistical methods relying on artificial neural networks (ANNs) have permitted the generation of a series of high-resolution continuous monthly maps for ocean surface CO_2 partial pressures (pCO_2) (e.g., Landschützer et al., 2013; Sasse et al., 2013; Nakaoka et al., 2013; Zeng et al., 2014). Although differing in their details (see e.g., Rödenbeck et al., 2015 for an overview), these products typically have a nominal spatial resolution of 1-degree and monthly temporal resolution. By filling in the spatial and temporal gaps, these products greatly facilitate the calculation of the air-sea CO_2 exchange, as they do not require separate assumptions about the surface ocean pCO_2 in areas lacking data. Such methods are also well suited to resolve spatial gradients, and they also permit to determine seasonal and inter-annual variations and trends in pCO_2 (e.g., Landschützer et al., 2014, 2015, 2016; Zeng et al., 2014). Because of the small relative contribution of the coastal ocean to the total oceanic surface area and the relatively coarse spatial resolution of the ANN-based surface ocean pCO_2 products so far, they are not well suited to resolve the high spatio-temporal variations of the surface ocean pCO_2 fields along the shelves.

Reproducing the complex seasonal dynamics of the CO_2 exchange at the air-water interface in the coastal ocean is of particular importance considering that they often have large intra-annual variability (Signorini et al., 2013). For instance, in temperate climates, it is common for continental shelf waters to turn from CO_2 sinks for the atmosphere during spring to CO_2 sources during summer (Shadwick et al., 2010; Cai, 2011; Laruelle et al., 2014, 2015). Shelf waters are also typically characterized by small-scale physical features such as coastal currents, river plumes and eddies inducing sharp biogeochemical fronts (Liu et al., 2010) that markedly influence the spatial patterns of the pCO_2 fields (e.g., Turi et al., 2014).

To resolve the high spatial and temporal variability in air-sea CO_2 exchange over the global shelf region, the two step artificial neural network method developed by Landschützer et al. (2013) is modified here for the specific conditions that prevail in these environments. Our calculations are performed at a much finer resolution of 0.25 degree and new environmental drivers such as sea ice cover are used at high latitudes to account for the potentially significant role of sea-ice in the CO_2 exchange (Bates et al., 2006; Vancoppenolle et al., 2013; Parmentier et al., 2013; Moreau et al., 2016; Grimm et al., 2016). The definition of the coastal/open oceanic boundary varies strongly from one study to the other (Walsh, 1988; Laruelle et al., 2013), with a potentially large impact on the shelf CO_2 budget (Laruelle et al., 2010). Here, we use a very wide definition for this boundary (i.e., 300km width or 1000m depth) to secure spatial continuity between our new shelf pCO_2 product and those already existing for the open ocean (Landschützer et al., 2013, 2016; Rödenbeck et al., 2015). Our approach leads to the first continuous and monthly resolved pCO_2 maps over the 1998-2015 period across the global shelf region, permitting us to study the seasonal dynamics of these regions in relationship to that of the adjacent open ocean.

2. Methods

The method used in this study is a modified version of the SOM-FFN method developed by Landschützer et al. (2013) to calculate monthly-resolved pCO_2 maps of the Atlantic Ocean at a 1 degree resolution and later applied to the entire global open ocean (Landschützer et al., 2014). The reconstruction of a continuous pCO_2 field involves establishing numerical relationships between pCO_2 and a number of independent environmental predictors that are known to control its variability both in time and space. The first step of the method relies on the use of a neural network clustering algorithm (Self Organizing Map, SOM) to define a discrete set of biogeochemical provinces characterized by similar relationships between the independent environmental variables and a monthly resolved pCO_2 field. The second step consists in deriving non-linear and continuous relationships between pCO_2 and some or all of the aforementioned independent variables using a feed-forward network (FFN) method, within each biogeochemical province created by the SOM. The method is extensively documented in Landschützer et al. (2013, 2014) but the specific modifications introduced in this study to better simulate the characteristics of the shelves, the choice of environmental drivers and their data sources as well as the definition of the geographic extent of this analysis are described in the following sections. Figure 1 summarizes the different steps involved in the calculations of the SOM-FFN.

2.1. Data Sources and processing

All the datasets used in our calculations were converted from their original spatial resolutions to a regular 0.25 degree resolution grid. The temporal resolution of all datasets is monthly (i.e., 216 months over the entire period), except for the bathymetry that is assumed constant over the course of the simulations and wind speed whose original resolution is 6 hours. For the latter, monthly averages are calculated for each grid cell to generate monthly values. SST and SSS maps were taken from the Met Office's EN4, which consists of quality controlled subsurface ocean temperature and salinity profiles and their objective analyses (Good et al., 2009). The bathymetry was extracted from the global ETOPO2 database (US Department of Commerce, 2006). The sea ice concentrations was taken from the global 25 km resolution monthly data product compiled by the NSIDC (National Snow and Ice Cover Data; Cavalieri et al., 1996). Wind speed data were extracted from ERA-Interim reanalysis (Dee et al., 2011). The chlorophyll surface concentrations were extracted from the monthly 9 km resolution SeaWIFS data product prior to 2010 and from MODIS for later years (NASA, 2016). The list of all data products used in the calculations as well as the transformations applied to produce monthly 0.25 degrees resolution forcing files are summarized in table 1.

Finally, the surface ocean pCO_2 were taken from the gridded SOCATv4 product (Sabine et al., 2013; Bakker et al., 2016) while those used for the validation stem from the LDEOv2015 database (Takahashi et al., 2016)., With our definition of the coastal zone, SOCATv4 contains ~8 10^6 data points and LDEO ~5.6 10^6,, with over 70% of the data shared with SOCATv4. Because of this significant overlap between both data products, we created two entirely independent datasets by randomly assigning each of those common data point to either database to insure that each data only belongs to one dataset. The resulting datasets are named SOCAT* and LDEO*, respectively, with the former being used for training and the latter for validation. Prior to the creation of both datasets, all data from SOCAT were converted from fCO_2 (fugacity of CO_2 in water) to pCO_2 using the formulation reported in Takahashi et al. (2012). The data densities of SOCAT* and LDEO* are shown on Fig. 2 and reveal a heterogeneous spatial coverage. Northern temperate shelves are generally well covered, especially in the North Atlantic. In this region, the data density is better in SOCAT* than LDEO* thanks to the addition of many European cruises in the SOCAT database. On the other hand, equatorial regions remain under-sampled, especially in the Indian Ocean. Because of the difficulty of sampling in waters seasonally covered in ice, Polar Regions are very unevenly represented in SOCAT* and LDEO*. Luckily, some areas, such as some parts of

Antarctica and the Bering Sea do contain enough data to train and validate the SOM-FFN. Overall SOCAT* contains roughly 40% more data than LDEO*.

2.2. Modifications of the SOM-FFN method

The specific characteristics of the continental shelves motivated a number of modifications of the global ocean SOM-FFN method, including a 16-fold increase in spatial resolution from 1 degree to 0.25 degree, the addition of new environmental variables as biogeochemical predictors, and a shortening of the simulation period to the period 1998 through 2015. All these modifications are detailed here below.

The higher resolution of 0.25°×0.25° results in over 2 million grid cells that help to better track the global coastline and its complex geomorphological features (Crossland et al., 2005; Liu, 2010). It is also common along Eastern and Western boundary currents to find continental shelves as narrow as 10-20 km, i.e., an extension that is significantly smaller than a single cell at 1-degree resolution. Additionally, biogeochemical fronts associated with river plumes, coastal currents and upwelling are characterized by spatial scales of the order of tens of kilometers or even smaller (Wijesekera et al., 2003). The chosen resolution is also identical to the gridded coastal pCO_2 product from the SOCAT initiative (Sabine et al, 2013, Bakker et al., 2014).

The definition of the geographic extent of the shelf region excludes estuaries and other inland water bodies, but uses a wide limit for the outer continental shelf that encapsulates all current definitions of the coastal ocean. This approach facilitates future integration with existing global ocean data products (e.g., Landschützer et al., 2016; Rödenbeck et al., 2015) and model outputs, which typically struggle to represent the shallowest parts of the ocean

(Bourgeois et al., 2016). The outer limit used here is given by whichever point is the furthest from the coast: either 300 km distance from the coastline (which roughly corresponds to the outer edge of territorial waters (Crossland et al., 2005)) or the 1000 m isobaths (Laruelle et al., 2013). The resulting domain (Fig SI B) covers 77 million km^2, more than twice the surface area generally attributed to the coastal ocean (Walsh et al., 1998; Liu et al., 2010; Laruelle et al., 2013).

The predictor variables for the SOM-FFN networks were chosen based on a set of trial-and-error experiments with the selection criteria being the quality of fit, i.e., the best reconstruction of the available observations. The first step of the SOM-FFN calculations, i.e., the self-organizing map-based clustering (SOM) relies on the assignment of the surface ocean data to biogeochemical provinces sharing common spatio-temporal patterns of sea-surface temperature (SST), sea-surface salinity (SSS), bathymetry, rate of change in sea ice coverage, wind speed and observed pCO_2. Chlorophyll a is not included in the list of environmental factors used to generate the biogeochemical provinces because of the incomplete data coverage at high latitude in winter due to cloud coverage. Both the use of wind speed and the rate of change in monthly sea ice concentration are novelties compared to the set-up of Landschützer et al. (2013). The latter is calculated from the gridded monthly sea ice concentration field of Cavalieri et al. (1996). It allows accounting for the complex processes occurring in melting and forming sea ice that are known to strongly influence the dynamics of the carbon within sea-ice covered areas (Parmentier et al., 2013). This first step is performed without any data normalization of the datasets, as this permits us to give more weight to the pCO_2 data. Based on a series of simulations using different numbers of biogeochemical provinces, we found that a clustering of the data into 10 biogeochemical provinces minimized the average deviation between simulated and observed pCO_2 (see below) while insuring that at least 1000 different grid cells can be used for validation against LDEO* in each province.

In the second step of the estimation procedure, i.e., the application of the feed-forward network method (FFN), SST, SSS, bathymetry, sea-ice concentration and chlorophyll a are used as predictors to establish the non-linear relationships between these predictors and the target pCO_2 (for data sources, see below). Similar to the SOM in step one, the selected variables not only comprise proxies representing the solubility and biological pumps of the coastal ocean, but also yield the best fit to the data. These calculations are done iteratively using a sigmoid activation function on an incomplete dataset in order to perform an assessment on the remaining data after each iteration, until an optimal relationship is found. Additionally, as performed in Landschützer et al. (2015), the output pCO_2 data were smoothed using the spatial and temporal mean of each point's neighboring pixels both in time and space within the 3 pixel neighborhood domain. This operation is performed iteratively and does not significantly alter the results, but it ensures smoother transitions in the pCO_2 field at the boundaries between the provinces. This smoothing method yielded good results for the open Southern Ocean where marked pCO_2 fronts are also observed (Landschützer et al., 2015) and was deemed relevant here due to the potentially strong pCO_2 gradients characterizing the shelves.

Another change from the most recent global ocean SOM-FFN application (Landschützer et al., 2016) is the different temporal extension of the simulation period, which covers the period from 1998 through 2015, instead of 1982 through 2011. This overall shortening was necessary because one of environmental driver, i.e., chlorophyll data derived from SeaWIFS, only starts in September 1997 (NASA, 2016). Monthly chlorophyll data throughout the entire simulation period was preferred here over the use of a monthly climatology as done in Landschützer et al. (2016) to better capture inter-annual variability. At the same time, we have been able to extend the coastal product by 4 years to the end of 2015.

2.3. Model training and evaluation

We evaluated the coastal SOM-FFN product using the root mean squared error (RMSE) metric, calculated as the difference between estimated and observed pCO_2. During the early development stage, preliminary simulations were performed using only data from SOCAT v2.0 (Pfeil et al., 2013, Sabine et al. 2013) to train the FFN algorithm. Each simulation was carried out using different subsets of environmental predictors extracted from the complete set (SST, SSS, bathymetry, sea ice concentration and chlorophyll a). The results obtained were then compared to the more complete dataset of SOCAT*, which contain 40% more shelf pCO_2 measurements from 1998 through 2015 (Bakker et al., 2016). This process allowed, for each province, to calculate the RMSE for several combinations of independent predictor variables for the pCO_2. Next, the combinations of predictors displaying the lowest RMSE were kept for the final simulations, which then used all data from SOCAT*. Thus, the pCO_2 calculations in each province potentially rely on a different set of predictors (Table 1).

The coastal SOM-FFN results are validated through a comparison with the LDEO* dataset through the calculation of residuals and RMSE. Additionally, a model-to-model comparison is also performed with the global ocean results of Landschützer et al. (2016) in the regions where the domains overlap. To perform this latter analysis, the coastal high resolution coastal pCO_2 product generated here was aggregated to a regular monthly 1° resolution to match the grid used by Landschützer et al. (2016).

Finally, the ability of the coastal SOM-FFN to capture seasonal variations is assessed by comparing the cell-average simulated monthly pCO_2 to monthly means for cells extracted from the LDEO* database. The cells retained for this analysis are all those for which the average for each month could be calculated from measurements performed on at least three different years.

3. Results and discussion

3.1. Biogeochemical provinces

Despite the fact that the SOM is not given any prior knowledge regarding space and time, the spatial distribution of the 10 biogeochemical provinces is mostly controlled by latitudinal gradients and distance from the coast (Figure 3; high-resolution monthly maps are also available in the supplementary information (SI)). Although the exact spatial extent of each province varies from one month to the other following the seasonal variations of the environmental forcing parameters, each province roughly corresponds to one type of climatological setting. Nevertheless, because of these spatial migrations, most cells belong to different provinces depending on the month (see figure SI B). These seasonal migrations are mostly driven by changes in temperature, sea-ice cover, pCO_2 and, to a lesser degree, salinity. P1, P2 (Province 1, etc.) and P4 are three of the largest provinces, covering a total of $35.7 \cdot 10^6$ km^2 and representing warm tropical regions with bottoms at shallow to intermediate depths. During summer, the spatial coverage of P4 expands north- and southward as a consequence of warming. P2 represents tropical regions with deeper bottom depths. P1 and P2 display less seasonal changes in their spatial distribution than P4 due to weaker seasonal temperature changes. P3 and P6, which cover a combined $9.2 \cdot 10^6$ km^2, are found in the Southern

Hemisphere and correspond to sub-polar and temperate regions, respectively. Their spatial distributions are subject to marked latitudinal migrations throughout the year as a result of the large amplitude changes in seasonal temperature observed in mid-latitude coastal waters (Laruelle et al., 2014). Similarly, P7, correspond to temperate Northern Hemisphere waters and display marked seasonal changes including the shelves of the Norwegian basin in summer and most of the Mediterranean Sea in winter. P5, P8, P9 and P10 together cover $22.7 \cdot 10^6$ km^2. These provinces are partly (seasonally) covered by sea-ice with an average spatial ice cover over the study period of 57%, 39%, 54% and 46% for P5, P8, P9 and P10, respectively. P5 represents the shelves of Antarctica all year round. P8 includes large fractions of the enclosed seas at higher northern latitudes such as the Baltic Sea and Hudson Bay while P9 (only $2.9 \cdot 10^6$ km^2) represents permanently deep and cold polar regions. P5 and P10 represent most of the polar shelves (P5 for the Antarctic and P10 for the Arctic) and are covered in sea ice at levels of 57% and 46%, respectively. The regions experiencing most notable shifts in province allocation during the year include the northern Polar Regions as well as the temperate narrow shelves of the Atlantic and Pacific, particularly Western Europe and Eastern North America and Eastern Asia (see Fig. SI B).

3.2. Performance of the coastal SOM-FFN

The mean climatological pCO_2 estimated by the coastal SOM-FFN for annually and seasonally averaged conditions are reported in Figure 3. Before briefly analysing the main spatial and temporal variability of the pCO_2 fields (section 3.3), we evaluate here the overall performance of our interpolation method globally and at the level of each province, including its ability to capture the seasonal cycle.

3.2.1. Comparison with training data (SOCAT*)

Within each province, the pCO_2 simulated by the coastal SOM-FFN are compared to the measurements extracted from SOCAT v4.0 (table 2). Globally, the average difference between observed and simulated pCO_2 is almost null (overall bias = 0.0 µatm) and the absolute bias is lower than 4 µatm in all ten provinces. The average RMSE over all provinces of 32.9 µatm is comparable with those reported for other statistical reconstructions of coastal pCO_2 fields summarized by (Chen et al., 2016), although none of these studies were performed at global scale and many rely on different statistical approaches often using remote sensing data. This RMSE is about twice that achieved for the open ocean (Landschützer et al., 2014) reflecting the larger spatiotemporal variability in the coastal ocean, as well as more complex processes governing that variability. Considering these complexities, achieving at the global scale RMSE in the same range as those reported for regional coastal studies is quite good.

Significant variations in both bias and RMSE can be observed between provinces (table 2). P1 and P3 have the best fit between simulated and observed pCO_2 with RMSE lower than 20 µatm. In 5 provinces that cover a cumulated surface area of 31.2 10^6 km^2 (P1, P2, P3, P6 and P9) RMSE's do not exceed 25 µatm. In P8 however, the maximum RMSE is found with a value of 46.8 µatm. Overall, the performance of the SOM-FFN deteriorates for provinces regularly covered by sea-ice ice (P5, P8-10) in which data coverage is relatively low (RMSE> 34 µatm). This trend is consistent with the spatial distribution of the average residual errors between the pCO_2 field generated by the model and pCO_2 data extracted SOCAT* (Fig. 4a). The residuals are obtained by subtracting the observed values from model output in each grid cell for every month where observations are available. Thus, positive values correspond to cells where the simulated pCO_2 overestimates the field data, while negative values represent cells where the simulated pCO_2 underestimates the field data. The bulk of the residuals fall in the -20 to 20 µatm range in temperate and tropical regions, except for very shallow regions tha are under the influence of a large river such as the Mississippi. There, the SOM-FFN often underestimates the observed pCO_2. There also exist coastal areas where the SOM-FFN underestimates the observed pCO_2 such as the Nova Scotia, the South Western coast of England or the shelves of California and Morocco. The complex hydrodynamics of those regions (some of them being characterized as upwelling regions) may explain the weaker performance of the SOM-FFN. At high latitudes, the performance of the model deteriorates somewhat. For example, the Bering Sea both contains cells with very high (>50 µatm) and very low average residuals (<-50 µatm).

3.2.2. Evaluation with LDEO* data

The comparison of our results with the data from LDEO* yields a global bias of 0.0 µatm (calculated as the average difference between observed and SOM-FFN estimated pCO_2) for the entire shelf domain. However, the spread is relatively large with an average RMSE of 39.2 µatm. This average RMSE is 19% larger than the one obtained when comparing the SOM-FFN results with the SOCAT* dataset, which has been used to train the model. A province-based analysis reveals strong differences in the calculated RMSEs, ranging from 20 µatm to 53 µatm (Table 2, LDEO*). A review of various statistical models used to generate continuous global ocean pCO_2 maps, including some using remote sensing data and algorithms, reports RMSE or uncertainties typically varying within the 10-35 µatm range (Chen et al., 2016) with outliers as high as 50 µatm in the Mississippi delta (Lohrenz and Cai, 2006). This report also shows that open ocean estimates generally yields RMSE lower than 17 µatm, in agreement with Landschützer et al. (2014), whereas coastal estimates are associated with much higher uncertainties. This is likely because these coastal regions have complex biogeochemical dynamics and high frequency variability that cannot be fully captured with the current generation of data interpolation techniques using the limited available predictor data.

In our simulations, the province averaged biases are larger than those calculated with SOCAT* but their absolute value remains small and never exceed 3.9 µatm (P8). Provinces P1, P2, P3 and P6 have RMSE < 30 µatm, which compares with the most robust pCO_2 regional coastal estimates from the literature (Chen et al., 2016). Together, these 4 provinces account for 37% of our domain. P4, P5 and P9 display RMSE comprised between 33 µatm and 38 µatm for P4 and P9, respectivelly. Overall, these 7 provinces covering the entire tropical and temperate latitudinal bands as well as some subpolar regions account for >72% of the shelf surface area and yield RMSE of less than 38 µatm and absolute biases of less than 2.3 µatm. Provinces in the polar regions (P5, P7, P8 and P10) overall display larger deviations with respect to the LDEO* dataset, but the absolute value of their biases never exceeds 3.9 µatm. Their RMSE all fall in the 51-53 µatm range. This suggests a significantly lower performance of the SOM-FFN in regions partly covered in sea-ice. This can be attributed to the limited number of available data points and their very heterogeneous distribution in time and space, as well as to the very limited range of variation of some of the controlling variable such as temperature and salinity. The relatively good performance of the model in tropical region might be partly attributed to the relatively small seasonal variations in pCO_2 within these areas. The residuals calculated by subtracting the SOM-FFN results from LDEO* are very similar to those obtained by subtracting the SOM-FFN results from SOCAT* (Fig. 4b).

The residual errors have a nearly Gaussian distribution for every biogeochemical province with the exception of province P8 (Fig. 5). In this case, the distribution has not only the highest spread, but is also skewed toward high values.

In order to evaluate the contribution of the newly added predictors compared to the oceanic set up of the SOM-FFN (Landschützer et al., 2013), the model was also trained without wind speed and sea ice cover. The RMSE obtained with those simulations (Table 4) are significantly higher than those obtained using all predictors (Table 3). However, the overall bias remain small. The results of those simulations are presented in the table below and allow to quantify how the addition of new predictors affects the performance of the model. For instance, it can be noticed that the global RMSE increases significantly (from 39.2 to 48 μatm in the comparison with LDEO* when chlorophyll, sea ice and wind speed are not taken into account and from 39.2 to 45 μatm when only sea ice and wind speed are not taken into account). This deterioration of the performance of the model, however, is not evenly affecting all provinces. Provinces located at high latitudes (i.e. P8, P9 and P10) perform significantly worse without the inclusion of wind speed and sea ice.

Finally, while the use of residuals and RMSE provide valid quantitative assessment of the model performance, it does not provide insights regarding its ability to reproduce the seasonal pCO_2 cycle. To address this issue, Figure 7 displays observed mean monthly pCO_2 extracted from LDEO* and calculated by the coastal SOM-FFN for the 40 locations where the LDEO* database has the most data (>40 month). The error bars associated with the observations reflect the inter-annual variability. Overall, the coastal SOM-FFN captures the timing of the seasonal pCO_2 cycle in most locations well with pCO_2 minima and maxima occurring at the same time in our results and in the uninterpolated LDEO* data. The pCO_2 maximum generally taking place in early summer is accurately captured by the coastal SOM-FFN. In terms of amplitudes in the pCO_2 signal, the coastal SOM-FFN and the LDEO* data reveal primarily how different the seasonal pCO_2 cycle is from one region to the other, with very low amplitude (<40 µatm) in some sub-tropical areas, amplitudes > 100 µatm at high Northern and Southern latitudes, and sometimes very sharp increases during summer like off the coast of Japan. In most regions, the SOM-FFN-based reconstructions are able to capture these variations and predict seasonal amplitudes comparable to those observed in the data. However, in cells for which the difference between observed and simulated seasonal pCO_2 amplitude is larger than 20%, the coastal SOM-FFN tends to systematically underestimate the amplitude of the seasonal pCO_2 cycle. This limitation of our model might result from the often short time scales associated with the continental influences in near-shore locations, which are not captured by the environmental predictors used in our calculation. It may also be the result of very short-term events that are aliased in our monthly average calculations.

3.2.3. Comparison with global SOM-FFN

The comparison of our coastal SOM-FFN results with those of Landschützer et al. (2016) for the overlapping grid cells (Table 3) reveals significant differences between both interpolated data products with a RMSE between 24 and 32 µatm for most provinces except P7, P9 and P10 (53, 55 and 37 µatm, respectively). These RMSE values are comparable, but slightly lower than those obtained for the comparison with the LDEO* database, in line with those observed with the SOCAT* database. The differences (coastal SOM-FFN minus global SOM-FFN), however, are much larger than those observed between our results and the LDEO* database and highlight the current knowledge gap regarding the mean state and variability of the transition zone. They range from -17.9 to 11.7 µatm from one province to the other but only amount to -0.6 µatm when considering the cells from all provinces at once. The overlaping cells used for the comparison with Landschützer et al. (2016) are mostly located over 100km away from the coastline and therefore the open ocean as well as our new shelf ocean data set are constrained by fairly different data because all the 'shelf' cells from the open ocean data product have a pCO_2 calculated by a model calibrated mostly for conditions representative of the open ocean. Overall, the occurrence of large residuals in the shallowest cells of our calculation domain in our results (Fig. 2) suggest that the very nearshore processes controlling the CO_2 dynamics likely are the most difficult to reproduce at the global scale. However, the added value of performing our simulations at the spatial resolution of 0.25° is exemplified by the ability of our model to capture the plumes of larges rivers such as the Amazon, where pCO_2 is significantly lower than that of the surrounding waters (Cooley et al., 2007; Ibanez et al., 2015).

3.3. Spatial and temporal variability of the coastal pCO_2

3.3.1 Spatial variability

Figure 4a presents the annual average pCO_2 estimated by the coastal SOM-FFN, representing the mean over 1998 through 2015 period (monthly climatological maps are shown in Fig. SI A). High annual mean values of pCO_2, close to or above atmospheric levels, are estimated around the equator up to the tropics. This is consistent with previous studies that identified tropical and equatorial coastal regions as weak CO_2 sources for the atmosphere (Borges et al., 2005; Cai, 2011; Laruelle et al., 2010; 2014). A hotspot of very high pCO_2 emerges from our analysis around the Arabian Peninsula, extending into the eastern Mediterranean Sea as well as into the Red Sea and the Persian Gulf. These regions are poorly monitored and it remains difficult to assess if pCO_2 values in excess of 450 µatm are realistic or not, but the limited body of available literature suggests that very high pCO_2 are indeed observed in these regions (Ali, 2008; Omer, 2010). The very high temperature and salinity conditions observed in the Red Sea, in particular, reduce the CO_2 solubility and induce very high pCO_2 conditions. However, these predicted pCO_2 lie outside of the range used for the training of the SOM-FFN (typically 200-450 µatm) and should thus be considered with caution. Along the oceanic coast of the Arabian Peninsula, the SOM-FFN predicts pCO_2 ranging from 365 to 390 µatm all year round and thus does not capture the well-known increase in pCO_2 resulting from the monsoon driven summer upwelling in the region (Sarma, 2003; Takahashi et al., 2009).

In both hemispheres, pCO_2 values in the 325 to 370 µatm range are generally reconstructed at temperate latitudes, i.e., up to 50°N and 50°S, respectively. The northern high latitudes generally have very low pCO_2 values, down to 300 µatm and below, a result that is consistent with the Arctic shelves contributing a large proportion (up to 60%) of the global coastal carbon sink (Bates and Mathis, 2009; Cai, 2011; Laruelle et al., 2014). Several hotspots of pCO_2 with values as high as 450 µatm can be observed nevertheless north of 70°N, most notably along the eastern coast of Siberia in winter (see Fig. SI P), which displays a large zone characterized by pCO_2 > 400 µatm centred on the mouth of the Kolyma River. Such high pCO_2 values have been punctually observed in Arctic coastal waters (Anderson et al., 2009) and could result from the discharge of highly oversaturated riverine waters. But, overall, pCO_2 measurements over Siberian shelves are rare. Thus, our results should be considered with caution in this region because of the scarcity of data to train and validate the coastal SOM-FFN. It should also be noted that the vast majority of this high pCO_2 region is covered by sea ice (Fig. 4b&c) and, although the model estimates pCO_2 values over the entire domain, only ice-free (or partially ice-free) cells will contribute to the CO_2 exchange across the air-sea interface (Bates and Mathis, 2009; Laruelle et al., 2014).

3.3.2. Temporal variability

The reconstructed pCO_2 field is also subject to large seasonal variations (see figures SI P&A). To explore these variations further, Figure 8 reports seasonal-mean latitudinal profiles of pCO_2 for continental shelves neighbouring the Eastern Pacific, Atlantic, Indian and Western Pacific, respectively. The analysis excludes continental shelves at latitudes higher than 65 degrees, because a large fraction of these shelves are seasonally covered by sea ice. The latitudinal pCO_2 profiles reveal that, in most regions, highest and lowest pCO_2 values are observed during the warmest and coldest months, respectively. This trend is particularly pronounced at temperate latitudes where the seasonal pCO_2 amplitude can reach 60µatm and is exemplified by regions such as the western Mediterranean Sea or the eastern coast of America, which become supersaturated in CO_2 compared to the atmosphere during the summer months. However, there are a few other regions, where the lowest pCO_2 is found in the summer, such as the Baltic Sea (Thomas and Schneider, 1999). Around the equator, the magnitude of the seasonal variations in pCO_2 is limited and does not exceed 30 µatm.

Although the general latitudinal trend of the annual mean pCO_2 is similar across all continental shelves, significant differences in the seasonality can be observed across the largest ocean basins. In particular, most of the East Pacific shelves, except for latitudes north of 55°N, display limited seasonal change in pCO_2 (typically below 30 µatm) while the West Pacific shelves have seasonal pCO_2 amplitudes that can exceed 50 µatm in temperate regions and 100 µatm at high latitudes (above 55° N). Along the Atlantic shelves, the seasonal signal is more pronounced in the north compared to the south, in agreement with Laruelle et al. (2014). Overall, the North Pacific (north of 55°N) displays the most pronounced seasonal change in pCO_2 with a difference of 80 µatm between summer and winter. In the Indian Ocean, the seasonal dynamics of pCO_2 is partly regulated by seasonal upwelling induced by the Monsoon (Liu et al., 2010). In this basin north the equator, April, May and June are the months having the highest pCO_2 and the seasonal variations do not exceed 30 µatm. In contrast, the seasonal cycle is quite pronounced in the Indian Ocean south of the equator (~50 µatm).

Latitudinal profiles of SST (Fig 8, bottom) are similar in all coastal oceans with minimal seasonal variations around the equator and amplitudes as large as 20°C at temperate latitudes. The comparison between the seasonal pCO_2 and SST profiles allows us to assess the contribution of temperature-induced changes in CO_2 solubility to the seasonal pCO_2 variations in the continental shelf waters. However, other factors such as seasonal upwelling and biological activity also strongly influence coastal pCO_2 and contribute to the complexity of the seasonal pCO_2 profiles. To quantify the effect of temperature on seasonal variations of pCO_2, the latter is normalized to the mean temperature at different latitudes in each oceanic basin (Fig. 8) using the formula proposed by Takahashi et al. (1993):

$$pCO_{2(SSTmean)} = pCO_{2,obs} \times exp\big(0.0423 \times (T_{mean} - T_{obs})\big) \qquad (1)$$

where $pCO_{2(SSTmean)}$ is the temperature normalized pCO_2, $pCO_{2,obs}$ is the observed pCO_2 at the observed temperature T_{obs}, and T_{mean} is the yearly mean temperature at the considered location. In sea-water, an increase in water temperature induces a decrease in gas solubility which leads to a higher water pCO_2. Thus, comparing $pCO_{2(SSTmean)}$ with observed pCO_2

monthly values provides a quantitative estimate of the influence of seasonal temperature change on the seasonality of pCO_2.

For most latitudes and oceanic basins, pCO_2 is minimum in late winter or early spring, i.e., at the time when $pCO_{2(SSTmean)}$ has its maximum. pCO_2 also generally displays a maximum in summer, while $pCO_{2(SSTmean)}$ reaches its minimum then (Fig. 9). The amplitude of the changes in $pCO_{2(SSTmean)}$ is quite consistent across oceans and about 2 to 3 times larger than that of pCO_2. Between 45°N and 60° N, the variations in $pCO_{2(SSTmean)}$ largely exceed 100 µatm (up to 220 µatm at 60° N in the West Pacific). In these regions, the magnitude of the seasonal temperature changes is also maximum and reaches 20° C between winter and summer (Fig. 5). A seasonal signal in pCO_2 with a minimum in late winter or spring when $pCO_{2(SSTmean)}$ is maximal can also be identified. However, the magnitude of the seasonal variations in pCO_2 is significantly smaller than those of $pCO_{2(SSTmean)}$, suggesting that other processes such as biological uptake or transport/mixing partly offsets the temperature effect on solubility. In the subpolar western Pacific shelves (60° N), a second pronounced dip in pCO_2 following a weaker one in spring is observed in summer, which suggests the occurrence of a pronounced summer biological activity taking up large amounts of CO_2. This would also explain the sharp increase in pCO_2 in the following month, as a result of the degradation of organic matter synthesized during the summer bloom. Although this region is also the one subjected to the strongest seasonal temperature, the amplitude of the seasonal $pCO_{2(SSTmean)}$ which reaches 220µatm suggests that non thermal processes drive most of the seasonal pCO_2 variations in the regions. At 20° N, the amplitude of the changes in both pCO_2 and $pCO_{2(SSTmean)}$ are lower than at higher latitudes. pCO_2 varies by ~30µatm between summer and winter in all oceanic basin while the seasonal variations in $pCO_{2(SSTmean)}$ are more pronounced in the Pacific

(~60µatm) than in the Atlantic or the Indian Oceans. In the Southern Hemisphere, the seasonal variations in pCO_2 are not as pronounced as in the Northern Hemisphere suggesting that the changes induced by the solubility pump are compensated by biological activities. At 10°S and 30°S, the seasonal variations in pCO_2 rarely exceed 30 µatm in either basin with a minimum observed around August.

4. Summary

This study presents the first global high-resolution monthly pCO_2 maps for continental shelf waters at an unprecedented 0.25° spatial resolution. We show that when tailored for the specific conditions of shelf systems, the SOM-FFN method previously employed in the open ocean is capable of reproducing well-known and well-observed features of the pCO_2 field in the coastal ocean. Our continuous shelf product allows, for the first time, to analyze the dominant spatial patterns of pCO_2 across all ocean basins and their seasonality. The data product associated to this manuscript consists of a netcdf file containing the pCO_2 for ice-free cells at a 0.25° spatial resolution for each of the 216 month of the simulation period (from January 1998 to December 2015). 12 maps representing mean pCO_2 fields calculated for each month over the simulation period are also provided. This data product can be combined with wind field products such as ERA-interim (Dee, 2010; Dee et al., 2011) or CCMP (Atlas et al., 2011) to compute spatially and temporally resolved air-sea CO_2 fluxes across the global shelf region, including the Arctic. Maps including pCO_2 for ice covered cells are also available but should be treated with care because the dynamics of CO_2 fluxes through sea ice are still poorly understood and air-sea gas transfer velocities in partially sea ice covered areas cannot be predicted from classical wind speed relationships (Lovely et al. 2015)

5. Data availability

The version 4 of the SOCAT database (Bakker et al., 2016) can be downloaded from www.socat.info/upload/SOCAT_v4.zip. The observation-based global monthly gridded sea surface pCO_2 product is provided by Landschützer, et al. (2015; doi: 10.3334/CDIAC/OTG.SPCO2_1982_2011_ETH_SOM-FFN.), was downloaded from http://cdiac.ornl.gov/ftp/oceans/SPCO2_1982_2011_ETH_SOM_FFN and is now available at: https://www.nodc.noaa.gov/ocads/oceans/SPCO2_1982_2015_ETH_SOM_FFN.html. The LDEOv2015 database (Takahashi et al., 2015; doi: 10.3334/CDIAC/OTG.NDP088(V2015)) was downloaded from http://cdiac.ornl.gov/oceans/LDEO_Underway_Database/. The global atmospheric reanalysis ERA-interim datasets (Dee et al., 2011, http://doi.wiley.com/10.1002/qj.828) are accessible on the European Centre for Medium-Range Weather Forecasts (ECMWF) website. SST and SSS were extracted from the Met Office's EN4 data set (Good et al., 2009; doi:10.1002/2013JC009067). The bathymetry used is the global ETOPO2 database (US Department of Commerce, 2006), which can be downloaded from http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html. The sea ice concentrations are derived from the global 25 km resolution monthly data product compiled by the NSIDC (National Snow and Ice Cover Data; Cavalieri et al., 1996).

6. Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

G. G. Laruelle and B Delille are postdoctoral researcher and research associate, respectively, of F.R.S.-FNRS. The Surface Ocean CO_2 Atlas (SOCAT) is an international effort, supported by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS), and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), in order to deliver a uniformly quality-controlled surface ocean CO_2 database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT. The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 643052 744 (C-CASCADES project). NG acknowledges support by ETH Zürich. PL is supported by the Max Planck Society for the Advancement of Science.

References

Ali, E.: The Inorganic Carbon Cycle in the Red Sea, Master's thesis, University of Bergen, 2008.

Anderson, L. G., Jutterström, S., Hjalmarsson, S., Wåhlström, I., and Semiletov, I. P.: Out-gassing of CO_2 from Siberian Shelf seas by terrestrial organic matter decomposition, Geophys. Res. Lett., 36, L20601, doi:10.1029/2009GL040046, 2009.

Antonov, J. I., Seidov, D., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson D. R.: in World Ocean Atlas 2009, Volume 2: Salinity, NOAA Atlas NESDIS, vol. 69, edited by S. Levitus, U.S. Gov. Print. Off., Washington, D. C., 2010.

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K. and Gombos, D.: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bulletin of the American Meteorological Society, 92(2), 157-174, 2011.

Bakker, D. C. E., Pfeil, B., Smith, K., Hankin, S., Olsen, A., Alin, S. R., Cosca, C., Harasawa, S., Kozyr, A., Nojiri, Y., O'Brien, K. M., Schuster, U., Telszewski, M., Tilbrook, B., Wada, C., Akl, J., Barbero, L., Bates, N. R., Boutin, J., Bozec, Y., Cai, W.-J., Castle, R. D., Chavez, F. P., Chen, L., Chierici, M., Currie, K., De Baar, H. J. W., Evans, W., Feely, R. A., Fransson, A., Gao, Z., Hales, B., Hardman-Mountford, N. J., Hoppema, M., Huang, W.-J., Hunt, C. W., Huss, B., Ichikawa, T., Johannessen, T., Jones, E. M., Jones, S., Jutterstrom, S., Kitidis, V., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Manke, A. B., Mathis, J. T., Merlivat, L., Metzl, N., Murata, A., Newberger, T., Omar, A. M., Ono, T., Park, G.-H., Paterson, K., Pierrot, D., Ríos, A. F., Sabine, C. L., Saito, S., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Sieger, R., Skjelvan, I., Steinhoff, T., Sullivan, K. F., Sun, H., Sutton, A. J., Suzuki, T., Sweeney, C., Takahashi, T., Tjiputra, J., Tsurushima, N., Van Heuven, S. M. A. C., Vandemark, D., Vlahos, P., Wallace, D. W. R., Wanninkhof, R., Watson, A. J.: An update to the Surface Ocean CO2 Atlas (SOCAT version 2). Earth System Science Data 6: 69-90. doi:10.5194/essd-6-69-2014, 2014.

Bakker, D. C. E. et al. (92 authors) : A multi-decade record of high-quality fCO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383-413, doi:10.5194/essd-8-383-2016, 2016

Bates, N. R., Moran, S. B., Hansell, D. A., and Mathis, J. T.: An increasing CO_2 sink in the Arctic Ocean due to sea-ice loss. Geophys. Res. Lett., 33(23), L23609, doi: 10.1029/2006GL027028, 2006.

Bates, N. R., and Mathis, J. T.: The Arctic Ocean marine carbon cycle: Evaluation of air-sea CO2 exchanges, ocean acidification impacts and potential feedbacks, Biogeosciences, 6, 2433–2459, doi:10.5194/bg-6-2433-2009, 2009.

Bauer, J. E., Cai, W.-J., Raymond, P. A., Bianchi, T. S., Hopkinson, C. S., and Regnier, P. A. G.: The changing carbon cycle of the coastal ocean, Nature, 504, 61–70, doi:10.1038/nature12857, 2013.

Borges, A. V., Delille, B., and Frankignoulle, M.: Budgeting sinks and sources of CO2 in the coastal ocean: Diversity of ecosystems counts, Geophys. Res. Lett., 32, L14601, doi:10.1029/2005GL023053, 2005.

Bourgeois, T., Orr, J. C., Resplandy, L., Terhaar, J., Ethé, C., Gehlen, M., and Bopp, L.: Coastal-ocean uptake of anthropogenic carbon, Biogeosciences, 13, 4167-4185, doi:10.5194/bg-13-4167-2016, 2016.

Cai, W. J.: Estuarine and coastal ocean carbon paradox: CO2 sinks or sites of terrestrial carbon incineration?, Annu. Rev. Mar. Sci., 3, 123–145, 2011.

Cavalieri, D. J., Parkinson, C. L., Gloersen, P., and Zwally, H.: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, years 1990–2011, NASA DAAC at the Natl. Snow and Ice Data Cent., Boulder, Colo. (Updated yearly.), 1996.

Chen, C.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 Research II, 56 (8-10), 578-590, 2009.

Chen, C. T. A., Huang, T. H., Chen, Y. C., Bai, Y., He, X., and Kang, Y.: Air-sea exchanges of CO2 in the world's coastal seas, Biogeosciences, 10, 6509–6544, doi:10.5194/bg-10-6509-2013, 2013.

Chen, S., Hu, C., Byrne, R. H., Robbins, L. L., and Yang, B.: Remote estimation of surface pCO2 on the West Florida Shelf, Continental Shelf Research, 128, 10–25, 2016.

Cooley, S. R., V. J. Coles, A. Subramaniam, and P. L. Yager (2007), Seasonal variations in the Amazon plume-related atmospheric carbon sink, Global Biogeochem. Cycles, 21, GB3014, doi:10.1029/2006GB002831.

Crossland, C. J., Kremer, H. H., Lindeboom, H. J., Marshall Crossland, J. I., and LeTissier, M. D. A. (Eds.): Coastal Fluxes in the Anthropocene, Global Change – The IGBP Series: 232 pp, Berlin, Heidelberg, Springer-Verlag, Germany, 2005.

Dee, D.P.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc., 137, pp.553–597, 2010.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hòlm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Doney, S. C.: The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry, Science 328(5985), 1210-1216, doi:10.1126/science.1185198, 2010.

Good, S. A., M. J. Martin and N. A. Rayner, 2013. EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research: Oceans, 118, 6704-6716, doi:10.1002/2013JC009067

Gruber, N.: Ocean biogeochemistry: Carbon at the coastal interface, Nature, 517, 148–149, doi:10.1038/nature14082, 2015.

Grimm, R., Notz, D., Glud, R.N., Rysgaard, S. and Six, K.D.: Assessment of the sea-ice carbon pump: Insight from a three-dimensional ocean-sea-ice-biogeochemical model (MPIOM/HAMOCC). Elementa: Science of the Anthropocene, 4:000136, doi: 10.12952/journal.elementa.000136, 2016.

Ibánhez, J. S. P., D. Diverrès, M. Araujo, and N. Lefèvre (2015), Seasonal and interannual variability of sea-air CO2 fluxes in the tropical Atlantic affected by the Amazon River plume, Global Biogeochem. Cycles, 29, 1640–1655, doi:10.1002/2015GB005110.

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T., 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.

Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, Global Biogeochemical Cycles, 28, 927–949, doi:10.1002/2014GB004853, 2014.

Landschützer, P., Gruber, N., Haumann, F. A. Rödenbeck, C. Bakker, D.C.E. , van Heuven, S. Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T., Tilbrook, B. and Wanninkhof, R.: The reinvigoration of the Southern Ocean carbon sink, Science, 349, 1221-1224. doi: 10.1126/science.aab2620, 2015.

Landschützer, P., Gruber, N. Bakker, D.C.E.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochemical Cycles, 30, doi:10.1002/2015GB005359, 2016

Laruelle, G. G., Dürr, H. H., Slomp, C. P., and Borges, A. V.: Evaluation of sinks and sources of CO_2 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.

Laruelle, G. G., Dürr, H. H., Lauerwald, R., Hartmann, J., Slomp, C. P., Goossens, N., and Regnier, P. A. G.: Global multi-scale segmentation of continental and coastal waters from the watersheds to the continental margins, Hydrol. Earth Syst. Sci., 17, 2029–2051, doi:10.5194/hess-17-2029-2013, 2013.

Laruelle, G. G., Lauerwald, R., Pfeil, B., and Regnier, P.: Regionalized global budget of the CO2 exchange at the air-water interface in continental shelf seas, Global Biogeochemical Cycles, 28, 1199–1214, doi:10.1002/2014GB004832, 2014.

Laruelle, G. G., Lauerwald, R., Rotschi, J., Raymond, P. A., Hartmann, J., and Regnier, P.: Seasonal response of air–water CO2 exchange along the land–ocean aquatic continuum of the northeast North American coast., Biogeosciences, 12, 1447-1458, doi:10.5194/bg-12-1447-2015, 2015.

Liu, K.-K., Atkinson, L., Quinones, R., and Talaue-McManus, L. (Eds.): Carbon and Nutrient Fluxes in Continental Margins, Global Change – The IGBP Series, 3, Springer-Verlag Berlin Heidelberg, 2010.

Lohrenz, S. E., and Cai, W.-J.: Satellite ocean color assessment of air-sea fluxes of CO2 in a river-dominated coastal margin, Geophys. Res. Lett., 33, L01601, doi:10.1029/2005GL023942, 2006.

Lovely, A., Loose, B., Schlosser, P., McGillis, W., Zappa C., Perovich, D., Brown, S., Morell, T., Hsueh, D., and Friedrich, R.: The Gas Transfer through Polar Sea ice experiment: Insights into the rates and pathways that determine geochemical fluxes. J. Geophys. Res. Ocean. 120:8177–8194, 2015.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson D. R.: World Ocean Atlas 2009, Volume 1: Temperature, NOAA Atlas NESDIS, vol. 69, edited by S. Levitus, U.S. Gov. Print. Off., Washington, D. C., 2010.

Moreau, S., Vancoppenolle, M., Bopp, L., Aumont, O., Madec, G., Delille, B., Tison, J.-L., Barriat, P.-Y. and Goosse, H.: Assessment of the sea-ice carbon pump: Insights from a three-dimensional ocean-sea-ice-biogeochemical model (NEMO-LIM-PISCES). Elementa: Science of the Anthropocene, 4:000122, doi: 10.12952/journal.elementa.000122, 2016.

Nakaoka, S., Telszewski, M., Nojiri, Y., Yasunaka, S., Miyazaki, C., Mukai, H., and Usui, N.: Estimating temporal and spatial variation of ocean surface pCO2 in the North Pacific using a self-organizing map neural network technique, Biogeosciences, 10, 6093-6106, doi:10.5194/bg-10-6093-2013, 2013.

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (Dataset Release 2016): MODIS-Aqua chlorophyll Data; NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, 2016.

Omer, W. M. M.: Ocean acidification in the Arabian Sea and the Red Sea. Master's thesis, University of Bergen, 2011.

Parmentier, F.-J. W., Christensen, T. R., Sørensen, L. L., Rysgaard, S., McGuire, A. D., Miller, P. A., and Walker, D. A.: The impact of lower sea-ice extent on Arctic greenhouse-gas exchange, Nature Climate Change, 3, 195–202, doi:10.1038/nclimate1784, 2013.

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, J., Manke, A., Metzl, N., Sabine, C. L., Akl, J., Alin, S. R., Bates, N., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W.-J., Chavez, F. P., Chen, A., Cosca, C., Fassbender, A. J., Feely, R. A., González-Dávila, M., Goyet, C., Hales, B., Hardman-Mountford, N., Heinze, C., Hood, M., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannessen, T., Jones, S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Nakadate, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G.-H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A., Ríos, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. S. S., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Tjiputra, J., Vandemark, D., Veness, T., Wanninkhof, R., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H.: A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT), Earth System Science Data 5: 125-143.

doi:10.5194/essd-5-125-2013, 2013.

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., Laruelle, G. G., Lauerwald, R., Luyssaert, S., Andersson, A. J., Arndt, S., Arnosti, C., Borges, A. V., Dale, A. W., Gallego-Sala, A., Goddéris, Y., Goossens, N., Hartmann, J., Heinze, C., Ilyina, T., Joos, F., LaRowe, D. E., Leifeld, J., Meysman, F. J. R., Munhoven, G., Raymond, P. A., Spahni, R., Suntharalingam, P. and Thullner, M.: Anthropogenic perturbation of the carbon fluxes from land to ocean. Nature Geoscience, 6, doi:10.1038/ngeo1830, 2013.

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G.-H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shutler, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO2 Mapping intercomparison (SOCOM), Biogeosciences, 12, 7251-7278, doi:10.5194/bg-12-7251-2015, 2015.

Sabine, C. L., et al. (76 authors): Surface Ocean CO2 Atlas (SOCAT) gridded data products, Earth System Science Data , 5, 145 153, doi:10.5194/essd-5-145-2013, 2013

Sasse, T. P., McNeil, B. I., and Abramowitz, G.: A new constraint on global air-sea CO_2 fluxes using bottle carbon data, Geophys. Res. Lett., 40, 1594–1599, doi:10.1002/grl.50342, 2013.

Sarma, V. V. S. S., Monthly variability in surface pCO2 and net air-sea CO2 flux in the Arabian Sea, J. Geophys. Res., 108 (C8), 3255, doi:10.1029/2001JC001062, 2003.

Shadwick, E. H., Thomas, H., Comeau, A., Craig, S. E., Hunt, C. W., and Salisbury, J. E.: Air-Sea CO2 fluxes on the Scotian Shelf: seasonal to multi-annual variability, Biogeosciences, 7, 3851–3867, doi:10.5194/bg-7-3851-2010, 2010.

Signorini, S. R., Mannino, A., Najjar Jr., R. G., Friedrichs, M. A. M., Cai, W.-J., Salisbury, J., Wang, Z. A., Thomas, H., and Shadwick, E.: Surface ocean pCO2 seasonality and sea-air CO2 flux estimates for the North American east coast, J. Geophys. Res.-Oceans, 118, 5439–5460, doi:10.1002/jgrc.20369, 2013.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S. C.: Seasonal variation of CO2 and nutrients in the high-latitude surface oceans: A comparative study. Global Biogeochemical Cycles, 7(4), 843–878, 1993.

Takahashi, T., Sutherland, S., and Kozyr A.: Global ocean surface water partial pressure of CO2 database: Measurements performed during 1957–2011 (Version 2011). ORNL/CDIAC-160, NDP-088(V2011), Carbon Dioxide Information Analysis Center, Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn., 2012.

Takahashi, T., Sutherland, S. C., and Kozyr, A.: Global Ocean Surface Water Partial Pressure of CO2 Database: Measurements Performed During 1957-2015 (Version 2015). ORNL/CDIAC-160, NDP-088(V2015). Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, doi: 10.3334/CDIAC/OTG.NDP088(V2015), 2016.

Thomas, H., and Schneider, B.: The seasonal cycle of carbon dioxide in Baltic Sea surface waters, J. Mar. Syst, 22, 53-67, 1999.

Turi, G., Lachkar, Z., and Gruber, N.: Spatiotemporal variability and drivers of pCO2 and air–sea CO2 fluxes in the California Current System: an eddy-resolving modeling study, Biogeosciences, 11, 671-690, doi:10.5194/bg-11-671-2014, 2014.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Geophysical Data Center. 2006. 2-minute Gridded Global Relief Data (ETOPO2v2).

http://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html. Accessed 26 Dec 2008.

Vancoppenolle M., Meiners, K. M., Michel, C., Bopp, L., Brabant, F., Carnat, G., Delille, B., Lannuzel, D., Madec, G., Moreau, S., Tison, J.-L., and van der Merwe, P.: Role of sea ice in global biogeochemical cycles: Emerging views and challenges, Quaternary Science Reviews, 79, 207-230, doi:10.1016/j.quascirev.2013.04.011, 2013.

Wanninkhof, R., Park, G.-H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., McKinley, G. A., Lenton, A., Le Quéré, C., Heinze, C., Schwinger, J., Graven, H., and Khatiwala, S.: Global ocean carbon uptake: magnitude, variability and trends, Biogeosciences, 10, 1983-2000, doi: 10.5194/bg-10-1983-2013, 2013.

Walsh, J. J.: On the nature of continental shelves, Academic Press, San Diego, New York, Berkeley, Boston, London, Sydney, Tokyo, Toronto, 1988.

Wijesekera, H. W., J. S. Allen, and P. A. Newberger, Modeling study of turbulent mixing over the continental shelf: Comparison of turbulent closure schemes, J. Geophys. Res., 108(C3), 3103, doi:10.1029/2001JC001234, 2003

Yasunaka, S., Murata, A., Watanabe, E., Chierici, M., Fransson, A., van Heuven, S., Hoppema, M., Ishii, M., Johannessen, T., Kosugi, N., Lauvset, S. K., Mathis, J. T., Nishino, S., Omar, A. M., Olsen, A., Sasano, D., Takahashi, T., and Wanninkhof, R.: Mapping of the air-sea CO2 flux in the Arctic Ocean and its adjacent seas: Basin-wide distribution and seasonal to interannual variability. Polar Science, 10(3):323-334, doi:10.1016/j.polar.2016.03.006, 2016.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S.: A global surface ocean fCO2 climatology based on a feed-forward neural network, J. Atmos. Ocean Technol., 31, 1838-1849, 2014.

Table 1: Datasets used to create the environmental forcing files. The original spatial and temporal resolution and the main manipulations applied for their use in the SOM_FFN are also reported.

Predictor	dataset	resolution	reference	Manipulation
SST	EN4	0.25°, daily	Good et al., 2013	Monthly average
SSS	EN4	0.25°, daily	Good et al., 2013	Monthly average
Bathymetry	ETOPO2	2 minutes	US Department of Commerce, 2006	Aggregation to 0.25°
Sea ice	NSIDC	0.25°, monthly	Cavalieri et al., 1996	Monthly rate of change in sea ice coverage
Chlorophyll a	SeaWifs, MODIS	9km, monthly	NASA, 2016	Aggregation to 0.25°
Wind speed	ERA	0.25°, 6hours	Dee et al., 2011	Monthly average

Table 2: List of the biogeochemical provinces, their geographic distribution and the environmental predictors used to calculate surface ocean pCO_2. SSS stands for sea surface salinity, SST for sea surface temperature, Bathy for bathymetry, Ice for sea-ice cover, Chl for chlorophyll concentration and Wind for wind speed.

Province	**SSS**	**SST**	**Bathy**	**Ice**	**Chl**	**Wind**
P1	**X**	**X**	**X**		**X**	**X**
P2	**X**	**X**	**X**		**X**	**X**
P3	**X**	**X**	**X**		**X**	**X**
P4	**X**	**X**	**X**		**X**	**X**
P5	**X**	**X**	**X**	**X**	**X**	**X**
P6	**X**	**X**	**X**	**X**	**X**	**X**
P7	**X**	**X**	**X**	**X**	**X**	**X**
P8	**X**	**X**	**X**	**X**		**X**
P9	**X**	**X**	**X**	**X**		**X**
P10	**X**	**X**	**X**	**X**		**X**

Table 3: Root mean squared error between observed and calculated pCO_2 in the different biogeochemical provinces. The SOM-FFN results are compared to data extracted from the LDEO database (Takahashi et al, 2014) and the overlapping cells from the Landschützer et al. (2016) pCO_2 climatology.

			SOCAT*		**Landschützer 2016**		**LDEO**	
Province	**Surface Area (km^2)**	**Ice Cover (%)**	**Bias (µatm)**	**RMSE (µatm)**	**Bias (µatm)**	**RMSE (µatm)**	**Bias (µatm)**	**RMSE (µatm)**
P1	8.2 10^6	0	0.0	19.1	2.0	27.2	2.0	20.5
P2	10.9 10^6	0	0.2	24.7	9.3	24.2	1.3	27.2
P3	4.4 10^6	0	-0.3	16.1	2.2	37.9	2.3	22.7
P4	16.6 10^6	0	-0.2	31.2	8.0	21.1	-1.6	33.0
P5	7.5 10^6	57.1	0.0	34.2	11.5	30.9	-1.4	38.0
P6	4.8 10^6	0	0.0	24.3	6.8	18.1	1.3	27.9
P7	9.3 10^6	0.0	0.1	37.2	0.7	23.5	-0.2	52.5
P8	3.3 10^6	38.5	0.2	46.8	13.9	70.1	3.9	51.4
P9	2.9 10^6	54.3	-0.1	23.0	-5.2	42.5	-2.5	33.4
P10	9.0 10^6	45.8	0.0	35.7	-9.7	50.9	1.6	53.1
	76.9 10^6		0.0	32.9	3.9	34.7	0.0	39.2

Table 4: Biases and root mean squared error (RMSE) between observed and calculated pCO_2 using only SST, SSS and bathymetry (STB) or SST, SSS, bathymetry and chlorophyll (STBC) as predictors.

	SOCAT*				LDEO*			
Province	**Bias (µatm)**		**RMSE (µatm)**		**Bias (µatm)**		**RMSE (µatm)**	
	STB	STBC	STB	STBC	STB	STBC	STB	STBC
P1	0.0	-0.2	20.8	21.0	2.4	2.0	21.7	21.5
P2	-0.1	0.1	26.9	27.8	0.5	0.8	29.0	29.6
P3	0.0	-0.5	22.7	21.3	3.0	2.3	27.1	26.8
P4	0.0	-0.2	33.0	33.0	-1.7	-2.3	33.8	33.8
P5	0.2	0.1	52.7	42.2	-1.7	-0.9	56.9	44.5
P6	0.0	0.1	26.8	26.5	-0.5	0.6	28.9	28.0
P7	0.4	0.3	44.3	44.1	1.2	0.3	59.3	58.8
P8	0.1	0.4	82.6	80.0	9.1	9.0	56.3	58.5
P9	0.1	0.9	34.7	36.5	-2.6	-2.8	39.8	41.8
P10	-0.3	0.7	49.8	49.5	-3.9	-3.0	76.5	75.4
Global	0.1	0.2	43.9	42.4	0.0	0.0	48.0	45.0

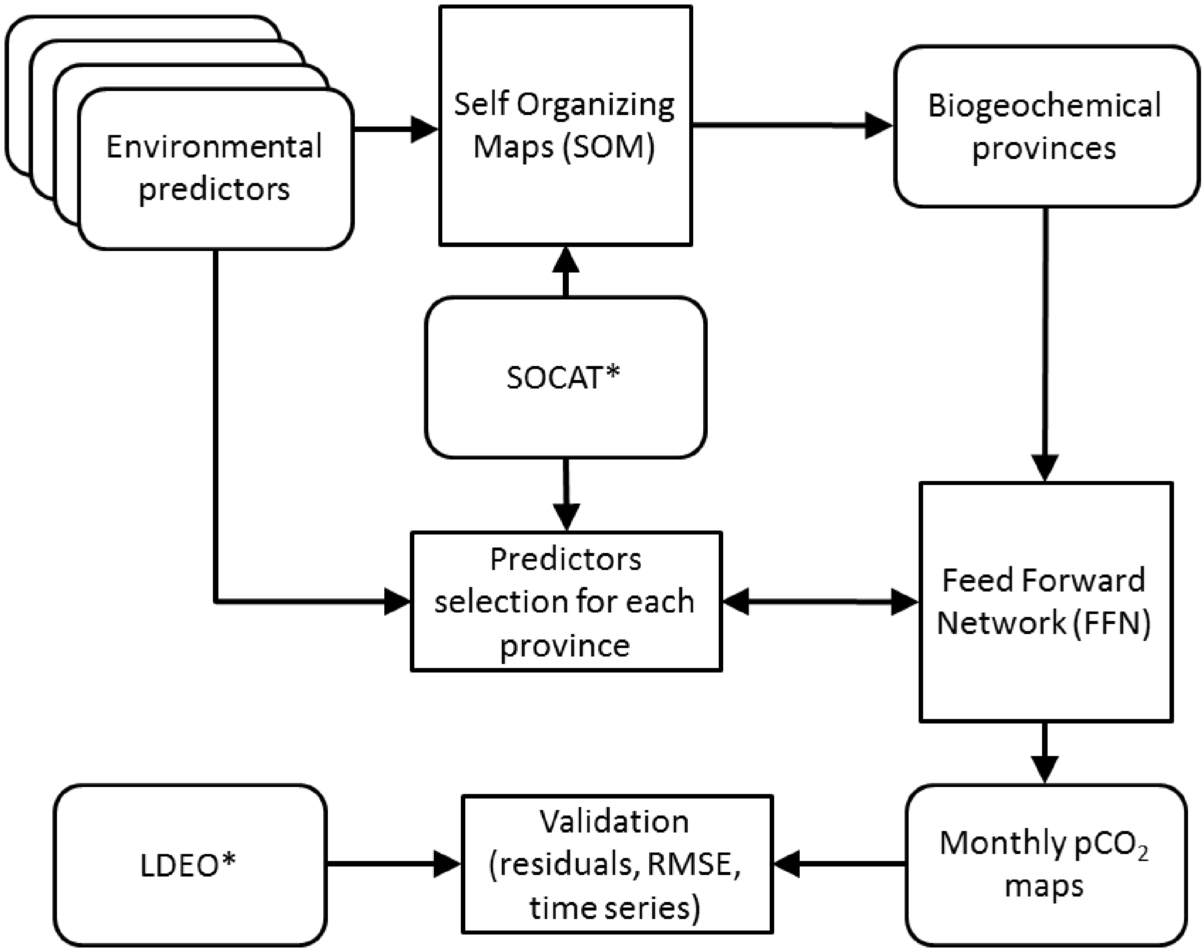

Figure 1: Schematic scheme of the different steps involved in the SOM-FFN artificial neural network calculations leading to continuous monthly pCO_2 maps over the 1998-2015 period.

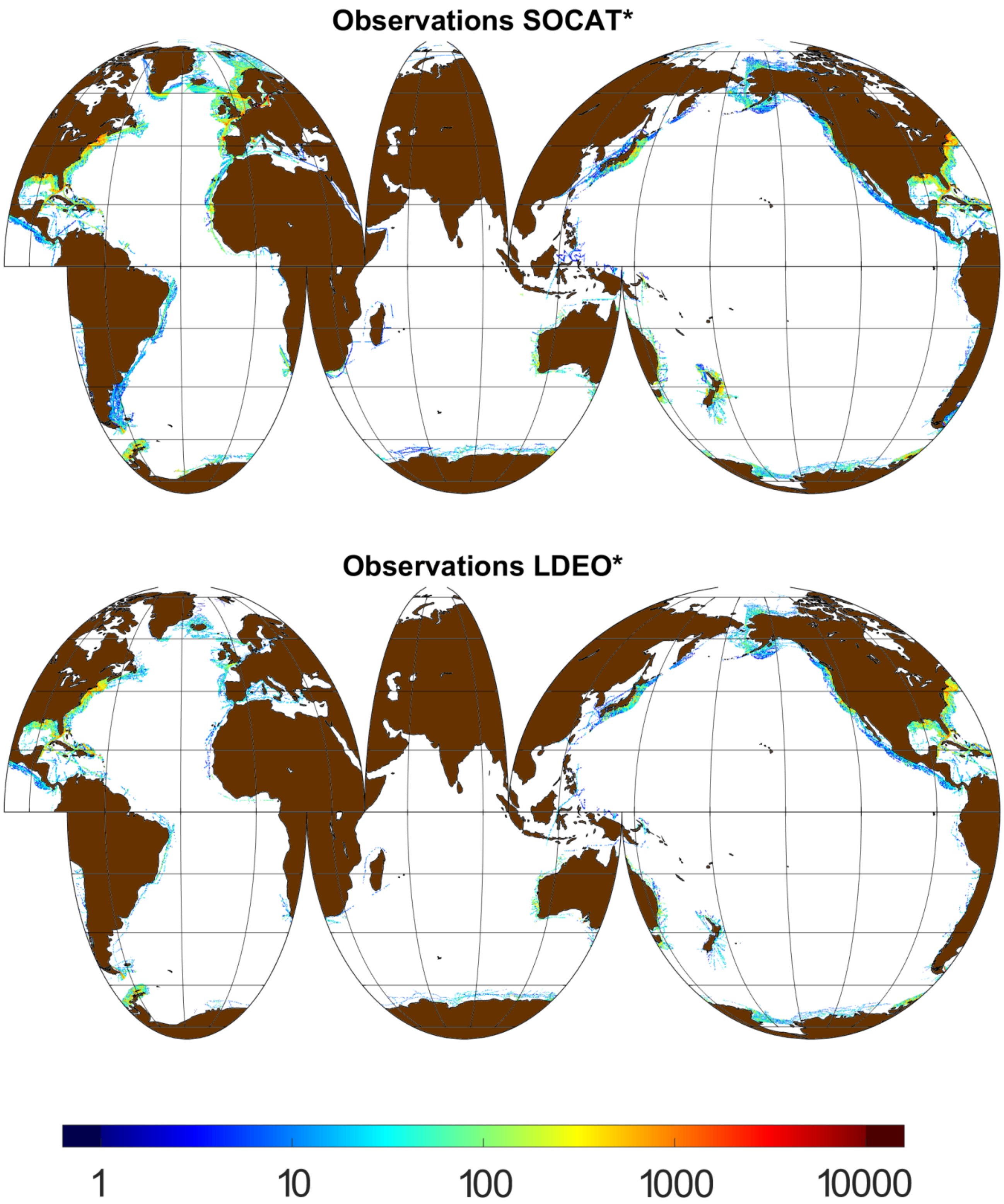

Figure 2: Number of observations contained in each 0.25° grid cell of the SOCAT* (top) and LDEO* (bottom) databases.

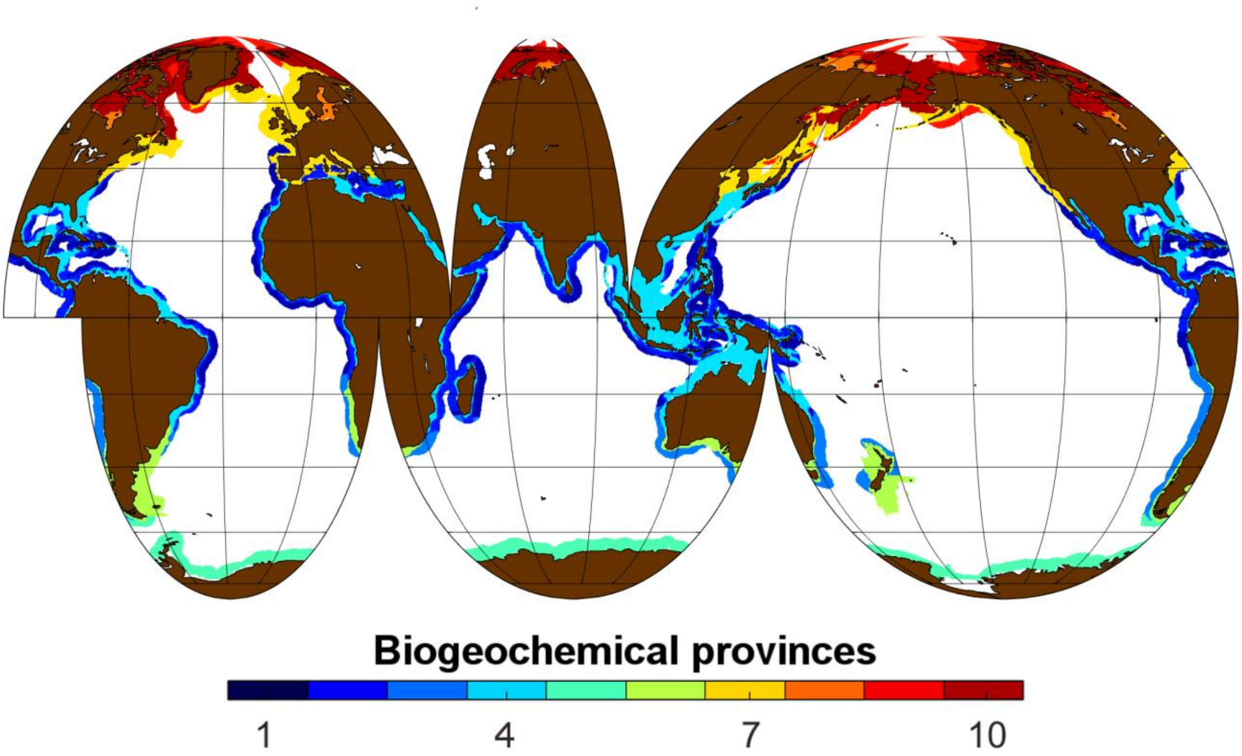

Figure 3: Map of the 10 different biogeochemical provinces generated by the artificial neural network method SOM-FFN.

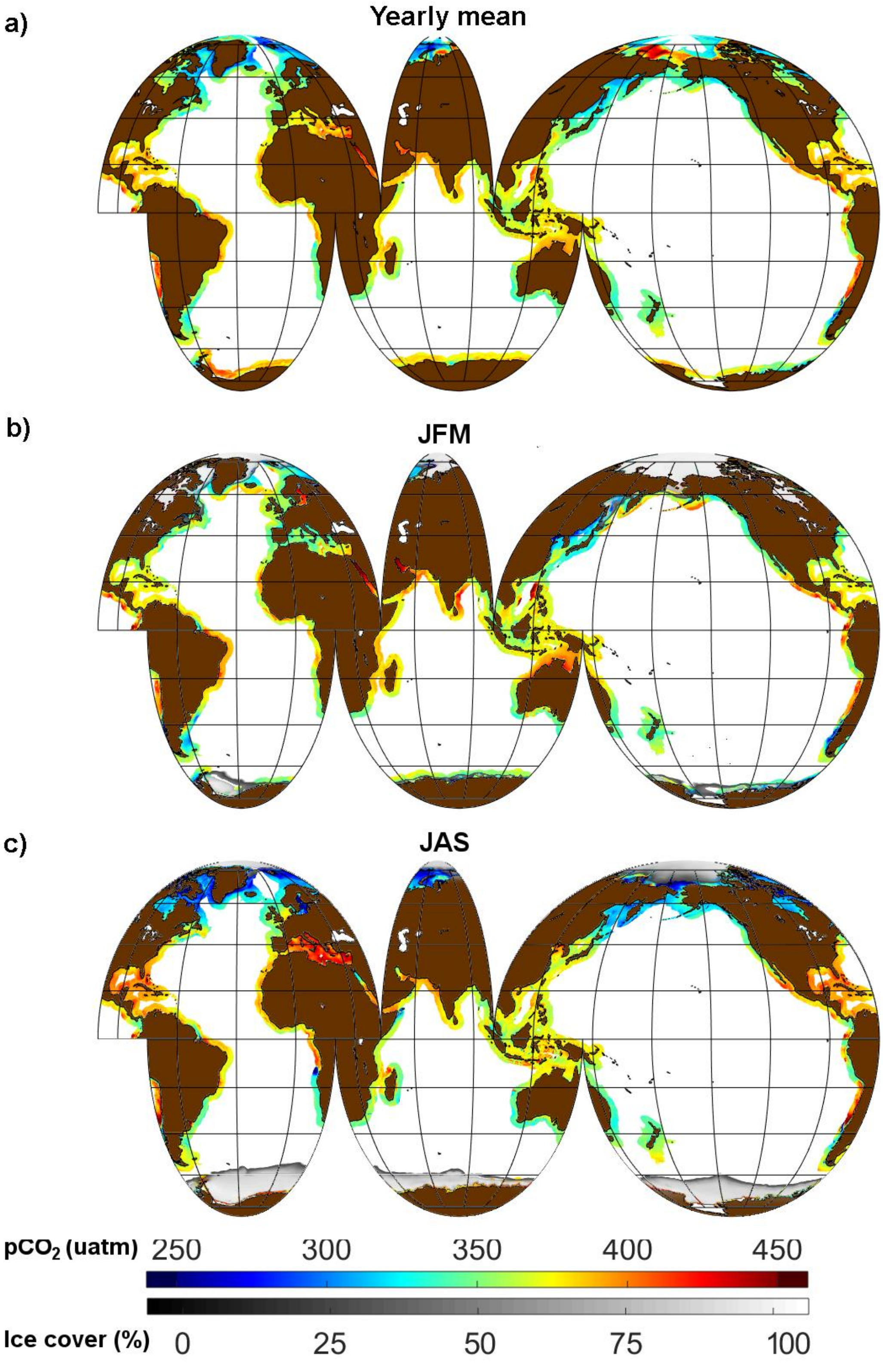

Figure 4: Climatological mean pCO_2 for (a) the long-term averaged pCO_2 (rainbow color scale) and sea-ice coverage (black-white color scale). The long-term average pCO_2 corresponds to roughly the nominal year 2006, as the average was formed over the full analysis period from 1998 through 2015; (b) the months of January, February and March; and (c) the months of July, August and September.

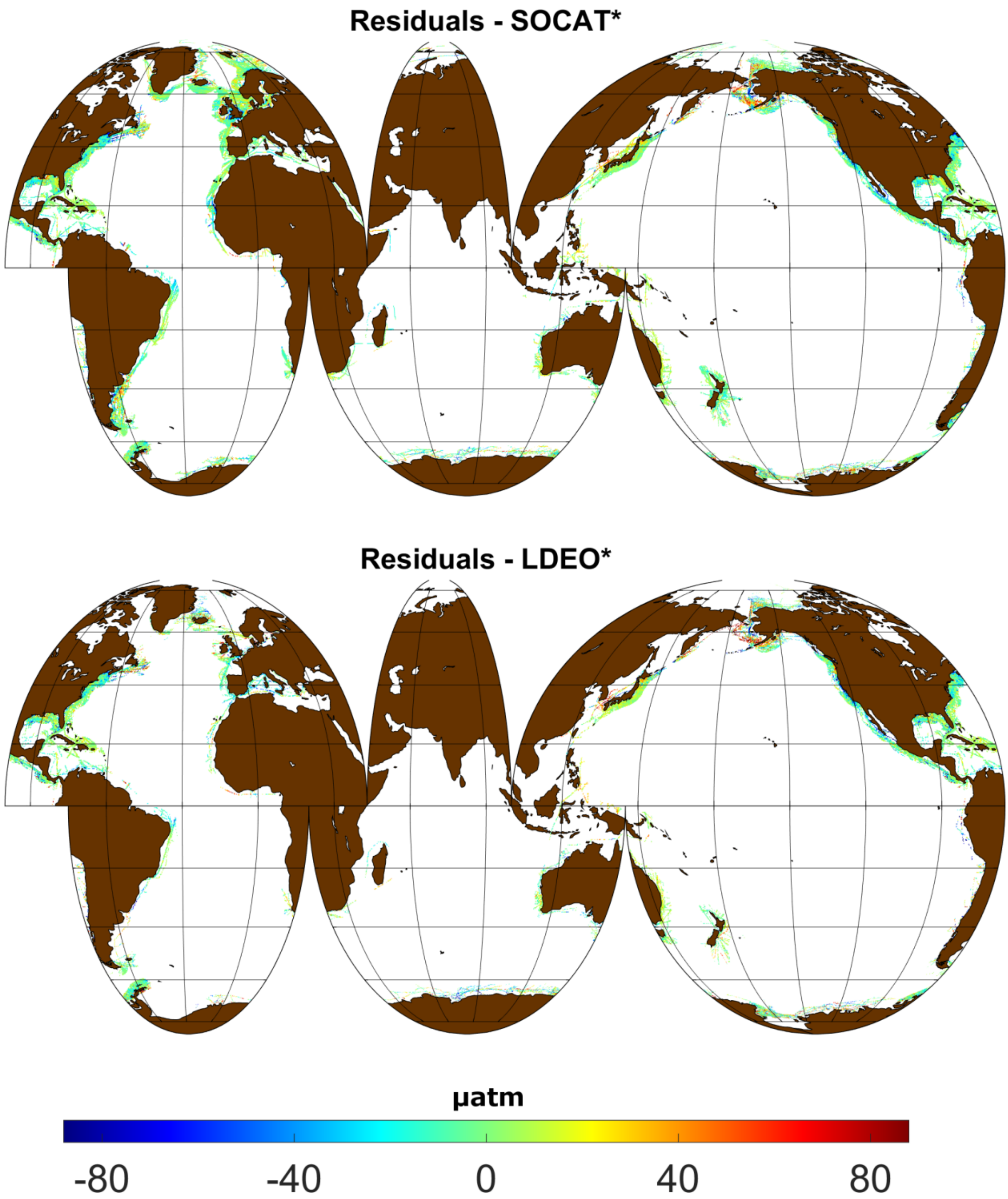

Figure 5: Mean residuals calculated as the difference between the SOM_FFM pCO_2 outputs and pCO_2 observations from SOCAT* (top) and LDEO* (bottom).

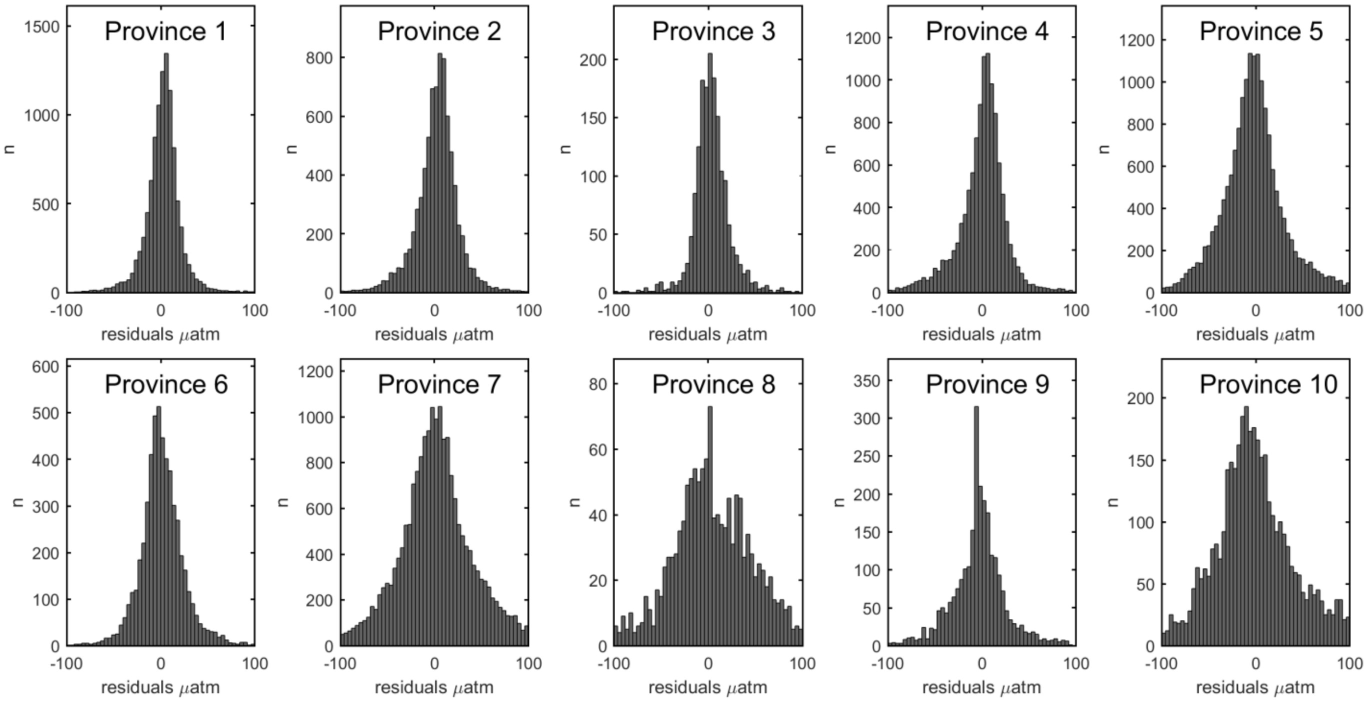

Figure 6: Histograms reporting the distribution of residuals between observed (LDEO*) and computed (SOM_FFN) pCO_2 in each biogeochemical province.

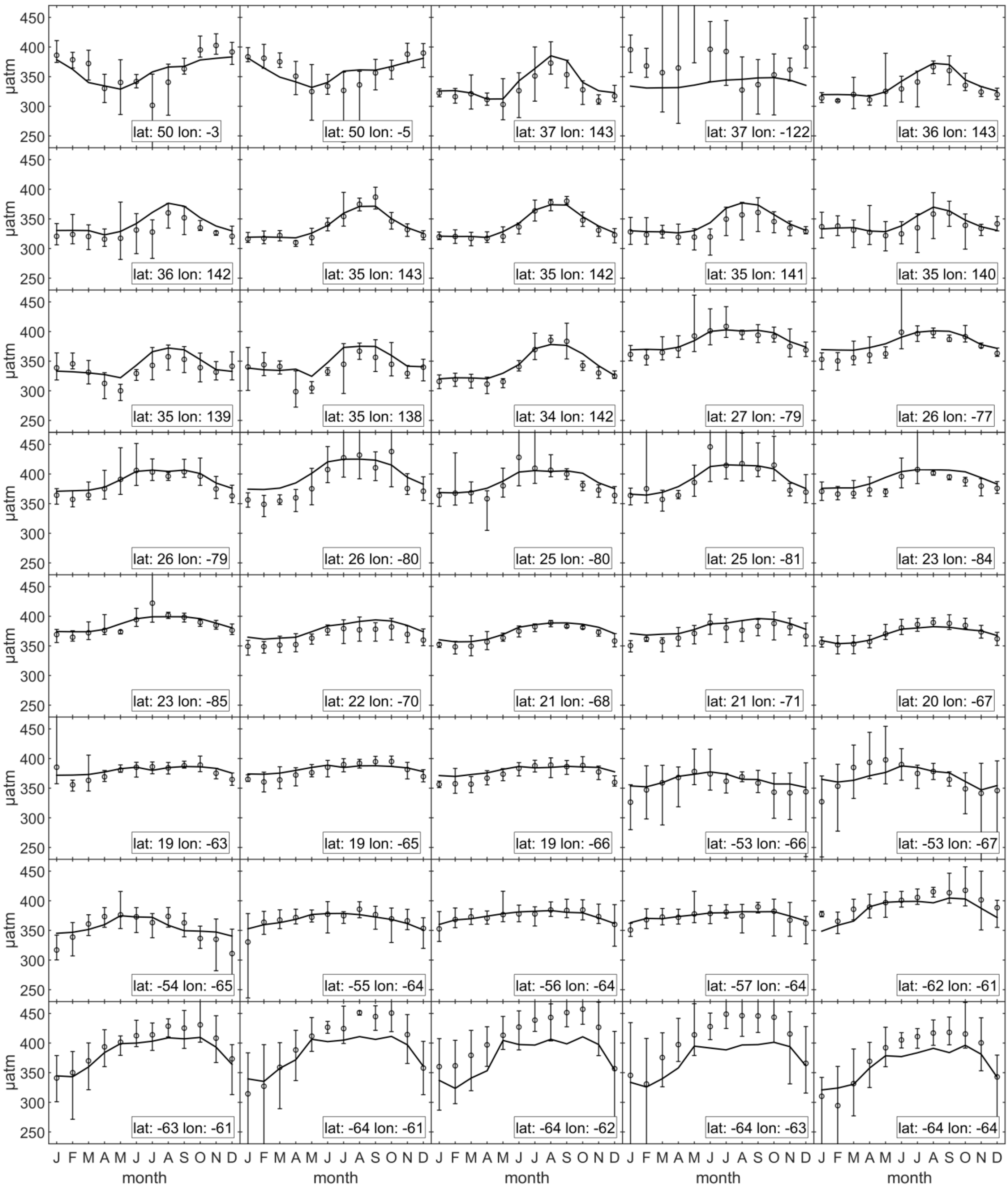

Figure 7: Climatological monthly mean pCO_2 extracted from the LDEO* database (points) and generated by the artificial neural network (lines) for grid cells having more than 40 months of data. The error bars associated with the data represent the inter-annual variability, reported as the highest and lowest recorded values for a given month at a given location.

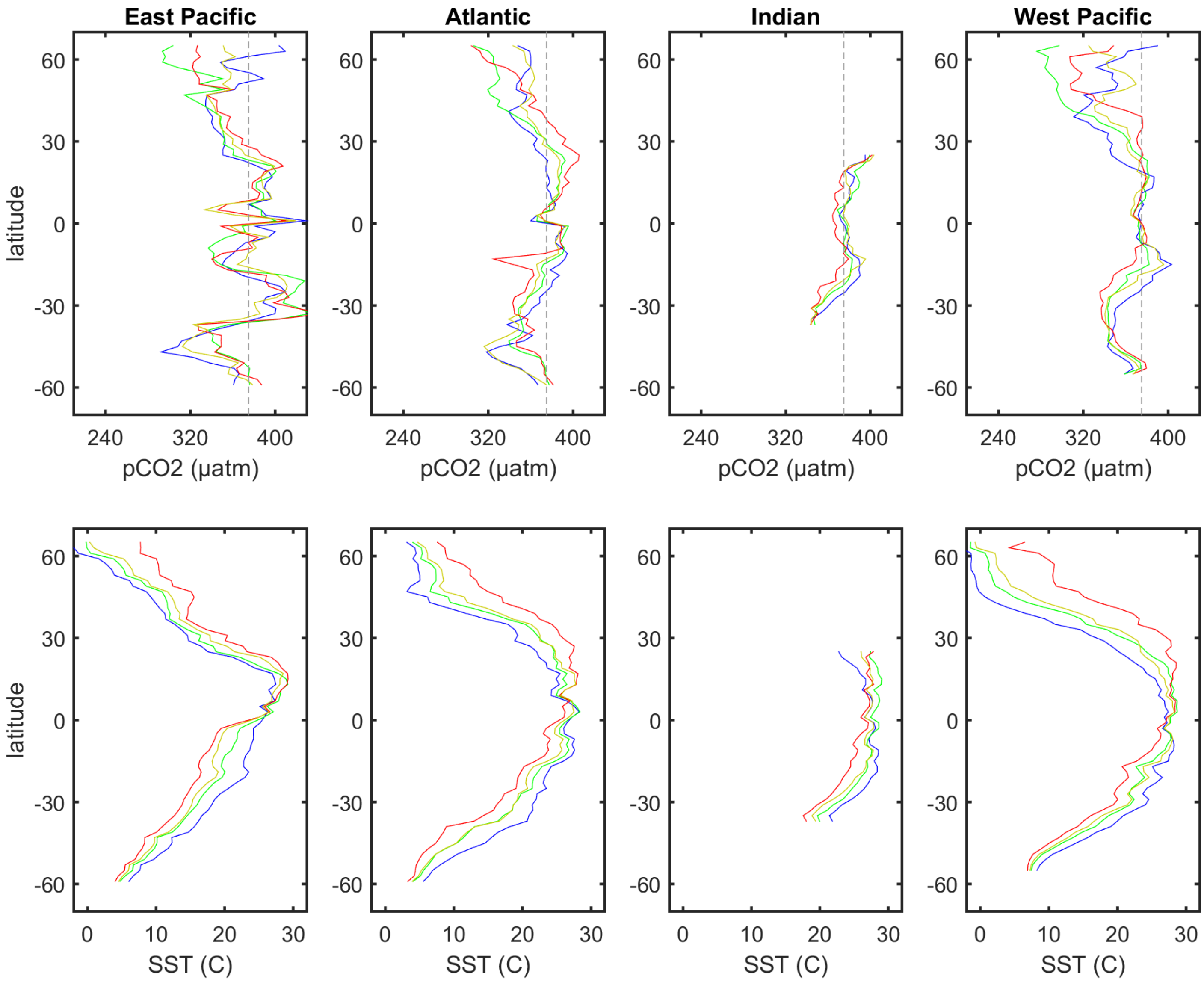

Figure 8: Seasonal-mean latitudinal profiles of pCO_2 (top) and SST (bottom) for the continental shelves surrounding 4 oceanic basins. Blue lines: averages over the months of January, February and March; green lines: averages over the months of April, May and June; red lines: averages over the months of July, August and September; yellow lines: averages over the months of October, November and December. The dashed line in the top panels represents the average atmospheric pCO_2 for year 2006.

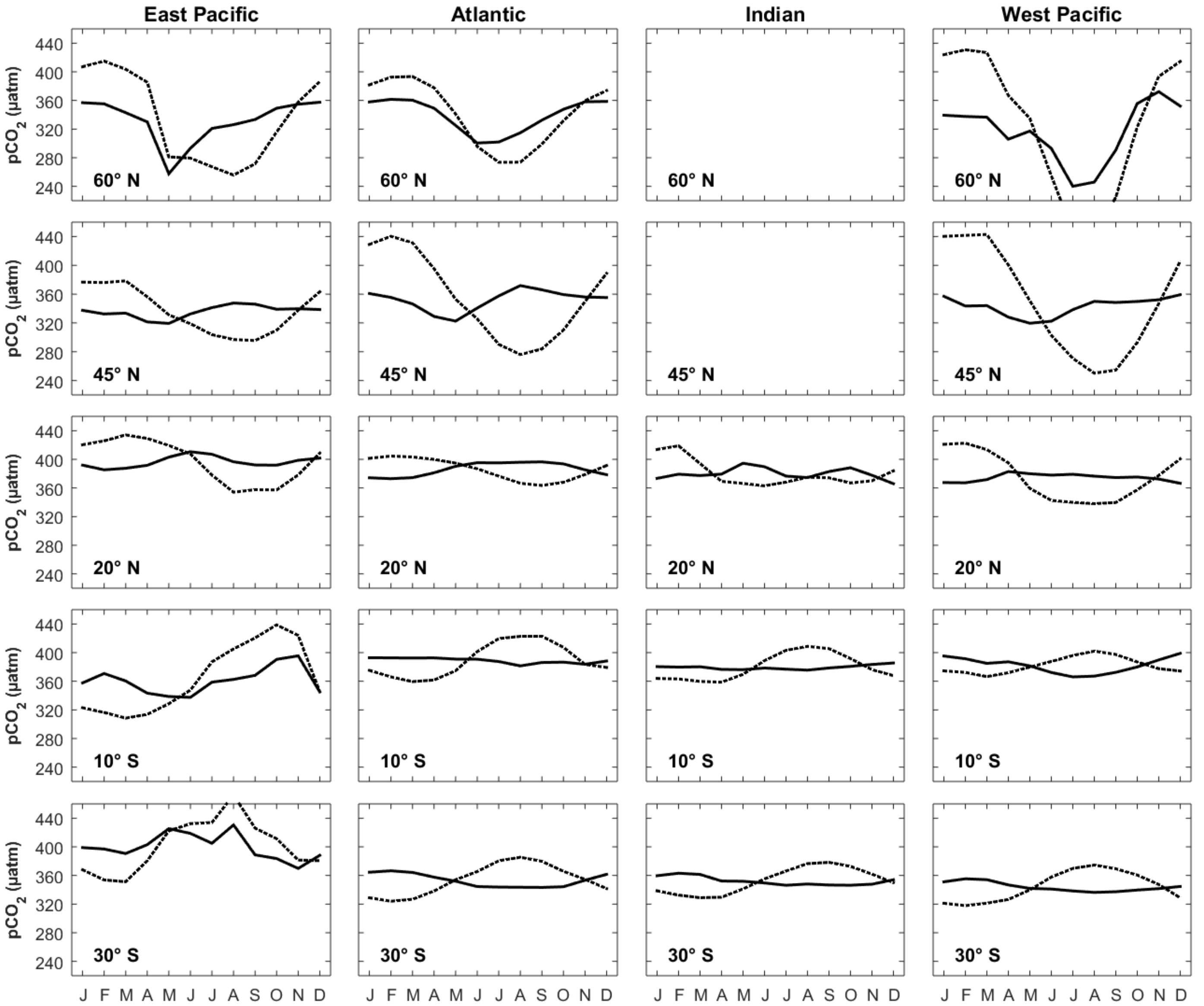

Figure 9: Seasonal cycle of observed (continuous lines) and temperature normalized pCO_2 ($pCO_{2(SSTmean)}$ dashed lines) at 5 different latitudes in 4 oceanic basins.