Biogeosciences Discuss., 10, C4785–C4815, 2013
www.biogeosciences-discuss.net/10/C4785/2013/
© Author(s) 2013. This work is distributed under
the Creative Commons Attribute 3.0 License.

Interactive comment on "A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink" *by* P. Landschützer et al.

P. Landschützer et al.

p.landschutzer@uea.ac.uk

Received and published: 6 September 2013

We would like to thank G. A. McKinley for the detailed review of our manuscript. The methodology presented in the manuscript is indeed intended to be used for further studies, e.g. a global analysis. We therefore very much appreciate the constructive suggestions and comments, which we will answer point by point below.

Major Comments

Reviewers Comment: 1. There are many parts to this analysis, and the reader

needs to understand the robustness of the final product. The RMSE to the data that goes into the NN is not sufficient. A more rigorous approach of using only part of the data to train the SOM and NN, and then comparing the resulting predictions to actual data that has been reserved from the training is needed. In the NN description in the appendix, there is reference to a "validation data set", but this is part of the NN methodology (as far as it is explained). Instead, I would like to see residuals (as in Figure 3) of 2006-2007 $p\mathrm{CO}_2$ to algorithms calculated with NN training only over the 1998-2005 period (or another reasonable choice of years). This would be a much better test of the results. What does an RMSE of 10 uatm really mean when the data to which you are comparing was used in the training of the NN?

Authors response: We agree that the validation with observations used within the method provides only a partial assessment of the robustness of the final product. Recognizing this, we emphasized the validation with independent timeseries products in the submitted version, but we clearly can do better and therefore added a section where we evaluate our results with new data from SOCAT v2. Notwithstanding these additional evaluations, we still think that the careful analysis of the residuals is an important part of the manuscript, particularly since residuals are often non-randomly distributed hinting at potential problems of the fits. E.g., we demonstrate in Table 2 that the unequal temporal distribution of the data does not lead to major hidden biases, as one could expect the method to fit the data better in the later years when the majority of observations were taken.

The suggested evaluations are good tests, but they have the disadvantage that they reduce the number of available observations for the building of the neural network model, hence impacting the entire estimate, not only the years excluded. Furthermore, the unequal distribution of data in time and space provides a major challenge with regards to the proposed tests. Here, we propose another evaluation. Since June this year an updated version of the SOCAT database (Version2) is available with additional observations within our study period that were not included in the version 1.5 we used. In total, the updated database includes 3065 new data points (roughly 15%) within our study region and period. This offers an opportunity for additional independent validation, without changing our estimate, i.e. we can compare the results presented with these data. Figure 1 in this response document, suggested being included as new Fig. 3 in the manuscript, showing the residuals in space and time equivalent to Fig. 2 in the manuscript.

Overall the RMSE between our Neural Network estimate and the new data is 22.83 μatm with a bias of 4.85 μatm.

We now introduce the new data at the end of section 2.1 (Data) as follows:

"In order to validate our results we use 3065 additional data points within our study region an period from the updated SOCAT v2 database (Bakker et al. 2013) which were not included in version v1.5 and therefore constitute independent data."

We then suggest adding a new paragraph and attached Figure 1 at the end of section 3.2 (Validation with independent observations) as follows:

"As a last test we use data from the recently updated SOCAT v2 database (Bakker et al. 2013), which provides new independent data points within our study period to validate the results. A total of 3065 gridded observations, spread over the entire Atlantic Ocean have been added for our study region from 1998-2007 representing 15% of the total amount of data used to train our network. Figure 3 shows the temporal mean and standard deviation of the residuals, similar to Figure 2. The largest misfit between our estimates and the additional observations can again be identified along the Gulf Stream and North Atlantic Current, confirming that our method fails to fully capture all variability within this region. Overall, the neural network estimates have a RMSE of 22.83 μatm and a bias of 4.85 μatm. When we exclude data north of 40°N, where we obtain the largest misfits, the results improve with a RMSE of 16.29

μatm and a mean difference of -1.12 μatm similar to the numbers obtained from the independent timeseries stations. This suggests that over most of the ocean, our method succeeds in predicting the observed $p\mathrm{CO}_2$ at any given time and place to within about 20 μatm, and a bias of a few μatm."

Reviewers Comment: 2. The paper also lacks discussion of the sensitivity of the results to input choices. There are many SST products, many MLD products, for example.

Authors response: We agree that the choice of products was subjective, and that it is a very good idea to discuss the sensitivity of our results to these choices. As it turns out, our basin-integrated results (decadal mean, trend, variability) are relatively insensitive to these data choices, confirming the robustness of our method. Also the effect on the provinces is fairly small since our provinces are mainly driven by the $p\mathrm{CO}_2$ climatology (discussed in point c below). In response, we have included a sensitivity discussion in our results section. This is explained below for each point a-c.

Reviewers Comment: a. What happens to these results if other choices are made?

Authors response: We now included 4 sensitivity runs, where we replace originally used products.

We suggest adding the following paragraph at the end of section 2.1 (Data) introducing the sensitivity runs:

"In order to evaluate the sensitivity of the results with regard to the chosen data product, we further performed 4 sensitivity runs, namely i) SR1 (sensitivity run 1) where we replace the SODA sea surface salinity with the World Ocean Atlas 2009 (Antonov et al. 2010) sea surface salinity climatology, ii) SR2, where we replace the ECCO2 MLD product with the de Boyer Montegut (de Boyer Montegut et al., 2004) MLD climatology, iii) SR3, where we use the SODA (Carton and Giese, 2008) sea surface temperature and iv) SR4, where we exclude chlorophyll a as an input parameter."

In section 3.4 ("Decadal mean $p\mathrm{CO}_2$ and air-sea CO_2 flux") we suggest adding the following paragraph:

"Comparing the decadal mean flux of -0.45±0.15 Pg C · yr^{-1} to the results derived from the sensitivity runs SR1-4 reveals that the choice of products does not significantly influence the long term mean result. The decadal mean fluxes from the sensitivity runs range from -0.41±0.14 (SR2) up to -0.48±0.16 Pg C · yr^{-1} (SR4) and are therefore well within the estimated uncertainty range."

In section 3.6 ("CO_2 trends and inter-annual variability") we suggest adding the following paragraph after page 8817 line3 (We would like to note here that trends are the linear fit to 12 month running average of our estimates [as described in the original manuscript] and uncertainties are the standard deviation of the fit. Furthermore, we discovered a typo in the original manuscript. The estimated $p\mathrm{CO}_2$ trend in the Atlantic Ocean is 1.46 and not 1.26 as falsely stated). This has been added following the comment below:

"The sensitivity runs reveal that trends estimates are barely influenced by the choice of the input data product, with the exception of SR2. While $p\mathrm{CO}_2$ and flux trends are statistically indistinguishable between our neural network estimate (1.46±0.76 μatm · yr^{-1}) and SR1-4 (1.42±0.59, 1.25±0.48, 1.48±0.77 and 1.37±0.73 μatm · yr^{-1} respectively), this is not always true for the fluxes. Here, SR2 reveals a flux trend (-0.26±0.03 Pg C · yr^{-1} · decade^{-1}) outside the uncertainty interval of our Neural Network estimate (-0.15±0.04 Pg C · yr^{-1} · decade^{-1})."

In section 3.6 we suggest adding the following to the paragraph p8818 line 3:

"This Atlantic Ocean low variability is further confirmed by the sensitivity runs, ranging from ±0.03 Pg C · yr^{-1} (SR1, SR2) to ±0.04 Pg C · yr^{-1} (SR3, SR4), indicating that our result is not sensitive with regards to the data choice."

Reviewers Comment: b. What happens if you do not use the surface chlorophyll? This is quite important to justify as it limits the results to post-1998.

Authors response: We show in point a (above) that the main basin-wide findings of this study remain statistically indistinguishable if we remove chlorophyll from the set of predictor variables. However, this does not imply that biology does not influence the variability of the Atlantic Ocean Carbon sink from 1998-2007. It simply implies that chlorophyll is not that critical for determining the large-scale mean properties of surface ocean $p\mathrm{CO}_2$ within our model. In response, we re-phrased P8806 line 21-24:

"Our analysis is restricted to the time period from 1998 to 2007 due to the temporal limitations of the data we chose for our study. No satellite chlorophyll data are available before the 1997 launch of the SeaWiFS mission, and the CO_2 observations in SOCAT v1.5 extend to the year 2007."

And p8819 line 18-21:

"Our results show that the main findings are statistically indistinguishable from those derived without chlorophyll a (SR4), indicating the possibility to expand the analysis period back in time in future studies. However, chlorophyll a is a simple, but important proxy representing the relation between biology and $p\mathrm{CO}_2$ and our results provide no evidence that chlorophyll can be neglected when considering longer timescales"

Reviewers Comment: c. What happens if you do not use the Takahashi $p\mathrm{CO}_2$ climatology in the biome definition?

Authors response: We state in the manuscript at page 8821 lines 13-18 that:

"We forced the relative weights of the input data toward the $p\text{CO}_2$ data, in order to minimize the variance of $p\text{CO}_2$ within each biogeochemical province. We do this by only log-normalizing MLD as input. As a consequence the range between the lowest and highest value of $p\text{CO}_2$ is one order of magnitude larger than that for SST, and about another order of magnitude larger than the remaining input parameters (log(MLD), SSS)."

As the Takahashi et al. (2009) $p\text{CO}_2$ product is a climatology, this means that the variance within this product stems from the seasonality. Hence, when forcing the SOM weights towards the variance of the $p\text{CO}_2$ product this results in provinces that follow the seasonal cycle of the $p\text{CO}_2$ product. The advantage we gain from this is that the seasonal signal stays as small as possible within each province and therefore (in an ideal case) the only variance left within a province stems from other sources of variability, e.g. inter-annual signals. With "in an ideal case" we mean here, that 16 regimes are not able to capture the full seasonality and there is some left-over signal of the seasonal cycle. We acknowledge that there are several ways to define biomes or provinces. The conventional ideas of biomes are steady regions, and we therefore tried to avoid the term biome to avoid confusion. We have tried several settings to define our regimes prior to submitting this manuscript, even some where we tried to keep the regimes steady. However, we found that we obtain the best fit (based on the independent observations) when the Takahashi et al. (2009) climatology is included.

When excluding the $p\text{CO}_2$ climatology, the fits to BATS and ESTOC data worsen considerably to a RMSE of 22.35 μatm with a bias of 11.04 μatm and a RMSE of 20.92 μatm and a bias of -15.75 μatm respectively. This compares to a RMSE of 17.53 μatm and a bias of 7.56 μatm at BATS and a RMSE of 16.85 μatm and a bias of -8.06 μatm at ESTOC in the current configuration (pate 8810 lines 19-22).

Reviewers Comment: 3. The use of Takahashi $p\text{CO}_2$ climatology in both biome definition and in the validation of results (Figure 8) is another significant concern. To what degree are your results influenced by the assumptions of the Takahashi $p\text{CO}_2$ climatology?

Authors response: This appears to be a misunderstanding since we never aimed to validate our results with the Takahashi et al. (2009) $p\text{CO}_2$ climatology. Rather we wanted to compare our results with theirs, simply because the Takahashi $p\text{CO}_2$ climatology is to date the most commonly used $p\text{CO}_2$ estimate and it behoves each new estimate to establish how well it compares to the current standard.

Reviewers Comment: 4. Overall, there is a lack of clarity in presentation of the methodology. Some examples: the figures that intend to present validation are only discussed with a sentence or two, the appendices are not directly referred to, acronyms and other terms are not defined. More detail on this is given with the "minor comments" below.

Authors response: We will comment on each point below.

Minor Comments

Reviewers Comment: Page 1 Content: "was also increasing" Comment: Replace with "increased"

Authors response: This has now been corrected to "increased".

Reviewers Comment: Page 3 Content: "Here, we overcome most of these limitations by presenting a new neural network- based approach, which determines the non-linear relationships between the surface ocean pCO$_2$ observations and a set of independent observations to produce basin- wide sea surface maps of pCO$_2$ on a monthly basis." Comment: This statement is over-confident. Please comment on the limitations of this method.

Authors response: We agree that limitations of the method should be stated in this paragraph. We suggest including them on page 8804 line 14: "... in the Atlantic Ocean. Our method relies on the assumption that the Atlantic Ocean carbon sink and its variability can be estimated as a function of proxy variables, which are subjectively chosen. We further rely on ocean carbon measurements in order to establish a correct relationship. We therefore benefit from the recent publication ..."

Reviewers Comment: Page 4, Comment: Thank you for including the fCO$_2$ to pCO$_2$ equation, rather than just stating "we converted"

Authors response: You are welcome.

Reviewers Comment: Page 4 Content: "For SST, we use the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation (OI) sea surface temperature v.2 (Reynolds et al., 2002), for CHL the SeaWiFS mapped chlorophyll (SeaWiFSProject, http://oceancolor.gsfc.nasa.gov/cgi/l3), for MLD the mixed layer depth data from the Estimating the Circulation and Climate of the Ocean, Phase II (ECCO2) project (Menemenlis et al., 2008), for SSS the Simple Ocean Data Assimilation (SODA) sea surface salinity data (Carton and Giese, 2008) and for xCO2,atm the monthly atmospheric CO$_2$ from GLOBALVIEWCO2

(2011). Furthermore, the monthly pCO$_2$ climatology of Takahashi et al. (2009) is used as an additional input parameter for defining the biogeochemical provinces. Due to their strongly skewed distribution, mixed layer depth (MLD) and chlorophyll a (CHL) were log-transformed before use as predictor values." Comment: Discussion of the sensitivity to these choices is needed somewhere in the text. In the work of my group, we have found significant sensitivity to choices such as the MLD climatology - for example, the choice of ECCO vs de Boyer MLD vs ARGO climatological MLD can influence our biomes significantly. How do such choices impact these results?

Authors response: We have addressed this issue in point 2 of the major comments section.

Reviewers Comment: Page 4 Comment: The use of the word "binned" might be reconsidered if it is really averaging up to a larger spatial scale. Unless you are including the original number of data points in further calculations, "binned" might be the wrong choice of word. "Averaged" should suffice.

Authors response: We have re-phrased this sentence (page 8806 line 25-27) to: "Data with an original resolution finer than the required 1°×1° were averaged onto the desired grid, whereas input data with a coarser resolution were interpolated using a bilinear interpolation."

Reviewers Comment: Page 5 Content: "Input vectors with empty vector elements were removed from the datasets" Comment: Thus in regions of very little pCO$_2$ data, your fits will be much less robust. How is this dealt with, and how much impact does it have? The terminology here of "input vectors" and

"targets" needs explanation.

Authors response: The removal of empty vector elements refers to all input data, not only those who have co-located observations. E.g. the original products all come from different resolutions, therefore the land-sea masks, or ice-masks are different. This leads to $1^\circ \times 1^\circ$ pixels that are occupied by certain proxies, but are empty in others. By removing empty vector elements we only use those pixels that are occupied by all proxy data and in the case of the feed-forward input data set (FINP) which are co-located to the observations.

We do not agree that our fits are automatically less robust where only a few observations exist. This is very much dependent on the region and the biogeochemical complexity of the region (see e.g., the response to major comment 1, where the residuals in the South Atlantic are smaller than those in the Gulf Stream and North Atlantic Current regions, despite having more observations there). If the input and target variables are sufficient to reconstruct the $p\mathrm{CO}_2$ relationship within one province, the results may be very robust.

We have revised this paragraph (page 8807 line 1-5), as the terminology was not entirely clear. We further thought that the employed datasets (Table 1) should be introduced here. The paragraph now reads:

"In the next step the monthly $1^\circ \times 1^\circ$ input data are rearranged into 3 major data sets. Each of these data sets consists of input vectors ($\boldsymbol{p}_n$) where the input data are organized as row vector elements, for example SST, log(MLD), SSS, and $p\mathrm{CO}_{2,Takahashi}$ for the self-organizing map input (SINP) dataset, sampled at the same space-time point (Table 1). Two of these sets, SINP and the feed-forward network input 2 set (FINP2) are global sets and do not have a corresponding target dataset (Table 1). Input vectors with empty vector elements, e.g. where no salinity data are available, were removed from these data sets. The third major set, the feed-forward network input set

(FINP), consists only of input vectors where corresponding SOCAT v1.5 observations, or targets (t), are available, i.e. they are subsampled in time at the locations where observations are available. In order to train the feed-forward network, two sub sets of the FINP set are created, namely the actual training (FITR) set and a validation (FIVAL) set (Table 1, Appendix A2)."

Reviewers Comment: Page 5 Content: "Where no chlorophyll a satellite data are available, due to cloud cover, we estimate the sea surface $p\mathrm{CO}_2$ only with the remaining input parameters." Comment: What percentage of the cases? How is this spatially biased? Does it impact results?

Authors response: Overall this applies to about 22% of all pixels and it mainly concerns the high latitudes in winter. Regarding the effect on the results we refer to our response on major point 2. Chlorophyll concentrations in winter, where most of the values are missing, tend to be low and do not play such an important role, compared to other seasons, e.g. spring. We agree that this issue has to be mentioned in the statement above. We therefore adjusted this sentence (page 8807 line 10-11) to: "Where no chlorophyll a satellite data are available, due e.g., to cloud cover or lack of sufficient light, we estimate the sea surface $p\mathrm{CO}_2$ with the remaining input parameters. This applies to about 22% of all pixels and mainly concerns the high latitudes oceans in winter. This is not ideal, however chlorophyll concentrations tend to be low and photosynthesis only has a minor effect on $p\mathrm{CO}_2$ in the cold winter months."

Reviewers Comment: Page 5 Content: "We use a self-organizing map (SOM) method (Kohonen, 1987, 2001) to partition the global ocean into 16 regimes of similar patterns, i.e., biogeochemical provinces. The choice of 16 provinces represents a subjectively determined optimum between too many regions with too little data and a high degree of correlation between the provinces, and too

few regions with a lot of data, but too high variance in the data. The monthly SST, log(MLD), SSS, and climatological $p\mathrm{CO}_2$ data of Takahashi et al. (2009) were used as input for the SOM(see Table 1). We chose not to include chlorophyll, i.e., log(CHL), due to missing values from cloud cover. Details on the SOM method can be found in the Appendix." Comments: A better description of how the SOM works is needed here, and also include reference to the relevant appendix. Included in here is the $p\mathrm{CO}_2$ data as analyzed to a climatology by Takahashi et al. (2009). Thus, the comparison to these data in section 3 does not appear to be fair. You have already wrapped these data in. Please explain. These data are monthly resolution or greater. Are the provinces moving by month when used as input for the feed-forward method?

Authors response: We suggest to adjust this paragraph (page 8807 line 17-25) to briefly explain the SOM in the main text body:

"We use a self-organizing map (SOM) method (Kohonen, 1987, 2001) to partition the global ocean into 16 biogeochemical provinces, characterized by all data observations having a similar relationship among all input variables of the SINP data set, i.e., climatological $p\mathrm{CO}_2$ as well as the independent variables SST, log(MLD) and SSS. The provinces change in shape from one month to the next and further change slightly between years. A SOM is a neural network based cluster algorithm that can detect regularities within the provided input data and then learns to group them together. Similar input data, arranged as input vectors, are identified via their Euclidean distance towards the nodes (or neurons) of the network. The choice of ..."

We suggest to change the last sentence of this paragraph (page 8807 line 24-25) to provide the reference to the relevant section of the appendix: "Details on the SOM method can be found in the Appendix A1."

Again, the comparison in section 3 does not aim to validate our results. We simply

included this comparison as we were convinced that readers might be interested how our product compares to this widely used product.

Reviewers Comment: Page 5 Content: "Despite their strong seasonal dynamics in space (Fig. 1a) and time (Fig. 1b), the estimated biogeochemical provinces exhibit a coherent large-scale behavior, reflecting the well known oceanic structures such as the gyres, the equatorial regions, and the high-latitude North Atlantic." Comment: There is inadequate discussion of the figures. Please guide the reader through the regions in figure 1a. Please explain the significance of figure 1b, which tells you that some pixels occupy as many as 9 regions over the course of the period. Few stay in the same one the whole time. Is this sensible or just statistical jibberish?

Authors response: We thought that the combination of figure 1 a and b gives a good idea on how dynamic the biogeochemical provinces are and how the dynamics work. However, we concede that the discussion of the plots was sometimes too short in the main text. We therefore added at the end of paragraph page 8807 line 26-27 and page 8808 line 1-2:

"We do not provide any additional time or space information to the SOM, hence the regions are strongly influenced by the temporal variability of the input data, in particular the seasonal variability within the climatological $p\mathrm{CO}_2$, and are therefore not static, unlike conventional provinces or biomes. Despite their strong seasonal dynamics in space (Fig. 1a) and time (Fig. 1b), the estimated biogeochemical provinces exhibit a coherent large-scale behaviour, reflecting the well known oceanic structures such as the gyres, the equatorial regions, and the high-latitude North Atlantic. Figure 1a shows the mode of the provinces, i.e., the province each pixel mainly belongs to from 1998-2007 and figure 1b shows the number of shifting provinces per pixel. These provinces vary in time and space mainly in accordance with the variability of the climatological

pCO$_2$. In the tropics, and the high latitude North Atlantic, the climatological pCO$_2$ vary little seasonally and therefore the provinces remain fairly steady, with only minimal province shifts. In contrast, the gyre regions of both hemispheres exhibit much larger seasonal variability, hence pixels there undergo many more province changes. We find the largest shifts along the Gulf Stream, where certain regions change their province association up to 10 times."

Reviewers Comment: Page 5 Content: "As a second step we use a feed-forward network (FFN) method to reconstruct the non-linear relationship between our input variables and the target, i.e., pCO$_2$, separately for each of the 16 biogeochemical provinces. The FFN method is a type of back- propagation network method that is capable of approximating any function with a finite number of discontinuities (Demuth et al., 2008). The established relationship is further used to predict the pCO$_2$ for each point in time and space where no observations are available." Comment: Please also explain the FFN with some more detail here and also refer to the appendix explicitly.

Authors response: We suggest adding the following to give a more detailed explanation of the FFN method in the main text body on page 8808 line 3-9:

"As a second step we use a feed-forward network (FFN) method to reconstruct the non-linear relationship between our input variables and the target, i.e., pCO$_2$, separately for each of the 16 biogeochemical provinces. The FFN method is a type of back- propagation network method that is capable of approximating any function with a finite number of discontinuities (Demuth et al., 2008). Similar to multi linear regressions, a feed-forward network adjusts coefficients to establish a relationship between inputs and targets. The adjustment of the coefficients follows an iterative process. The first iteration includes an initial guess, where the coefficients are randomly initialized, the estimates are computed and compared to the target observations. From there

on the network goes backwards (hence the name backpropagation) and automatically re-adjusts the coefficients with the aim to reduce the mean squared error between estimates and targets. For each iteration, only a random subset of the data is used to train the network, while the remaining data are used for validation. The updating process of the coefficients is repeated until the network estimates derived from the validation set no longer improve significantly relative to the targets. The established relationship is further used to predict the pCO$_2$ for each point in time and space where no observations are available. This process is explained in more detail in Appendix A2."

We further refer now directly to the relevant appendix at the end of the following paragraphs on page 8808 line 16:

"Details on the settings used for the FFN can be found in the Appendix A2."

and on page 8808 line 21:

"More details are provided in the Appendix A3."

Reviewers Comment: Page 5 Content: "FINP" Comment: Define acronym prior to use.

Authors response: The acronym is now defined in the revised paragraph in section 2.1 page 8807 line 1-5 (see comment above regarding the terminology on page 5)

Reviewers Comment: Page 5 Content: "Due to the temporal and spatial variability of the regimes and the heterogeneous distribution of the pCO$_2$ data, large differences exist in the number of observations within the different provinces." Comment: What are the implications of this heterogeneity? What is the sensitivity to these settings? These things need discussion in the main text.

Authors response: In section 3.1, we actually tested the network outputs and discussed potential issues arising from heterogeneity. We stated on page 8809 lines 21-25 that:

"In conclusion, the residuals indicate that the combined SOM-FFN method fulfils most tests for a good fit and does not contain any major hidden biases. In particular, there is no indication of a substantial degeneration of the fits as a function of data density, neither in time nor in space. Regions with high spatial or temporal variability are the least well fitted, while the fits for most of the open ocean are very good."

We do however agree that we can provide more clarity by adding:

"Due to the temporal and spatial variability of the provinces and the heterogeneous spatiotemporal distribution of the $p\mathrm{CO}_2$ data, large differences exist in the number of observations within the different provinces. However, our neural network fit does not show degeneration as a function of the data density, as shown in section 3.1 for the temporal distribution and the spatial heterogeneity of the data does not lead to any major hidden bias."

Reviewers Comment: Page 6, Lack of independent data outside the North Atlantic subtropics should be noted as a challenge for your validation.

Authors response: Considering the new data provided by the SOCAT v2 dataset we believe that the proposed statement is now obsolete.

Reviewers Comment: Page 6, Please provide more text explanation for the residuals presented in Figure 3. There are some large residuals, causing concern with respect to the very low estimated errors RMSE of 10 and bias of -.10). Is this analysis based on the raw or 1x1 monthly data? Are these the same data that went into the NN analysis?

Authors response: The analysis is based on the $1^\circ\times1^\circ$ monthly data, and the residuals are computed from the difference between our $1^\circ\times1^\circ$ monthly estimates and the $1^\circ\times1^\circ$ monthly gridded observations. These are the $p\mathrm{CO}_2$ data that went into the neural network analysis. There are indeed some "extreme" residuals (up to 100 μatm difference). However, the box and whiskers plots in Figure 3 show that these are single outliers and the majority of the residuals stay close to 0 μatm, hence the low RMSE. These "extreme" residuals come mostly from near shore areas or from frontal regions.

We have now expanded the final sentence of the caption of Figure 3 for more clarity regarding the symbols:

"The upper plot in each panel shows the residuals, shown as a box-and-whiskers plot. The red line in the box show the median, the blue box indicates the 25 and 75 percentiles and red plusses mark residuals outside this interval. The lower plot shows the relative number of observations within each bin"

We suggest to add the following text at p8809 line 16 at the beginning of the paragraph dealing with Figure 3:

"To test the impact of the inhomogeneous distribution of the neural network input data and $p\mathrm{CO}_2$ observations, we show the residuals, calculated as the difference between the neural network $p\mathrm{CO}_2$ estimates and the gridded SOCAT v1.5 $p\mathrm{CO}_2$ observations (Fig. 3)."

We further suggest to add the following at the end of the paragraph (p8809 line 20):

"Figure 3 further shows that large residuals, most of which stem from regions characterized by strong horizontal $p\mathrm{CO}_2$ gradients, are independent of the data density."

Reviewers Comment: Page 7 Content: "Given the overall small bias and the low RMSE between the two very different methods to interpolate the data, it

appears that the long-term mean surface ocean $p\text{CO}_2$ can be very robustly estimated from the available observations." Comment: Again, what is the impact of including the climatology in the definition of your biomes?

Authors response: This question is dealt with in the major comments (item 2) above.

Reviewers Comment: Page 8 Content: "To determine the drivers behind the seasonal cycles, we split the long-term mean seasonal cycle at each grid cell into a thermal and into a non-thermal component (Takahashi et al., 2002; Sarmiento and Gruber, 2006)." Comment: Which sst data is used? Please state the equation applied.

Authors response: We use the same SST product as has been used for the neural network training. We suggest adjusting the above sentence for clarity and state the equation applied (page 8814 line 11-13):

"To determine the drivers behind the seasonal cycles, we split the long-term mean seasonal cycle at each grid cell into a thermal and into a non-thermal component (Takahashi et al., 2002, equations 1 and 2; Sarmiento and Gruber, 2006), i.e., assuming a 4% change in $p\text{CO}_2$ per unit change in SST), and employing the same SST product used for the network training."

Reviewers Comment: Page 8 Content: "seasonal cycles of the thermally and non-thermally driven partial pressures tend to cancel each other (Fig. 9), consistent with previous analyses (Takahashi et al., 2002; Sarmiento and Gruber, 2006)." Comment: Include in fig 9 plots of Takahashi climatology thermal and non-thermal cycles.

Authors response: We have now included the equivalent Figures (Figure 2 attached to this document) for the Takahashi climatology as Figures 9c and 9d in the original manuscript.

Reviewers Comment: Page 9 Content: "-1 i.e., 1.80 muatm/yr versus 1.90 muatm/yr" Comment: Include uncertainty on these trends, and on all others discussed here, similarly for the interannual variability.

Authors response: We have now added uncertainties based on the standard deviation of the linear fit to all trend estimates, e.g. 1.80$\pm$0.77 μatm $\cdot$ yr^{-1} and 1.90$\pm$0.34 μatm $\cdot$ yr^{-1}. The numbers we report regarding the inter-annual variability (e.g. 0.02 PgC $\cdot$ yr^{-1} for the South Atlantic) are standard deviations of the IAV, hence we added a $\pm$ sign to all estimates in the text, e.g. $\pm$0.02 PgC $\cdot$ yr^{-1} for the South Atlantic.

Reviewers Comment: Page 9 Content: "1.46 muatm/yr for the non-thermal component, while the thermal driven trend is on average 0.37muatm/yr)." Comment: Again, need to quote uncertainty. Is the thermal trend distinguishable from 0?

Authors response: We added again uncertainties based on the standard deviation of the linear fit to our estimates. For the thermal driven trend in the North Atlantic we compute 0.37$\pm$1.47 μatm $\cdot$ yr^{-1} and in the South Atlantic 0.19$\pm$0.79 μatm $\cdot$ yr^{-1}. Similar for the non-thermal trend we compute in the North Atlantic 1.46$\pm$1.75 μatm $\cdot$ yr^{-1} and in the South Atlantic 0.76 $\pm$1.30μatm $\cdot$ yr^{-1}. Hence none of these trends are statistically distinguishable from zeo.

We therefore re-phrased the sentence page 8816 line 25-27:

"Similar to the North Atlantic, the non-thermal component of the $p\text{CO}_2$ with an average trend of 0.76±1.30 μatm · yr^{-1} appears to be stronger compared to 0.19±0.79 μatm · yr^{-1} of the thermal component. However, given their uncertainty, we cannot statistically distinguish both trends from zero."

Reviewers Comment: Page 9 All trends calculated here are initiated in 1998, which follows a very strong ENSO event. Fay and McKinley 2013 indicate that this choice of start year influences pCO$_2$ trends around the world's oceans, presumably due to ENSO's global influence. What happens if you shift your trend calculation to 1999 or to 2000? Presumably, there would be significant change in the results.

Authors response: We agree that the start and end year have a significant influence on the calculated trends. In the manuscript we therefore stated (page 8817 line 4-7):

"It is not possible to conclude from our data whether the 10-yr trends we identify are part of a longer term trend (Schuster et al., 2009) or whether they are part of a decadal time-scale variability (Thomas et al., 2008; Gruber, 2009; McKinley et al., 2011). The most recent study by (McKinley et al., 2011) suggests the latter to be the case, ..."

An in-depth trend analysis with variable start and end years, as done by McKinley et al. (2011) and Fay and McKinley (2013), would go far beyond the aim of this manuscript, but we do agree that the relevant papers dealing with this issue need to be cited in this context. We therefore suggest re-phrasing the above to:

"It is not possible to conclude from our data whether the 10-yr trends we identify are part of trends that exceed the 10 years of our analysis, or whether they are part of a decadal time-scale variability (Thomas et al., 2008; Gruber, 2009; McKinley et al., 2011). The most recent studies by McKinley et al., (2011) and Fay and McKinley (2013) suggest the latter to be the case. The authors show that short term trends on timescales similar to this study are strongly influenced by the chosen start and end

year and strongly reflect climate mode signals such as the ENSO signal, which are likely to effect the trends calculated in our analysis. However, reported 50-yr ..."

Reviewers Comment: Page 10 Content: "The most recent study by (McKinley et al., 2011) suggest the latter to be the case, but reported 50-yr trends in heat storage (Levitus et al., 2012) and interior ocean oxygen changes in the North Atlantic (Stendardo and Gruber, 2012) indicate that the North Atlantic and in particular its subpolar gyre has been subject to multi-decadal changes. " Comment: Variability can occur on mutlidecadal timescales. But you imply that this would be a "trend". What do you mean, specifically, by a longterm trend? If you mean the long-term response of the ocean to anthropogenic climate warming, please state so. If you mean something other than this, please state that clearly.

Authors response: We agree with the reviewer that our 10-yr linear trend estimates were computed over a relatively short period. We do not imply that any 10-yr trend we determine is part of a longer-term trend (>10 years), but instead could be a persistent part of some low frequency variability.

Reviewers Comment: Page 10 Content: "-0.39±0.13PgC/yr in 2001 up to -0.56±0.18PgC/yr" Comment: These are not formally distinguishable by any reasonable measure, given their uncertainty. You cannot quote them as evidence of interannual variability.

Authors response: We do agree that this statement was misleading and that these numbers are not statistically significantly different. We therefore suggest re-phrasing this statement and avoid the use of the terms "substantial" in the previous sentence

(8817 line 21 to 23):

"Integrating our monthly air-sea CO_2 flux estimates for each year over the Atlantic Ocean reveals the largest annual mean flux differences during the second half of our study period (Fig. 11a), where annual mean fluxes range from -0.39±0.13 Pg C $\cdot$ yr^{-1} in 2001 up to -0.56±0.18 Pg C $\cdot$ yr^{-1} in 2006."

Reviewers Comment: Page 11 Content: "would be beneficial to extend the study period to further investigate responses to climate modes such as the NAO and to investigate multi-decadal variabilities. Currently however, we are limited to 1997 since no basin-wide chlorophyll a measurements are available before and chlorophyll a is a simple, but important proxy representing the re- lation between biology and pCO$_2$." Comment: How much more error do you get without chl in your method? Are you saying that you cannot get reasonable results without satellite chl? If so, please state more clearly.

Authors response: We commented on this issue in the major comments 2a and 2b above. Here we would like to point out that this statement is not related to any error analysis. Chlorophyll is the only proxy representing the effect of biology within our method.

Reviewers Comment: Page 11 General comment on text - There is mixed terminology throughout the paper with regard to the strength of the sink/trends. "significant", "stronger", "slower increase", "lower undersatution". More consistent verbiage would make the message more clear.

Authors response: To be more consistent with our verbiage we suggest using the terms "stronger" and "weaker" and withdrawing from the usage of "slower increase"

and "lower undersaturation". The term "significant" is now only used when this is not contradicted by the given uncertainty. We suggest withdrawing from the use of "lower undersaturation" in "having a trend towards lower undersaturation over" on page 8815 line 25-26 and replace it with "having a trend towards a stronger increase of the sea surface pCO$_2$"

We suggest replacing "Trends for the South Atlantic show a slower increase ..." (page 8816 line 21) with "Trends for the South Atlantic show a weaker increase ..." We suggest replacing "... on basin scale by slower increasing trends elsewhere ..." (page 8819 line 10-11) with "... on basin scale by weaker trends elsewhere ..."

Reviewers Comment: Page 11-14 Appendix general comment - The Appendix is quite difficult to follow. The lack of definition of many terms (e.g. Weight matrix, "winner", distance function) is a significant problem with making the methodology make sense. Please define terminology, and otherwise carefully proofread to enhance clarity.

Authors response: We do agree that the appendix was very technical and hard to follow. We therefor re-phrased section A1 to provide more explanation to the terms used:

We adjusted page 8820 paragraph 1 and it reads now:

"A map with 16 neurons was chosen, organized on a 2 dimensional 4×4 point hexagonal grid. This means that the input data are clustered into 16 neurons, which represent the 16 biogeochemical provinces. The term neuron refers to a processing unit, which consist of a weight vector, where each element of the weight vector corresponds to one input parameter. In our case each weight vector consists of 4 elements (SST, log(MLD), SSS, pCO$_{2,Takahashi}$), representing its co-ordinates and the distance between 2 neurons is calculated via a distance function. These processing units are initially spread over a 2 dimensional field, in our case in a hexagonal formation, forming a single layer of neurons. Our experience has shown, however, that the choice of neuron topology does not have a significant effect on the final province distribution. The use of neurons, their initialization and their distance relation describes the biggest difference towards other clustering algorithms, e.g. k-means clustering. For our study, the Euclidean distance between a neurons weight vector and the input vectors of the SINP dataset was used for the distance function. The weight matrix ($\mathbf{W}_{m=16,n=4}$), which is formed by the 16 neurons with their 4 vector elements, was randomly initialized."

We added following sentence at page 8820 line 15:

"... of the input vector $\boldsymbol{p}_n^j$. The smallest element of the distance vector, i.e. the shortest distance element, marks the distance towards the closest neuron, called the winning neuron. The neuron i, gets updated ..."

We edited page 8821 line 13 onwards:

"We forced the relative weights of the input data toward the climatological $p\mathrm{CO}_2$ data, in order to minimize the variance of $p\mathrm{CO}_2$ within each biogeochemical province. To do so, we did not normalize our input data, with the exception of MLD, which we log-transformed (Table 1). As a consequence, the range between the lowest and highest value of $p\mathrm{CO}_2$ is one order of magnitude larger than that for SST, and about another order of magnitude larger than that for the remaining input parameters (log(MLD), SSS)."

Reviewers Comment: Page 12 Content: "As a consequence, the biogeochemical provinces follow the seasonal pattern of the $p\mathrm{CO}_2$ climatology, meaning that the seasonality of $p\mathrm{CO}_2$ at any given location will be mostly determined by the seasonal changes of the biogeochemical provinces and to a lesser degree by the seasonal cycle of the input data in the second state." Comment: Again, you have imposed the $p\mathrm{CO}_2$ climatology on your results. How, then, can you justify

your comparison/validation in figure 8?

Authors response: This issue is discussed above for major comment 3.

Reviewers Comment: Page 19, Put your results all the way to right, and next to the RECCAP best estimate

Authors response: We agree that for a comparison this is more convenient. We have now put our results to the right, next to the best estimate of the RECCAP Atlantic study (Table 3).

Reviewers Comment: Page 20, Figure 1a - If the provinces are changing boundaries please identify what year/month this map is for.

Authors response: We have stated in the figure caption that Figure 1a represents the "province number of the mode". Hence it is the mode of all years and months for each pixel. We thought that in combination with Figure 1b these plots best represent the structure and dynamics of the provinces. Figure 1 is now introduced in Section 2.2 (paragraph from page 8807 line 26 to page 8808 line 2), as explained above in response to the final Page 5 Comment.

Reviewers Comment: Page 20, Why in 2a, the NAC region is somewhat filled in, but in 2b there is a much larger blank area? Please explain.

Authors response: We calculate the standard deviation of the residuals within each pixel only when there are at least 2 observations available in time. Hence, if there is only 1 observation available in time, this pixel will have a mean error (a) but no standard deviation (b) and the pixel is empty.

To clarify this issue we added the following sentence to the caption of Figure 2:

"Pixels that have a value in (a) but not in (b) indicate where only 1 observation in time is available."

Reviewers Comment: Page 22, In 5b, and 5a, what is going on between Greenland and iceland with very high pCO$_2$ and high efflux? Looks wrong. Why does the method allow this feature to be retained? Does it impact results?

Authors response: The high pCO$_2$ between Greenland and Iceland (at around 60°N in Figure 5) stems from the high pCO$_2$ values measured during the winter months in the subpolar gyre. This feature is present in the original observations (www.socat.info), as well as in previous studies (see e.g. Takahashi et al. (2009, Figure 13, or Watson et al. (2009), Figure 3). Figure 2 shows that this region is well sampled and the misfit between our estimate and the observations is rather small. We are therefore fairly confident that this feature is real and not an artefact of the method. Surely this has an impact on the sink strength of the North Atlantic, but due to the small surface area, the overall (or basin-wide) effect is small. Furthermore, GLOBALVIEW atmospheric CO$_2$ shows strong seasonality, with high atmospheric CO$_2$ in high latitudes of the northern hemisphere in winter. This counterbalances the high surface water pCO$_2$ and retains an ocean sink rather than providing a source.

Reviewers Comment: Page 23, Presumably a zonal average. This needs to be stated in the caption. Ditto for following figures.

Authors response: Correct. Figures 7-9 are zonal averages. This has now been added to the figure captions:

Figure 7: "Hovmöller plot of the zonally averaged long term mean ..."

Figure 8: "Zonally averaged difference in the surface ..."

Figure 9: "Zonally averaged mean seasonal cycle of the ..."

Reviewers Comment: Page 25: Why are the green triangles not all on the black line? Doesn't the black line come from the same data that the green triangles comes from, thus shouldn't all the green triangles fall on the black line?

Authors response: In Figure 12 the difference between the green triangles and the black lines is due to the spatial difference. The black line represents the spatial average of the entire 10×10 degree grid box (average of 100 pixels), whereas the green triangles only represent the spatial average of those pixels which have co-located observations. Depending on the location within the boxes and the number of observations, differences may occur. Hence one could also expect the green triangles to be closer to the red triangles (representing the spatial average of all observations within these boxes). However, due to the model misfit, there is a difference between the red and green triangles.

To clarify this issue we re-phrased the figure caption to:

"... The black line shows the spatial average pCO$_2$ within each 10°×10° box. Red triangles illustrate the average sea surface measured pCO$_2$ within each box where observations are available and the green triangles represent the average of the neural network pCO$_2$ of those 1°×1° pixels which have co-located pCO$_2$ observations in SOCAT v1.5. (c) ..."

Interactive comment on Biogeosciences Discuss., 10, 8799, 2013.

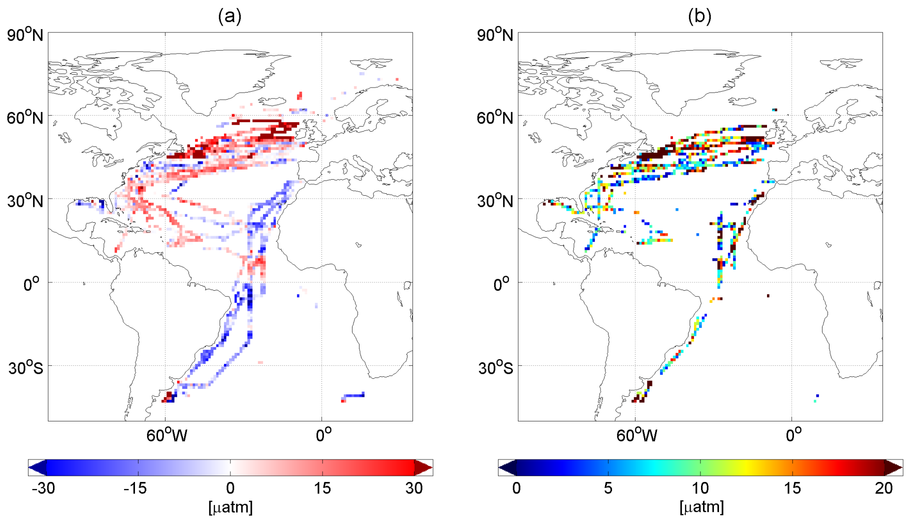

Fig. 1. (a) Temporal mean residuals and (b) standard deviation of the residuals in μatm between neural network estimates and independent data points obtained from the SOCAT v2 gridded observations.

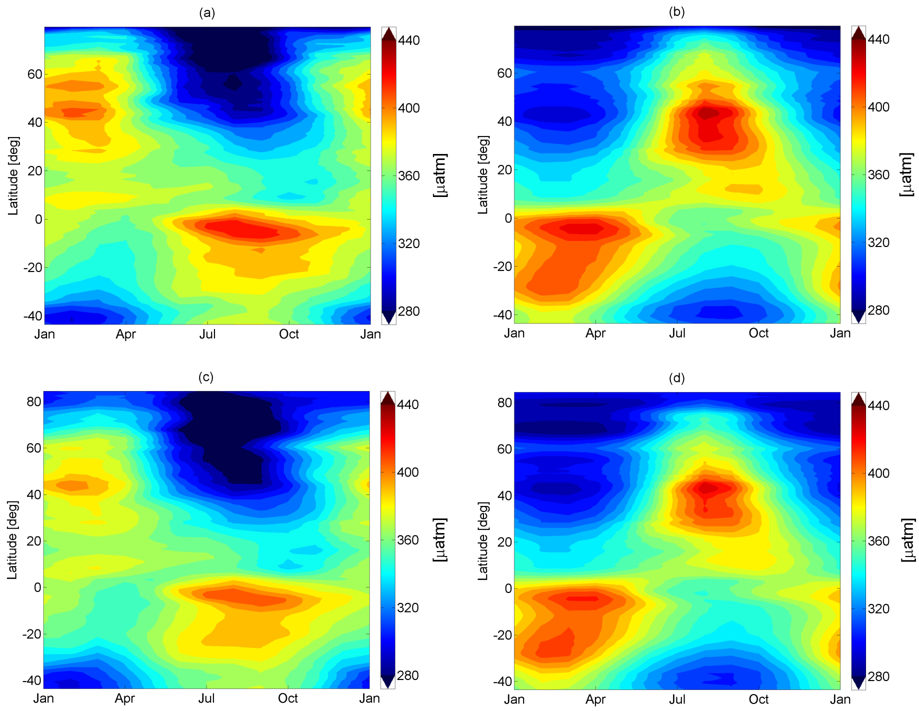

Fig. 2. ... Figure (c) and (d) show the non-thermal and thermal component respectively for the Takahashi et al. (2009) climatology. The decadal mean pCO2 has been added to both components