A two-step neural network model for the reconstruction of surface ocean pCO$_2$ over the Global Ocean.

Anna Denvil-Sommer$^1$, Marion Gehlen$^1$, Mathieu Vrac$^1$, Carlos Mejia$^2$

$^1$Laboratoire des Sciences du Climat et de l’Environnement (LSCE), Institut Pierre Simon Laplace (IPSL), CNRS/CEA/UVSQ/Univ. Paris-Saclay, Orme des Merisiers, Gif-Sur-Yvette, 91191, France

$^2$Sorbonne Université, CNRS, IRD, MNHN, Institut Pierre Simon Laplace (IPSL), Paris, 75005, France

Correspondence to: Anna Denvil-Sommer (anna.sommer.lab@gmail.com)

Abstract.

A new Feed-Forward Neural Network (FFNN) model is presented to reconstruct surface ocean partial pressure of carbon dioxide (pCO$_2$) over the global ocean. The model consists of two steps: (1) reconstruction of pCO$_2$ climatology and (2) reconstruction of pCO$_2$ anomalies with respect to the climatology. For the first step, a gridded climatology was used as the target, along with sea surface salinity and temperature (SSS and SST), sea surface height (SSH), chlorophyll $a$ (Chl), mixed layer depth (MLD), as well as latitude and longitude as predictors. For the second step, data from the Surface Ocean CO$_2$ Atlas (SOCAT) provided the target. The same set of predictors was used during step 2 augmented by their anomalies. During each step, the FFNN model reconstructs the non-linear relationships between pCO$_2$ and the ocean predictors. It provides monthly surface ocean pCO$_2$ distributions on a 1°x1° grid for the period 2001-2016. Global ocean pCO$_2$ was reconstructed with a satisfying accuracy compared to independent observational data from SOCAT. However, errors are larger in regions with poor data coverage (e.g. Indian Ocean, Southern Ocean, subpolar Pacific). The model captured the strong interannual variability of surface ocean pCO$_2$ with reasonable skills over the Equatorial Pacific associated with ENSO (El Niño Southern Oscillation). Our model was compared to three pCO$_2$ mapping methods that participated in the Surface Ocean pCO$_2$ Mapping intercomparison (SOCOM) initiative. We found a good agreement in seasonal and interannual variability between the models over the global ocean. However, important differences still exist at the regional scale, especially in the Southern Hemisphere and in particular, the Southern Pacific and the Indian Ocean, as these regions suffer from poor data-coverage. Large regional uncertainties in reconstructed surface ocean pCO$_2$ and sea-air CO$_2$ fluxes have a strong influence on global estimates of CO$_2$ fluxes and trends.
1. Introduction.

The global ocean is a major sink of excess CO\textsubscript{2} emitted to the atmosphere since the beginning of the industrial revolution. In 2011, the best estimate of the ocean inventory of anthropogenic carbon (C\textsubscript{am}) amounts to 155 ± 30 PgC or 28% of cumulated total CO\textsubscript{2} emissions attributed to human activities since 1750 (Ciais et al., 2013). Between 2000 and 2009, the yearly average ocean C\textsubscript{am} uptake was 2.3 ± 0.7 PgC yr\textsuperscript{-1} (Ciais et al., 2013). However, these global estimates hide substantial regional and inter-annual fluctuations (Rödenbeck et al., 2015), which need to be quantified in order to track the evolution of the Earth’s carbon budget (e.g. Le Quéré et al., 2018).

Until recently, most estimates of inter-annual sea-air CO\textsubscript{2} flux variability were based on atmospheric inversions (Peylin et al., 2005, 2013; Rödenbeck et al., 2005) or global ocean circulation models (Orr et al., 2001; Aumont and Bopp, 2006; Le Quéré et al., 2010). However, models tend to underestimate the variability of sea-air CO\textsubscript{2} fluxes (Le Quéré et al., 2003), while atmospheric inversions suffer from a still sparse network of atmospheric CO\textsubscript{2} measurements (Peylin et al., 2013). These approaches are increasingly complemented by data-based techniques relying on in situ measurements of CO\textsubscript{2} fugacity or partial pressure (e.g. Bitting et al., 2018; Landschützer et al., 2016; Rödenbeck et al., 2014, 2015; Takahashi et al., 2002, 2009, Landschützer et al., 2013; Schuster et al., 2013; Nakaoka et al., 2013; Fay et al., 2014). These techniques rely on a variety of data-interpolation approaches developed to provide estimates in time and space of surface ocean pCO\textsubscript{2} (Rödenbeck et al., 2015) such as statistical interpolation, linear and non-linear regressions, or model-based regressions or tuning (Rödenbeck et al., 2014, 2015). These methods, their advantages and disadvantages are compared and discussed in Rödenbeck et al. (2015). This intercomparison did not allow identifying a single optimal technique but rather pleaded in favour of exploiting the ensemble of methods.

Artificial neural networks (ANN) have been widely used to reconstruct surface ocean pCO\textsubscript{2} (open ocean: Lefèvre et al., 2005; Friedrich and Oschlies, 2009b; Telszewski et al., 2009; Landschützer et al., 2013; Nakaoka et al., 2013; Zeng et al. 2014; Bitting et al., 2018; coastal region: Laruelle et al., 2017). ANN fill the spatial and temporal gaps based on calibrated non-linear statistical relationships between pCO\textsubscript{2} and its oceanic and atmospheric drivers. The existing products usually present monthly fields with a 1ºx1º spatial resolution and capture a large part of temporal-spatial variability. Methods based on ANN are able to represent the relationships between pCO2 and a variety of predictor combinations, but
they are sensitive to the number of data used in the training algorithm and can generate artificial variability in regions with sparse data coverage (Bishop, 2006).

This study proposes an alternative implementation of a neural network applied to the reconstruction of surface ocean pCO$_2$ over the period 2001-2016. It belongs to the category of Feed Forward Neural Networks (FFNN) and consists of a two-step approach: (1) the reconstruction of monthly climatologies of global surface ocean pCO$_2$ based on data from Takahashi et al. (2009), and (2) the reconstruction of monthly anomalies (with respect to the monthly climatologies) on a 1°x1° grid exploiting the Surface Ocean CO$_2$ Atlas (SOCAT) (Bakker et al., 2016). The model is easily applied to the global ocean without any boundaries between the ocean basins or regions. However, as mentioned before, it is still sensitive to the observational coverage. This limitation is partly overcome by the two-step approach as the reconstruction of monthly climatologies draws on a global ocean gridded climatology (Takahashi et al., 2009), thereby keeping FFNN output close to realistic values. Furthermore, the reconstruction of monthly climatologies during the first step allows taking into account a potential change in seasonal cycle in response to climate change when applied to time slices or to model output providing the drivers, but no carbon cycle variables.

The remainder of this paper is structured as follows: section 2 introduces datasets used during this study and describes the neural network; section 3 presents results for its validation and qualification, as well as a comparison to three mapping methods part of the Surface Ocean pCO$_2$ Mapping intercomparison (SOCOM) exercise (Rödenbeck et al., 2015). Results and perspectives are summarized in the last section.

2. Data and method.

2.1. Data.

The standard set of variables known to represent physical, chemical and biological drivers of surface ocean pCO$_2$ – mean state and variability – (Takahashi et al., 2009; Landschützer et al., 2013) were used as input variables (or predictors) for training the FFNN algorithm. These are sea surface salinity (SSS), sea surface temperature (SST), mixed layer depth (MLD), chlorophyll $a$ concentration (CHL), atmospheric CO$_2$ mole fraction ($x_{CO_2,atm}$). Based on Rodgers et al. (2009) who reported a strong correlation between natural variations in dissolved inorganic carbon (DIC) and sea surface height (SSH), SSH was added as a new driver to this list. First tests suggested that the inclusion of SSH does not significantly improve the accuracy of reconstructed pCO$_2$ at global scale. At basin and regional
scale, however, adding SSH improves the spatial pattern of reconstructed pCO$_2$ and the accuracy of our method.

For the first step, the reconstruction of monthly climatologies, the Takahashi et al. (2009) monthly pCO$_2$ gridded climatology (1°x1°) was used as the target. The original climatology was constructed by an advection-based interpolation method on a 4°x5° grid. It was interpolated on the 1°x1° SOCAT grid which is also the resolution of the final output for the FFNN.

For the second step, the target is provided by the observational database SOCAT v5 (Bakker et al., 2016). We used a gridded version of this dataset that was derived by combining all SOCAT data collected within a 1°x1° box during a specific month. SOCAT v5 represents global observations of sea surface fugacity of CO$_2$ (fCO$_2$) over the period 1970 to 2016. It includes data from moorings, ships and drifters. These data are distributed irregularly over the global ocean with 188274 gridded measurements over the Northern Hemisphere and 76065 over the Southern Hemisphere. In order to ensure a satisfying spatial and temporal data coverage, we limited the reconstruction to the period 2001-2016, which represents ~77% of the database (Fig. 1(a)).

The following formula is used to convert fCO$_2$ to pCO$_2$ (Körtzinger et al., 1999):

$$fCO_2 = pCO_2 \exp \left( \frac{p}{RT} \left( B + 2 \delta \right) \right), \quad (1)$$

where fCO$_2$ and pCO$_2$ are in μatm, p is the total pressure (Pa), R=8.314 JK$^{-1}$ is the gas constant, T is the absolute temperature (K). Parameter B (m$^3$mol$^{-1}$) is estimated as: $B = (-1636.75 + 12.0408 T – 3.27957 \times 10^{-2} T^2 + 3.16528 \times 10^{-5} T^3) \times 10^{-6}$. The parameter δ is the cross virial coefficient (m$^3$mol$^{-1}$): $\delta = (57.7 – 0.118T) \times 10^{-6}$. The total pressure is from the Jena database (6h, 5°x5°) based on the NCEP reanalysis (Källnay et al., 1996) (http://www.bgc-jena.mpg.de/CarboScope/?ID=s).

Monthly global reprocessed products of physical variables from ARMOR3D L4 distributed through the Copernicus Marine Environment Monitoring Service (CMEMS) (0.25°x0.25°) (http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MULTIOBS_GLO_PHY_REP_015_002) were used for SSS, SST and SSH (Guinehut et al., 2012). The GlobColour project provided monthly CHL distributions at 1°x1° resolution (http://www.globcolour.info/products_description.html). For MLD, daily data from the “Estimating the Circulation and Climate of the Ocean” (ECCO2) project Phase II (Cube 92), at 0.25°x0.25° resolution (Menemenlis et al., 2008) were used. For xCO$_2$ atmospheric, the 6h data from Jena CO$_2$ inversion s76_v4.1 on a 5°x5° grid were selected (http://www.bgc-jena.mpg.de/CarboScope/?ID=s). Finally, an ice mask based on daily “Operational Sea Surface
Temperature and Sea Ice Analysis” (OSTIA) with a gridded 0.05ºx0.05º resolution (Donlon et al., 2011) was applied.

MLD and CHL were log-transformed before their use in the FFNN algorithm because of their skewed distribution. In regions with no CHL data (high latitudes in winter) log(CHL) = 0 was applied. It does not introduce discontinuities since log(CHL) is close to zero in the adjacent region.

All data were averaged or interpolated on a 1ºx1º grid and, depending on the resolution of the dataset, averaged over the month. It is worth noting that all datasets have to be normalized (i.e. centered to zero-mean and reduced to unit standard deviation) before their use in the FFNN algorithm, for example:

\[ SSS_n = \frac{SSS - \bar{SSS}}{std\,|SSS|} \]

Normalization ensures that all predictors fall within a comparable range and therefore avoids giving more weight to predictors with large variability ranges (Kallache et al., 2011).

As surface ocean pCO$_2$ also varies spatially, geographical positions (lat, lon) after conversion to radians were included as predictors. In order to normalize (lat, lon) the following transformation is proposed:

\[ \text{lat}_n = \sin|\text{lat} \ast \pi/180^\circ| \]
\[ \text{lon}_{n,1} = \sin|\text{lon} \ast \pi/180^\circ| \]
\[ \text{lon}_{n,2} = \cos|\text{lon} \ast \pi/180^\circ| \]

Two functions $\sin$ and $\cos$ for longitudes are used to preserve its periodical 0 to 360 degrees behavior and thus to consider the difference of positions before and after the 0º longitude. For step 2, data required for training were co-located at the SOCAT data positions that are used as a target for the FFNN model. Details are provided in the next section.

2.2. Method.

a) Network configuration and evaluation protocol

In this work, we use Keras, a high-level neural network Python library (“Keras: The Python Deep Learning library”, Chollet, 2015; https://keras.io) to build and train the FFNN models. The identification of an optimal configuration is the first step in the FFNN model building. This includes: the choice of number and size of hidden layers (i.e., intermediate layers between input and output
layers), connection type, activation functions, loss function and optimization algorithm, as well as the learning rate and other low-level parameters. Based on a series of tests and their statistical results (RMSE, correlation, bias) a hyperbolic tangent was chosen as an activation function for neurons in hidden layers, and a linear function for the output layer. As optimization algorithm, the mini-batch gradient descent or RMSprop was used (adaptive learning rates for each weight, Chollet, 2015; Hinton et al., 2012). The number of layers and neurons depends on the problem. For totally connected layers (i.e., a neuron in a hidden layer is connected to all neurons in the precedent layer and connects all neurons in the next one), that is the case here, it is enough to have only one single hidden layer but two or more can help the approximation of complex functions (or complex relationships between the input and the output of the problem).

The number of the FFNN layers and the number of neurons depends on one side on the complexity of the problem: the more layers and neurons, the better the accuracy of the output. However, the size also depends on the number of patterns (data) used for training. The empirical rule advises to have a factor of 10 between the number of patterns (data) and the number of connections, or weights to adjust (in line with Amari et al. (1997), we use a factor of 10 that necessitates a cross-validation to avoid overfitting). This limits the size, the number of parameters and incidentally the number of neurons, of the FFNN. This empirical rule was followed in this study.

(1) Step 1: reconstruction of monthly climatologies

FFNN reconstructs a normalized monthly surface ocean pCO$_2$ climatology as a nonlinear function of normalized SSS, SST, SSH, Chl, MLD climatologies and geographical position (longitude, latitude):

$$pCO_{2,n} = \{SSS_n, SST_n, SSH_n, Chl_n, MLD_n, lon_n, lat_n\}$$

Surface ocean pCO$_2$ from Takahashi et al. (2009) provided the target. The dataset was divided into 50% for FFNN training and 25% for its evaluation. This 25% did not participate in the training. This set is used to monitor the performance of the training process and to drive its convergence. The remaining 25% (each 4$^{th}$ point) of the dataset were used after training for the FFNN model validation. More details about the FFNN training process can be found in Rumelhart et al. (1986) and Bishop (1995).

Validation and evaluation datasets were chosen quasi-regularly in space and time to take into account all regions and seasonal variability. In order to improve the accuracy of the reconstruction, the model was applied separately for each month. We have developed a FFNN model with 5 layers (3 hidden layers). 12 models with a common architecture were trained. Tests with one model for 12 months
showed a slight decrease in accuracy (not presented here). About 17,500 data were available for each month to train the model, resulting in monthly FFNN models with about 1,856 parameters.

(2) Step 2: reconstruction of anomalies

During the second step, normalized \( pCO_2 \) anomalies were reconstructed as a nonlinear function of normalized SSS, SST, SSH, Chl, MLD, \( xCO_2 \) and their anomalies, as well as geographic position:

\[
pCO_{2,\text{anom},n} = \{SSS_n, SST_n, SSH_n, Chl_n, MLD_n, xCO_{2,n}, SSS_{anom,n}, SST_{anom,n}, SSH_{anom,n}, Chl_{anom,n}, MLD_{anom,n}, xCO_{2,anom,n}, lon_{n,1}, lon_{n,2}, lat_n\} \tag{3}
\]

Surface ocean \( pCO_2 \) anomalies computed as the differences between collocated \( pCO_2 \) values based on SOCAT observations and monthly \( pCO_2 \) climatologies reconstructed during the first step provided the targets:

\[
pCO_{2,\text{anom}} = pCO_{2,\text{SOCAT}} - pCO_{2,\text{clim,FFNN}} \tag{4}
\]

The set of target data was again divided into 50% for the training algorithm, 25% for evaluation and 25% for model validation. As in step (1) the model was trained separately for each climatological month. There were thus 12 models sharing a common architecture but trained on different data. At this step, in order to increase the amount of data during training and to introduce information on the seasonal cycle, the model was trained using as a target \( pCO_2 \) data from the month in question as well as those from the previous and following month during the entire period 2001-2016. Figures 1 (b) and 1 (c) show an example of data distribution for the sole months of January over the period 2001-2016 (Fig. 1 (b)) and for the three months time-window December-January-February 2001-2016 used in the training algorithm of the January FFNN model (Fig. 1 (c)). In this particular example, the choice of three months provided a better cover of the region and doubled the number of data at high latitudes.

K-fold cross-validation was used for the evaluation and the validation of the FFNN architecture. Cross-validation relied on K=4 different subsampling of the dataset to draw 25% of independent data for validation (Fig. S1). Each sampling fold was tested on 5 runs of the FFNN for each month. Each of these 5 runs is characterized by different initial values that are chosen randomly. From these 5 results, the best was chosen based on root-mean-square-error (RMSE), \( r^2 \) and bias.

The final model architecture at step 2 had 3 layers (1 hidden layer). About 10,000 samples were available for training for each month, thus, a model with 541 parameters was developed. Note that a higher number of parameters did not show a significant improvement of accuracy.
b) Reconstruction of surface ocean $pCO_2$

The previous section presented the development of the “optimal” architecture of a FFNN model for the reconstruction of global surface ocean $pCO_2$, and the estimation of its accuracy. This FFNN model was used to provide the final product for scientific analysis and comparison with other mapping approaches. In order to provide the final output, the selected FFNN architecture is trained on all available data: 100% of data for training, 100% for evaluation and 100% for validation. The network was executed 5 times (different initial values) and the best model was selected based on validation results considering root-mean-square-error (RMSE), $r^2$ and bias computed between network output and SOCAT derived surface ocean $pCO_2$ data. The final model output is referred to as the LSCE-FFNN product.

2.3. Computation of sea-air $CO_2$ fluxes.

Sea-air $CO_2$ flux $f$ was calculated following Rödenbeck et al. (2015) as:

$$f = k \rho L \left| pCO_2 - pCO_2^{atm} \right| \quad (5)$$

where $k$ is the piston velocity estimated according to Wanninkhof (1992):

$$k = \Gamma u^2 \left( \frac{Sc}{Sc^{ref}} \right)^{-0.5} \quad (6)$$

The global scaling factor $\Gamma$ was chosen as in Rödenbeck et al. (2014) with the global mean $CO_2$ piston velocity equaling to 16.5 cm/h. $Sc$ corresponds to the Schmidt number estimated according to Wanninkhof (1992). The wind speed was computed from 6-hourly NCEP wind speed (Kalnay et al., 1996). $\rho$ is seawater density in (5) and $L$ is the temperature-dependent solubility (Weiss, 1974). $pCO_2$ corresponds to the surface ocean $pCO_2$ output of the mapping method. $pCO_2^{atm}$ was derived from the atmospheric $CO_2$ mixing ratio fields provided by the Jena inversion s76_v4.1 (http://www.bgc-jena.mpg.de/CarboScope/).

3. Results.

3.1. Validation.

The subset of data used for network validation, that is 25% of the total, represents independent observations as they did not participate in training during model development (see 2.2a). The skill of the FFNN to reconstruct monthly climatologies of surface ocean $pCO_2$, was assessed by comparing collocated reconstructed $pCO_2$ and corresponding values from Takahashi et al. (2009). The global
climatology was reconstructed with a satisfying accuracy during step 1 with a RMSE of 0.17 μatm and $r^2$ of 0.93. Model output of step 2 was assessed by K-fold cross-validation as presented before: $K=4$ different subsets of independent data were drawn from the dataset and the network was run 5 times on each subset. From these 20 results the best one was chosen based on RMSE, $r^2$ and mean absolute error (MAE) (the bias is presented in Table S1). The combination of the four best model output was used for the statistical analysis summarized in Table 1. Metrics were computed over the full period (2001-2016) and with reference to SOCAT observations (independent data only). At the global scale, the analysis yielded a RMSE of $\sim 17.97$ μatm, while the MAE was 11.52 μatm and $r^2$ was 0.76. These results are comparable to those obtained by Landschützer et al. (2013) for the assessment of a surface ocean pCO$_2$ reconstruction based on an alternative neural network-based approach. The RMSE between SOCAT data and the climatology of pCO$_2$ from Takahashi et al. (2009) equals 41.87 μatm, larger than errors computed for the regional comparison between FFNN and SOCAT (Table 1). We also estimated the RMSE for the case of 100% data used for training. It equals 14.8 μatm and confirms the absence of overfitting.

Figure 2 (a) shows the time mean difference between the estimated pCO$_2$ and pCO$_2$ from SOCAT v5 data used for validation $\text{mean}_{i,j} |pCO_{2,i,j,\text{FFNN}} - pCO_{2,i,j,\text{SOCAT}}|$. Large differences occurred at high latitudes, in equatorial regions, along the Gulf Stream and Kuroshio currents – the regions with strong horizontal gradients of pCO$_2$. Moreover, the standard deviation of residuals (Figure 2 (b)) in these regions was larger indicating that the model fails to accurately reproduce the temporal variability. The reduced skill of the model in these regions reflects the poor data coverage along with a strong seasonal variability (e.g. Southern Ocean) and/or high kinetic energy (e.g. Southern Ocean, Kuroshio and Gulf Stream currents) (Fig. 1 (a)). At the scale of ocean regions, (Table 1) the largest RMSE and MAE were computed for the Pacific Subpolar ocean (RMSE = 34.77 μatm, MAE = 23.12 μatm), while the lowest correlation coefficient was obtained for the equatorial Atlantic Ocean ($r^2 = 0.57$). These low scores directly reflect low data density and are to be contrasted with those obtained over regions with better data coverage (e.g. Subtropical North Pacific: RMSE = 15.86 μatm, MAE = 9.9 μatm, $r^2 = 0.77$ or Subpolar Atlantic: RMSE = 22.99 μatm, MAE = 15.04 μatm, $r^2 = 0.76$). Despite large time mean differences computed over the eastern Equatorial Pacific, scores are satisfying at the regional scale indicating error compensation by improved scores over the western basin (RMSE = 15.73 μatm, MAE = 10.33 μatm, $r^2 = 0.79$). Scores are low in the Southern Hemisphere (Table 1) and time mean differences are large (Fig. 2 (a)) reflecting sparse data coverage (Fig. 1 (a)).
3.2. Qualification.

This section presents the assessment of the final time series of reconstructed surface ocean pCO$_2$. The time series was computed using the best monthly models as described in section 2.2, as well as 100% of data for learning, evaluation and validation.

Results of the LSCE-FFNN mapping model were compared to three published mapping methods which participated in the “Surface Ocean pCO2 Mapping Intercomparison” (SOCOM) exercise presented in Rödenbeck et al. (2015) (http://www.bgc-jena.mpg.de/SOCOM/). These methods are: (1) Jena-MLS oc_v1.5 (Rödenbeck et al., 2014), a statistical interpolation scheme (data-driven mixed-layer scheme; principal drivers used in parametrisation: ocean-internal carbon sources/sinks, SST, wind speed, mixed-layer depth climatology, alkalinity climatology); (2) JMA-MLR (updated version up to 2016) (Iida et al., 2015), based on multi-linear regressions with SST, SSS and Chl a as independent variables, and (3) ETH-SOMFFN v2016 (Landschützer et al., 2014), a two-step neural network model with SST, SSS, MLD, Chl a, xCO$_2$ as drivers. The time series of pCO$_2$ and sea-air CO$_2$ flux ($f$) were assessed over 17 biomes defined by Fay and McKinley (2014) (Fig. 3, Table 2). These biomes were derived based on coherence in SST, Chl a, ice fraction, maximum MLD and represent regions of coherent biogeochemical dynamics.

We followed the protocol and diagnostics proposed in Rödenbeck et al. (2015) for the comparison of the mapping methods between each other, respectively to observations. The following diagnostics were computed: (1) the relative interannual variability (IAV) mismatch R$_{iav}$ (in %) and (2) the amplitude of interannual variations. The relative interannual variability (IAV) mismatch R$_{iav}$ (in %) is the ratio of the mismatch amplitude M$_{iav}$ of the difference between the model output and observations (its temporal standard deviation) and the mismatch amplitude M$_{iav,benchmark}$ of the “benchmark”. The latter was derived from the mean seasonal cycle of the corresponding model output where the trend of increasing yearly atmospheric pCO$_2$ was added (see details in Rödenbeck et al., 2015). It corresponds to a climatology corrected for increasing atmospheric CO$_2$, but without interannual variability.

$$R_{iav} = \frac{M_{iav}}{M_{iav,benchmark}} \times 100\%,$$

where

$$M_{iav} = \text{std} \left( \text{mean} \left( pCO_{2,\text{Model}} - pCO_{2,\text{SOCAT}} \right) \right),$$

$$M_{iav,benchmark} = \text{std} \left( \text{mean} \left( D_{season} \right) \right).$$
where “mean” is a mean over the region and year and
\[
D_{\text{season}} = \left| pCO_{2,SS} + \text{trend} \left( CO_{2,\text{atm}} \right) \right| - pCO_{2,\text{SOCAT}} ,
\]
pCO_{2,SS} is the seasonal cycle of pCO_2 from the corresponding mapping method. CO_{2,\text{atm}} estimates from xCO_2 Jena CO_2 inversion s76_v4.1 were used.

R^{\text{ia}} provides information on the capability of each method to reproduce the IAV compared to observations: a smaller R^{\text{ia}} stands for better fit compared to the reference. The amplitude of the interannual variations (A^{\text{ia}}) of sea-air flux of CO_2 (its 2-month running mean) is estimated as the temporal standard deviation over the period.

3.2.1. Interannual variability.

The time series of globally averaged surface ocean pCO_2 over the period 2001-2016 are presented in Figure 4 for LSCE-FFNN and the three other models. Surface ocean pCO_2 (μatm) varied between the 4 mapping methods in the range of ±7 μatm (Fig. 4 (a)). Modeled pCO_2 values were at the lower end for ETH-SOMFFN and JMA-MLR, while LSCE-FFNN and Jena-MLS13 computed higher values. The same behavior was found for 12-month running mean time series (Fig. 4 (b)). Figure 4 (c) shows the 12-month running mean of the difference between computed pCO_2 and SOCAT data (model – SOCAT) over the globe. JMA-MLR mostly underestimated observed pCO_2 with a strong interannual variability of the misfit, especially at the end of the period with up to -5 μatm. The difference between ETH-SOMFFN output and SOCAT data fluctuated in the range of ±1 μatm, with an increase in amplitude up to -2 μatm from 2010 onward. Jena-MLS13 overestimated observations with the difference in the range of 0-1 μatm. The difference between LSCE-FFNN and SOCAT varies around zero between -0.7 and 1 μatm.

The model was assessed next at biome scale. Results for all biomes are presented in the supplementary material (Fig. S2, S3, S4). Two biomes with contrasting dynamics are discussed hereafter in greater detail: (1) the Equatorial East Pacific (biome 6) characterized by a strong IAV of surface ocean pCO_2 and sea-air CO_2 fluxes in response to ENSO, the El Niño Southern Oscillation (Feely et al., 1999; Rödenbeck et al., 2015), and (2) the North Atlantic Permanently Stratified biome (biome 11) with a well-marked seasonal cycle, but little IAV (Schuster et al., 2013). Results for these biomes are presented in Figure 5.
Biome 6 is relatively well-covered by observations and represents a key region for testing the skill of the model to reproduce the observed strong IAV linked to ENSO. El Niño events are characterized by positive SST anomalies, reduced upwelling and decreased surface ocean pCO$_2$ values. These episodes could be identified in all model time series (Fig. 5 (a)) with reduced pCO$_2$ levels in 2004/2005 and 2006/2007 (weak El Niño), 2002/2003 and 2009/2010 (moderate El Niño), and 2015/2016 (strong El Niño). JMA-MLR (blue curve) tended to underestimate pCO$_2$ during weak El Niño events. It was underestimated during the La Niña 2011-2012 event by Jena-MLS13. LSCE-FFNN and ETH-SOMFFN, both based on a neural network approach yielded similar results despite differences in network architecture and predictor datasets.

Data coverage is particularly high over Biome 11 (Fig. 5 (b), (d), (f)). The seasonal cycle in this biome is dominantly driven by temperature. Modeled seasonal variability showed a good agreement across the ensemble of methods (Fig. 5(b)) with an increase in spring-summer and a decrease in autumn-winter. However, the amplitude can be different by up to 10 μatm between different models. The seasonal amplitude of pCO$_2$ computed by JMA-MLR increased from smaller values at the beginning of the time series to higher ones in the middle of the period 2005-2012. The variability of seasonal amplitude was the highest for Jena-MLS13 in line with the 12-month running mean time series (Fig. 5 (d)). Again, similar seasonal amplitude and year-to-year variability of surface ocean pCO$_2$ were obtained with LSCE-FFNN and ETH-SOMFFN (Fig. 5 (b), (d)). The yearly pCO$_2$ mismatch (Fig. 5 (f)) shows that observed surface ocean pCO$_2$ was underestimated by JMA-MLR at the beginning and at the end of the period by up to -6 μatm, and overestimated during 2007-2011 by up to 8 μatm. Jena-MLS13 shows mostly positive differences in the range 0-2 μatm over the full period. LSCE-FFNN and ETH-SOMFFN vary around zero and between -2 – 2 μatm, being close to each other.

3.2.2. Sea-air CO$_2$ flux variability.

Sea-air exchange of CO$_2$ was estimated using the same gas exchange formulation (4) and wind data speed (6-hourly NCEP wind speed) for each mapping data (Rödenbeck et al., 2005). It is worth noting that the sea-air flux is sensitive to the choice of the wind speed dataset (Roobaert et al., 2018).

Figure 6 (a) presents the global 12-month running mean of the sea-air CO$_2$ flux for four mapping methods. All models showed an increase in CO$_2$ uptake in response to increasing atmospheric CO$_2$ levels, albeit with a strong between-model variability in multi-annual trends. There is less agreement
between the methods compared to reconstructions of surface ocean pCO$_2$ variability (Fig. 4 (b)). This
results from the contribution of uncertainties in sea-air CO$_2$ flux estimations over regions with poor
data-coverage (mostly in the South Hemisphere: South Pacific, South Atlantic, Indian Ocean, South
Ocean; see Fig. S5). Nevertheless, the relative IAV mismatch was less than 30% for all methods (Fig. 6
(b)), suggesting a reasonable fit to observational data. The relative IAV mismatch is, however, a global
score and it is biased towards regions with good data coverage (Rödenbeck et al., 2015). The time
series reconstructed in this study is too short to capture decadal variations and in particular the
strengthening of the sink from 2000 onward (Landschützer et al., 2016). LSCE-FFNN computed a
slowdown of ocean CO$_2$ uptake between 2010 and 2013 with a flux of $\sim$1.8 GtC yr$^{-1}$ compared to $\sim$2.2
GtC yr$^{-1}$ for ETH-SOMFFN. A leveling-off was also found for JMA-MLR, albeit shifted in time. In
general, the amplitudes of reconstructed CO$_2$ fluxes across all four methods agreed within 0.2-0.36
PgC/yr. The weighted mean of IAV (horizontal line in Fig. 6 (b)) computed from the four methods
included here was 0.25 PgC/yr. This value is close to the one of Rödenbeck et al. (2015) for the
complete ensemble of SOCOM models (0.31 PgC/yr) estimated for the period 1992-2009. The largest
amplitude was obtained for ETH-SOMFFN, $\sim$0.35 PgC/yr. On the other hand, LSCE-FFNN has the
smallest amplitude with 0.21 PgC/yr. Jena-MLS13 and JMA-MLR lie very close to the weighted mean
value with 0.26 PgC/yr and 0.22 PgC/yr, respectively. The weighted mean and the dispersion of
individual models around it, reflect the period of analysis (2001-2015, ETH-SOMFFN output provided
up to 2015) and the total number of models contributing to it (see for comparison Rödenbeck et al.,
2015). As such it does not provide information on the skill of any particular model.

The interannual variability of reconstructed sea-air CO$_2$ fluxes (12-month running mean) showed a
good agreement for biome 6 (East Pacific Equatorial, Fig. 7 (a)). A small discrepancy was found at the
beginning of the period. A strong increase was computed by Jena-MLS13 for 2010-2014 that was also
identified on pCO$_2$ variability (Fig. 5 (a)). Despite this, Jena-MLS13 has a low relative R$^{IAV}$ (26%),
which confirms a tendency mentioned in Rödenbeck et al. (2015) that mapping products with a small
relative IAV mismatch show larger amplitude. LSCE-FFNN and ETH-SOMFFN yielded comparable
results (Fig. 7 (a), (c)) with relative IAV mismatches of 46% and 53%, respectively, and with
amplitudes $\sim$ 0.03 PgC/yr. Interannual variability reproduced by JMA-MLR falls within the range of
the other models (Fig. 7 (c)), but with a R$^{IAV}$ of $\sim$68%.

Reconstructed sea-air CO$_2$ fluxes over the North Atlantic Subtropical Permanently Stratified region
(biome 11) show large between model differences in amplitudes and variability. The two models based
on a neural network show again a good agreement with $R^{\text{AV}}$ of 17% for LSCE-FFNN and 20% for ETH-SOMFFN. Jena-MLS13 produced a strong seasonal variability (Fig. 7 (b)) up to 0.06 PgC/yr, and small $R^{\text{AV}}$ of ~11%. **Contrary to the other approaches**, JMA-MLR did not reproduce a decrease in sea-air CO$_2$ at the middle of the period by up to 0.02 PgC/yr (Fig. 7 (b)). The model is characterized by a $R^{\text{AV}}$ of 46% and an amplitude of 0.013 PgC/yr.

### 3.3.3. Sea-air CO$_2$ flux trend.

The long-term trend of sea-air CO$_2$ fluxes is dominantly driven by the increase in atmospheric CO$_2$ (see Fig. S7). On shorter time scales, such as for the period 2001-2016, the interannual variability at regional scales reflects natural modes of climate variability and local oceanographic dynamics (Heinze et al., 2015).

Figure 8 shows the significant linear trends ($p_{\text{val}} = 0.05$) of sea-air CO$_2$ fluxes for LSCE-FFNN (a), Jena-MLS13 (b), ETH-SOMFFN (c) and JMA-MLR (d). A total (averaged over the globe) negative trend was computed for all models, albeit with large regional contrasts, and LSCE-FFNN falls within the range: Jena-MLS13, -0.0012 PgC/yr/yr (-0.0028 PgC/yr/yr, total value without significant t-test, Fig. S8); LSCE-FFNN, -0.00087 PgC/yr/yr (-0.0032 PgC/yr/yr); JMA-MLR, -0.0013 PgC/yr/yr (-0.0037 PgC/yr/yr); ETH-SOMFFN, -0.0025 PgC/yr/yr (-0.0059 PgC/yr/yr). LSCE-FFNN computed negative trends over most of the Atlantic basin, Indian Ocean and South of 40°S, which contrasts with decreasing fluxes over the Pacific and locally in the Antarctic Circumpolar Current. At first order, this broad regional pattern is found in all models. Regional maxima and minima are, however, more pronounced in Jena-MLS13 (Fig. 8 (b)) and ETH-SOMFFN (Fig. 8 (c)), while a patchy distribution at sub-basin scale is diagnosed for JMA-MLR.

The agreement in sign of computed linear trends from four models is presented in Fig. 9 (total linear trends without significant t-test). Over most of the ocean, all four models show very close sea-air CO$_2$ tendency. In the Indian Ocean (biome 14), on the other hand a positive trend was computed for JMA-MLR (0.0004 PgC/yr/yr, and with t-test: 0.00006 PgC/yr/yr) while the three other models present a negative trend. The differences between models were also found in the Pacific Ocean, especially the Southern Pacific. In the Eastern Equatorial Pacific region (biome 6) a total significant negative trend is presented by all models. All models reproduced a maximum in the southern part of biome 6 but they disagree about its amplitude and spatial distribution. Almost everywhere over the Atlantic Ocean the
mapping methods produced the same sign of linear trend (Fig. 9). Only in the eastern part of the subtropical North Atlantic Jena-MLS13 gave a positive linear trend of fCO$_2$ (Fig. 8 (b)).

According to LSCE-FFNN, the global ocean took up in average 1.55 PgC/yr between 2001-2015. This estimate is consistent with results from the other three models (Table 3) (see Table S2 for estimations per biomes). The spread between individual models falls in the range of the error reported in Landschützer et al. (2016), ±0.4-0.6 PgC/yr. Per biome, estimates of CO$_2$ sea-air fluxes provided by LSCE-FFNN are similarly in good agreement with those derived from the other models.

4. Summary and conclusion.

We proposed a new model for the reconstruction of monthly surface ocean pCO$_2$. The model is applied globally and allows a seamless reconstruction without introducing boundaries between the ocean basins or biomes. Our model relies on a two-step approach based on Feed-Forward Neural Networks (LSCE-FFNN). The first step corresponds to the reconstruction of a monthly pCO$_2$ climatology. It allows to keep the output of the FFNN close to the observed values in regions with poor data cover. At the second step, pCO$_2$ anomalies are reconstructed with respect to the climatology from the first step. The model was applied over the period 2001-2016. Validation with independent data at global scale indicated a RMSE of 17.57 μatm, $r^2$ of ~0.76 and an absolute bias of 11.52 μatm. In order to assess the model further, it was compared to three different mapping models: ETH-SOMFFN (self-organizing maps + neural network), Jena-MLS13 (statistical interpolation), JMA-MLR (linear regression) (Rödenbeck et al., 2015). Network qualification followed the protocol and diagnostics proposed in Rödenbeck et al. (2015).

Reconstructed surface ocean pCO$_2$ distributions were in good agreement with other models and observations. The seasonal variability was reproduced satisfyingly by LSCE-FFNN, the yearly pCO$_2$ mismatch varied around zero, and relative IAV mismatch was 7%. LSCE-FFNN proved skillful in reproducing the interannual variability of surface ocean pCO$_2$ over the Eastern Equatorial Pacific in response to ENSO. Reductions in surface ocean pCO$_2$ during El Niño events were well reproduced. The comparison between reconstructed and observed pCO$_2$ values yielded a RMSE of 15.73 μatm, $r^2$ of 0.79 and an absolute bias of 10.33 μatm over the Equatorial Pacific. The relative IAV misfit in this region was ~17%. Despite an overall good agreement between models, important differences still exist at the regional scale, especially in the Southern Hemisphere and in particular, the Southern Pacific and the Indian Ocean. These regions suffer from poor data-coverage. Large regional uncertainties in
reconstructed surface ocean pCO\textsubscript{2} and sea-air CO\textsubscript{2} fluxes have a strong influence on global estimates of CO\textsubscript{2} fluxes and trends.

**Code and data availability.**

Python code for pCO\textsubscript{2} climatology reconstruction, 1\textsuperscript{st} step of LSCE-FFNN model, python code for reconstruction of pCO\textsubscript{2} anomalies, 2\textsuperscript{nd} step of LSCE-FFNN model, are provided at the end of supplementary material.

Time series of reconstructed surface ocean pCO\textsubscript{2} and CO\textsubscript{2} fluxes are distributed through the Copernicus Marine Environment Monitoring Service (CMEMS), http://marine.copernicus.eu/services-portfolio/access-to-products/, search keyword: MULTIOLS.

**Author contribution.**

ADS, MG, MV and CM contributed to the development of the methodology and designed the experiments, ADS carried them out. ADS developed the model code and performed the simulations. ADS prepared the manuscript with contributions from all co-authors.

**Acknowledgments.**

The authors would like to thank the two referees, Christian Rödenbeck and Luke Gregor, for their helpful comments and questions, as well as Frederic Chevallier and Gilles Reverdin for their suggestions. This study was funded by the AtlantOS project (EU Horizon 2020 research and innovation program, grant agreement no. 2014-633211). MV also acknowledges funding by the CoCliServ and EUPHEME projects (ERA4CS program).

**References**


Williams, N.L., Juranek, L.W., Feely, R.A., Johnson, K.S., Sarmiento, J.L., Talley, L.D., Dickson, A.G.,
from biogeochemical Argo floats equipped with pH: An uncertainty analysis, Global Biogeochemical

climatology based on a feed-forward neural network, J. Atmos. Ocean Technol., 31, 1838–1849,
Figure 1: Spatial distribution of SOCAT data (number of measurements per grid point): (a) - period 2001-2016; (b) - all months of January for period 2001-2016; (c) - all months of December-January-February for period 2001-2016.
Figure 2: Time mean differences (μatm) between monthly LSCE-FFNN \( p\text{CO}_2 \) and SOCAT \( p\text{CO}_2 \) data used for evaluation of the model over the period 2001-2016 (a) and its std (b).

Figure 3: Map of biomes (after Rodenbeck et al. (2015); and Fay and McKinley (2014)) used for comparison. See table 2 for biome names.
Figure 4: Global oceanic pCO₂: black - LSCE-FFNN, blue - JMA, red - Jena, green - ETH-SOMFFN; (a) – global average monthly time series, (b) – global 12-month running mean average, (c) - yearly pCO₂ mismatch (difference of mapping methods and SOCAT data).
Figure 5: Surface ocean pCO$_2$: Equatorial East Pacific (biome 6) (left) and Subtropical Permanently Stratified North Atlantic (biome 11) (right): black – FFNN, blue – JMA, red – Jena, green – ETH-SOMFFN; (a), (b) – monthly time series averaged over biome; (c), (d) – 12-month running mean averaged over biome; (e), (f) – yearly pCO$_2$ mismatch (difference of mapping methods and SOCAT data).
Figure 6: (a) – Interannual global ocean sea-air CO₂ flux (12-month running mean); (b) – amplitude of interannual CO₂ flux plotted against the relative IAV mismatch amplitude. The weighted mean is given as a horizontal line.
Figure 7: Global ocean interannual sea-air CO$_2$ flux (12-month running mean): (a) Equatorial East Pacific (biome 6) and (b) Subtropical Permanently Stratified North Atlantic (biome 11). Amplitude of interannual CO$_2$ flux plotted against the relative IAV mismatch amplitude: (c) Equatorial East Pacific (biome 6) (left) and (d) Subtropical Permanently Stratified North Atlantic (biome 11). The weighted mean is given as a horizontal line.
Figure 8: Significant (p_val = 0.05) linear trend of fCO$_2$ for common period 2001-2015: (a) – LSCE-FFNN; (b) – Jena-MLS13; (c) – ETH-SOMFFN; (d) – JMA-MLR.
Table 1: Statistical validation of LSCE-FFNN. Comparison between reconstructed surface ocean pCO$_2$ and pCO$_2$ values from SOCAT v5 database not used in the training algorithm for the period 2001-2016 over the global ocean (except for regions with ice-cover) and for large oceanographic regions. In round brackets: number of measurements per region.

Figure 9: Agreement between four mapping methods in their linear trend of sea-air CO$_2$ flux. Color-bar represents the number of products that have the same sign of linear trend.
<table>
<thead>
<tr>
<th>Model</th>
<th>Latitude boundaries</th>
<th>RMSE (μatm)</th>
<th>r²</th>
<th>MAE (μatm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFNN Global</td>
<td></td>
<td>17.97</td>
<td>0.76</td>
<td>11.52</td>
</tr>
<tr>
<td>Arctic (150)</td>
<td>76ºN to 90ºN</td>
<td>22.05</td>
<td>0.54</td>
<td>17.1</td>
</tr>
<tr>
<td>Subpolar Atlantic (21903)</td>
<td>49ºN to 76ºN</td>
<td>22.99</td>
<td>0.76</td>
<td>15.04</td>
</tr>
<tr>
<td>Subpolar Pacific (4529)</td>
<td>49ºN to 76ºN</td>
<td>34.77</td>
<td>0.65</td>
<td>23.12</td>
</tr>
<tr>
<td>Subtropical Atlantic (41331)</td>
<td>18ºN to 49ºN</td>
<td>17.28</td>
<td>0.69</td>
<td>11.27</td>
</tr>
<tr>
<td>Subtropical Pacific (41867)</td>
<td>18ºN to 49ºN</td>
<td>15.86</td>
<td>0.77</td>
<td>9.9</td>
</tr>
<tr>
<td>Equatorial Atlantic (7300)</td>
<td>18ºS to 18ºN</td>
<td>17.27</td>
<td>0.57</td>
<td>11.44</td>
</tr>
<tr>
<td>Equatorial Pacific (27092)</td>
<td>18ºS to 18ºN</td>
<td>15.73</td>
<td>0.79</td>
<td>10.33</td>
</tr>
<tr>
<td>South Atlantic (3002)</td>
<td>44ºS to 18ºS</td>
<td>17.81</td>
<td>0.63</td>
<td>12.28</td>
</tr>
<tr>
<td>South Pacific (12934)</td>
<td>44ºS to 18ºS</td>
<td>13.52</td>
<td>0.63</td>
<td>9.36</td>
</tr>
<tr>
<td>Indian Ocean (2871)</td>
<td>44ºS to 30ºN</td>
<td>17.25</td>
<td>0.62</td>
<td>11.6</td>
</tr>
<tr>
<td>Southern Ocean (16334)</td>
<td>90ºS to 44ºS</td>
<td>17.4</td>
<td>0.58</td>
<td>11.92</td>
</tr>
</tbody>
</table>

Table 2: Biomes from Fay and McKinley (2014) used for time series comparison (Fig. 3)

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Omitted) North Pacific Ice</td>
</tr>
<tr>
<td>2</td>
<td>Subpolar Seasonally Stratified North Pacific</td>
</tr>
<tr>
<td>3</td>
<td>Subtropical Seasonally Stratified North Pacific</td>
</tr>
<tr>
<td>4</td>
<td>Subtropical Permanently Stratified North Pacific</td>
</tr>
<tr>
<td>5</td>
<td>Equatorial West Pacific</td>
</tr>
<tr>
<td>6</td>
<td>Equatorial East Pacific</td>
</tr>
<tr>
<td>7</td>
<td>Subtropical Permanently Stratified South Pacific</td>
</tr>
<tr>
<td>8</td>
<td>(Omitted) North Atlantic Ice</td>
</tr>
<tr>
<td>9</td>
<td>Subpolar Seasonally Stratified North Atlantic</td>
</tr>
<tr>
<td>10</td>
<td>Subtropical Seasonally Stratified North Atlantic</td>
</tr>
<tr>
<td>11</td>
<td>Subtropical Permanently Stratified North Atlantic</td>
</tr>
<tr>
<td>Region</td>
<td>Latitude boundaries</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Global</td>
<td></td>
</tr>
<tr>
<td>Arctic</td>
<td>76°N to 90°N</td>
</tr>
<tr>
<td>Subpolar Atlantic</td>
<td>49°N to 76°N</td>
</tr>
<tr>
<td>Subpolar Pacific</td>
<td>49°N to 76°N</td>
</tr>
<tr>
<td>Subtropical Atlantic</td>
<td>18°N to 49°N</td>
</tr>
<tr>
<td>Subtropical Pacific</td>
<td>18°N to 49°N</td>
</tr>
<tr>
<td>Equatorial Atlantic</td>
<td>18°S to 18°N</td>
</tr>
<tr>
<td>Equatorial Pacific</td>
<td>18°S to 18°N</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>44°S to 18°S</td>
</tr>
<tr>
<td>South Pacific</td>
<td>44°S to 18°S</td>
</tr>
<tr>
<td>Indian Ocean</td>
<td>44S to 30N</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>90°S to 44°S</td>
</tr>
</tbody>
</table>

Table 3: Mean of sea-air CO$_2$ flux (PgC/yr) over the Global Ocean and per regions for period in common (2001-2015). Averages over the period 2001-2009 are presented between brackets. The last column presents a comparison to best estimates from Schuster et al. (2013) for the Atlantic Ocean (1990 – 2009).
We would like to thank Astrid Kerkweg for her comments about the journal requirements.

We added in the title of the manuscript a number of model version: line 1.

The program codes of model are added at the end of in supplementary materials: lines 474-476 in the manuscript.

Time series of reconstructed surface ocean pCO$_2$ and CO$_2$ fluxes are distributed through the Copernicus Marine Environment Monitoring Service (CMEMS), http://marine.copernicus.eu/services-portfolio/access-to-products/, search keyword: MULTIOBS: lines 478-480 in the manuscript.
We thank Dr Christian Rödenbeck for his comments and suggestions. The manuscript was revised to address each point.

The authors present a method to interpolate temporally and spatially discrete surface-ocean pCO2 observations into a gridded field, using non-linear regression against various environmental drivers. Compared to similar methods, two specific features are (a) estimating seasonality separately based on data also outside the calculation period, thereby improving its data constraint, and (b) the addition of SSH to the set of predictors. This is an interesting contribution to the ensemble of existing interpolation methods, enlarging the range of plausible outcomes. The manuscript nicely describes the details of the method, and presents evaluation metrics, also in the context of other methods from the SOCOM ensemble. I clearly recommend to publish this manuscript. Below some suggestions mainly to further improve clarity at a few places.

Thank you very much for your positive evaluation. We provide hereafter a detailed point-by-point reply.

Christian Rödenbeck:
The method is very similar to CARBONES-NN (not published in the peer-reviewed literature but described in the SOCOM paper Rödenbeck et al., 2015), developed at LSCE as well. I therefore assume that the presented method builds on and supersedes CARBONES-NN. Is this correct? If so, this is an interesting piece of information to users of the SOCOM ensemble and should be mentioned in this paper. Further, I would find it fair to give credit to the authors of CARBONES-NN (to my knowledge, Abdou Kane and Philippe Peylin).

The present study (by Denvil-Sommer et al.) reflects our current vision of this particular estimation problem. It is part of a long-standing activity at LSCE with focus on the surface ocean carbon system. Abdou Kane and Philippe Peylin’s work, the development of the CARBONES-NN model, is an earlier example. The model has been largely redesigned to turn into the one presented in the current submission. Hence, the new model presented in this paper replaces the former one. Philippe Peylin is of course fully aware of our latest efforts, but his work has now other priorities.

Christian Rödenbeck: You mention the inclusion of SSH, but do not discuss it. How does it influence, and possibly improve, the results?

Authors: First tests suggested that the inclusion of SSH does not significantly improve the accuracy of reconstructed pCO2 at global scale. We added a corresponding sentence in the manuscript (lines 96-99). At basin and regional scale, however, adding SSH improves the spatial pattern of reconstructed pCO2 and the accuracy of our method. The full assessment of SSH will be part of a follow-up study.

C.R.: How much do the data outside the period improve the estimates of seasonality from the 1st step?
A.: The climatology by Takahashi et al. (2009) is based on data obtained from the early 1970s to 2007. Coastal pCO2 measurements and values obtained in the Eastern Equatorial Pacific are excluded. It represents average open ocean pCO2 conditions and is centred on the year 2000. A way to estimate the contribution of data outside the reconstruction period to estimates of seasonality would consist in computing the climatology over the period covered by the reconstruction only. A sensitivity analysis could also be carried out subsampling a numerical model. However, since Takahashi et al. excluded years with strong anomalies, we expect the influence of data outside of the reconstruction period to be small.
C.R.: Couldn’t step 1 also be fitted against the SOCAT data (folded into a climatological year as done by Takahashi et al., 2009) as well, rather than against the already-interpolated climatology? This would also bring in the data from the most recent years not yet in the Takahashi et al. (2009) climatology.

A.: This could indeed by an alternative approach to step 1, which could be tested in the future. The use of an interpolated climatology allows however to benefit from a full data coverage (a larger data set, as stated in lines 76-77), which would not be the case with SOCAT.

C.R.: Title: While there is no question that the authors are fully free to choose a name for their method, it seems that they intended to use a ”SOCOM-style” name. Therefore I’d like to point out that all names in the SOCOM paper are of the form INSTITUTION- METHOD, not the other way round.

A.: We followed the reviewer’s suggestion and changed the name of the model to LSCE-FFNN in all manuscript.

C.R.: Line 78: ”On a larger data set” - please clarify what this means

A.: The use of a gridded climatology during the first step allows benefitting from a global and continuous data coverage. It provides more data for training and validation. It was clarified in the text, line 76.

C.R.: Lines 79-81 and 446: Unclear statement - isn’t it the 2nd step (not the 1st) which takes care of changes?

A.: The 1st step reconstructs the climatological seasonal variability. The 2nd step adds interannual variability to the signal if present. It contributes information on anomalies with respect to the climatological state.

C.R.: Line 114: If you took surface pressure from the s76_v4.1 run, it is in fact from NCEP (Kalnay et al.). This should be mentioned.

A.: It was modified and mentioned in the text, lines 117-118.

C.R.: Lines 141, 142, 191: The "1" and "2" are awkward, and should rather be part of the subscript.

A.: It was modified, lines 146, 147, 199.

C.R.: Line 174: Isn’t the 1st step using climatologies of the driver variables? Please clarify. Also, is it correct that the 1st step uses un-normalized drivers (rather than SST_n etc.)?

A.: Yes, driver climatologies are used during step 1. All data have to be normalized for both steps of the model. It was clarified in the text, lines 181-182.

C.R.: Line 191: I assume that the "n" in the subscript to pCO$_2$ is not correct, is it?

A.: It is correct. “n” stands for “normalized”. All variables are normalized prior to their use in the model. It was clarified, lines 197-198.

C.R.: Line 197: Maybe say “for each climatological month”.

A.: Corrected, line 205.


A.: Yes, it is randomly. It was modified in the text, line 218.

C.R.: Line 212: What does ”final model” mean? Is it the model of step 2?
A.: Yes it is step 2. We have clarified it, line 221.

C.R.: Line 236: "(4)" should probably be "(5)".
A.: Corrected, line 237.

C.R.: Line 238: Is it from run s76_v4.1 as well? Please mention.
A.: Yes, it is. We have clarified it, line 246.

C.R.: Line 244: Isn't "did not participate" in contradiction to Sect 2.2 b) which says "100% used for training"?
A.: We distinguish between the development of the model presented in section 2.2 a) and its application to pCO$_2$ reconstruction in section 2.2 b). It is stated in the manuscript as follows: “The previous section presented the development of a FFNN model for the reconstruction of global surface ocean pCO$_2$, and the estimation of its accuracy”. During development, the accuracy of the method was evaluated based on 25% of data that did not participate in training. We added a sentence for clarification, line 253.

C.R.: Line 283: Though the Jena scheme indeed uses SST etc in its parameterizations, the sentence can be misunderstood as meaning regression drivers. Maybe just remove the list of drivers here.
A.: We prefer to provide a list of variables used. We have added that these variables are used in its parameterizations, line 297.

C.R.: Line 285: I feel "combined" is confusing and can be removed.
A.: We agree and removed it, line 298.

C.R.: Line 286: Please mention whether you used the same versions of these methods as used in the SOCOM paper, or whether you used updated versions. If updated, the respective version IDs should be given.
A.: We have added the number of versions used, lines 296, 298, 300.

C.R.: Line 286-287: The sentence "Qualification...(2015)." is repeated in the next paragraph and may be deleted here. I feel the next sentence "The time series..." should start a new paragraph.
A.: We agree, the sentence was deleted, line 299.

C.R.: Line 392: I guess "small IAV" should be "small relative IAV mismatch", isn’t it?
A.: It is. It was corrected, lines 409-410.

C.R.: Line 394, 296, 403, 455: The figures after the decimal point are certainly meaningless and should be removed.
A.: We agree, it was corrected, lines 408, 411, 413, 421.

C.R.: Lines 412, 425: It is unclear to me what "total negative trend" means.
A.: It is a linear trend averaged over the globe. It was clarified in the text, line 431.

C.R.: Line 413 and below: "PgC/yr" is the unit of a flux, not of a flux trend. Please clarify the unit (it should be something like "PgC/yr/yr" or "PgC/yr/decade").
A.: Yes, indeed, it is PgC/yr/yr. This has been corrected, lines 433-435.
C.R.: Lines 421ff: Are trends like 0.0004PgC/yr/yr statistically significant given the IAV? If not, it is not actually appropriate to call them "positive".

A.: We added results according with t-test (p-val is 0.05), lines 433-435. Figure 8 was changed according with t-test, previous Figure 8 was removed to supplementary material, Figure S8.

C.R.: Figures: In my print-out, the annotations and labels of most figures are much too small to be readable.

A.: The figures were changed, Fig. 4-7 and Fig. S2-S6.

C.R.: Fig 9, color bar: In this figure, the numbers should be within the colors, not between them, otherwise the meaning is ambiguous.

A.: Done.

C.R.: Fig S7 caption: "according" probably means "agreement"?

A.: Corrected.
We would like to thank Luke Gregor for his positive and helpful review.

We have answered his comments below.

Luke Gregor: The study introduces a new method that calculates pCO2 as climatology + anomaly. I think that this is a novel approach and is definitely relevant to the community. The authors do a good job of explaining the method and comparing it with past studies. I really like the approach and the method seems to perform well relative to the other methods in the SOCOM ensemble. I enjoyed reading the manuscript; however, there are many typos scattered throughout the document. I have noted most of these with track changes in the PDF document using Adobe Reader (attached as "supplement"). Below are some general comments, then repeated mistakes that can be fixed with find/replace.

Authors: Thank you for the positive assessment of our study. We apologize for the typos.

L.G.: The authors limit their study from 2001 to 2016. I realize that data is sparse before 2001, but it would be interesting to see how the model fares. It is useful for the community to know if this method (and in general the climatology-regression approach) can predict in data-sparse periods. It seems to be good at predicting in data-sparse regions.

A.: Thank you for your suggestion. We agree that the extension of the time series is a priority for future developments.

L.G.: It is really good that the authors use quasi-regular spacing to define the train-test validation splits, but I would like to know more about how they defined these splitting points. On this point, I also think it would be useful to have a measure of how much a model overfits. The authors could compare the test RMSE with the RMSE of the model trained with 100% of the data. If the latter is much smaller than the 50:25:25 splits, then there is overfitting. This is important for when the method is compared with other gap-filling methods.

A.: To be used by the FFNN algorithm data have to be formatted as a list “latitude by latitude”. Data are read “line by line” with each line corresponding to a latitude. Figure S1 presents the method of selection that was applied. This approach ensures that information from almost all regions and months are used for training.

The RMSE for the case when 100% of data were used for training amounts 14.8 μatm, only slightly smaller than the one computed for cross-validation (17.97 μatm). We added this information to the manuscript, lines 267-269.

L.G.: Regarding the figures, I struggle to see the difference between the brown and the black lines as I’m slightly colourblind. I would like to see another colour for the Jena-MLS method.

A.: Colour was changed to red.

L.G.: “Methods based on ANN are able to represent the large class of pCO2-driver relationships, but they are sensitive to the number of data used in the training algorithm and can generate artificial variability in regions with sparse data coverage.”. - I find the first part of the sentence a bit unclear. Not sure what is meant by "large class of pCO2-driver relationships". The second part of the sentence requires a reference.

A.: “Large class of pCO2-driver relationships“ refers to a large variety of combinations of drivers/predictors. The first part of the sentence was modified to “Methods based on ANN are able to represent the relationship between pCO2 and a variety of predictor combinations (e.g. pCO2=f(SSS,SST,SSH) or pCO2=f(SSS,SST,xCO2,CHL,MLD)), lines 63-64.

A reference was added, line 66.
L.G.: “The model is easily applied to the global ocean without any boundaries between the ocean basins or regions”. - This amazes me! Good results with one domain. A great advantage to the method!
A.: Thank you!

L.G.: “Based on Rodgers et al. (2009) who reported a strong correlation between natural variations in dissolved inorganic carbon (DIC) and sea surface height (SSH), SSH was added as a new driver to this list”. - I like the addition of SSH to the list, but then why not extend the predictions to 1993 (SSH limiting) or 1998 (Chl-a). I agree with Reviewer 1 that this can be discussed more.
A.: In this study, the priority was given to data coverage for the development of the model. The period retained (2001-2016) represents 77% of the SOCAT dataset. The extension of the period of reconstruction will be considered in future studies.

L.G.: I would have hoped that the climatology approach would be better at estimating pCO$_2$ for periods and regions where data is sparse.
A.: It is difficult to assess the skill of the model over regions with sparse data coverage based on observations only. An alternative approach would consist in subsampling a numerical model with the temporal and spatial coverage of real observations. These pseudo-data will be used for reconstructing surface ocean pCO$_2$. The comparison between reconstructed and modelled pCO$_2$ distributions will provide an upper estimate of the accuracy.

L.G.: “Monthly global observed physics reprocessed products distributed through the Copernicus Marine Environment Monitoring Service (CMEMS) (0.25°x0.25°) (http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MULTIOBS_GLO_PHY_REP_015_002) were used for SSS, SST and SSH”. - ARMOR3D L4? Reference: Guinehut et al 2012?
A.: Yes, it is the ARMOR3D L4 data. We added the reference Guinehut et al. (2012), lines 120 and 124.

L.G.: “For MLD, daily data from the “Estimating the Circulation and Climate of the Ocean” (ECCO2) project Phase II, at 0.25°x0.25° resolution (Menemenlis et al., 2008) were used”. - Please add the type of data product.
A.: The product name is ECCO2 and the release name is Cube 92. The information was added to the manuscript, line 127.

L.G.: Lignes 140-142 - Converted to radians first.
A.: Yes, of course. The information was added, line 143.

L.G.: “There is a well-known empirical rule advising to have a factor of 10 between number of patterns (data) and number of connections, or weights to adjust”. - Please add a reference.
COMMENT: Landschutzer 2013 used 30. Are there any signs of overfitting to the training data? i.e. is the training RMSE score much lower than the test RMSE score?
A.: Landschutzer et al. (2013) based their choice on Amari et al. (1997). A factor of 30 allows avoiding cross-validation. However, we believe that a problem such as the reconstruction of the time and spatial distribution of surface ocean pCO$_2$ requires cross-validation. In line with Amari et al. (1997), we use a factor of 10 that necessitates a cross-validation to avoid overfitting. We added information in the text, line 175-177.
Validation and evaluation data sets were chosen quasi-regularly in space and time to take into account all regions and seasonal variability. This is good, but I would like to know more about how the quasi-regular sampling was done.

As mentioned above, to be used by the FFNN algorithm data have to be formatted as a list “latitude by latitude”. Data are read “line by line” with one line per latitude. This approach ensures that information from almost all regions and months are used for training.

In order to improve the accuracy of the reconstruction, the model was applied separately for each month. So in effect, 12 models were trained? This is mentioned later in the text. Move that point here.

The corresponding sentence was added, lines 191-192.

We have developed a FFNN model with 5 layers (3 hidden layers). Sentence is unclear in this context. Put earlier.

We placed this sentence earlier in the manuscript, line 191.

There were thus 12 models sharing a common architecture but trained on different data. This could move to the first step to clarify the point a bit earlier. I have noted where this point could be moved to.

We have added a sentence to the description of the first step. We kept the sentence as it corresponds to the 2nd step of the model and we feel that it is needed for clarification, line 192.

At this step, in order to increase the amount of data during training and to introduce information on the seasonal cycle, the model was trained using as a target pCO$_2$ data from the month in question as well as those from the previous and following month during the entire period 2001-2016. My feeling is that having monthly models would really reduce the number of data in the Southern Ocean during winter. Really very little data. I have made a comment about this for Figure 1.

We agree that the Southern Ocean is particularly under-sampled with a bias towards summer months. However, the use of a single model for the reconstruction of surface ocean pCO$_2$ over the full period will not significantly improve the accuracy in the Southern Ocean, but it will reduce it over the global ocean.

Figures 1 (b) and 1 (c) show an example of data distribution for the sole months of January over the period 2001-2016 (Fig. 1 (b)) and for the three months time-window December-January-February 2001-2016 used in the training algorithm of the January FFNN model (Fig. 1 (c)). I would like to see this for July rather. This would really highlight the missing data for the Southern Ocean. Also, the figure is very small. I would prefer if the subplots were vertically stacked (single column).

The aim of these figures is to show a possible distribution of data and an impact of data from two neighboring months. We did not want to focus on a particular month and discuss associated problems. Figures were vertically stacked.

Each sampling was tested…” “sampling fold” would be a lot clearer.

We changed the text accordingly, line 217.

different initial values that are chosen automatically”. How are these chosen automatically? There must be an initialisation or a grid of hyper-parameters passed to the model. Please clarify.

The values are chosen randomly.
L.G.: “Note that a higher number of parameters did not show a significant improvement of accuracy (not shown).” - Remove (not shown). Makes me want to see this.

COMMENT: This is good. shows that over-training is being minimised.

A.: Done, line 223.

L.G.: “In order to provide the final output, the selected FFNN architecture is trained on all available data: 100% of data for training, 100% for evaluation and 100% for validation”. - I would like to see RMSE from validation data pre- and post-training, but only necessary in the supplementary materials. The authors can then state that their method is or is not prone to overfitting. This would show if there is overtraining.

A.: Thank you for the suggestion. We agree that it would be interesting to present these values to provide additional information on the model development. We have noted it for the next model release. The RMSE for the case of 100% of training data is 14.8 uatm. The information was added to the manuscript, lines 267-269.

L.G.: Ligne 225: FFNN-LSCE

A.: Corrected.

L.G.: “The global climatology was reconstructed with a satisfying accuracy during step 1 with a RMSE of 0.17 μatm and $r^2$ of 0.93”. - This is very low! might be worthwhile commenting on. This shows that pCO$_2$ deviation from the climatology contributes nearly all the error in estimates!

A.: The high scores during step 1 underline the capacity of our FFNN approach to reconstruct the surface ocean pCO$_2$ when there is a good data coverage. However, it is hard to conclude whether the 1$^{st}$ or the 2$^{nd}$ step has more or less contribution to total model accuracy. As you and Christian Rödenbeck mentioned, the estimation of the climatology during step 1 includes data outside of the reconstruction period, which might influence final results. This needs more investigation using numerical models, for example. However, we expect that the influence of the data outside of the period will not significantly modify our results.

L.G.: Ligne 253: absolute bias - Rather refer to this as the mean absolute error and I would also like to see the bias.

A.: We have found that total bias is difficult to explain as it varies a lot over the global ocean and its small value does not mean that the biases are small everywhere, often it results from compensation. We added a table presenting the bias to the Supplementary material.

“Absolute bias” was replaced by “mean absolute error” (MAE), lines 260, 263, 278, 279, 282, 283, 285.

L.G.: “The RMSE between SOCAT data and the climatology of pCO$_2$ from Takahashi et al. (2009) equals 41.87 μatm, larger than errors computed for the regional comparison between FFNN and SOCAT (Table 1)”. - Is this given as a normalisation/comparison - I think this is quite useful.

A.: Indeed, this is given as a comparison. The comparison stresses the skill of the model to reconstruct interannual anomalies. An error between reconstructed pCO$_2$ and SOCAT larger than between the climatology by Takahashi et al. (2009) and SOCAT would indicate that the model failed the capture the interannual variability.

L.G.: Ligne 269: “…Subtropical Pacific…” - assuming NH, but better to clarify.

A.: Done, 282.
L.G.: “Despite large time mean differences computed over the eastern Equatorial Pacific, scores are satisfying at the regional scale indicating error compensation by improved scores over the western basin”. - I’m glad this is spoken about. Please add a number to explain "satisfying".

A.: Done, line 285.

L.G.: Ligne 273: “…South hemisphere…” - Southern Hemisphere, each should be capitalised.

A.: Done, lines 29, 109, 286, 481.

All typos were corrected, some phrases were modified without change of its meanings:

lines 37, 39, 43, 46, 54, 60, 69, 103, 104, 109, 111, 117, 157, 162, 163, 168, 172-175, 186, 215, 226, 232, 244, 258-259, 263, 274, 301, 325, 330-331, 335, 338, 379, 381, 385, 408, 419, 420, 427, 465, 466, 480, 489-495, 504-508, 512-513, 527-529, 567-568; Captions for Fig. 1, 2, 4, 5, 6, 7, 8. Order of regions names in tables 1, 2, 3.