



# A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach

Daniel Broullón<sup>1</sup>, Fiz F. Pérez<sup>1</sup>, Antón Velo<sup>1</sup>, Mario Hoppema<sup>2</sup>, Are Olsen<sup>3</sup>, Taro Takahashi<sup>4</sup>†, Robert M. Key<sup>5</sup>, Toste Tanhua<sup>6</sup>, J. Magdalena Santana-Casiano<sup>7</sup>, and Alex Kozyr<sup>8</sup>

<sup>1</sup>Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello 6, 36208 Vigo, Spain

<sup>2</sup>Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research,  
Postfach 120161, 27515 Bremerhaven, Germany

<sup>3</sup>Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research,  
Allégaten 70, 5007 Bergen, Norway

<sup>4</sup>Lamont–Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA

<sup>5</sup>Atmospheric and Oceanic Sciences, Princeton University, 300 Forrester Road,  
Sayre Hall, Princeton, NJ 08544, USA

<sup>6</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

<sup>7</sup>Instituto de Oceanografía y Cambio Global, IOCG, Universidad de Las Palmas de Gran Canaria,  
Las Palmas de Gran Canaria, Spain

<sup>8</sup>NOAA National Centers for Environmental Information, 1315 East-West Hwy Silver Spring, MD 20910, USA

†deceased

**Correspondence:** Daniel Broullón (dbroullon@iim.csic.es)

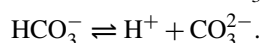
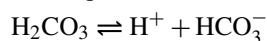
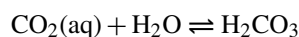
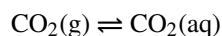
Received: 14 February 2020 – Discussion started: 11 March 2020

Revised: 18 June 2020 – Accepted: 2 July 2020 – Published: 5 August 2020

**Abstract.** Anthropogenic emissions of CO<sub>2</sub> to the atmosphere have modified the carbon cycle for more than 2 centuries. As the ocean stores most of the carbon on our planet, there is an important task in unraveling the natural and anthropogenic processes that drive the carbon cycle at different spatial and temporal scales. We contribute to this by designing a global monthly climatology of total dissolved inorganic carbon (TCO<sub>2</sub>), which offers a robust basis in carbon cycle modeling but also for other studies related to this cycle. A feedforward neural network (dubbed NNGv2LDEO) was configured to extract from the Global Ocean Data Analysis Project version 2.2019 (GLODAPv2.2019) and the Lamont–Doherty Earth Observatory (LDEO) datasets the relations between TCO<sub>2</sub> and a set of variables related to the former’s variability. The global root mean square error (RMSE) of mapping TCO<sub>2</sub> is relatively low for the two datasets (GLODAPv2.2019: 7.2 μmol kg<sup>-1</sup>; LDEO: 11.4 μmol kg<sup>-1</sup>) and also for independent data, suggesting that the network does not overfit possible errors in data. The ability of NNGv2LDEO to capture the monthly variability of TCO<sub>2</sub> was testified through the good reproduction of the seasonal cycle in 10 time series stations spread over different regions of the ocean (RMSE: 3.6 to 13.2 μmol kg<sup>-1</sup>). The climatology was obtained by passing through NNGv2LDEO the monthly climatological fields of temperature, salinity, and oxygen from the World Ocean Atlas 2013 and phosphate, nitrate, and silicate computed from a neural network fed with the previous fields. The resolution is 1° × 1° in the horizontal, 102 depth levels (0–5500 m), and monthly (0–1500 m) to annual (1550–5500 m) temporal resolution, and it is centered around the year 1995. The uncertainty of the climatology is low when compared with climatological values derived from measured TCO<sub>2</sub> in the largest time series stations. Furthermore, a computed climatology of partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>) from a previous climatology of total alkalinity and the present one of TCO<sub>2</sub> supports the robustness of this product through the good correlation with a widely used pCO<sub>2</sub> climatology (Landschützer et al., 2017). Our TCO<sub>2</sub> climatology is distributed through the data repository of the Spanish National Research Council (CSIC; https://doi.org/10.20350/digitalCSIC/10551, Broullón et al., 2020).

## 1 Introduction

The ocean is the major carbon reservoir of the Earth. Most of this carbon occurs as dissolved inorganic carbon (TCO<sub>2</sub>, also known as DIC or C<sub>T</sub>; Ciais et al., 2013; Tanhua et al., 2013). Three species make up TCO<sub>2</sub>: dissolved CO<sub>2</sub> (generally considered to be the sum of the dissolved CO<sub>2</sub> itself, CO<sub>2</sub>(aq), and carbonic acid, H<sub>2</sub>CO<sub>3</sub>), bicarbonate ion (HCO<sub>3</sub><sup>-</sup>), and carbonate ion (CO<sub>3</sub><sup>2-</sup>). The relative concentrations of these species with respect to each other determine the seawater pH (Zeebe and Wolf-Gladrow, 2001). The seawater CO<sub>2</sub> chemistry system can be represented as a set of chemical equilibria reactions that describe the speciation of the various ions of TCO<sub>2</sub> as follows:



Since the Industrial Revolution, the concentration of TCO<sub>2</sub> in the global ocean has increased, generally to a certain depth level (depending on the particular processes in each ocean area) due to the entry of CO<sub>2</sub> into the seawater from the atmosphere (Sarmiento and Gruber, 2002; Doney et al., 2009; Vázquez-Rodríguez et al., 2009; Bates et al., 2012; Sallée et al., 2012; Khatiwala et al., 2013). The uptake is driven by the increasing partial pressure of CO<sub>2</sub> (*p*CO<sub>2</sub>) in the atmosphere relative to the ocean, generated by the anthropogenic emissions of CO<sub>2</sub> that cause an annual net flux of this gas into the ocean (Le Quéré et al., 2018). Accompanying the change in TCO<sub>2</sub>, the pH and carbonate ion concentration have been declining because of the anthropogenic process previously mentioned, these changes being reflected in the proportions of the chemical species of TCO<sub>2</sub> (Kleypas and Langdon, 2000; Orr et al., 2005). These changes in seawater chemistry framed in the ocean acidification process can negatively influence various processes involving marine organisms such as calcification, growth, and survival (Orr et al., 2005; Fabry et al., 2008; Hendriks et al., 2010; Hoegh-Guldberg and Bruno, 2010; Kroeker et al., 2013).

In addition to the secular trends driven by the uptake of anthropogenic CO<sub>2</sub>, ocean TCO<sub>2</sub> varies both temporally and spatially as a consequence of several natural processes. This variability may reach values of 15 % of the mean TCO<sub>2</sub> value in the ocean (Lee et al., 2000). The processes that increase TCO<sub>2</sub> are net flux of CO<sub>2</sub> from the atmosphere to the ocean, organic matter remineralization, and the dissolution of calcium carbonate (CaCO<sub>3</sub>). The processes that reduce TCO<sub>2</sub> are net flux of CO<sub>2</sub> from the ocean to the atmosphere, primary production, and calcification. Advection and mixing also influence the variability of TCO<sub>2</sub> in these two ways (Sabine et al., 2002). In the surface ocean, the main vari-

ables influencing the variability of TCO<sub>2</sub> are temperature and salinity (Weiss et al., 1982; Lee et al., 2000; Wu et al., 2019), through the modification of the solubility of CO<sub>2</sub>, affecting the seawater *p*CO<sub>2</sub> (which is almost instantaneous) and thus the air–sea CO<sub>2</sub> flux, which eventually drives the change in TCO<sub>2</sub> over time. Nutrients and oxygen can also reflect the processes that modify the concentration of TCO<sub>2</sub> through their consumption and release, like during the cycling of organic matter (Körtzinger et al., 2001; Bauer et al., 2013). From products generated with measured data (Key et al., 2004; Takahashi et al., 2014; Lauvset et al., 2016) and in modeling studies (e.g., Doi et al., 2015), it is known that the global surface distribution of TCO<sub>2</sub> follows a zonal gradient: there is a reduction in its concentration from the poles to the Equator, reflecting the processes that control its variability. Key et al. (2004) emphasize that this distribution is associated with the distribution pattern of nutrients. Recently, Wu et al. (2019) found that the distribution of surface-salinity-normalized TCO<sub>2</sub> (nDIC) has two main drivers: temperature and upwelling. At depth, the variation shown in almost any measured profile of TCO<sub>2</sub> mainly reflects the remineralization of organic matter and, to a lesser extent, the dissolution of CaCO<sub>3</sub> (Millero, 2007), resulting in an increase in TCO<sub>2</sub> from the surface to intermediate depths.

Understanding the distribution and variability of TCO<sub>2</sub> in the ocean and its secular trends driven by anthropogenic carbon uptake is needed to assess the magnitude and possible impacts of ocean acidification. It is also necessary for the evaluation of numerical models that include the carbon cycle and their estimates of past, current, and future ocean carbon cycle behavior (e.g., Yool et al., 2013; Aumont et al., 2015; Butenschön et al., 2016; Le Quéré et al., 2016; Goris et al., 2018). Seasonality of TCO<sub>2</sub> and the horizontal and vertical variability underscore the necessity to design a climatology with both monthly and spatial resolutions according to the processes that influence this variable on a global scale. The existing climatologies of TCO<sub>2</sub> do not include all these characteristics collected together. Key et al. (2004) and Lauvset et al. (2016) built an annual climatology at 33 depth levels using interpolation techniques with data from the Global Ocean Data Analysis Project version 1 (GLODAPv1; Key et al., 2004) and version 2 (GLODAPv2; Key et al., 2015; Olsen et al., 2016), respectively. Takahashi et al. (2014) published a monthly climatology for the surface ocean computed from climatologies of *p*CO<sub>2</sub> and total alkalinity (*A<sub>T</sub>*). Other studies used the covariability between TCO<sub>2</sub> and other more commonly measured variables discussed above for mapping and gap-filling via empirical regressions and neural networks. Lee et al. (2000) used temperature and nitrate to compute surface nDIC with an area-weighted error of ±7 μmol kg<sup>-1</sup>. Sauzède et al. (2017) and Bittig et al. (2018) trained neural networks with GLODAPv2 data to compute TCO<sub>2</sub> over the depth range 0–8000 m with

an accuracy of  $\pm 9$  and  $\pm 7.1 \mu\text{mol kg}^{-1}$ , respectively. The input variables used in those studies were location, pressure, temperature, salinity, dissolved oxygen, and time.

In the present study, we introduce the use of neural networks for going one step further in the design of a climatology. We have generated a climatology of TCO<sub>2</sub> with a resolution consistent with that of the climatology of A<sub>T</sub> of Broullón et al. (2019): horizontal resolution of  $1^\circ \times 1^\circ$ , 102 depth levels between 0 and 5500 m, and a monthly (0–1500 m) and annual (1550–5500 m) temporal resolution. The availability of global databases containing variables of the seawater CO<sub>2</sub> system with more and more data – e.g., GLODAPv2.2019; the Lamont–Doherty Earth Observatory (LDEO) database (Takahashi et al., 2017); the Surface Ocean CO<sub>2</sub> Atlas (SOCAT; Bakker et al., 2016) – and the great ability of the neural networks to interpolate as shown in other climatological studies about CO<sub>2</sub> system variables (Landschützer et al., 2014; Broullón et al., 2019) show the appropriateness of this approach for generating a global monthly climatology covering more than the surface ocean.

## 2 Methodology

### 2.1 Neural network design

A feedforward neural network was configured to compute TCO<sub>2</sub> in the global ocean and to create a global climatology based on the good results previously obtained with this method in similar studies (e.g., Broullón et al., 2019). Briefly, a neural network of this type (Fig. S1a in the Supplement) is used to extract relationships between a set of input variables and a target one through a training process. At this stage, the inputs are passed through different parallel layers composed of a tunable number of neurons to reach values as closest as possible to the target ones (Fig. S1a). Initially, all inputs enter each neuron of the first layer, where they are multiplied by different weights depending on the neuron they go to. Inside the neurons (Fig. S1b), the results of the previous operation are summed, and a bias is added. The obtained value inside each neuron is passed through an activation function, which yields an output. The outputs of each neuron in each layer go to the following layer undergoing the same process described to this point. In the last layer, which is composed of one neuron, a unique value for the target variable is calculated for each pair consisting of inputs and a target. This value is compared to the desired one, and the difference between both values is backpropagated through the entire network in order to adjust the weights and biases and to start the processes again and reach an accurate output value after multiple iterations. A complete description of the most common algorithms used to backpropagate and minimize the errors can be found in Rumelhart et al. (1986), Levenberg (1944), and Marquardt (1963).

The method used here is equivalent to that fully described by Broullón et al. (2019) for A<sub>T</sub>. In addition to the target vari-

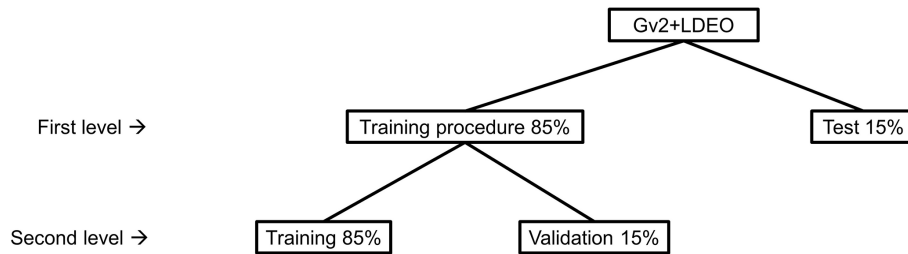
able (TCO<sub>2</sub> instead of A<sub>T</sub>), the main changes in the present study compared to that of Broullón et al. (2019) are the inclusion of the input variable “year”, accounting for the anthropogenic increase in the TCO<sub>2</sub> pool, and the use of the pCO<sub>2</sub> database from the LDEO (Takahashi et al., 2017) in addition to the extended GLODAPv2.2019 (Olsen et al., 2019) to enable more robust TCO<sub>2</sub> estimates in the surface ocean. Similar to Broullón et al. (2019), the neural networks were trained using the Levenberg–Marquardt method (Levenberg, 1944; Marquardt, 1963) through the *trainlm* function (detailed in Beale et al., 2018) in MATLAB. The splitting of the database used in the present study (see Sect. 2.2) into the sets needed for training and testing the network is depicted in Fig. 1. The data were randomly associated with each dataset to capture (training) and evaluate (test) all possible variability. The input variables are temperature, salinity, phosphate, nitrate, silicate, oxygen, sample position, and year (Fig. S1a). The number of neurons tested in the unique hidden layer to find the best neural network was 16, 32, 64, 128, and 256. Ten networks were trained for each number of neurons. The criteria for selecting the final number of neurons are based on a trade-off between the root mean square error (RMSE; between the measured TCO<sub>2</sub> and that estimated by the neural network) on the one hand and the generalization of the network (to prevent overfitting, maintaining a similar error in the training and in the test sets) on the other hand. Furthermore, an additional criterion based on the influence of each input variable on the TCO<sub>2</sub> extracted with the connection weight approach (Olden and Jackson, 2002) was followed to ensure that biogeochemical input variables have a larger influence on the TCO<sub>2</sub> estimates than the input variables related to sample position for selecting a proper network. The influence of each input variable on the computed TCO<sub>2</sub> was obtained from Eq. (1):

$$C_i = \sum_{k=1}^H w_{ik} \cdot w_k, \quad (1)$$

where  $C_i$  is the relative importance of the input variable  $i$ ,  $H$  the number of neurons in the hidden layer,  $w_{ik}$  the weight of the connection between the variable  $i$  and the neuron  $k$  of the hidden layer, and  $w_k$  the weight of the connection between the neuron  $k$  of the hidden layer and output layer.

### 2.2 Data

We included the LDEO database version 2016 (Takahashi et al., 2017; [https://www.nodc.noaa.gov/ocads/oceans/LDEO\\_Underway\\_Database](https://www.nodc.noaa.gov/ocads/oceans/LDEO_Underway_Database), last access: 13 November 2017) because it contains significantly more data in the surface layer than GLODAPv2.2019. Since the higher variability in the surface layer may lead to high errors in modeling variables of the seawater CO<sub>2</sub> system (e.g., Carter et al., 2018; Bittig et al., 2018; Broullón et al., 2019), including the LDEO database should force the network to reach a more



**Figure 1.** Division of the complete database in the datasets needed to train the neural network. The percentages in each level are relative to the number of data in the previous one. Data in the datasets of the first level are always the same for each network. Data in the sets of the second level are randomly associated with each set for each network to find the best network weights because of the different starting points in the error weight space of the training process (see also Broullón et al., 2019).

robust fit. The idea is that these additional data probably have more different relationships between input variables and TCO<sub>2</sub> to help the neural network to adequately capture spatiotemporal variability. The  $p\text{CO}_2$ , temperature, and salinity data from the LDEO were monthly averaged for each year in a  $1^\circ \times 1^\circ$  grid. The points where the standard deviation (SD) of the averaged  $p\text{CO}_2$ , temperature, and salinity were greater than  $\pm 20 \mu\text{atm}$ ,  $1.5^\circ\text{C}$ , and  $0.5$ , respectively, were discarded since the objective is to capture the monthly variability, and therefore an extremely high submonthly variability could lead to errors. To obtain TCO<sub>2</sub> values from the LDEO data, an additional variable of the CO<sub>2</sub> system is necessary, for which we take  $A_T$  computed using the neural network NNGv2 of Broullón et al. (2019). The input variables required by NNGv2 were obtained from (1) temperature and salinity from the LDEO; (2) filtered oxygen from the World Ocean Atlas version 2013 (WOA13; see Broullón et al., 2019); and (3) phosphate, nitrate, and silicate computed with CANYON-B (Bittig et al., 2018) using the previous variables as inputs. Finally, TCO<sub>2</sub> was calculated from this  $A_T$  and the averaged  $p\text{CO}_2$  using the MATLAB version of the CO<sub>2</sub>SYS program (van Heuven et al., 2011); we used the dissociation constants of Mehrbach et al. (1973; as refit by Dickson and Millero, 1987) and the borate dissociation constant of Dickson (1990). Note that we used this software and set of constants for all seawater CO<sub>2</sub> chemistry calculations in the present study. The TCO<sub>2</sub> calculated this way and the associated input variables were used as a part of the training and testing data for the neural networks created here. The final number of data points derived from the LDEO was 54572.

To represent interior ocean conditions, the GLO-DAPv2.2019 database (Olsen et al., 2019) was added to the LDEO dataset for training and testing the neural network. Only samples which had data for all input variables and TCO<sub>2</sub> were used. This database was included in two ways: (1) only samples where all variables passed the second quality control ( $n = 287\,953$ ; Olsen et al., 2016; Olsen et al., 2019; hereafter abbreviated Gv2QC) and (2) all samples ( $n = 321\,647$ ; hereafter abbreviated Gv2). Therefore, two neural

network options were trained and tested: NNGv2QCLDEO and NNGv2LDEO, respectively.

### 2.3 Comparison of methods

We compared our method with CANYON-B of Bittig et al. (2018), where TCO<sub>2</sub> values were also computed from multiple input variables. Both methods are based on neural networks but with certain differences as summarized in Table 1.

An error analysis was carried out in the same areas for which this was done by Broullón et al. (2019) for  $A_T$  and in several depth ranges (0–50, 50–200, 200–500, and 500–1000 m and 1000 m–bottom) for the two methods (our method and CANYON-B) and for the two datasets (Gv2QC and LDEO). The Gv2QC database was analyzed in this section instead of Gv2 because in the designing of CANYON-B only quality-controlled data were included. The analysis of CANYON-B using the LDEO dataset is useful to evaluate the validity of the approach followed by converting  $p\text{CO}_2$  to TCO<sub>2</sub> since CANYON-B has not been trained with this dataset.

Computed  $p\text{CO}_2$  from  $A_T$  and TCO<sub>2</sub> derived from neural networks was also evaluated in the LDEO dataset to assess the adequacy of including this dataset in our approach and to assess the ability of NNGv2 of Broullón et al. (2019) and the present TCO<sub>2</sub> neural network to compute other variables of the seawater CO<sub>2</sub> system. Furthermore, we compared the magnitude of the errors with the ones obtained by Landschützer et al. (2014), in which  $p\text{CO}_2$  is computed directly with a neural network, to evaluate the accuracy of our computed  $p\text{CO}_2$ .

### 2.4 Validation

In addition to the ability to compute TCO<sub>2</sub> using the Gv2 and LDEO test sets, the neural network has been tested using independent data from 10 ocean time series, located in different regions of the world ocean (data were obtained from [https://www.nodc.noaa.gov/ocads/oceans/time\\_series\\_moorings.html](https://www.nodc.noaa.gov/ocads/oceans/time_series_moorings.html), last access: 4 June 2019): Hawaii Ocean

**Table 1.** Differences between the methods used in the present study and in CANYON-B (Bittig et al., 2018).

|                    | Bittig et al. (2018)                              | This study  |
|--------------------|---|---|
| Training technique | Bayesian regularization                           | Levenberg–Marquardt   |
| Input variables    | Temperature, salinity, oxygen, position, and time | Temperature, salinity, oxygen, phosphate, nitrate, silicate, position, and time |
| Datasets           | GLODAPv2 (Olsen et al., 2016)                     | GLODAPv2.2019 (Olsen et al., 2019)<br>LDEOv2016 (Takahashi et al., 2017)        |

Time-series (HOT ALOHA and HOT ALOHA SURFACE; Dore et al., 2009), Bermuda Atlantic Time-series Study (BATS; Bates et al., 2012), European Station for Time-series in the Ocean at the Canary Islands (ESTOC; González-Dávila et al., 2010), Iceland Sea time series (ICELAND; Olafsson et al., 2010), Irminger Sea time series (IRMINGER; Olafsson et al., 2010), Kyodo North Pacific Ocean time series (KNOT; Wakita et al., 2010), K2 (Wakita et al., 2010), Ocean Weather Station Mike (OWS; Gislefoss et al., 1998), and Kerguelen Islands in the Indian sector of the Southern Ocean (KERFIX; Jeandel et al., 1998). CANYON-B was also used to compute TCO<sub>2</sub> in the time series to show the differences between that method and ours. The TCO<sub>2</sub> values were obtained by feeding the neural networks with the measured values of the input variables at each time series. The data from these time series allow us to test the ability of the neural network to reconstruct not only the seasonal variability of TCO<sub>2</sub> at the various locations and depths sampled but also its long-term trends. For the trend analyses, the measured and estimated TCO<sub>2</sub> values were deseasonalized following Bates et al. (2014).

As an additional test, the measured *p*CO<sub>2</sub> or the *p*CO<sub>2</sub> calculated from measured TCO<sub>2</sub> and *A*<sub>T</sub> at the time series stations was compared with *p*CO<sub>2</sub> calculated from the neural-network-generated values of *A*<sub>T</sub> and TCO<sub>2</sub>. This provides insight into the combined performance of the NNGv2 of Broullón et al. (2019) and the neural network designed in the present study. Furthermore, we compared the magnitude of the errors to that obtained by Landschützer et al. (2014) for some of the time series.

## 2.5 Climatology of TCO<sub>2</sub>

We used the selected network, based on the results of the analyses described above, to construct a climatology of TCO<sub>2</sub>. Climatologies of the input variables were passed through the network to obtain the climatological fields of TCO<sub>2</sub>. The spatiotemporal resolution of the product is determined by that of the climatologies used as inputs: 1° × 1° horizontal resolution, 102 upper depth levels of the WOA13, and monthly (for 0–1500 m depth) to annual (for 1550–5500 m depth) temporal resolution. Temperature and salinity climatologies were obtained from objectively analyzed WOA13 fields (Locarnini et al., 2013; Zweng et al., 2013; <https://www.nodc.noaa.gov/OC5/woa13/woa13data.html>, last ac-

cess: 6 February 2017). Oxygen, phosphate, nitrate, and silicate climatologies were taken from Broullón et al. (2019; <https://doi.org/10.20350/digitalCSIC/8644>, last access: 1 August 2019). These climatologies of nutrients were created using the objectively analyzed climatologies of temperature, salinity, and oxygen (Garcia et al., 2014; oxygen climatology from WOA13 has been filtered by applying a fifth-order one-dimensional median filter through the depth dimension; see Broullón et al., 2019) from the WOA13 in CANYON-B (Bittig et al., 2018). As a year input is needed, we decided to center the TCO<sub>2</sub> climatology around 1995 based on the time distribution of the data used to create the WOA13 climatologies: the World Ocean Database 2013 (Boyer et al., 2013).

The computed climatological values were compared with those from measured data to assess the uncertainty of the climatology since the WOA13 does not offer an uncertainty field with the objectively analyzed climatologies. Unfortunately, only two locations have enough measured data to calculate a pure climatological value of TCO<sub>2</sub> for each month: HOT ALOHA and BATS. The measured values were monthly averaged at several depth levels, and the anthropogenic carbon as calculated by Lauvset et al. (2016) was added or subtracted to correct the data to the reference year of the climatology according to

$$\text{TCO}_2^{\text{year}_2} = \text{TCO}_2^{\text{year}_1} - C_{\text{ant}2002} \left[ (1 + 0.0191)^{(\text{year}_1 - 2002)} - (1 + 0.0191)^{\text{year}_2 - 2002} \right], \quad (2)$$

where TCO<sub>2</sub><sup>year<sub>2</sub></sup> is the TCO<sub>2</sub> corrected for year<sub>2</sub>, which is the reference year of the climatology; TCO<sub>2</sub><sup>year<sub>1</sub></sup> is the TCO<sub>2</sub> measured in year<sub>1</sub>, C<sub>ant2002</sub> is the anthropogenic carbon for 2002; and 0.0191 is the annual increase rate derived from the scaling factor determined by Gruber et al. (2019) for the global ocean between 1994 and 2007.

We compared our climatology with previously published climatologies of TCO<sub>2</sub>. The monthly surface climatology created by Takahashi et al. (2014) was used to assess the spatiotemporal differences in the surface layer. The annual climatology of Lauvset et al. (2016) was used to evaluate the spatial differences in the deeper parts of the ocean. For the comparisons, the climatologies of Takahashi et al. (2014) and Lauvset et al. (2016) were adjusted to the year 1995, subtracting the anthropogenic carbon (C<sub>ant</sub>) of Lauvset et al. (2016) as in Eq. (2).

Finally, a surface climatology of  $p\text{CO}_2$  was computed from the TCO<sub>2</sub> climatology of the present study and the  $A_T$  climatology of Broullón et al. (2019) to assess the potential of computing climatologies of other variables of the seawater CO<sub>2</sub> system. For comparison, the updated monthly  $p\text{CO}_2$  climatology from Landschützer et al. (2016, 2017) was used. The values between 1981 and 2010 were averaged to obtain the climatological year 1995. The variable selected from Landschützer et al. (2017) was that labeled as `spco2_raw` (sea surface  $p\text{CO}_2$ ) in the `netCDF` file.

It should be noted that the RMSE and the bias were obtained for all the comparisons, the last statistic being computed as the difference between the measured (or computed by the method to compare) TCO<sub>2</sub> and the one obtained with the neural network of the present study.

### 3 Results

#### 3.1 Neural network analysis

Following the established criteria to obtain the optimal number of neurons, the configuration with 128 neurons in the hidden layer was selected. From the 10 networks trained with this number of neurons for each approach (NNGv2LDEO and NNGv2QCLDEO), the ones with the lowest influence of the position input variables were selected. These two networks present a similar RMSE in both training and test datasets, showing there is no overfitting. Because in Gv2QC, both NNGv2LDEO and NNGv2QCLDEO produce the same global RMSE ( $6.1 \mu\text{mol kg}^{-1}$ ), it is likely that the Gv2 dataset contains high-quality measurements, and the possible errors in the non-QC data of this dataset are clearly avoided by the network; otherwise NNGv2LDEO should have a higher RMSE in the test dataset than NNGv2QCLDEO because of an overfitting of the errors in the Gv2 dataset. The same holds for the LDEO dataset. The network properly fitted TCO<sub>2</sub> derived from the LDEO dataset since it does not significantly increase the global RMSE relative to a network only trained with Gv2. Therefore, we decided to continue with NNGv2LDEO only since it has fitted more relationships between variables (e.g., Gv2 has more data points than Gv2QC in the Mediterranean Sea), providing a more robust fitting. For this network, the influence of each input variable on the computed TCO<sub>2</sub> is depicted in Fig. S2. The position variables together (latitude, clongitude, slongitude, and depth) have no more than 30 % influence, allowing biogeochemical variables to be the main ones responsible for the variability of TCO<sub>2</sub>. Furthermore, the input variable year has an influence lower than 5 %. This is probably responsible for capturing the positive interannual trend due to the TCO<sub>2</sub> increase derived from anthropogenic emissions of CO<sub>2</sub> to the atmosphere (see Sect. 3.2).

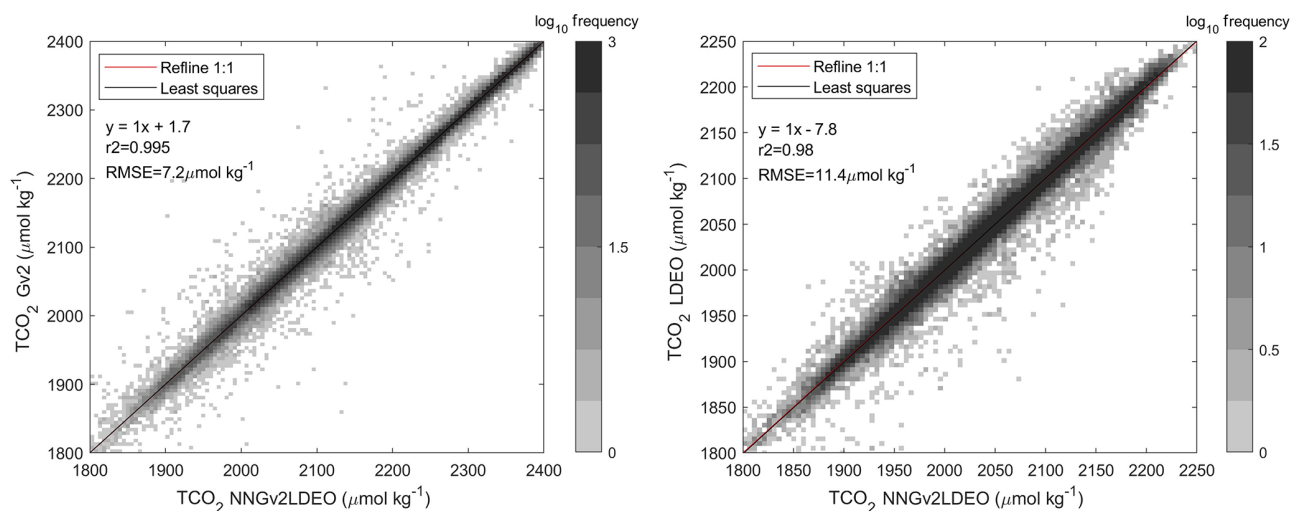
The global RMSE is quite low for the Gv2 dataset and for the LDEO dataset (Fig. 2). The measured and the computed data are highly correlated (Fig. 2), and the bias is negli-

ble in both datasets. The higher RMSE in the LDEO dataset likely results from the higher variability of TCO<sub>2</sub> in the surface layer and from uncertainties in its calculation from  $p\text{CO}_2$ .

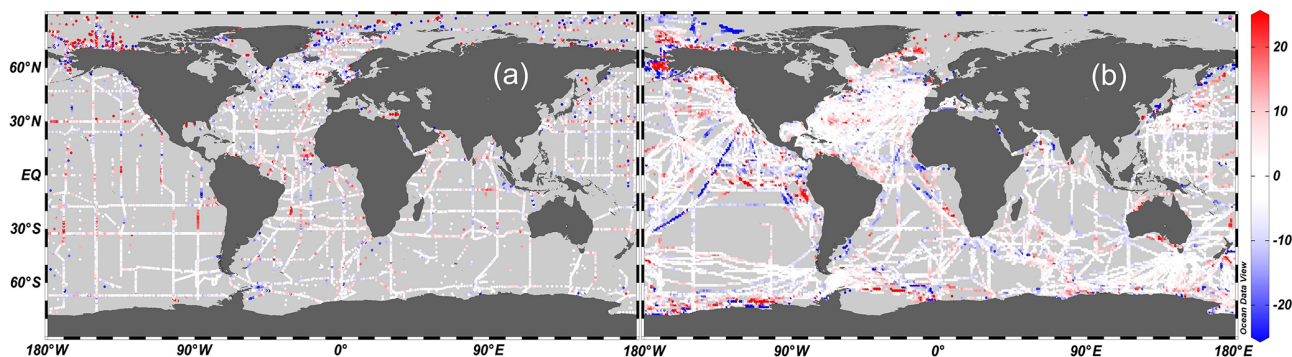
The RMSE by area and depth for NNGv2LDEO and CANYON-B in Gv2QC is shown in Table 2. The highest errors for the two methods are in the 0–50 m layer for the Gv2QC dataset and the LDEO dataset. These errors get smaller with increasing depth for all areas, and the depth-weighted RMSE of the two methods is not significantly different below 50 m. In the LDEO dataset, NNGv2LDEO produces a lower error than CANYON-B, except for two areas: East GIN (Greenland, Iceland, and Norwegian) Seas and the Bengal Basin (Table 2), although there are only 9 and 13 data points, respectively, in each area. Interestingly, CANYON-B is able to reproduce the TCO<sub>2</sub> data derived from the complete LDEO dataset with a lower error than the one it obtains for the complete Gv2QC dataset in the surface ocean (RMSE LDEO:  $16.4 \mu\text{mol kg}^{-1}$ ; RMSE Gv2QC, 0–5 m:  $17.8 \mu\text{mol kg}^{-1}$ ), supporting the approach of computing reliable TCO<sub>2</sub> values from the  $p\text{CO}_2$  of the LDEO dataset and the  $A_T$  computed with NNGv2 (Broullón et al., 2019) since CANYON-B was not trained with the LDEO database. A similar result was obtained for NNGv2LDEO but with a higher difference between the two errors (RMSE LDEO:  $11.4 \mu\text{mol kg}^{-1}$ ; RMSE Gv2QC, 0–5 m:  $17.1 \mu\text{mol kg}^{-1}$ ). Finally, the surface RMSE towards LDEO data of NNGv2LDEO is clearly lower than that of CANYON-B. This shows the value of including  $p\text{CO}_2$ -derived surface TCO<sub>2</sub> among the training data, through which there are more fitted relations in our new method.

For data from Gv2 where no QC was performed for at least one of the variables used in the present study (Gv2noQC), the RMSE also decreases with increasing depth (< 50 m:  $22.5 \mu\text{mol kg}^{-1}$ ; 50–200 m:  $9.8 \mu\text{mol kg}^{-1}$ ; 200–500 m:  $7 \mu\text{mol kg}^{-1}$ ; 500–1000 m:  $5.4 \mu\text{mol kg}^{-1}$ ; > 1000 m:  $5.4 \mu\text{mol kg}^{-1}$ ). Thus, the error in Gv2noQC is similar to that in the areas with the highest error in Gv2QC (Table 2; except in Beaufort Sea, where the error is considerably higher). However, the higher error in Gv2noQC is mainly caused by the samples located in the Arctic Ocean since cruises in the Atlantic and Pacific oceans are modeled with a very low error. Therefore, using Gv2noQC does not imply the introduction of low-quality data in our study; otherwise the network would not compute TCO<sub>2</sub> with low errors in Gv2QC because of an overfitting of the possible low-quality data that Gv2noQC could contain.

In general, the highest differences between measured and estimated TCO<sub>2</sub> occurs in the high-latitude surface oceans (Figs. 3 and 4). In Gv2, 40 % of the samples with differences beyond  $\pm 3\text{RMSE}$  (3 times RMSE; threshold selected to refer to samples with large residuals) are at latitudes greater than 70° N. In the LDEO dataset, 39 % of the samples with differences beyond  $\pm 3\text{RMSE}$  are from latitudes south of 70° S. These samples where RMSE is high are 7.5 % of the total



**Figure 2.** Regression of TCO<sub>2</sub> computed using NNGv2LDEO and TCO<sub>2</sub> in Gv2 and the LDEO dataset. The graph is divided into pixels. The color of each pixel is determined by the number of points inside it. Note the logarithmic scale of the pixels accounting for the large number of data.



**Figure 3.** Differences between (a) Gv2 TCO<sub>2</sub> and NNGv2LDEO TCO<sub>2</sub> (0–30 m) and (b) LDEO TCO<sub>2</sub> and NNGv2LDEO TCO<sub>2</sub> (0 m). This figure was made with Ocean Data View (Schlitzer, 2016).

north of 70° N in Gv2 and 42 % of the total south of 70° S in the LDEO dataset. The samples with low salinities have the highest errors (Fig. 4). A total of 41.5 % of the samples in Gv2 and 43 % in the LDEO dataset with differences beyond  $\pm 3\text{RMSE}$  have salinities below 33. Furthermore, in the LDEO dataset, the number of samples with residuals beyond  $\pm 3\text{RMSE}$  increases with increasing SD of both  $p\text{CO}_2$  and salinity in the monthly averaging in each pixel in the LDEO subset (Fig. S3). This result shows the difficulty of modeling areas with a high submonthly variability in  $p\text{CO}_2$  and salinity and supports the exclusion of the averaged LDEO data with a high SD since it could cause the network to interpret the submonthly variability as monthly variability (note that the purpose of this study is to capture the monthly variability).

Like for modeling  $A_T$  (Takahashi et al., 2014; Broullón et al., 2019), the Arctic Ocean is one of the regions with the highest RMSE of neural-network-estimated TCO<sub>2</sub>. The

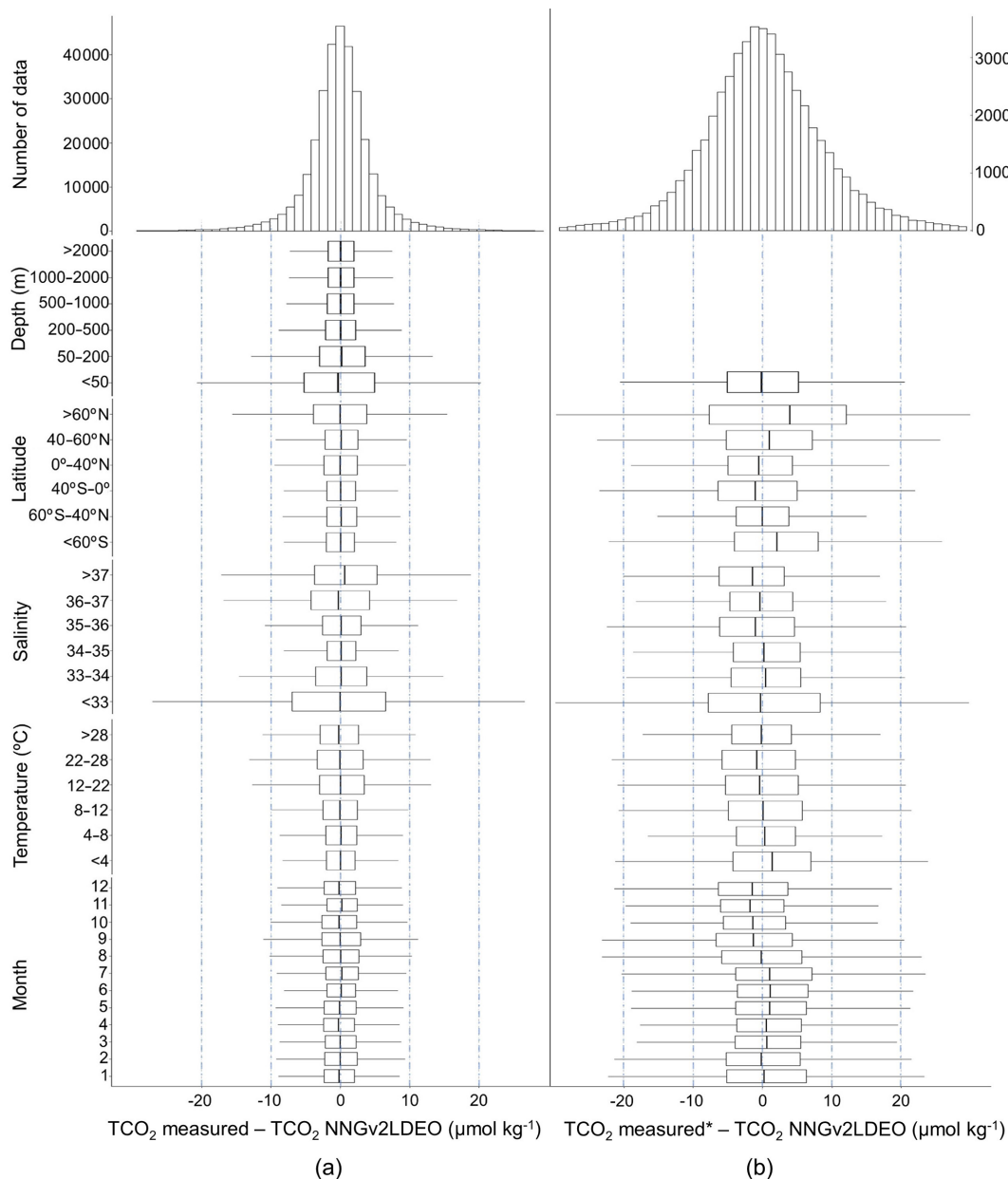
major Arctic rivers contribute TCO<sub>2</sub> concentrations ranging between 400 and 3600  $\mu\text{mol kg}^{-1}$  (estimated by Tank et al., 2012), derived mainly from carbonated rocks in the watersheds. Other areas like the Okhotsk Sea also show a high RMSE (Table 2 and Fig. 3), probably because of the high riverine input of TCO<sub>2</sub> (Watanabe et al., 2009). An input variable accounting for the contribution of the rivers to the TCO<sub>2</sub> pool would improve the neural network performance in areas like these, but it is not available.

The errors of the  $p\text{CO}_2$  computed in the LDEO dataset with TCO<sub>2</sub> from NNGv2LDEO and  $A_T$  from NNGv2 (Broullón et al., 2019) are similar to the errors obtained by Landschützer et al. (2014) for the SOCAT database in some of the areas (10–16  $\mu\text{atm}$ ; Table 2). This result shows the potential of computing  $p\text{CO}_2$  values with neural networks trained for other variables of the seawater CO<sub>2</sub> system, at least in some ocean regions. The global error of the  $p\text{CO}_2$  in the LDEO dataset is clearly higher than that obtained by

**Table 2.** RMSE (bias) by area and depth for TCO<sub>2</sub> and pCO<sub>2</sub>, computed with CANYON-B and NNGv2LDEO in Gv2QC and LDEO datasets. For each depth range, the RMSE (bias) in each area was weighted by the contribution of its data to the total. Units are micromoles per kilogram ( $\mu\text{mol kg}^{-1}$ ) for TCO<sub>2</sub> and microatmospheres ( $\mu\text{atm}$ ) for pCO<sub>2</sub>.

| Areas defined in<br>Takahashi et al. (2014) | TCO <sub>2</sub> |          |           |          |           |          |            |          |           |          | pCO <sub>2</sub> |          |         |         |
|---|------------------|----------|-----------|----------|-----------|----------|------------|----------|-----------|----------|------------------|----------|---------|---------|
|   | 0–50 m           |          | 50–200 m  |          | 200–500 m |          | 500–1000 m |          | > 1000 m  |          | 0 m              |          |         |         |
|   | LDEO (0 m)       |          | Gv2QC     |          | Gv2QC     |          | Gv2QC      |          | Gv2QC     |          | LDEO             |          |         |         |
|   | NNGv2LDEO        | CANYON-B | NNGv2LDEO | CANYON-B | NNGv2LDEO | CANYON-B | NNGv2LDEO  | CANYON-B | NNGv2LDEO | CANYON-B | NNGv2LDEO        | CANYON-B |         |         |
| West GIN Seas                               | 16.8             | 21.2     | 14.2      | 15.3     | 6.4       | 6.7      | 5.5        | 6.8      | 4.1       | 5        | 4                | 4.1      | 28.6    | 34.9    |
|   | (7.6)            | (18.1)   | (−1.7)    | (0.7)    | (−0.6)    | (0.3)    | (−1.1)     | (−0.5)   | (0.9)     | (1.3)    | (0.8)            | (0.8)    | (17.2)  | (35.2)  |
| East GIN Seas                               | 11.1             | 9.2      | 10.2      | 11.4     | 5.9       | 6.2      | 4.6        | 5.2      | 3.5       | 4        | 3.9              | 3.7      | 17.3    | 15      |
|   | (14.1)           | (5.6)    | (1.6)     | (1.3)    | (0.5)     | (2.9)    | (−0.8)     | (1.7)    | (−0.6)    | (2.3)    | (−0.6)           | (0.1)    | (22.4)  | (9.4)   |
| High Arctic                                 | 13.3             | 32.4     | 20.2      | 24.1     | 11.4      | 12.3     | 6.5        | 6.6      | 6.8       | 7.6      | 6.1              | 6.2      | 43      | 79.1    |
|   | (−8.3)           | (13.7)   | (−1.1)    | (4.4)    | (−0.1)    | (1.7)    | (0.1)      | (−0.2)   | (−0.1)    | (0.6)    | (−0.4)           | (−1.1)   | (−22.8) | (22.6)  |
| Beaufort Sea                                | 29.7             | 42       | 54        | 53.1     | 14.5      | 13.3     | 8.1        | 9.2      | 7.5       | 8.4      | 6.5              | 7.2      | 58.7    | 135.6   |
|   | (0.8)            | (10)     | (0.1)     | (1.3)    | (−1.7)    | (1.8)    | (0.1)      | (−1)     | (2)       | (−0.5)   | (1.1)            | (−2.1)   | (6.7)   | (5.3)   |
| Labrador Sea                                | 5.3              | 7.5      | 10.5      | 11.6     | 5.4       | 6.6      | 4.1        | 4.3      | 3.5       | 3.7      | 2.7              | 3.1      | 10.7    | 13.6    |
|   | (0.6)            | (2.7)    | (−0.9)    | (1.3)    | (−0.2)    | (1.6)    | (−0.6)     | (1.4)    | (−0.1)    | (1.5)    | (0.2)            | (1.6)    | (1)     | (4.5)   |
| Subarctic Atlantic                          | 11.7             | 14.6     | 9         | 11.1     | 4.5       | 5.4      | 4.3        | 4.8      | 4.2       | 4.6      | 4                | 4.3      | 21.9    | 26.7    |
|   | (3.3)            | (11.5)   | (−2.3)    | (1.5)    | (0.1)     | (1.9)    | (0.1)      | (0.3)    | (0.1)     | (0.4)    | (−0.2)           | (0)      | (7.8)   | (23.2)  |
| North Atlantic Drift                        | 11.1             | 12.9     | 13.3      | 14.5     | 9.9       | 10.4     | 7.8        | 7.8      | 4.3       | 4.5      | 3.5              | 3.6      | 21      | 24.7    |
|   | (0.5)            | (5.7)    | (−1.3)    | (0.5)    | (0.7)     | (0.5)    | (0)        | (−0.4)   | (0.2)     | (0.1)    | (0.3)            | (−0.1)   | (1.4)   | (10.3)  |
| Central Atlantic                            | 7.9              | 9.6      | 15.8      | 14.9     | 6.6       | 6.5      | 5.2        | 5.1      | 4.6       | 4.5      | 4.3              | 4.4      | 13.5    | 17.4    |
|   | (−0.3)           | (−1.2)   | (0)       | (0.3)    | (0.2)     | (1)      | (−0.5)     | (0.5)    | (−0.1)    | (0)      | (0)              | (0)      | (−0.3)  | (−2)    |
| South Atlantic Transition Zone              | 7.7              | 13.8     | 7.2       | 7.8      | 5.4       | 5.7      | 5.7        | 4.7      | 5         | 4.8      | 4.3              | 4.2      | 14.6    | 25      |
|   | (−1.3)           | (−2.1)   | (−0.9)    | (0.3)    | (0.8)     | (0.1)    | (1.1)      | (0.7)    | (−0.1)    | (1)      | (0.5)            | (0.3)    | (−2.7)  | (−5.2)  |
| Antarctic Atlantic                          | 11.8             | 19.2     | 8.6       | 10       | 4.6       | 5.4      | 3.5        | 3.8      | 3.1       | 3.1      | 3.1              | 3.1      | 25.6    | 41.5    |
|   | (1.4)            | (20.9)   | (−1.6)    | (0.5)    | (0)       | (0.4)    | (−0.1)     | (0.3)    | (0.1)     | (0.3)    | (0)              | (0.3)    | (4.9)   | (50.9)  |
| Kuroshio Alaska Gyre                        | 10.9             | 12.3     | 8.5       | 12.2     | 6.4       | 8.1      | 5          | 5.2      | 4.5       | 4.3      | 3.7              | 3.9      | 20.7    | 23.7    |
|   | (1.6)            | (1.5)    | (−0.9)    | (2.4)    | (0.9)     | (0.3)    | (0.3)      | (1.1)    | (0.6)     | (0.3)    | (0.4)            | (0.3)    | (3.7)   | (3.4)   |
| North Central Pacific                       | 26.3             | 34.5     | 9.6       | 15       | 6.8       | 8.3      | 4.2        | 4.7      | 4         | 4.1      | 3.4              | 3.8      | 46.7    | 56.6    |
|   | (−3.6)           | (−9.6)   | (0.2)     | (3.4)    | (0.5)     | (0.4)    | (0.3)      | (0.5)    | (−0.3)    | (0.3)    | (0.3)            | (0)      | (−0.4)  | (−7)    |
| Okhotsk Sea                                 | –                | –        | 23.1      | 16.4     | 11.3      | 6.8      | 6.3        | 5.1      | 5.2       | 3.4      | 4.1              | 3.5      | –       | –       |
|   | –                | –        | (0.9)     | (1.6)    | (−1.2)    | (−0.7)   | (−2.3)     | (−1.6)   | (−4)      | (−1.3)   | (1.2)            | (1.9)    | –       | –       |
| Central Tropical North Pacific              | 8                | 9.7      | 7.9       | 8.8      | 7.2       | 7.2      | 4.9        | 5        | 4.3       | 4.5      | 3.6              | 3.8      | 14.2    | 17.2    |
|   | (−1.3)           | (−3.2)   | (−0.9)    | (0.5)    | (0.5)     | (1.2)    | (−0.6)     | (0.2)    | (−0.4)    | (0.5)    | (−0.2)           | (0.2)    | (−2.1)  | (−5.6)  |
| Tropical East North Pacific                 | 11.1             | 14.5     | 10.9      | 13.8     | 5.9       | 8.5      | 2.6        | 3.4      | 2.1       | 2.1      | 2.4              | 2.1      | 20.8    | 28.2    |
|   | (−0.1)           | (−4.5)   | (0.9)     | (−1.4)   | (0.3)     | (2.5)    | (0.3)      | (1.7)    | (0)       | (0.6)    | (−0.2)           | (−0.3)   | (0.2)   | (−8.8)  |
| Panama Basin                                | 12.5             | 17.4     | 10.2      | 9.5      | 6.5       | 3.9      | 4          | 5.8      | 3.8       | 3.2      | 4.2              | 4.3      | 25.8    | 38.8    |
|   | (−0.7)           | (0.7)    | (−3.4)    | (1.5)    | (−2.7)    | (−6.7)   | (2.3)      | (−1.2)   | (0.8)     | (1.4)    | (0)              | (2.8)    | (−0.3)  | (1.3)   |
| Central South Pacific                       | 10.1             | 12.9     | 10.3      | 10.9     | 8.9       | 9.4      | 4.4        | 4.5      | 3.8       | 3.8      | 3.3              | 3.5      | 18.6    | 24.3    |
|   | (−2.1)           | (−3.4)   | (1.2)     | (−0.7)   | (0)       | (0.2)    | (−0.1)     | (0.8)    | (−0.1)    | (−0.4)   | (0)              | (−0.1)   | (−3)    | (−5)    |
| East Central South Pacific                  | 10.7             | 15.4     | 10.6      | 15.2     | 6.9       | 7.5      | 4.1        | 2.8      | 3.8       | 3.5      | 3.3              | 3        | 24.1    | 34.2    |
|   | (−1)             | (−0.1)   | (0.6)     | (1.4)    | (1.2)     | (1.3)    | (0.4)      | (0.1)    | (−0.5)    | (−0.2)   | (−0.6)           | (0.3)    | (−1.2)  | (0.5)   |
| Subpolar South Pacific                      | 6.9              | 7.9      | 5.8       | 7.7      | 5         | 5.4      | 2.9        | 2.8      | 4.7       | 4.8      | 4.1              | 4.4      | 13.7    | 16      |
|   | (1.2)            | (−0.9)   | (−0.7)    | (2.4)    | (0.4)     | (0.6)    | (0.9)      | (1.7)    | (−1)      | (1.2)    | (0.7)            | (1)      | (2.4)   | (−2.2)  |
| Antarctic Pacific                           | 19.5             | 29.3     | 8.3       | 7        | 3.6       | 4.1      | 2.7        | 3.3      | 2.6       | 2.9      | 2.6              | 2.1      | 34.1    | 53.8    |
|   | (1.9)            | (4.9)    | (−1.4)    | (0.5)    | (−0.5)    | (0.5)    | (−0.1)     | (0.7)    | (0.3)     | (0.3)    | (0.5)            | (−0.1)   | (7.7)   | (14.9)  |
| Main North Indian                           | 10.8             | 13       | 10.5      | 12.9     | 8.1       | 7.8      | 3.2        | 3.3      | 2.4       | 2.6      | 3.1              | 3.7      | 19.8    | 23.5    |
|   | (−1.8)           | (−7.7)   | (2.8)     | (1.7)    | (−0.1)    | (0.5)    | (0.7)      | (0.9)    | (−0.4)    | (−0.4)   | (−0.2)           | (0.2)    | (−2.8)  | (−12.7) |
| Red Sea                                     | 18.3             | 20.9     | 12        | 16.8     | 9.4       | 8.7      | 7.6        | 7.9      | 7.4       | 5.7      | 3.3              | 7.2      | 28      | 32.3    |
|   | (−13.9)          | (−16.7)  | (−4.3)    | (−3.5)   | (0.2)     | (−3.7)   | (−4.3)     | (−5.1)   | (1)       | (−4.4)   | (−1.3)           | (−1.1)   | (−21)   | (−25.5) |
| Bengal Basin                                | 6                | 3.7      | 9.8       | 7.4      | 7.4       | 6.4      | 2          | 2.1      | 1.9       | 2.1      | 2                | 2.2      | 10.7    | 6.7     |
|   | (1.1)            | (−5.5)   | (−0.2)    | (2)      | (0.3)     | (1.3)    | (0.4)      | (1.1)    | (0.4)     | (−0.2)   | (−0.6)           | (−0.4)   | (1.4)   | (−10.2) |
| Main South Indian                           | 8.1              | 10.4     | 9.1       | 10       | 7.1       | 6.9      | 3.8        | 3.8      | 4.2       | 4.5      | 3.4              | 3.8      | 14      | 17.7    |
|   | (−0.2)           | (0.4)    | (−0.1)    | (1)      | (0.3)     | (0.6)    | (−0.1)     | (0.9)    | (0.1)     | (1.1)    | (0.1)            | (0.1)    | (−0.1)  | (0.9)   |
| South Indian Transition                     | 8                | 9.4      | 5.3       | 5.3      | 4.1       | 4.4      | 4          | 3.7      | 3.9       | 3.4      | 3.4              | 3.7      | 16      | 18.8    |
|   | (−0.4)           | (0.7)    | (−1.7)    | (−0.4)   | (0.8)     | (0.8)    | (0)        | (0.5)    | (−0.2)    | (0.3)    | (−0.6)           | (−0.9)   | (−0.7)  | (1.9)   |
| Antarctic Indian                            | 9.7              | 11.8     | 6         | 6.8      | 4.1       | 4.7      | 3          | 3.3      | 2.5       | 2.6      | 2.7              | 2.6      | 23.2    | 28.7    |
|   | (1.1)            | (5.9)    | (−1.4)    | (0.2)    | (0.4)     | (0.6)    | (0.1)      | (0.3)    | (−0.3)    | (0.3)    | (0.3)            | (0.5)    | (3.9)   | (15.3)  |
| Circumpolar Southern Ocean                  | 15.9             | 24.5     | 7.6       | 8.2      | 4.4       | 4.9      | 3.1        | 3.4      | 2.9       | 2.9      | 2.9              | 2.7      | 30.7    | 48.9    |
|   | (0.9)            | (9.8)    | (−1.6)    | (0.4)    | (−0.2)    | (0.3)    | (−0.1)     | (0.2)    | (0)       | (0.2)    | (0.1)            | (0.2)    | (4.7)   | (25.7)  |
| Weighted                                    | 11.1             | 15       | 11        | 12.1     | 6.6       | 7        | 4.3        | 4.5      | 4         | 4.1      | 3.5              | 3.7      | 21.1    | 29.6    |
|   | (0)              | (2.8)    | (−0.5)    | (0.8)    | (0.2)     | (0.8)    | (−0.1)     | (0.5)    | (−0.1)    | (0.3)    | (0.1)            | (0.2)    | (1.3)   | (7.5)   |





**Figure 4.** Histograms and box plots of differences between measured and neural-network-computed TCO<sub>2</sub> in (a) Gv2 and (b) LDEO. \* TCO<sub>2</sub> computed from measured  $p\text{CO}_2$  and neural-network-derived  $A_T$ .

Landschützer et al. (2014) for the SOCAT dataset (22 vs. 12  $\mu\text{atm}$ , respectively), although the critical areas are mainly the same (Fig. S4): equatorial Pacific upwelling system, Arctic and subarctic waters around the Alaska Peninsula, the Southern Ocean, the Gulf Stream, and the North Atlantic Current. At this point, the following should be considered: (1) the  $p\text{CO}_2$  computed in the present study is derived from  $A_T$  and TCO<sub>2</sub> and not from specific modeling for  $p\text{CO}_2$ , and therefore it contains errors associated with this computation ( $\sim 6 \mu\text{atm}$ ; Millero, 1995) and the neural network estimates of  $A_T$  and TCO<sub>2</sub>; (2) the present study includes the Arc-

tic region where the highest errors occur (Table 2; Beaufort Sea and High Arctic areas); and (3) there is a longer temporal range in the present study (1973–2016). The analysis of Landschützer et al. (2014) in the LDEO dataset for data that differ from SOCAT shows a global error higher than the one obtained in the present study for all LDEO data between 1998 and 2011 (25.9 vs. 21.3  $\mu\text{atm}$ , respectively). The error between 40° S and 40° N is similar in the two studies (Landschützer et al., 2014: 16.5  $\mu\text{atm}$ ; NNGv2LDEO: 16.4  $\mu\text{atm}$ ). Although it is not the main objective of this work, these last two results show how NNGv2LDEO and NNGv2 (Broullón

et al., 2019) have the potential to compute  $p\text{CO}_2$  values between 40° S and 40° N with similar errors as the method with the lower error in the  $p\text{CO}_2$  modeling to obtain a climatology and with lower errors in high latitudes for the LDEO dataset, even taking into account the inclusion of the critical area of the Arctic in the computation of the error of the  $p\text{CO}_2$  from the present study (it is not included in Landschützer et al., 2014) and the higher number of data from high latitudes in the present study (15 479 vs. 3799).

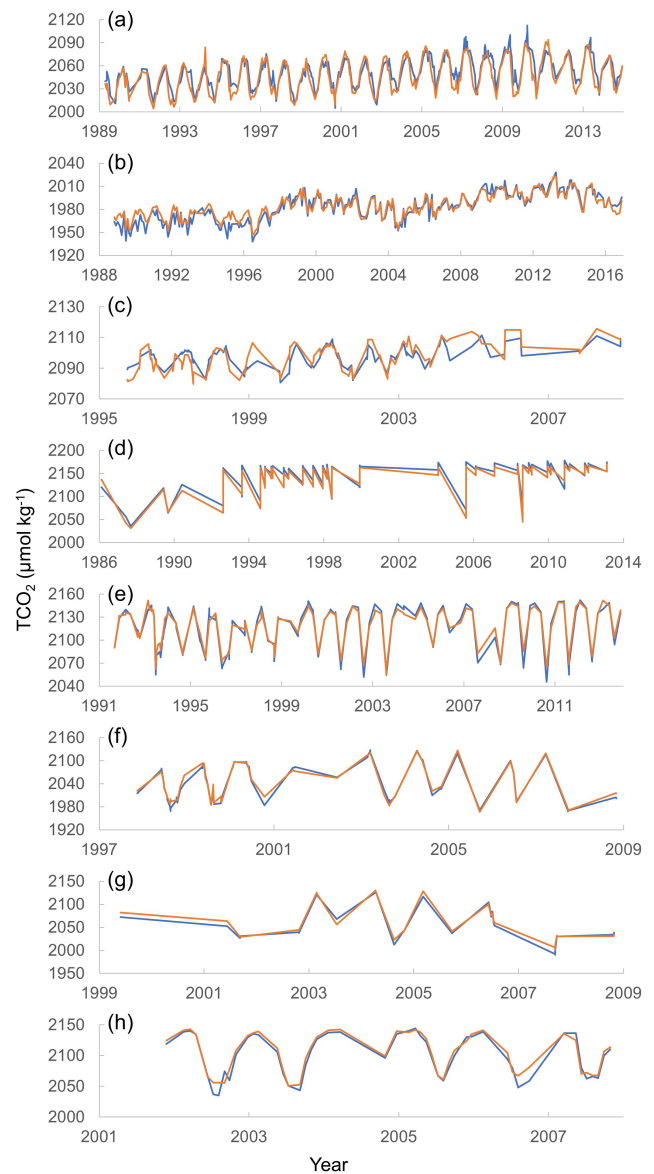
### 3.2 Time series validation

The good generalization of the network in the test dataset containing data from Gv2 and the LDEO dataset with a similar RMSE as the one reached in the training set is also evidenced through independent time series data (Table 3). Except for KERFIX, where the number of data points is very low and Olsen et al. (2019) suggested an adjustment to the original data of  $-39 \mu\text{mol kg}^{-1}$ , TCO<sub>2</sub> computed using NNGv2LDEO and CANYON-B at the time series locations is characterized by low errors and biases (Table 3). NNGv2LDEO computes TCO<sub>2</sub> with a lower RMSE and bias than CANYON-B for most of the time series stations (Table 3). CANYON-B reaches a lower RMSE in HOT ALOHA SURFACE and ESTOC than NNGv2LDEO, but the bias is considerably higher in these time series for CANYON-B.

The seasonal variability is well captured by NNGv2LDEO, showing its great potential to design a monthly climatology. In the surface layer, where the seasonal variability is the highest, the computed values are strongly correlated with the measured TCO<sub>2</sub> in all the time series (Fig. 5). In addition, the high correlation holds for all depths (Table S1). The TCO<sub>2</sub> computation with a low error in these time series located in different oceanic regimes as well as in the areas of Table 2 shows the good performance of NNGv2LDEO in almost any region of the ocean.

Assessing the potential of neural networks to obtain values of other variables of the seawater CO<sub>2</sub> system in the time series,  $p\text{CO}_2$  calculated with  $A_T$  from NNGv2 (Broullón et al., 2019) and TCO<sub>2</sub> from NNGv2LDEO compared quite well with  $p\text{CO}_2$  as measured or calculated from  $A_T$  and TCO<sub>2</sub> at the time series stations (Table 4). Except for BATS, the  $p\text{CO}_2$  obtained in the present study has a lower error than that reported by Landschützer et al. (2014; Table 4). In contrast, the bias in the present study is higher, except for ESTOC. Considering the error involved in the calculation of  $p\text{CO}_2$  from  $A_T$  and TCO<sub>2</sub> ( $\sim 6 \mu\text{atm}$ ; Millero, 1995), and the error in the computed  $A_T$  and TCO<sub>2</sub> with the neural networks (Table 4), our results demonstrate again the ability of NNGv2 and NNGv2LDEO to calculate other variables of the seawater CO<sub>2</sub> system with a relatively low error.

Using NNGv2LDEO, it is also possible to reproduce the secular trends in TCO<sub>2</sub>. Using seasonal detrending to enhance the multiannual changes, similar trends in the longer time series are found for the measured TCO<sub>2</sub> and the neural-



**Figure 5.** Measured (blue line) and computed (orange line) TCO<sub>2</sub> with NNGv2LDEO for the depth range 0–15 m (0–30 m in panel b) for several time series. (a) BATS, (b) HOT ALOHA SURFACE, (c) ESTOC, (d) ICELAND, (e) IRMINGER, (f) KNOT, (g) K2 and (h) OWS.

network-computed TCO<sub>2</sub> (Table 5). The same holds for  $p\text{CO}_2$  (Table 5), although at the IRMINGER site the trend obtained from the neural-network-generated data is significantly lower than that from measured data. The neural networks seem to capture the anthropogenic influence in the seawater CO<sub>2</sub> system and thus the ocean acidification process (Fig. 6). Furthermore, using NNGv2LDEO increases the number of TCO<sub>2</sub> data where the various inputs were measured but not TCO<sub>2</sub> itself. This allows for the evaluation of high-frequency changes (Fig. 6) and for the calculation of

**Table 3.** RMSE and bias between measured and computed TCO<sub>2</sub> concentrations in several time series. The comparison was done using only water samples where all the input variables for NNGv2LDEO and the TCO<sub>2</sub> were measured in the same water sample.

| Time series    | Location         | Time period | <i>n</i> | NNGv2LDEO                           |                                     | CANYON-B                            |                                     |
|----------------|------------------|-------------|----------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
|                |                  |             |          | RMSE<br>( $\mu\text{mol kg}^{-1}$ ) | Bias<br>( $\mu\text{mol kg}^{-1}$ ) | RMSE<br>( $\mu\text{mol kg}^{-1}$ ) | Bias<br>( $\mu\text{mol kg}^{-1}$ ) |
| BATS           | 31.7° N, 64.2° W | 1988–2014   | 4121     | 7.7                                 | 0.1                                 | 7.7                                 | −0.6                                |
| HOT ALOHA      | 22.8° N, 158° W  | 1988–2017   | 4054     | 5.4                                 | −0.5                                | 5.1                                 | −2                                  |
| HOT ALOHA SURF | 22.8° N, 158° W  | 1988–2016   | 281      | 6.3                                 | −1.6                                | 5.8                                 | −5.1                                |
| ESTOC          | 29.3° N, 15.5° W | 1995–2008   | 1697     | 7.1                                 | 0.8                                 | 6.6                                 | 4.7                                 |
| ICELAND        | 68° N, 12.7° W   | 1985–2013   | 1322     | 5.4                                 | 5.6                                 | 6.9                                 | 5.3                                 |
| IRMINGER       | 64.3° N, 28° W   | 1991–2013   | 1086     | 4.8                                 | 3.3                                 | 7.5                                 | 6.6                                 |
| K2             | 47° N, 160° E    | 1999–2008   | 615      | 3.6                                 | 1.3                                 | 6.3                                 | 2.4                                 |
| KNOT           | 44° N, 155° E    | 1997–2008   | 1321     | 5.8                                 | −0.8                                | 7.2                                 | −1.9                                |
| OWS            | 66° N, 2° E      | 2001–2007   | 803      | 6.8                                 | −1                                  | 10.5                                | −4.7                                |
| KERFIX         | 50.4° S, 68.2° E | 1992–1994   | 38       | 13.2                                | 26.4                                | 13.1                                | 28.9                                |

**Table 4.** RMSE and bias between measured *p*CO<sub>2</sub> (and in some cases, computed from measured *A*<sub>T</sub> and TCO<sub>2</sub> in time series where *p*CO<sub>2</sub> was not measured) and computed *p*CO<sub>2</sub> with *A*<sub>T</sub> from NNGv2 (Broullón et al., 2019) and TCO<sub>2</sub> from NNGv2LDEO in several time series. The time period for *p*CO<sub>2</sub> from this study is the same as in Table 3. Consult Table 2 in Landschützer et al. (2014) for its time period. The depth range is 0–15 m. Only time series with more than 30 data points are included. RMSE and bias for computed *A*<sub>T</sub> with NNGv2 (Broullón et al., 2019) and TCO<sub>2</sub> with NNGv2LDEO are included to show the errors in the variables used to compute TCO<sub>2</sub>.

| Time series    | <i>p</i> CO <sub>2</sub>    |                             |                             |                             | <i>A</i> <sub>T</sub>         |                             | TCO <sub>2</sub>            |                             |
|----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------------|-----------------------------|-----------------------------|-----------------------------|
|                | NNGv2LDEO                   |                             | Landschützer et al. (2014)  |                             | NNGv2 (Broullón et al., 2019) |                             | NNGv2LDEO                   |                             |
|                | RMSE<br>( $\mu\text{atm}$ ) | Bias<br>( $\mu\text{atm}$ ) | RMSE<br>( $\mu\text{atm}$ ) | Bias<br>( $\mu\text{atm}$ ) | RMSE<br>( $\mu\text{atm}$ )   | Bias<br>( $\mu\text{atm}$ ) | RMSE<br>( $\mu\text{atm}$ ) | Bias<br>( $\mu\text{atm}$ ) |
| BATS           | 17.2                        | 9.7                         | 15.6                        | 0.4                         | 5.6                           | −1.7                        | 10.1                        | 4.4                         |
| HOT ALOHA SURF | 10.3                        | −3.6                        | 11.6                        | 0.1                         | 5.0                           | 0.9                         | 6.5                         | −1.6                        |
| ESTOC          | 10.6                        | 2.7                         | 14.5                        | −7.1                        | 2.6                           | −2.7                        | 5.3                         | −0.6                        |
| ICELAND        | 16                          | 14.8                        | –                           | –                           | 5.4                           | 0.7                         | 5.4                         | 5.4                         |
| IRMINGER       | 13.1                        | −1.8                        | 22.6                        | −1.1                        | 7.0                           | −0.4                        | 6.6                         | −1.1                        |
| K2             | 18.1                        | −3.2                        | 27.8                        | −0.2                        | 5.1                           | −0.5                        | 5.7                         | −2.4                        |
| KNOT           | 20.8                        | 8.6                         | –                           | –                           | 6.6                           | −7.3                        | 8.2                         | −2.5                        |

interannual trends with a low error (as temporal sampling biases are reduced).

### 3.3 Climatology

Using NNGv2LDEO we have demonstrated its ability to compute TCO<sub>2</sub> values with low errors and especially to capture the monthly variability of this variable. In addition, the climatologies of the input variables used to create the climatology of TCO<sub>2</sub> have been satisfactorily evaluated previously for the construction of an *A*<sub>T</sub> climatology (Broullón et al., 2019). Considering these results, a monthly climatology of TCO<sub>2</sub> is obtained by passing the input climatologies through NNGv2LDEO.

The spatial distribution of the surface annual mean climatology of TCO<sub>2</sub> (Fig. 7a) is similar to two recent climatologies: those of Takahashi et al. (2014) and Lauvset et al. (2016). The largest surface TCO<sub>2</sub> concentra-

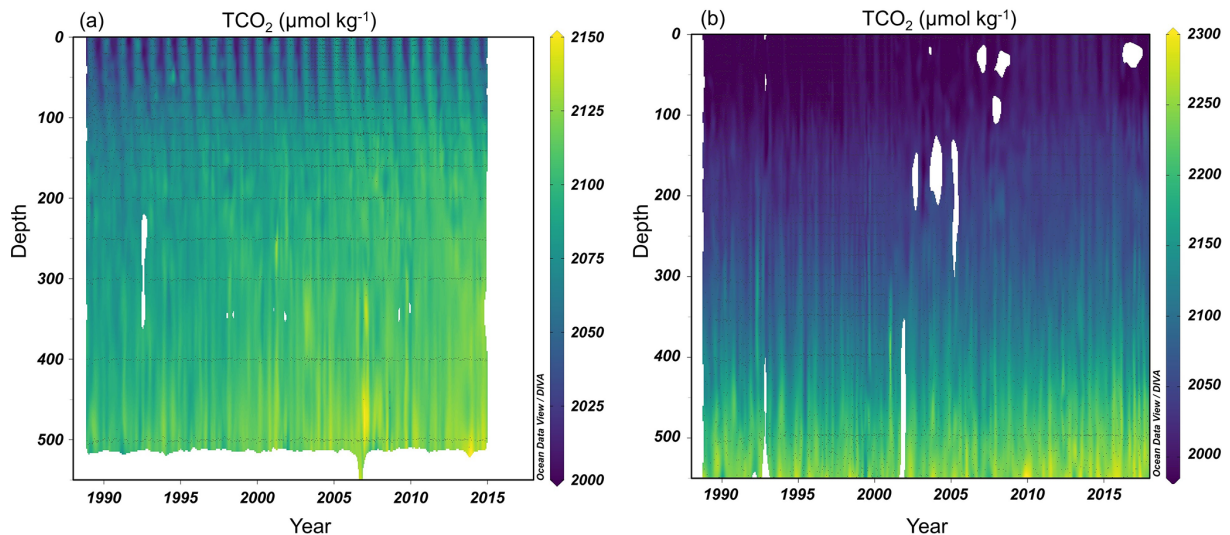
tions occur in the Southern Ocean, subpolar North Atlantic, Nordic Seas, and Mediterranean Sea (note that the latter is not included in these other climatologies). In general, surface TCO<sub>2</sub> decreases from high to low latitudes. The Indian and the Pacific oceans are characterized by lower concentrations of TCO<sub>2</sub> at higher latitudes than the Atlantic, the latter being the ocean with the highest surface TCO<sub>2</sub> by area. TCO<sub>2</sub> increases with depth in all oceans, in particular in the upwelling regions, where this increase is expanded eastwards with depth (Fig. 7b and video at <https://doi.org/10.20350/digitalCSIC/10551>, Broullón et al., 2020). Depending on the area, the values reach a maximum at certain intermediate depths, and below it the concentration gradually decreases or remains almost constant (Fig. S5).

The largest seasonal variability occurs at the surface at high latitudes, in the Pacific upwelling region, the equatorial African coasts, and in the area under influence of the Amazon River (Fig. 8a). At depth, the seasonal variability de-

**Table 5.** Long-term trends (seasonally detrended) of the measured and computed TCO<sub>2</sub> and pCO<sub>2</sub> from neural networks at time series locations in the depth range 0–15 m.

| Time series    | TCO <sub>2</sub> (μmol kg <sup>-1</sup> yr <sup>-1</sup> ) |          | pCO <sub>2</sub> (μatm yr <sup>-1</sup> ) |          |
|----------------|--|----------|---|----------|
|                | Measured   | Computed | Measured*                                 | Computed |
| BATS           | 1.2  | 1.1      | 1.8                                       | 1.7      |
| HOT ALOHA SURF | 1.7  | 1.3      | 1.8                                       | 1.4      |
| ICELAND        | 0.9  | 0.9      | 1.5                                       | 1.6      |
| IRMINGER       | 0.6  | 0.5      | 2.5                                       | 1.7      |

\* Computed from measured A<sub>T</sub> and TCO<sub>2</sub> in time series where pCO<sub>2</sub> was not measured.

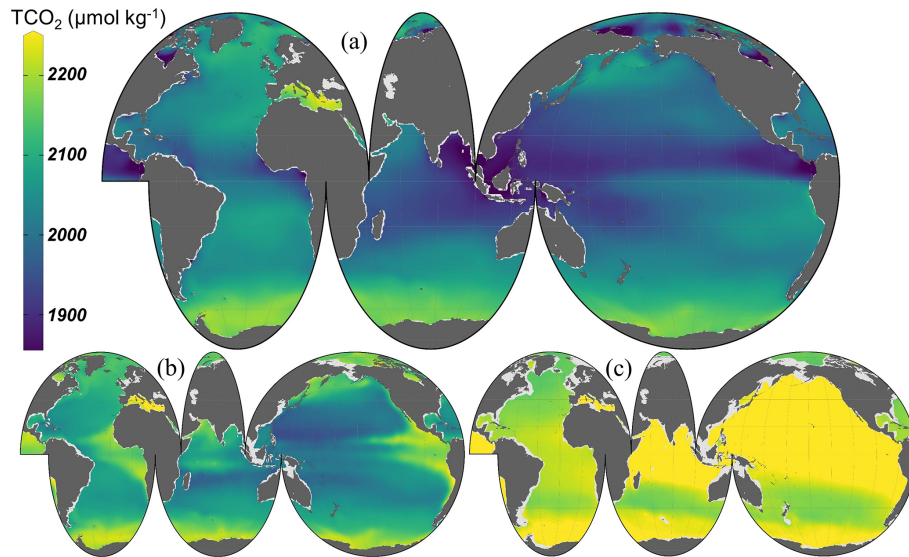
**Figure 6.** Time series of TCO<sub>2</sub> using NNGv2LDEO at (a) BATS and (b) HOT ALOHA locations. The water column shows a higher concentration of TCO<sub>2</sub> year by year. This figure was made with Ocean Data View (Schlitzer, 2016).

increases, except for the Pacific upwelling region, where it increases and moves progressively northward between 30 and 150 m (Fig. 8b). This increase is correlated with the high seasonal variability of the climatologies of nutrients, oxygen, and temperature at these depths. Czeschel et al. (2012) also showed an increase in the subsurface variability of oxygen from measured profiles. Similar increases also occur in the Indian Ocean north of 20° S between 50 and 100 m and in the equatorial Atlantic Ocean in the same depth range. At the 1500 m level, the seasonal variability is below 10 μmol kg<sup>-1</sup> in most of the ocean (Fig. 8c). This last result shows that an annual climatology below 1500 m is sufficient.

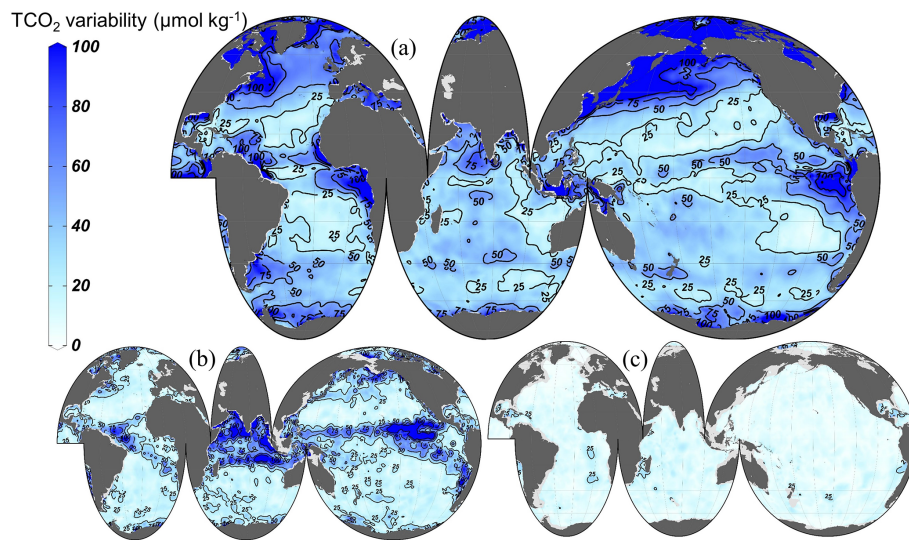
Although the surface patterns of the annual mean of the TCO<sub>2</sub> climatology are very similar to those of the other recent climatologies (Takahashi et al., 2014; Lauvset et al., 2016), differences do occur. The annual mean climatology of the present study is closest to that of Takahashi et al. (2014; Table 6). The largest differences between these two climatologies are located in the Arctic, North Pacific, Peru upwelling area, western South Pacific, and the area of influence of the Antarctic Circumpolar Current (Fig. S6a). The

Atlantic and the Indian oceans do not show significant differences. Our climatology shows more deviations from that of Lauvset et al. (2016), compared in the grid of Takahashi et al. (2014; Table 6). The highest differences are found in the North Pacific, around Antarctica, the Nordic Seas, the South and North Atlantic, and in several less localized areas around the oceans (Fig. S6b). When the climatology of Takahashi et al. (2014) is compared to that of Lauvset et al. (2016), the differences are even higher (Table 6), and the critical areas are the same of those of the previous comparison. Although it is clear that discrepancies between the three climatologies are derived from the different methods used, the higher similarity between ours and the one of Takahashi et al. (2014) is probably due to the influence of the same source used to create them, the World Ocean Atlas.

The comparison of our climatology with that of Lauvset et al. (2016) at the 33 depth levels of Lauvset et al. (2016) shows a reduction in the RMSE with depth. Between 0 and 1000 m, the RMSE is reduced from ~ 32 to 7 μmol kg<sup>-1</sup> (Table S2; note the higher RMSE at surface compared to the one obtained for the grid of Takahashi et al., 2014, because of



**Figure 7.** Annual mean climatology of TCO<sub>2</sub> at (a) 0 m, (b) 100 m, and (c) 1000 m. This figure was made with Ocean Data View (Schlitzer, 2016).



**Figure 8.** Seasonal amplitude of TCO<sub>2</sub> at (a) 0 m, (b) 100 m, and (c) 1500 m. The contour lines of 25, 50, 75, and 100  $\mu\text{mol kg}^{-1}$  are shown. This figure was made with Ocean Data View (Schlitzer, 2016).

the inclusion of areas which are not included in the latter's grid, and the difficulty of modeling TCO<sub>2</sub> in some areas, like the Arctic and the Mediterranean Sea). This reduction with depth is probably due to the reduction in the variability in most of the ocean below the surface. The surface values in Lauvset et al. (2016) are likely characteristic of months in which most of the sampling was carried out. Because of the lower variability of TCO<sub>2</sub> at depth, the values are closer to the annual mean, and therefore the two compared climatologies are more similar at depth than at surface depth levels. Below 1000 m, the differences between the two climatologies

are not significant, with an RMSE around  $5 \mu\text{mol kg}^{-1}$  and a bias around  $0.5 \mu\text{mol kg}^{-1}$  at each depth level (Table S2).

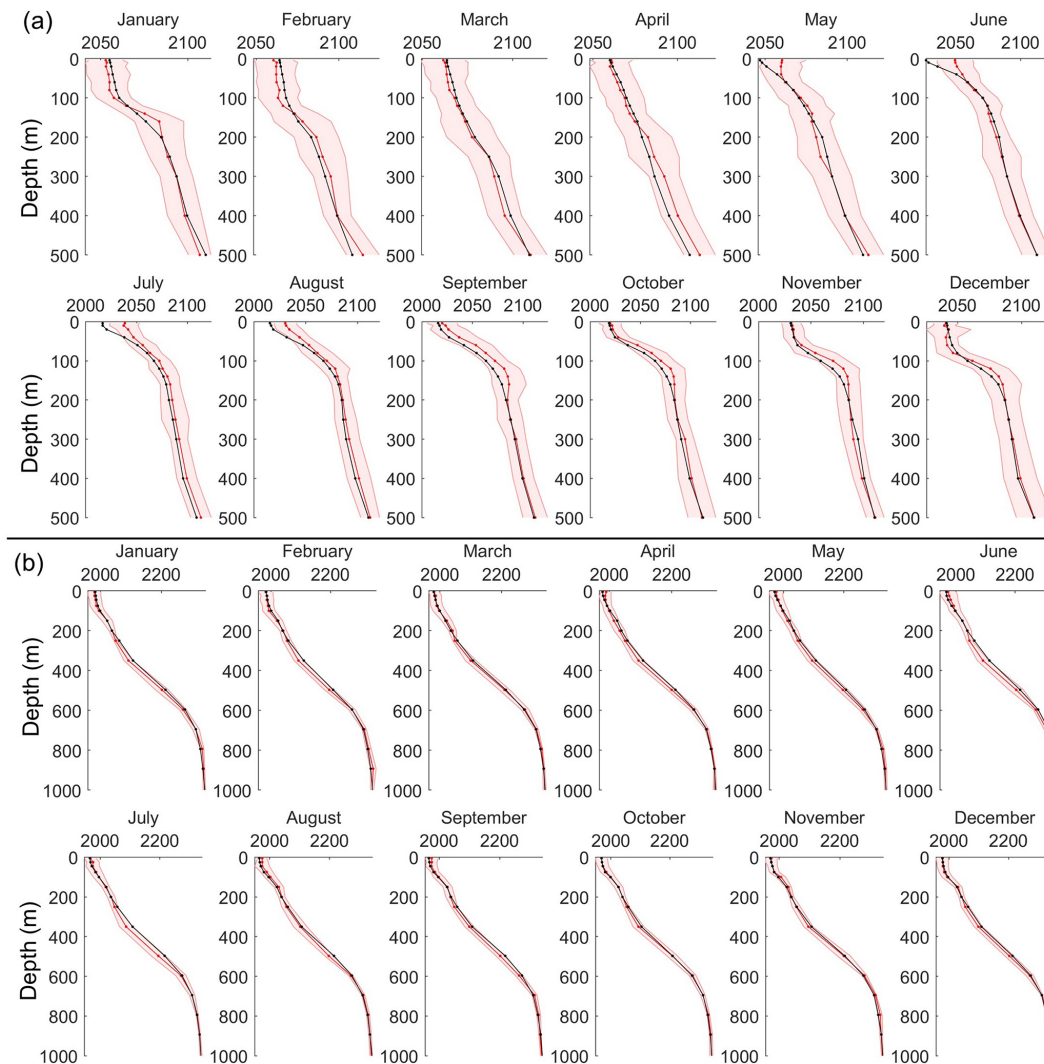
Our monthly climatology shows a high correspondence with that of Takahashi et al. (2014), although the RMSE values show that there are also large differences in certain areas (Table 7). These areas are mainly the same as those in the comparison of the annual mean climatologies, but some other small regions with high differences appear for each month all through the ocean (Fig. S7).

Unfortunately, the uncertainty of the TCO<sub>2</sub> climatology cannot be assessed globally and robustly. As Broullón et al. (2019) stated, the unavailability of an uncertainty field

**Table 6.** Comparison of four annual mean surface climatologies of TCO<sub>2</sub>. Numbers in the lower left corner represent RMSE. Numbers in the upper right corner represent  $r^2$ .

| RMSE ( $\mu\text{mol kg}^{-1}$ ) – $r^2$ | NNGv2LDEO | Lauvset et al. (2016)* | Takahashi et al. (2014) |
|--|-----------|------------------------|-------------------------|
| NNGv2LDEO                                | –         | 0.93                   | 0.97                    |
| Lauvset et al. (2016)*                   | 19.8      | –                      | 0.90                    |
| Takahashi et al. (2014)                  | 13.2      | 23.7                   | –                       |

\* The domain analyzed is the same as in Takahashi et al. (2014) for coherency reasons.



**Figure 9.** Comparison of the monthly climatological profiles of TCO<sub>2</sub> computed from measured data (red profile; shadow area is the SD of the averaged values at each depth level) and those from the TCO<sub>2</sub> climatology at the (a) BATS and (b) HOT ALOHA locations. The units on the  $x$  axis are micromoles per kilogram ( $\mu\text{mol kg}^{-1}$ ).

associated with the objectively analyzed WOA13 climatologies does not allow us to perform a proper global uncertainty assessment. Therefore, the analysis is relegated to the areas where repeated sampling of TCO<sub>2</sub> has been carried out monthly over a long period, that is, the HOT ALOHA and BATS time series stations. The climatology of TCO<sub>2</sub>

from NNGv2LDEO is consistent with the monthly climatological values at these two places (Fig. 9). In general, the profiles from the TCO<sub>2</sub> climatology are within the variability range (shadow area in Fig. 9) of the monthly averaged measured data for each depth level. In the upper 30 m of the water column, the climatology of TCO<sub>2</sub> differs from

**Table 7.** Comparison of the monthly TCO<sub>2</sub> climatology of Takahashi et al. (2014) and the one of the present study.

| Month     | RMSE<br>( $\mu\text{mol kg}^{-1}$ ) | Bias<br>( $\mu\text{mol kg}^{-1}$ ) | $r^2$ |
|-----------|-------------------------------------|-------------------------------------|-------|
| January   | 16                                  | 3.0                                 | 0.95  |
| February  | 16.7                                | 1.5                                 | 0.94  |
| March     | 15.8                                | 2.5                                 | 0.95  |
| April     | 17                                  | 2.6                                 | 0.95  |
| May       | 16.8                                | 2.5                                 | 0.95  |
| June      | 17.2                                | 3.2                                 | 0.95  |
| July      | 22.6                                | 4.0                                 | 0.92  |
| August    | 17.8                                | 3.4                                 | 0.95  |
| September | 15.5                                | 2.5                                 | 0.97  |
| October   | 15.6                                | 2.3                                 | 0.96  |
| November  | 15.7                                | 2.7                                 | 0.96  |
| December  | 17.6                                | 4.3                                 | 0.95  |

the measured BATS data from May to August. This difference is mainly explained by the surface error of the network shown for this time series in Fig. 5a, where the computed TCO<sub>2</sub> decreases from maximum to minimum sooner than the measured TCO<sub>2</sub>. For HOT ALOHA, the RMSE of the profiles of the TCO<sub>2</sub> climatology oscillates between 3.6 and 9.2  $\mu\text{mol kg}^{-1}$  with a mean value of 6.3  $\mu\text{mol kg}^{-1}$  (bias range:  $-3.8$  to 1.2  $\mu\text{mol kg}^{-1}$ ; mean bias:  $-1.4$   $\mu\text{mol kg}^{-1}$ ). At BATS, the RMSE is lower than for HOT ALOHA: 1.1 to 8.5  $\mu\text{mol kg}^{-1}$  with a mean value of 4.4  $\mu\text{mol kg}^{-1}$  (bias range:  $-1.4$  to 7.6  $\mu\text{mol kg}^{-1}$ ; mean bias:  $-2.5$   $\mu\text{mol kg}^{-1}$ ). Furthermore, the seasonal variability of the TCO<sub>2</sub> climatology is quite similar to that of the measured data at BATS and HOT ALOHA (Fig. S8). Although in other time series there are not enough measured data to obtain climatological values, these pseudoclimatological values also correlate very well with the TCO<sub>2</sub> climatology (data not shown). These results suggest that the climatology is robust in different oceanographic regimes and adequately captures the seasonal cycle of TCO<sub>2</sub>.

It has been demonstrated in this study that  $p\text{CO}_2$  and possibly other variables of the seawater CO<sub>2</sub> system can be computed from  $A_T$  and TCO<sub>2</sub> derived from neural networks with a relatively low error in different datasets (LDEO in Sect. 3.1 and time series in Sect. 3.2). The  $p\text{CO}_2$  climatology (Fig. S9) computed from the TCO<sub>2</sub> climatology of the present study and the  $A_T$  climatology of Broullón et al. (2019) is very similar to that of Landschützer et al. (2017). The differences between the annual mean climatology of the two studies are below 15  $\mu\text{atm}$  in most of the ocean (RMSE: 8.3  $\mu\text{atm}$ ; bias: 2.9  $\mu\text{atm}$ ;  $r^2$ : 0.82). The differences above this threshold are mainly located in the Pacific equatorial upwelling system, the eastern part of the South Pacific Gyre, Nordic Seas, Labrador Sea, Atlantic section of the Southern Ocean, Bay of Bengal, and the waters surrounding the eastern margin of Asia (Fig. S10). In most of these areas, both methods have the

greatest errors (Figs. 2 and 4 in Landschützer et al., 2014, and Fig. S4 of the present study).

On a monthly basis, the RMSE between the two  $p\text{CO}_2$  climatologies is between 13.6 and 15.6  $\mu\text{atm}$ , and the correlation is lower than for the annual mean comparison ( $r^2$ : 0.55–0.72 vs. 0.82). The areas with the higher differences are the same as in the annual comparison, but other small regions appear along the ocean month by month (Fig. S11). Furthermore, the seasonal variability in the two climatologies matches in a great extension of the ocean, although there are areas with notable differences (Fig. S12). In general, the  $p\text{CO}_2$  climatology is quite similar to that of Landschützer et al. (2017), and this result helps show that both the TCO<sub>2</sub> climatology of the present study and the  $A_T$  climatology of Broullón et al. (2019) are mostly robust and suggest that climatologies of other seawater CO<sub>2</sub> system variables can be confidently computed.

#### 4 Data availability

The climatologies of TCO<sub>2</sub> and  $p\text{CO}_2$  as well as NNGv2LDEO designed in this study are available at the data repository of the Spanish National Research Council (CSIC; <https://doi.org/10.20350/digitalCSIC/10551>; Broullón et al., 2020).

#### 5 Conclusions

We presented a tool for computing TCO<sub>2</sub> in the global ocean. Compared to previous methods, the uncertainties in such computations have been reduced. Including two updated datasets containing thousands of measurements of inorganic carbon variables across the ocean in the training of the neural network, we were able to capture a wide range of variability of TCO<sub>2</sub>. The low errors obtained in independent subsets at time series stations are further evidence of the potential of the network in computing TCO<sub>2</sub>.

Our global monthly climatology created with a neural network is the first that covers the oceans from the surface to the abyss at such temporal resolution. In addition to the accuracy of the network, the low uncertainty of the climatology in different regions and its usefulness in creating climatologies of other seawater CO<sub>2</sub> chemistry variables (i.e.,  $p\text{CO}_2$ ) show its robustness. Therefore, we present the global climatology of TCO<sub>2</sub> to the scientific community to complement the recently designed climatology of  $A_T$  by Broullón et al. (2019) for its use in the initialization and evaluation of models or any other analysis related to the carbon cycle.

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/essd-12-1725-2020-supplement>.

**Author contributions.** DB, FFP, and AV designed the study. The manuscript was written by DB and revised and discussed by all the authors. The dataset of the climatology and the neural network were created by DB.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** The authors want to thank the three referees, whose comments improved the study. The paper is dedicated to Taro Takahashi, who contributed immensely to the understanding of the ocean's role in the carbon cycle and passed away in December 2019.

**Financial support.** This research was supported by Ministerio de Educación, Cultura y Deporte (FPU grant no. FPU15/06026); Ministerio de Economía y Competitividad through the ARIOS (grant no. CTM2016-76146-C3-1-R) project, cofunded by the Fondo Europeo de Desarrollo Regional 2014–2020 (FEDER); and EU Horizon 2020 through the AtlantOS project (grant agreement no. 633211). Are Olsen was supported by the Norwegian Research Council through ICOS (grant no. 245927). Mario Hoppema was partly supported by the European Union's Horizon 2020 program under grant agreement no. 821001 (SO-CHIC).

**Review statement.** This paper was edited by Jens Klump and reviewed by three anonymous referees.

## References

- Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, *Geosci. Model Dev.*, 8, 2465–2513, <https://doi.org/10.5194/gmd-8-2465-2015>, 2015.
- Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibáñez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality *f*CO<sub>2</sub> data in version 3 of the Surface Ocean CO<sub>2</sub> Atlas (SOCAT), *Earth Syst. Sci. Data*, 8, 383–413, <https://doi.org/10.5194/essd-8-383-2016>, 2016.
- Bates, N. R., Best, M. H. P., Neely, K., Garley, R., Dickson, A. G., and Johnson, R. J.: Detecting anthropogenic carbon dioxide uptake and ocean acidification in the North Atlantic Ocean, *Biogeosciences*, 9, 2509–2522, <https://doi.org/10.5194/bg-9-2509-2012>, 2012.
- Bates, N., Astor, Y., Church, M., Currie, K., Dore, J., Gonaález-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santa-Casiano, M.: A Time-Series View of Changing Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO<sub>2</sub> and Ocean Acidification, *Oceanography*, 27, 126–141, <https://doi.org/10.5670/oceanog.2014.16>, 2014.
- Bauer, J. E., Cai, W. J., Raymond, P. A., Bianchi, T. S., Hopkinson, C. S., and Regnier, P. A. G.: The changing carbon cycle of the coastal ocean, *Nature*, 504, 61–70, <https://doi.org/10.1038/nature12857>, 2013.
- Beale, M. H., Hagan, T. M., and Demuth, H. B.: Deep Learning Toolbox™, User's Guide, Release 2018a, The MathWorks, Inc., Natick, Massachusetts, US, available at: [https://es.mathworks.com/help/pdf\\_doc/deeplearning/nnet Ug.pdf](https://es.mathworks.com/help/pdf_doc/deeplearning/nnet Ug.pdf), last access: 20 August 2018.
- Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO<sub>2</sub> Variables and Nutrient Concentrations From T, S, and O<sub>2</sub> Data Using Bayesian Neural Networks, *Front. Mar. Sci.*, 5, 328, <https://doi.org/10.3389/fmars.2018.00328>, 2018.
- Boyer, T. P., Antonov, J. I., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., O'Brien, T. D., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., and Zweng, M. M.: World Ocean Database 2013, edited by: Levitus, S. and Mishonov, A., NOAA Atlas NESDIS 72, Silver Spring, MD, 209 pp., <https://doi.org/10.7289/V5NZ85MT>, 2013.
- Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., González-Dávila, M., Jeansson, E., Kozyr, A., and van Heuven, S. M. A. C.: A global monthly climatology of total alkalinity: a neural network approach, *Earth Syst. Sci. Data*, 11, 1109–1127, <https://doi.org/10.5194/essd-11-1109-2019>, 2019.
- Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, M., Tanhua, T., Santana-Casiano, J. M., and Kozyr, A.: A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach (Dataset), DIGITAL.CSIC, <https://doi.org/10.20350/digitalCSIC/10551>, 2020.
- Butenschön, M., Clark, J., Aldridge, J. N., Allen, J. I., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., van Leeuwen, S., van der Molen, J., de Mora, L., Polimene, L., Saille, S., Stephens, N., and Torres, R.: ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, *Geosci. Model Dev.*, 9, 1293–1339, <https://doi.org/10.5194/gmd-9-1293-2016>, 2016.
- Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., and Takeshita, Y.: Updated methods for global locally interpolated estimation of alkalinity, pH,



- and nitrate, *Limnol. Oceanogr.-Meth.*, 16, 119–131, <https://doi.org/10.1002/lom3.10232>, 2018.
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. Le, Myneni, R., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, in: *Climate Change 2013 – The Physical Science Basis*, edited by: Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 465–570, 2013.
- Czeschel, R., Stramma, L., and Johnson, G. C.: Oxygen decreases and variability in the eastern equatorial Pacific, *J. Geophys. Res.-Oceans*, 117, C11019, <https://doi.org/10.1029/2012JC008043>, 2012.
- Dickson, A. G.: Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K, *Deep-Sea Res. Pt. I*, 37, 755–766, [https://doi.org/10.1016/0198-0149\(90\)90004-F](https://doi.org/10.1016/0198-0149(90)90004-F), 1990.
- Dickson, A. G. and Millero, F. J.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, *Deep-Sea Res. Pt. I*, 34, 1733–1743, [https://doi.org/10.1016/0198-0149\(87\)90021-5](https://doi.org/10.1016/0198-0149(87)90021-5), 1987.
- Doi, T., Osafune, S., Sugiura, N., Kouketsu, S., Murata, A., Masuda, S., and Toyoda, T.: Multidecadal change in the dissolved inorganic carbon in a long-term ocean state estimation, *J. Adv. Model. Earth Sy.*, 7, 1885–1900, <https://doi.org/10.1002/2015MS000462>, 2015.
- Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO<sub>2</sub> Problem, *Annu. Rev. Mar. Sci.*, 1, 169–192, <https://doi.org/10.1146/annurev.marine.010908.163834>, 2009.
- Dore, J. E., Lukas, R., Sadler, D. W., Church, M. J., and Karl, D. M.: Physical and biogeochemical modulation of ocean acidification in the central North Pacific, *P. Natl. Acad. Sci. USA*, 106, 12235–12240, <https://doi.org/10.1073/pnas.0906044106>, 2009.
- Fabry, V. J., Seibel, B. A., Feely, R. A., Fabry, J. C. O., and Fabry, V. J.: Impacts of ocean acidification on marine fauna and ecosystem processes, *ICES J. Mar. Sci.*, 65, 414–432, <https://doi.org/10.1093/icesjms/fsn048>, 2008.
- García, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., Reagan, J. R., and Johnson, D. R.: *World Ocean Atlas 2013, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*, edited by: Levitus, S. and Mishonov, A., NOAA Atlas NESDIS 75, Silver Spring, Maryland, United States, 27 pp., 2014.
- Gislefoss, J. S., Nydal, R., Slagstad, D., Sonninen, E., and Holmén, K.: Carbon time series in the Norwegian sea, *Deep-Sea Res. Pt. I*, 45, 433–460, [https://doi.org/10.1016/S0967-0637\(97\)00093-9](https://doi.org/10.1016/S0967-0637(97)00093-9), 1998.
- González-Dávila, M., Santana-Casiano, J. M., Rueda, M. J., and Llinás, O.: The water column distribution of carbonate system variables at the ESTOC site from 1995 to 2004, *Biogeosciences*, 7, 3067–3081, <https://doi.org/10.5194/bg-7-3067-2010>, 2010.
- Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., and Jeansson, E.: Constraining projection-based estimates of the future North Atlantic carbon uptake, *J. Climate*, 31, 3959–3978, <https://doi.org/10.1175/JCLI-D-17-0564.1>, 2018.
- Gruber, N., Clement, D., Carter, B. R., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., Lauvset, S. K., Monaco, C. Lo, Mathis, J. T., Murata, A., Olsen, A., Perez, F. F., Sabine, C. L., Tanhua, T., and Wanninkhof, R.: The oceanic sink for anthropogenic CO<sub>2</sub> from 1994 to 2007, *Science*, 363, 1193–1199, <https://doi.org/10.1126/science.aau5153>, 2019.
- Hendriks, I. E., Duarte, C. M., and Álvarez, M.: Vulnerability of marine biodiversity to ocean acidification: A meta-analysis, *Estuar. Coast. Shelf S.*, 86, 157–164, <https://doi.org/10.1016/J.ECSS.2009.11.022>, 2010.
- Hoegh-Guldberg, O. and Bruno, J. F.: The Impact of Climate Change on the World's Marine Ecosystems, *Science*, 328, 1523–1528, <https://doi.org/10.1126/science.1189930>, 2010.
- Jeandel, C., Ruiz-Pino, D., Gjata, E., Poisson, A., Brunet, C., Charriaud, E., Dehairs, F., Delille, D., Fiala, M., Fravallo, C., Carlos Miquel, J., Park, Y. H., Pondaven, P., Quéguiner, B., Razouls, S., Shauer, B., and Tréguer, P.: KERFIX, a time-series station in the Southern Ocean: A presentation, *J. Marine Syst.*, 17, 555–569, 1998.
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T. H.: A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), *Global Biogeochem. Cy.*, 18, 1–23, <https://doi.org/10.1029/2004GB002247>, 2004.
- Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishi, M., Perez, F. F., and Suzuki, T.: Global Ocean Data Analysis Project, Version 2 (GLODAPv2), ORNL/CDIAC-162, NDP-093, available at: [https://www.glodap.info/wp-content/uploads/2017/08/NDP\\_093.pdf](https://www.glodap.info/wp-content/uploads/2017/08/NDP_093.pdf) (last access: 31 July 2020) 2015.
- Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S., Gerber, M., Doney, S. C., Graven, H. D., Gruber, N., McKinley, G. A., Murata, A., Ríos, A. F., and Sabine, C. L.: Global ocean storage of anthropogenic carbon, *Biogeosciences*, 10, 2169–2191, <https://doi.org/10.5194/bg-10-2169-2013>, 2013.
- Kleypas, J. and Langdon, C.: Overview of CO<sub>2</sub>-induced changes in seawater chemistry, in: *Proc. 9th Int. Coral Reef Symp. Bali.*, 2, 2000.
- Körtzinger, A., Hedges, J. I., and Quay, P. D.: Redfield ratios revisited: Removing the biasing effect of anthropogenic CO<sub>2</sub>, *Limnol. Oceanogr.*, 46, 964–970, <https://doi.org/10.4319/lo.2001.46.4.0964>, 2001.
- Kroeker, K. J., Kordas, R. L., Crim, R., Hendriks, I. E., Ramajo, L., Singh, G. S., Duarte, C. M. and Gattuso, J.-P.: Impacts of ocean acidification on marine organisms: quantifying sensitivities and interaction with warming, *Glob. Change Biol.*, 19, 1884–1896, <https://doi.org/10.1111/gcb.12179>, 2013.
- Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, *Global Biogeochem. Cy.*, 28, 927–949, <https://doi.org/10.1002/2014GB004853>, 2014.
- Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, *Global Biogeochem. Cy.*, 30, 1396–1417, <https://doi.org/10.1002/2015GB005359>, 2016.
- Landschützer, P., Gruber, N., and Bakker, D. C. E.: An updated observation-based global monthly gridded sea surface pCO<sub>2</sub> and air-sea CO<sub>2</sub> flux product from 1982 through 2015 and its monthly climatology (NCEI Accession 0160558), Version 2.2, NOAA National Centers for Environmental Information, Dataset, available at: <https://www.nodc.noaa.gov/ocads/>

- oceans/SPCO2\_1982\_2015\_ETH\_SOM\_FFN.html (last access: 31 July 2020), 2017.
- Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2, *Earth Syst. Sci. Data*, 8, 325–340, <https://doi.org/10.5194/essd-8-325-2016>, 2016.
- Lee, K., Wanninkhof, R., Feely, R. A., Millero, F. J., and Peng, T.-H.: Global relationships of total inorganic carbon with temperature and nitrate in surface seawater, *Global Biogeochem. Cy.*, 14, 979–994, <https://doi.org/10.1029/1998GB001087>, 2000.
- Le Quéré, C., Buitenhuis, E. T., Moriarty, R., Alvain, S., Aumont, O., Bopp, L., Chollet, S., Enright, C., Franklin, D. J., Geider, R. J., Harrison, S. P., Hirst, A. G., Larsen, S., Legendre, L., Platt, T., Prentice, I. C., Rivkin, R. B., Sailley, S., Sathyendranath, S., Stephens, N., Vogt, M., and Vallina, S. M.: Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles, *Biogeosciences*, 13, 4111–4133, <https://doi.org/10.5194/bg-13-4111-2016>, 2016.
- Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneeth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregón, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, *Earth Syst. Sci. Data*, 10, 2141–2194, <https://doi.org/10.5194/essd-10-2141-2018>, 2018.
- Levenberg, K.: A method for the solution of certain non-linear problems in least squares., *Q. Appl. Math.*, II, 164–168, 1944.
- Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., Paver, C. R., Reagan, J. R., Johnson, D. R., Hamilton, M., and Seidov, D.: World Ocean Atlas 2013, Volume 1: Temperature, edited by: Levitus, S. and Mishonov, A., NOAA Atlas NESDIS 73, 40 pp., 2013.
- Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *J. Soc. Ind. Appl. Math.*, 11, 431–441, 1963.
- Mehrbach, C., Culberson, C. H., Hawley, J. E., and Pytkowicz, R. M.: Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure, *Limnol. Oceanogr.*, 18, 897–907, <https://doi.org/10.4319/lo.1973.18.6.0897>, 1973.
- Millero, F. J.: Thermodynamics of the carbon dioxide system in the oceans, *Geochim. Cosmochim. Ac.*, 59, 661–677, [https://doi.org/10.1016/0016-7037\(94\)00354-O](https://doi.org/10.1016/0016-7037(94)00354-O), 1995.
- Millero, F. J.: The Marine Inorganic Carbon Cycle, *Chem. Rev.*, 107, 308–341, <https://doi.org/10.1021/cr0503557>, 2007.
- Olafsson, J., Olafsdottir, S. R., Benoit-Cattin, A., and Takahashi, T.: The Irminger Sea and the Iceland Sea time series measurements of sea water carbon and nutrient chemistry 1983–2008, *Earth Syst. Sci. Data*, 2, 99–104, <https://doi.org/10.5194/essd-2-99-2010>, 2010.
- Olden, J. D. and Jackson, D. A.: Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks, *Ecol. Model.*, 154, 135–150, [https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9), 2002.
- Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., and Suzuki, T.: The Global Ocean Data Analysis Project version 2 (GLODAPv2) – an internally consistent data product for the world ocean, *Earth Syst. Sci. Data*, 8, 297–323, <https://doi.org/10.5194/essd-8-297-2016>, 2016.
- Olsen, A., Lange, N., Key, R. M., Tanhua, T., Álvarez, M., Becker, S., Bittig, H. C., Carter, B. R., Cotrim da Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jones, S. D., Jutterström, S., Karlsen, M. K., Kozyr, A., Lauvset, S. K., Lo Monaco, C., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Telszewski, M., Tilbrook, B., Velo, A., and Wanninkhof, R.: GLODAPv2.2019 – an update of GLODAPv2, *Earth Syst. Sci. Data*, 11, 1437–1461, <https://doi.org/10.5194/essd-11-1437-2019>, 2019.
- Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M. F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, <https://doi.org/10.1038/nature04095>, 2005.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, 323, 533–536, <https://doi.org/10.1038/323533a0>, 1986.
- Sabine, C. L., Key, R. M., Feely, R. A., and Greeley, D.: Inorganic carbon in the Indian Ocean: Distribution and dissolution processes, *Global Biogeochem. Cy.*, 16, 15–1–15–18, <https://doi.org/10.1029/2002GB001869>, 2002.
- Sallée, J.-B., Matear, R. J., Rintoul, S. R., and Lenton, A.: Localized subduction of anthropogenic carbon dioxide in the Southern Hemisphere oceans, *Nat. Geosci.*, 5, 579–584, <https://doi.org/10.1038/ngeo1523>, 2012.
- Sarmiento, J. L. and Gruber, N.: Sinks for Anthropogenic Carbon, *Phys. Today*, 55, 30–36, <https://doi.org/10.1063/1.1510279>, 2002.
- Sauzède, R., Claustre, H., Pasqueron de Fommervault, O., Bittig, H. C., Gattuso, J.-P., Legendre, L., and Johnson, K. S.: Estimates of water-column nutrients concentration and carbonate system parameters in the global ocean: A novel approach based on neural networks, *Front. Mar. Sci.*, 4, 128, <https://doi.org/10.3389/fmars.2017.00128>, 2017.
- Schlitzer, R., Ocean Data View, available at: <http://odv.awi.de> (last access: 21 May 2018), 2016.

- Takahashi, T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., and Ho, C.: Climatological distributions of pH,  $p\text{CO}_2$ , total  $\text{CO}_2$ , alkalinity, and  $\text{CaCO}_3$  saturation in the global surface ocean, and temporal changes at selected locations, *Mar. Chem.*, 164, 95–125, <https://doi.org/10.1016/j.marchem.2014.06.004>, 2014.
- Takahashi, T., Sutherland, S. C., and Kozyr, A.: Global Ocean Surface Water Partial Pressure of  $\text{CO}_2$  Database: Measurements Performed During 1957–2016 (Version 2016), ORNL/CDIAC-161, NDP-088(V2015), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, Dataset, Document available at: [https://www.nodc.noaa.gov/ocads/oceans/LDEO\\_Underway\\_Database/NDP-088\\_V2016.pdf](https://www.nodc.noaa.gov/ocads/oceans/LDEO_Underway_Database/NDP-088_V2016.pdf) (last access: 31 July 2020), 2017.
- Tanhua, T., Bates, N. R., and Körtzinger, A.: The Marine Carbon Cycle and Ocean Carbon Inventories, in *International Geophysics*, vol. 103, edited by: Siedler, G., Griffies, S., Gould, J., and Church, J., Academic Press, 787–815, 2013.
- Tank, S. E., Raymond, P. A., Striegl, R. G., McClelland, J. W., Holmes, R. M., Fiske, G. J., and Peterson, B. J.: A land-to-ocean perspective on the magnitude, source and implication of DIC flux from major Arctic rivers to the Arctic Ocean, *Global Biogeochem. Cy.*, 26, GB4018, <https://doi.org/10.1029/2011GB004192>, 2012.
- van Heuven, S., D. Pierrot, J. W. B. Rae, E. Lewis, and Wallace, D. W. R.: MATLAB Program Developed for  $\text{CO}_2$  System Calculations. ORNL/CDIAC-105b, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, available at: [https://cdiac.ess-dive.lbl.gov/ftp/co2sys/CO2SYS\\_calc\\_MATLAB\\_v1.1/?C=N;O=A](https://cdiac.ess-dive.lbl.gov/ftp/co2sys/CO2SYS_calc_MATLAB_v1.1/?C=N;O=A) (last access: 31 July 2020), 2011.
- Vázquez-Rodríguez, M., Touratier, F., Lo Monaco, C., Waugh, D. W., Padin, X. A., Bellerby, R. G. J., Goyet, C., Metzl, N., Ríos, A. F., and Pérez, F. F.: Anthropogenic carbon distributions in the Atlantic Ocean: data-based estimates from the Arctic to the Antarctic, *Biogeosciences*, 6, 439–451, <https://doi.org/10.5194/bg-6-439-2009>, 2009.
- Wakita, M., Watanabe, S., Murata, A., Tsurushima, N., and Honda, M.: Decadal change of dissolved inorganic carbon in the subarctic western North Pacific Ocean, *Tellus B*, 62, 608–620, <https://doi.org/10.1111/j.1600-0889.2010.00476.x>, 2010.
- Watanabe, Y. W., Nishioka, J., Shigemitsu, M., Mimura, A., and Nakatsuka, T.: Influence of riverine alkalinity on carbonate species in the Okhotsk Sea, *Geophys. Res. Lett.*, 36, L15606, <https://doi.org/10.1029/2009GL038520>, 2009.
- Weiss, R. F., Jahnke, R. A. and Keeling, C. D.: Seasonal effects of temperature and salinity on the partial pressure of  $\text{CO}_2$  in seawater, *Nature*, 300, 511–513, <https://doi.org/10.1038/300511a0>, 1982.
- Wu, Y., Hain, M. P., Humphreys, M. P., Hartman, S., and Tyrrell, T.: What drives the latitudinal gradient in open-ocean surface dissolved inorganic carbon concentration?, *Biogeosciences*, 16, 2661–2681, <https://doi.org/10.5194/bg-16-2661-2019>, 2019.
- Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, *Geosci. Model Dev.*, 6, 1767–1811, <https://doi.org/10.5194/gmd-6-1767-2013>, 2013.
- Zeebe, R. E. and Wolf-Gladrow, D.:  $\text{CO}_2$  in seawater: Equilibrium, Kinetics, Isotopes, in *Elsevier oceanography series*, edited by: Halpem, D., Elsevier Oceanography Series, 2001.
- Zweng, M.M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V., Boyer, T. P., Garcia, H. E., Baranova, O. K., Johnson, D. R., Seidov, D., and Biddle, M. M.: World Ocean Atlas 2013, Volume 2: Salinity, Levitus, S. (Ed.), Mishonov, A. (Technical Ed.), NOAA Atlas NESDIS 74, 39 pp., 2013.