https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

OceanSODA-ETHZ: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification

Luke Gregor[1] and Nicolas Gruber[1]

[1]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zürich, Switzerland

Correspondence: Luke Gregor (luke.gregor@usys.ethz.ch)

Abstract. Ocean acidification has altered the ocean's carbonate chemistry profoundly since preindustrial times, with potentially serious consequences for marine life. Yet, no long-term global observation-based data set exists that permits to study changes in ocean acidification for all carbonate system parameters over the last few decades. Here, we fill this gap and present a methodologically consistent global data set of all relevant surface ocean parameters, i.e., dissolved inorganic carbon (DIC), total alkalinity (TA), partial pressure of CO_2 (pCO_2), pH, and the saturation state with respect to mineral $CaCO_3$ (Ω) at monthly resolution over the period 1985 through 2018 at a spatial resolution of 1x1°. This data set, named OceanSODA-ETHZ, was created by extrapolating in time and space the surface ocean observations of pCO_2 (from the Surface Ocean CO_2 ATlas (SOCAT)) and total alkalinity (TA, from the Global Ocean Data Analysis Project (GLODAP)) using the newly developed Geospatial Random Cluster Ensemble Regression (GRaCER) method. This method is based on a two-step (cluster-regression) approach, but extends it by considering an ensemble of such cluster-regressions, leading to higher robustness. Surface ocean DIC, pH, and Ω were then computed from the globally mapped pCO_2 and TA using the thermodynamic equations of the carbonate system. For the open ocean, the cluster regression method estimates pCO_2 and TA with global near-zero biases and root mean squared errors of 12 μatm and 13 μmol kg^{-1}, respectively. Taking into account also the measurement and representation errors, the total error increases to 14 μatm and 21 μmol kg^{-1}, respectively. We assess the fidelity of the computed parameters by comparing them to direct observations from GLODAP, finding surface ocean pH and DIC global biases of near zero, and root mean squared errors of 0.023 and 16 μmol kg^{-1}, respectively. These errors are very comparable to those expected by propagating the total errors from pCO_2 and TA through the thermodynamic computations, indicating a robust and conservative assessment of the errors. We illustrate the potential of this new dataset by analyzing the climatological mean seasonal cycles of the different parameters of the surface ocean carbonate system, highlighting their commonalities and differences. The OceanSODA-ETHZ data can be downloaded from https://doi.org/10.25921/m5wx-ja34 (Gregor and Gruber, 2020).

1 Introduction

The oceans have taken up roughly one quarter of the anthropogenic CO_2 that has been released into the atmosphere since the start of the industrial era (Sabine et al., 2004; Gruber et al., 2019), lowering the increase in atmospheric CO_2 substantially. But

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

this buffering of anthropogenic climate change by the ocean comes with a substantial cost, i.e., ocean acidification(Doney et al., 2009). The uptake of anthropogenic CO_2 over the last 150 years has made the surface ocean more acidic with a decrease in the global mean pH from ∼8.2 around 1850 to ∼8.1 today (Feely et al., 2009; Jiang et al., 2019). This decrease in pH equates to a ∼30% increase in the concentration of the H^+ ions. Some of the anthropogenic CO_2 taken up from the atmosphere remains in the seawater as dissolved CO_2, thus increasing its partial pressure (pCO_2). In fact, surface pCO_2 tends to track the increase in atmospheric pCO_2 rather closely (e.g. Bates et al. 2014) owing to the order of one year timescale for the equilibration of CO_2 across the air-sea interface (Sarmiento and Gruber, 2006), which is smaller than the decadal timescale increase in atmospheric CO_2. While some of the added CO_2 stays as CO_2, the majority of is titrated away by the ocean's carbonate ion (Sarmiento and Gruber, 2006), leading to a substantial reduction in its concentration. This reduces the saturation state (Ω) with regard to mineral $CaCO_3$, where an Ω of < 1 leads to dissolution of $CaCO_3$.

These chemical changes, collectively described as ocean acidification, will have a profound impact on marine organisms, especially those that form shells made of calcium carbonate ($CaCO_3$) (Orr et al., 2005; Fabry et al., 2008; Doney et al., 2009; Bednaršek et al., 2019; Doney et al., 2020). Calcifying organisms living in high latitudes and subtropical and tropical upwelling regions, with their naturally low Ω and pH, may be particularly vulnerable, as these regions will be among the first to cross critical thresholds such as becoming undersaturated (Orr et al., 2005; Steinacher et al., 2009; Gruber et al., 2012; Franco et al., 2018; Fabry et al., 2009; Hauri et al., 2016; Negrete-García et al., 2019). However, marine organisms may be susceptible to changes even where Omega > 1 due to a shift in energetic requirements for shell formation (Orr et al., 2005; Pörtner and Farrell, 2008). For example, it is well known that corals start to decrease their calcification already at saturation states well above 3 (Gattuso et al., 1998). Ocean acidification will thus have a significant economic impact on fisheries and tourism through the impact on shellfish and corals respectively (Cooley and Doney, 2009; Doney et al., 2020).

At the global scale, most of what we know about the progression of ocean acidification in the recent decades has come from either models (Bopp et al., 2013; Kwiatkowski et al., 2020) or from the combination of model-based trends with observation-based climatologies (Feely et al., 2009; Jiang et al., 2019). A notable exception are the large number of studies that have analyzed the trends and variability of surface ocean pCO_2 (e.g. Landschützer et al. 2013, 2016; Rödenbeck et al. 2014; Denvil-Sommer et al. 2019; Gregor et al. 2019) and the effort of Lauvset et al. (2015) and Turk et al. (2017) to analyze long-term trends in pH and Omega respectively. But these studies remained limited to one single parameter. At the local to regional scale, a number of long-term timeseries have provided excellent insights into the processes and trends of ocean acidification across all carbonate system parameters (e.g. Bates et al. 2014), but no global comprehensive view of the historical development of ocean acidification based on observations exist. This is largely a consequence of the limited observations, although observational efforts have increased substantially in the recent decades, among others through efforts such as GOA-ON (Global Ocean Acidification Observing Network (Tilbrook et al., 2019). The OceanSODA (Satellite Oceanographic Datasets for Acidification) project (https://esa-oceansoda.org), which this study forms part of, aims to close this gap by linking satellite observations with in situ observations of the marine carbonate system.

In line with the goal of the OceanSODA project, we aim to develop a global, observation-based data set documenting the progression of ocean acidification over the recent decades. Such a data set will be crucial to put the current trends of ocean

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

acidification into the context of the changes over the last few decades. By also describing the level of variability in ocean acidification around the long-term trend, it will also help to better understand the challenges that marine organisms are facing. And it will permit us to explore in much more detail how ocean acidification has unfolded regionally, and potentially deviated from the simple model of it being slaved to the rise in atmospheric CO_2.

The well measurable parameters of the marine carbonate system are dissolved inorganic carbon (DIC), total alkalinity (TA), pH and the partial pressure of carbon dioxide (pCO_2). Very few measurement programs measure all of these parameters concurrently. In fact, the vast majority of the observational programs measure only one parameter, with pCO_2 being the most often measured one, followed by DIC, TA, and pH (Bakker et al., 2016; Olsen et al., 2016). Since two parameters are sufficient to fully describe the marine carbonate system, any combination of two will permit to fully reconstruct the entire carbonate system. But not all combination are equally suited, given the uncertainties in the measurements, the uncertainties in the coefficients of the carbonate chemistry, and the spatiotemporal coverage vis-à-vis the variability of these parameters.

We use here the pair pCO_2 and TA as the basis for our reconstruction for two reasons. First, these are the best observed parameters relative to their spatio-temporal variability, permitting us to develop better predictive models for the global surface ocean than possible for, e.g., DIC and pH. Second, detailed assessments of the internal consistency of the oceanic carbonate system have shown that pCO_2 and TA are a well suited pair to estimate pH, owing to the reliability of the measurements and the predictive accuracy (Bockmon and Dickson, 2015; Bakker et al., 2016; Raimondi et al., 2019). This is not the case if DIC was used instead of TA. Our choice is supported by Takahashi and Sutherland (2013) who developed the first seasonal climatology of all parameters of the surface ocean carbonate system using the same pair.

Measurements of pCO_2 are abundant compared to the other variables due to a well-established and robust underway sampling protocol that allows instruments to also be installed on non-scientific vessels under the Volunteer Observing Ship (VOS) program (Bakker et al., 2016; Pierrot et al., 2009). High quality pCO_2 data are also easily accessible thanks to SOCAT that consolidates underway pCO_2 observations and ensures the quality of observations (Bakker et al., 2016). Total alkalinity is not as widely measured as pCO_2 due to the fact that measurements are made discretely with bottle samples (Dickson et al., 2007). But, fortunately, TA is highly correlated with salinity on a global scale (r=0.96) making it a suitable variable for prediction with a < 10% error of the observed range (Lee et al., 2006; Olsen et al., 2016; Broullón et al., 2018). Further, the accessibility to TA measurements is made possible through the continued efforts of GLODAP Olsen et al. 2016). We discarded the option to use DIC instead of TA, even though DIC is slightly more often sampled than TA. This decision is based on the fact that DIC is more variable than TA, and also its correlation with salinity is much lower. As a result, it is much more difficult to develop predictive models. Oceanic pH is also not an option, since historically, it has been measured far less often than the other parameters. This is changing, since progress with reference materials and new sensors have permitted to grow the number of pH measurements in recent years tremendously, largely benefitting from the deployments on the biogeochemical Argo floats (Claustre et al., 2020).

The actual spatial and temporal coverage for any of these parameters is very low. Even for pCO_2, i.e., the parameter with the densest coverage, only about 1.4% of the global surface ocean has been sampled in any given month over the past 30 years (Bakker et al., 2016). Thus, the global-scale reconstruction of the progression of ocean acidification requires a very substantial

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

inter- and extrapolation effort. Advances in remote sensing (Land et al., 2019), and the increasing power and usability of machine learning techniques have permitted to address this challenge, leading to a proliferation of such efforts. But they vary greatly between the different parameters of the marine carbonate system.

By far the most established efforts are those that interpolate and extrapolate the ocean pCO_2 observations, as demonstrated by the inter-comparison project by Rödenbeck et al. (2015). Feed-forward neural networks (FFNN) have become one of the favored tools (Landschützer et al., 2013; Zeng et al., 2014; Denvil-Sommer et al., 2019), but other statistical and machine learning methods, such as Bayesian regression and tree-based regression, have also been used with similar success (Rödenbeck et al., 2014; Gregor et al., 2019). However, the specific implementation of the methods is what sets the assortment of methods apart. For example, the SOM-FFN method of Landschützer et al. (2013) and the CSIR-ML6 method of Gregor et al. (2019) (amongst others) first cluster the data based on a certain set of climatological predictors and then perform a regression on pCO_2 for each resulting cluster. An alternate approach, used by both the LSCE-FFNN (Denvil-Sommer et al., 2019) and NIES-FNN (Zeng et al., 2014) methods, is to include the positional coordinates, without the need for subsetting the data by clustering. Despite the differences in implementation and regression algorithms, the majority of methods achieve an accuracy of roughly 17.5 μatm when compared with SOCAT (Gregor et al., 2019). But each of these methods has its strengths and weaknesses: For example, the SOM-FFN and CSIR-ML6 methods are able to generalize estimates in data sparse regions due to information sharing within a cluster, but the methods suffer from discrete boundaries where clusters meet (Gregor et al., 2019). These discrete boundaries may introduce artefacts when applied to certain questions.

The extrapolation of TA onto a global grid is also well established (Gruber et al., 1996; Millero et al., 1998; Lee et al., 2006; Takahashi and Sutherland, 2013; Good et al., 2013; Carton et al., 2018; Bittig et al., 2018). The highly linear relationship between salinity and TA means that linear regressions have been able for quite some time to estimate TA with adequate accuracy. For example, Gruber et al. (1996) developed a globally applicable multi-linear regression model involving salinity and the conservative tracer PO (PO = $O_2 + 170 \cdot PO_4$, (Broecker and Peng, 1974)) and achieved a global RMSE of 11 μmol kg^{-1}. Lee et al. (2006) also used a MLR approach, but differentiated it regionally using salinity, temperature and spatial coordinates as independent variables. The same approach was followed by Takahashi et al. (1993). More recently, more nuanced and non-linear regression approaches have improved upon the MLR approaches (Sasse et al., 2013; Carter et al., 2018; Broullón et al., 2018; Bittig et al., 2018). For example, the Locally Interpolated Alkalinity Regression (LIARv2) still makes use of linear regression but interpolates the regression coefficients spatially from a fixed set of trained regression nodes located at every 5'th point (Carter et al., 2016). Sasse et al. (2013) used a self-organizing map approach coupled with a local linear optimizer (called SOMLO) and achieved a global RMSE of 9 μmol kg^{-1}. A similar RMSE was achieved by Broullón et al. (2018) using a neural network approach (NNGv2). In addition to these global regressions, several regionally specific regressions were developed (see Table 1 in Land et al. 2019).

In comparison, only very few efforts attempted to inter- and extrapolate DIC. Lee et al. (2000) were the first to produce a global map of DIC using a regression methodology. Concretely, they employed a regional multiple linear regression model similar to that used later to map TA. But their application was limited to the generation of a seasonal climatology. It was then not until Sasse et al. (2013) when the first global reconstruction of the temporal progression of DIC over multiple years was

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

published. They used the same SOMLO method as they had used for TA, creating global maps of DIC with a RMSE of 11 μmol kg^{-1}. More recently, Keppler et al. (2020) used the SOM-FFN method of Landschützer et al. (2013) to reconstruct DIC throughout the upper water column on a monthly basis, but they limited their discussion to the mean seasonal cycle.

Here, to map TA and pCO_2 to the globe, we will use a newly developed two step cluster-regression approach that is similar in design to the SOM FFN method (Landschützer et al., 2013, 2016) but extend it by using an ensemble of such cluster-regressions. This method, referred to as Geospatial Random Cluster Ensemble Regression (GRaCER), increases the robustness of the estimates considerably. It also removes the boundary problems inherent in all methods that use fixed regional boundaries. We apply the same methodology to TA and pCO_2, resulting in methodologically consistent global estimates of the two parameters, from which DIC, pH, and Ω can then be computed using the well-established thermodynamic models of the seawater carbonate system. These latter estimates can then be compared against the many available DIC and pH measurements, providing a large set of independent data to assess the fidelity of our estimates. This requires also a good understanding of the different sources of uncertainties, including those emanating from sampling and measurement, from the statistical modeling, and from the lack of representativeness, i.e., the fact that a local measurement is not representative for the large pixel (100 x 100 km and 1 month), that one models in our regressions).

The rest of the manuscript describes the data and methods used to calculate this data set for ocean acidification. The uncertainties of the predictions are assessed, followed by the presentation of the data with a focus on the seasonal cycle. Last, we discuss the implications of the uncertainties for the use of the derived marine carbonate system.

2 Methods

To reconstruct the global progression of all parameters of the surface ocean carbonate system over the last three decades (1985 through 2018), we follow the three steps depicted by the flow diagram in Figure 1. Concretely, we develop first a statistical model for the measured TA and pCO_2 using the newly developed GRaCER method. This method itself consists of two steps, i.e., a cluster step, where the target variables are clustered regionally, and a regression step, where for each cluster, a regression is evaluated. These two steps are repeated multiple times, creating an ensemble of models; Second we map these two quantities to the globe and over time using this ensemble of statistical models and global observations of the predictor variables; Third and last, we use a thermodynamic model of the seawater carbonate system to compute the remaining parameters of the surface ocean carbonate system, namely DIC, pH, and Ω. Along the way, we extensively evaluate and test each step with independent observations. We refer to the dataset with the evaluated and complete marine carbonate system as OceanSODA-ETHZ.

Next, we describe the concept of the GRaCER method, and then detail its implementation for pCO_2 and TA. This is followed by a description of the numerous types of data employed and how they were prepared. Lastly, we demonstrate how we used a thermodynamic model to derive the remaining parameters of the marine carbonate system.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

2.1 GRaCER Algorithm

The GRaCER algorithm builds conceptually on a series of cluster-regression algorithms that have been successfully used for the inter- and extrapolation of surface ocean pCO_2 (Sasse et al., 2013; Landschützer et al., 2013, 2016; Iida et al., 2015; Gregor et al., 2019). The main advantage of such a two-step approach is that the first **clustering** step organizes the variability regionally and temporally. This greatly enhances then the fidelity of the second step, i.e., the regression, as the size of the regression problem is reduced from the global domain to smaller, more homogeneous regions. A second advantage is that this clustering brings together regions with similar seasonality and similar co-variability with potential predictors, irrespective of the number of observations. This permits the regression to transfer information from spatially distant, but geochemically similar regions, making the inter and extrapolation more robust in data poor regions. The main innovation of the GRaCER algorithm relative to the previously used two-step approaches is its use of ensembles of cluster-regressions, i.e., the generation of a whole series of clusters and corresponding regressions, which permits to overcome the boundary problems that are inherent in all two-step approaches.

For the **clustering** step, we use monthly climatological data of pCO_2 and TA and related parameters (Figure 2a-c), to determine the main patterns of variability of the target variable and its co-variability with potential predictor variables. The mini-batch K-means implementation in the Python *Scikit-Learn* package is used to perform the clustering due to its computational efficiency and scalability with large data sets (Pedregosa et al., 2011). A user-defined number of cluster centers are initiated with the K-means++ algorithm, which randomly selects the location of the first cluster center, then iteratively selects a best-guess location for the remaining cluster centers. Thereafter, the algorithm minimizes the distance between cluster centers and data points in the variable space. Once the clusters have been defined for the climatological domain, the co-located training data are assigned to the monthly clusters.

The **Regression** is then performed individually for each of the clusters (Figure 2d-f). The GRaCER method does not use a prescribed regression method - rather the appropriate algorithm for the particular use case is implemented. Importantly, the algorithm must be able to scale appropriately to the size of the problem. For example, the training data set for TA is one 20^{th} of the size of the pCO_2 training data set, thus a more computationally expensive method can be used to predict TA.

The **Ensemble** members are created by performing the cluster-regression step multiple times. Creating an ensemble is possible due to the fact that each clustering instance is slightly different (Figure 2g-i). In practice, the spatial distribution of the clusters is similar, *i.e.* there is consistency in the typology of the clusters, particularly in regions where clusters are well defined (*e.g.* subtropical gyres and tropical eastern Pacific). However, there are regions that belong to different clusters, *i.e.* there is slight variance in the typology between ensemble members. The differences are due to the random initialization of the first cluster center in the K-means clustering step, and the fact that clustering variables for some regions have weak gradients in spatial auto-correlation resulting in weak association to a cluster. In practice, this means that the location of cluster boundaries vary between ensemble members, thus the ensemble mean does not have discrete boundaries (Figure 2j).

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

2.2 Algorithm Implementation

2.2.1 Total alkalinity

For the estimation of TA, we employ the support vector regression (SVR) regression method with 12 clusters and 16 ensemble members. The clustering is performed on climatological mean total alkalinity, sea-surface salinity (SSS), sea-surface temperature (SST) and nitrate (Table 1 and section 2.3 below). The optimal variables on which clustering should be performed were selected by assessing the regression scores of each combination of variables following the methodology of Gregor et al. (2019).

A similar exhaustive search was used for determining the number of clusters. The number of ensemble members was chosen by the number above which there is no longer an increase in performance, analogous to the number of trees in a Random Forest. Test data are a subset of years spaced three years apart starting in 1985. We ensure that the models are not overfitted by selecting hyper-parameters using K-fold cross validation (further details are in Section A3).

To regress and map TA, we use SSS, SST, silicic acid (Si), and $N^* = NO_3 - 16 \cdot PO_4$ (Gruber et al., 1996) as predictors. Our choice of SSS and SST as predictors is easily justified by these two variables accounting for the majority of TA variability (Lee et al., 2006; Carter et al., 2018). The addition of silicic acid and N^* as predictors is to account for seasonal changes in primary production that has an impact on TA (Wolf-Gladrow et al., 2007; Carter et al., 2018). Further, N^* expresses the zonal differences between and within the large ocean basins better than using simply nitrate or phosphate- an important consideration, since coordinates (i.e., latitude and longitude) are not given as predictors.

2.2.2 Partial pressure of CO_2

For the estimation of pCO_2, we use two regression methods, i.e., GBDT (gradient boosted decision trees) and FFNN (feed forward neural network). These are implemented with 21 clusters and 16 ensemble members (eight each). The number of clusters is at the upper end of the range compared with the number of clusters used by the MPI-SOMFFN or CSIR-ML6 methods. But testing has shown that additional clusters are required to account for the additional complexity by the inclusion of the coastal, Arctic and Mediterranean seas.

Clustering is performed on climatological values of pCO_2, SST, mixed layer depth and Chlorophyll-a, with additional weighting given to pCO_2. Details of the regression method, and of the hyper-parameter selection are given in section A3. Test data are selected as every 5^{th} year starting in 1985, and validation data for early stopping is selected using the same approach starting in 1987, where the latter is used for early stopping to reduce over-fitting and keep model complexity within bounds.

The regression and mapping is performed with the following variables as predictors: SST, SSS, the logarithm of Chlorophyll-a, the logarithm of mixed-layer depth, the meridional and zonal components of the surface winds, the sine and cosine of $\frac{\text{day of year} \cdot \pi}{365 \cdot 180}$, and the atmospheric dry-air mixing ratio (xCO_2). These predictors are the same as used by Gregor et al. (2019) and various combinations of these methods have been used by previous approaches (Landschützer et al., 2014; Denvil-Sommer et al., 2019).

It is important to note that the predictors are proxies for the spatio-temporal changes in pCO_2 and do not necessarily explain the physical mechanism by which changes in pCO_2 are driven. For example, an increase in sea-surface temperature in the

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

Table 1. Variables used as the clustering features and predictor variables for regression. Details about these data are given in the text. Note that clustering features are all resampled to monthly climatologies. $^{\dagger}TA_{map}$ is the exception where a single mapped time step is used. The first column for regression shows the target variable. All machine learning models use the same variables to train and predict the final estimates with the exception of $^{\ddagger}SSS$, where ungridded GLODAP data is used to train and SODA salinity is used to predict. All other references to SSS refer to the gridded SODA product.

Clustering	Clustering features (monthly)
TA	$^{\dagger}TA^{map}$, SSS, SST, N^*
pCO_2	pCO_2^{map}, SST, Chl-a, MLD_{clim}
Regression & Mapping	Predictors
TA^{GLODAP}	$^{\ddagger}SSS_{SODA}^{GLODAP}$, SST, Si, N^*
pCO_2^{SOCAT}	xCO_2^{atm}, SST, SSS, Chl-a, MLD_{clim}, u-wind, v-wind

subtropics results in an increase pCO_2 as shown by Takahashi et al. (1993); Lefèvre and Taylor (2002). In contrast, surface warming in the Southern Ocean can be a proxy for stratification that reduces outcropping of high CO_2 waters (Landschützer et al., 2015; Gregor et al., 2018). Similarly, changes in SSS and MLD also capture the distribution and processes that drive changes in surface pCO_2, such as stratification and mixing. However, the climatological MLD product used here does not capture interannual variability in stratification and mixing. We thus include the two surface wind components that, along with SST, are a proxy for wind-driven mixing and upwelling. Chlorophyll is also an important driver of pCO_2 on a local scale, particularly in the high latitude regions where high primary productivity results in rapid uptake of pCO_2 Bakker et al. (2008); Gregor et al. (2018). Lastly, xCO_2 is included to account for the close tracking of oceanic pCO_2 to atmospheric CO2 concentrations (Bates et al., 2014).

2.3 Data

Data are used to develop the two-step GRaCER model, i.e., clustering and regression, and to evaluate the estimates. Table 1 provides an overview of all data employed and the purposes for which they are used and Table 2 shows the corresponding source of the data. We describe each data set by parameter and use.

2.3.1 Data for clustering

For the clustering of TA, we used the mapped product of total alkalinity (TA_{map}) from the GLODAPv2 (Lauvset et al., 2016). We repeat the temporally averaged mapped TA to create a monthly data set over which clustering can be performed. We thus assume that the spatial variability of TA is larger than the seasonal variability. This is backed by Takahashi and Sutherland (2013) and Broullón et al. (2018) who found that the seasonal variability of TA for the majority of the ocean was more than a factor of 10 smaller than the spatial variability.

For the clustering step of pCO_2, we use four data-based products resampled and gridded to a monthly by $1\times1^{\circ}$ resolution (pCO_2^{map}), namely LDEO by Takahashi et al. (2014), MPI-SOMFFN by Landschützer et al. (2016), Jena-MLS by Rödenbeck

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

Table 2. Data sources used in this study. *C3S (2017) is short for Copernicus Climate Change Service (C3S) (2017).

Product name	Variable	Abbrev	Reference
SOCAT	pCO_2^{SOCAT}	pCO_2	Bakker et al. (2016)
MPI-SOMFFN	Gap-filled pCO_2	pCO_2^{map}	Landschützer et al. (2016)
Jena-MLS			Rödenbeck et al. (2014)
LSCE-FFNN			Denvil-Sommer et al. (2019)
LDEO			Takahashi et al. (2014)
GLODAP	Total Alkalinity (in-situ)	TA	Olsen et al. (2016)
	Total Alkalinity (mapped)	TA^{map}	Lauvset et al. (2016)
OSTIA v2	Sea surface temperature	SST	Merchant et al. (2019)
	Sea-ice fraction	ICE	
SODA v3.4.2	Salinity	SSS	Carton et al. (2018)
	Mixed Layer Depth	MLD	Holte et al. (2017)
ERA5	Sea-level pressure	Pres	*C3S (2017)
	U-component of wind	U	
	V-component of wind	V	
NOAA: ATM	Mole fraction of CO_2^{atm}	xCO_2^{atm*}	Dlugokencky et al. (2019)
Globcolour	Chlorophyll-a	Chl-a	Maritorena et al. (2010)
WOA	Phosphate	PO_4	Boyer et al. (2013)
	Nitrate	NO_3	
	Silicate	Si	

et al. (2014), and LSCE-FFNN by Denvil-Sommer et al. (2019). It may seem tautological to use other machine learning estimates, but these data are just used to create regional clusters, i.e., they are not used in the regression step. Relative to previous two-step approaches Landschützer et al. (2016); Denvil-Sommer et al. (2019), which used just the LDEO product, we expanded on this by including three more estimates. In doing so we make the implicit assumption that these estimates are a better representation of the pCO_2 monthly climatology than the LDEO climatology alone.

SSS is from the Simple Ocean Data Assimilation (SODA) analysis (Carton et al., 2018) and SST from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) v2 product (Merchant et al., 2019). N^* is calculated using monthly climatologies of nitrate and phosphate from the World Ocean Atlas updated in 2018 (Boyer et al., 2013). We use the the monthly climatology of density-based mixed-layer depth (MLD) from Holte et al. (2017) that is estimated from Argo float profiles. A two-dimensional moving average filter is applied to the MLD to interpolate missing data and remove the noise introduced by interannual and sub-monthly variability.

Chlorophyll-a (Chl-a) is from the Globcolour project where a monthly climatology is calculated for the period from 1998 through 2018 (Maritorena et al., 2010). The missing data in the high latitudes during winter are filled with a 0.3 mg/m^{-3}, which is roughly the 20^{th} percentile of global chlorophyll-a. Lastly, we take the log transformation (base 10) of Chl-a.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

2.3.2 Data for regression and mapping

For the regression step of TA, the bottle measurements from the GLODAP v2 product are used as the target variable (Olsen et al., 2016). Following Lee et al. (2006), we select data shallower than 20 m in latitudes lower than 30° and shallower than 30 m at higher latitudes. The quality of TA measurements was historically not as rigorous as the SOCAT pCO_2 data due to the lack of reference standards before the mid-1990's (Bockmon and Dickson, 2015). However, most of the biases in the cruises were corrected based on calibration to deep samples, where it is assumed that interannual TA variability is negligible relative to the magnitude of the bias (Olsen et al., 2016). These bias corrections amount to ± 5 μmol kg^{-1} on average.

For the regression of pCO_2, we use SOCAT v2019 where only data with a SOCAT cruise quality flag of A to D and a WOCE quality flag of 2 are used. The fugacity of CO_2 (fCO_2) reported in SOCAT v2019 is converted to pCO_2 using:

$$pCO_2 = fCO_2 \cdot \exp(P_{\text{atm}}^{\text{surf}} \cdot \frac{B + 2 \cdot \delta}{R \cdot T^{\text{SOCAT}}}) \quad (1)$$

where P is atmospheric pressure at sea-level from the ERA5 reanalysis product (Copernicus Climate Change Service (C3S), 2017). B and δ are virial coefficients, R is the gas constant, and T^{SOCAT} is the ship intake temperature in °C (Dickson et al., 2007). In exploratory work for this study, we tested predicting $\Delta pCO_2 = pCO_2^{atm} - pCO_2$ instead of just pCO_2, but found that this did not produce credible results; for a more in depth discussion see Section A2.

The discrete measurements of pCO_2 and TA are resampled on a monthly grid (Jan 1985 through Dec 2018) with a spatial resolution of $1 \times 1°$ to match the predictors used in the mapping step.

Finally, outliers are removed from gridded pCO_2^{SOCAT} using the methods described in Section A1. In total, 2425 points are removed from the gridded pCO_2^{SOCAT} using these outlier removal approaches, equivalent to 0.85% of the original gridded data.

We use sea-surface temperature from OSTIA for both TA and pCO_2 regression (Merchant et al., 2019). The TA model is trained using *in situ* salinity from GLODAP v2 but salinity from SODA v4.3.2 is used for the mapping step (Carton et al., 2018). The N^* and silicic acid are the same as used in the clustering step, but are repeated for the number of years. Similarly, the mixed layer depth climatology described in section 2.3.1 is repeated for each year. We use the global mean of the mole fraction of CO_2 for the marine boundary layer (xCO_2^{mbl}) as a predictor in the regression as the correction for water vapor pressure may otherwise introduce co-variance with other predictors (*i.e.* SST, SSS). Missing data in the monthly Globcolour chlorophyll-a product is filled with climatological data described in section 2.3.1. The meridional and zonal components of the surface winds are averaged from the hourly output from the ERA5 reanalysis (Copernicus Climate Change Service (C3S), 2017).

2.3.3 Evaluation variables

The machine learning estimates of TA, pCO_2, and the computed DIC and pH are evaluated against data that are not used in the training or mapping step.

We use DIC and pH measurements from GLODAP v2.2019 (Olsen et al., 2016). Bockmon and Dickson (2015) report a measurement error of ± 5 μmol/kg for GLODAP DIC in an inter-laboratory comparison. Olsen et al. (2016) estimate the

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

measurement error for pH to be 0.005. To be consistent with TA, we select data shallower than 20 m in latitudes lower than 30° and shallower than 30 m at higher latitudes and resampled the data to monthly by $1° \times 1°$.

Three long term time series stations are used to provide direct independent comparisons against DIC and TA, namely: the Hawaii Ocean Time-series at 22.57°N, 158°W (HOT, Dore et al. 2009); the Bermuda Atlantic Time Series at 32°N, 64°W (BATS, Bates and Peters 2007); and the Irminger station for only DIC in the high northern Atlantic (64.3°N, 28°W, Olafsson et al. 2010). The accuracy for these measurements is reported to be below 2 μmol/kg for DIC and $\sim$4 μmol/kg for TA for all stations. We use the same depth constraints for the long term stations as for GLODAP, explained in the paragraph above. pCO$_2$ is also calculated from DIC and TA for HOT and BATS to provide an additional constraint.

Data present in the Lamont-Doherty Earth Observatory pCO$_2$ data set, but not in SOCAT are used to independently compare pCO$_2$. Takahashi et al. (2019) report an error estimate of $\pm 2.5 \mu$atm, but it must be added that some of the data unique to LDEO may be excluded from SOCAT due to stricter quality control criteria for of the latter, thus errors for the LDEO data are expected to be larger (Bakker et al., 2016).

Finally, we include Argo float measurements of pH from the Southern Ocean Carbon and Climate Observations and Modeling project (SOCCOM) (Johnson et al., 2017; Williams et al., 2017). Johnson et al. (2016) report a mean uncertainty of ± 0.019 for pH for the entire water column, though this is likely higher for the upper 30 m as the authors report lower errors for estimates below 50 m.

2.4 Computation of DIC, pH, and Ω

The remaining parameters of the marine carbonate system, i.e., DIC, pH, and Ω are computed using the Python version of *CO2SYS* (Humphreys et al., 2020) originally developed by Lewis et al. (1998). In addition to pCO$_2$ and TA, *CO2SYS* requires the input of sea-surface temperature, sea-surface salinity, pressure (assumed 0 dBar at the surface), phosphate and silic acid. We use the same data sources described in section 2.3.2 and Table 1. Climatologies of silicic acid and phosphate are repeated for each year, thus assuming no interannual variability. The dissociation constants by Dickson et al. (1990) for $K_{\mathrm{HSO_4}}$ and the total boron-salinity relationship by Uppström (1974) were used, as recommended by Orr et al. (2015) and Raimondi et al. (2019). For further details on the calculation and the full description of the marine carbonate system, see Dickson et al. (2007).

An important consideration in these calculations is the internal consistency of the marine carbonate system, i.e., the error due to uncertainties in the equations and coefficients that describe the marine carbonate system. Raimondi et al. (2019) pointed out that the pCO$_2$-TA pair has the lowest error in the calculation of pH (0.003 ± 0.008 pH units) using the dissociation constants by Mehrbach et al. (1973) as refitted by Dickson and Millero (1987). However, using the same pair and the same dissociation constants resulted in an estimate of Ω with respect to Aragonite that is very different from that computed using the DIC-TA pair. But since Raimondi et al. (2019) lacked direct measurements of Ω, it remains unclear which pair is actually better for Ω. We cannot resolve this here but need to acknowledge that this inconsistency adds some additional uncertainty to our computed Ω values.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

3 Uncertainty assessment

Any application of our data product requires a firm understanding of the uncertainties associated with each of the reported parameters of the surface ocean carbonate system. We first discuss the uncertainties associated with the statistically modeled quantities TA and $p\mathrm{CO}_2$, and then those with the computed parameters DIC, pH and Ω. Then, we will compare these propagated uncertainties with the uncertainty of the computed DIC and pH established by comparing these values with in situ observations. This provides a strong check on our ability to establish a full error budget.

3.1 Sources of uncertainty for TA and $p\mathrm{CO}_2$

We identify three sources of uncertainty that contribute to the total error for $p\mathrm{CO}_2$ and TA, namely the measurement (M), representation (R) and prediction (P) errors. Assuming independence of the three error sources, the total error (E) for the TA and $p\mathrm{CO}_2$ estimates can thus be expressed as the root of the squared sum of the three errors:

$$E^2 = M^2 + R^2 + P^2 \tag{2}$$

The **measurement error** reflects the combination of potential biases (systematic errors) from sampling and measurement as well as random errors associated with sampling and the imprecise nature of the measurement system. Since both TA and $p\mathrm{CO}_2$ are being measured against certified reference materials and have undergone extensive secondary quality control, we assume that they have no systematic error, i.e., that their bias is zero. We also assume the sampling error to be small, so that the measurement error M can be well approximated by the precision of the employed measurement methodology (Dickson et al., 2007).

The **representation error**, R, is a result of the fact that we develop our statistical model (GRaCER) on a grid that is in many places coarser in time and space than the typical scales of variability of TA and $p\mathrm{CO}_2$. As a result, any given observation may not be representative for the $1\times1^\circ$ by month grid cell used as a basis for our regression, leading to a bias in the estimated mean relative to the true spatial and temporal mean. This problem is particularly severe if the number of samples within any grid cell is low, and the spatio-temporal variability is high. This is often the case. For example, more than 90% of the data in the monthly gridded $p\mathrm{CO}_2$ product are based on a single day of sampling within the month. And with an order of magnitude fewer observations of TA, an even larger number of cells are populated by a single observation. Since we are lacking full knowledge of the spatial and temporal variability of TA and $p\mathrm{CO}_2$, we cannot fully quantify the representation error. Instead, we approximate it using the few regions where we have sufficient observations, or then using closely related parameters for which we have more observations. For simplicity, we make the assumption that this error is, on global average, normally distributed with a bias of 0.

The **prediction error**, P, is determined by the test scores from the evaluation of the statistical model vis-à-vis the independent test data. The test scores describe the error incurred in the prediction of the subset of data that is not used in the training step. This error includes also the propagated uncertainty associated with the predictor variables.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

We summarize these errors with mean biases and root mean squared error (RMSE). We thereby separate the coastal and open ocean regions using the COastal Segmentation and related CATchments (COSCATs) mask (Laruelle et al., 2013) in order to reflect their very different levels of spatio-temporal variability.

Table 3. Summary of the errors of total alkalinity and pCO_2 from the different error sources (see (2) separately evaluated for the open ocean and for coastal regions (defined by the COSCATs regions Laruelle et al. 2013). See text for details on how the different sources were quantified.

	Alkalinity (μmol kg^{-1})		pCO_2 (μatm)	
Uncertainty	Open ocean	Coastal	Open ocean	Coastal
Measurement	5 ($\leq$ 10)		2 ($\leq$ 5)	
Representation	16	34	7	17
Prediction	13	28	12	27
Total	21	45	14	32

3.1.1 Uncertainty for total alkalinity

We adopt a measurement error of TA of ± 5 μmol kg^{-1} based on the laboratory intercomparison by Bockmon and Dickson (2015). This is only half the measurement error of ± 10 μmol kg^{-1} or 0.5% reported by GLODAPv2. We consider this to be an overly conservative estimate, since Bockmon and Dickson (2015) pointed out that the majority of the laboratories involved in the round-robin exercise achieved an error of better $\pm 5 \mu$mol kg^{-1}. We thus opted for this lower value that is more representative of the majority of the data.

Owing to the sparseness of the TA observations, we cannot estimate the representation error directly. Instead, we use the fact that TA and salinity are very highly correlated. This permits us to determine the representation error for TA indirectly from an estimate of the representation error of sea-surface salinity. Concretely, we compare the test RMSE of TA predicted with GLODAPv2's *In-situ* salinity with the RMSE of TA predicted with the satellite based SODA salinity (see Table 1). Since the latter salinity is supposed to reflect the true time-space average over each grid cell, the difference between these two salinities is a direct estimate of the representation error for salinity. Consequently, the difference in TA from these two estimates is an estimate of the representation error for TA. The resulting estimates for the open and coastal ocean are summarized in Table 3.

The prediction error is based on the model's RMSE score calculated from test data and is listed in Table 3 and Figure 3a. The global mean prediction error for the open ocean amounts to 13 μmol kg^{-1}, with some regional differences. The prediction error is more than twice this number in the coastal regions, i.e., 28 μmol kg^{-1} (coastal regions are defined by the COSCATs regions Laruelle et al. 2013). We find especially high prediction errors, for example, in the highly dynamic Amazon outflow region or the Gulf of Maine in the northwestern Atlantic. However, in such regions, one can expect that part of the high prediction error is actually stemming from a representation error, as we are not using directly co-measured variables when we train our regression model.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

While the global bias of the TA product of OceanSODA-ETHZ is near zero (0.5 μmol/kg), confirming our assumption about the unbiased nature of our prediction error, this is not the case regionally. For example, OceanSODA-ETHZ tends to consistently overestimate TA in the southeastern Atlantic and underestimate TA in the southern Indian Ocean (Figure 3b). A seasonal breakdown of the biases into DJF and JJA reveals that the winter period of each hemisphere has biases in the high latitudes, though data paucity weakens this outcome.

A good check on the model prediction error is provided by comparing the estimated TA against independent observations. To this end, we use data from the Hawaii Ocean Time-series (HOT), the Bermuda Atlantic Time Series (BATS) and the Irminger station shown in Figures 4(*a,b,e,f,i,j*) and Table 4. For the period 1990-2018, the bias for BATS is 3 μmol kg^{-1} and for HOT -2 μmol kg^{-1} indicating that the method captures the interannual variability of TA well at these subtropical stations. Further, the mean seasonal cycle well is relatively well represented at HOT and BATS, being within one standard deviation of the interannual variability (Figure 4b,f). However, the results are not as good for the Irminger station in the Atlantic high latitudes ($\sim$65°N), where OceanSODA-ETHZ has a large negative bias (-10 μmol kg^{-1}) when compared to TA computed from pCO$_2$ and DIC measurements. OceanSODA-ETHZ also overestimates the weak seasonal cycle of TA at the Irminger station, contributing to the large bias that is particularly strong from December to May. The RMSE at Irminger station is 15 μmol kg^{-1}, less than 5 μmol kg^{-1} larger than the RMSE for HOT and BATS stations (10 μmol kg^{-1} respectively) owing to the small interannual and seasonal amplitude at Irminger. The RMSEs are thus smaller than the mean prediction error (13 μmol kg^{-1}) at the subtropical stations, yet exceeds this mean estimate at the high latitude station

3.1.2 Uncertainty for pCO$_2$

For the measurement error of pCO$_2$, we adopt a value of $\pm$2 μatm. This reflects the fact that 80% of the data we have used from SOCAT (flags A and B) have a precision better than that number. The remaining data we used (SOCAT flags C and D) have a measurement error of less than $\pm$5 μatm.

We estimate the representation error of pCO$_2$ on the basis of a spatio-temporal gradient analysis. To this end, we compare the pCO$_2$ in our regular grid that has a resolution of 1$\times$1° $\times$ 1 month, with the pCO$_2$ binned to a grid with twice this resolution, i.e., 0.5$\times$0.5$\times$ 15-days. In regions with high spatio-temporal coverage, the difference in the average of adjacent grid cells represents the potential change that can occur within the coarser 1$\times$1° $\times$ 1 month grid cell. The spatial and temporal gradients are calculated separately and we take the average of these two elements. Using this analysis, we estimate a representation error of our pCO$_2$ estimates of 7 μatm and 17 μatm for the open and coastal ocean respectively (Table 3).

From the RMSE of our test data, we estimate a pCO$_2$ prediction error of 12 μatm for the open ocean and 28 μatm for the coastal ocean (Table 3). Within the open ocean (Figure 3e), the eastern tropical Pacific has the highest RMSE, but this is also the region with the highest variance in the observations. The high RMSE for the coastal region stems primarily from coastal Antarctica as well as some coastal regions in the higher latitudes of the northern hemisphere. The former is due to large uncertainties in pCO$_2$ during the summer months when retreating ice and ensuing rapid net primary production result in large gradients (Bakker et al., 2008).

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

Table 4. Comparison of training and independent data sources with various methods for the open ocean region using the COSCATs coastal mask by Laruelle et al. (2013). GLODAP refers to the GLODAP v2 2019 data, HOT to the Hawaii Ocean Time-series, BATS to Bermuda Atlantic Time Series, SOCAT is the 2019 version of the Surface Ocean Carbon Atlas, SOCCOM is the pH measured by autonomous floats from the Southern Ocean Carbon and Climate Observations and Modeling project. Statistical outliers were excluded in the calculation of LDEO RMSE. OS-ETHZ is the OceanSODA-ETHZ data from this study, NNGv2 is from Broullón et al. (2018), LIARv2 from Carter et al. (2018), FFNNv2 from Denvil-Sommer et al. (2019), and SOMFFN from Landschützer et al. (2016). NNGv2 and LIARv2 predictions are made with SODA salinity and OSTIA sea surface temperature resulting in different estimates to the original publications (Broullón et al., 2018; Carter et al., 2018). Note that the full data set is used for OceanSODA-ETHZ, unlike Table 3 which presents the errors for test years.

		TA (μmol kg^{-1})			pCO_2 (μatm)		DIC (μmol kg^{-1})			pH	
		GLODAP	HOT	BATS	SOCAT	LDEO	GLODAP	HOT	BATS	GLODAP	SOCCOM
Bias	this study	0.5	-2.1	2.6	-0.4	0.1	0.5	-1.0	0.4	-0.001	0.009
	LIAR + FFNN	0.3	-3.1	0.2	0.5	0.6				0.001	0.013
	NNGv2	1.2	-3.2	4.3			2.3	2.2	-0.4		
	SOMFFN				0.4	0.4					
RMSE	this study	17.5	9.5	10.1	11.1	19.9	16.3	8.7	9.1	0.024	0.036
	LIAR + FFNN	18.0	8.8	8.8	13.1	19.6				0.023	0.037
	NNGv2	16.2	6.7	10.4			23.1	9.5	15.2		
	SOMFFN				11.7	21.4					
r^2	this study	0.91	0.58	0.13	0.82	0.45	0.93	0.77	0.76	0.67	0.047
	LIAR + FFNN	0.91	0.6	0.21	0.78	0.44				0.67	-0.043
	NNGv2	0.93	0.75	-0.1			0.82	0.71	0.38		
	SOMFFN				0.82	0.49					

The comparison between the regression estimated and observed pCO_2 reveal also some regional biases (Figure 3e), even though the global bias is very close to zero (-0.37 μatm). Some of the highest biases are found, again, in the eastern tropical Pacific, where strong gradients and large interannual variability may drive the observed juxtaposed biases. The large negative biases in winter in the Southern Ocean are likely driven by the paucity of data in this region (Gregor et al., 2019).

The time series comparisons show that the seasonal cycle is well represented at BATS and HOT with r^2 scores of 0.89 and 0.82 respectively (Figures 4d,h). Low biases (< 2 μatm in absolute terms) further demonstrate that pCO_2 estimates are reliable in the subtropics. The seasonal cycle is also well captured at the Irminger station in the high latitudes, but a lower r^2 score and larger bias (-8.0 μmol kg^{-1}) allude to the dampened amplitude of the seasonal cycle particularly in the winter months (Figure 4).

3.2 Uncertainties of the calculated parameters

We determine the uncertainties of the calculated parameters in two ways. First, we propagate the uncertainties of pCO_2 and TA through pyCO2SYS (Orr et al., 2018; Humphreys et al., 2020) onto the computed parameters DIC, pH, and Ω. This yields

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

an expected error that we refer to as a "bottom-up" total error estimate. Second, we obtain a "top down" error estimate, by comparing the calculated DIC and pH with independent measurements (see Table 4). We first describe the top-down estimates and then the bottom-up. For the top-down estimate, we use the GLODAP DIC and pH data as our independent test of the method's performance. This is because these data are not used in any way in our estimation.

Global estimates of DIC in OCEANSODA-ETHZ have a low bias when compared with in situ GLODAP measurements (0.5 μmol/kg) and are the same as the global TA bias (Table 4). The spatial distribution of the DIC biases shows a more nuanced picture, with large positive biases in the western equatorial Pacific and negative biases in the western equatorial Atlantic. The bias in the western equatorial Atlantic matches the negative bias in the pCO_2 in the same region; however, the source of the DIC bias in the Pacific is not clear from the pCO_2 and TA test data biases. The global mean of the top-down error estimate for DIC (16.3 μmol $^{-1}$ for the open ocean) is then perhaps a better reflection of the uncertainty as positive and negative values don't cancel each other out. The prediction errors for OceanSODA-ETHZ of DIC are lower across all independent data when compared with the NNGv2 (23.1 μmol kg^{-1}).

The comparison of the DIC time series data (BATS, HOT and Irminger stations) supports the findings of the global top-down estimates (Figure 6). The biases are relatively low for HOT and BATS (-1.0 and 0.4 μmol kg^{-1} respectively), but the bias is much larger at Irminger station (-6.9 μmol kg^{-1}) which is at a much higher latitude (64°N) compared to HOT and BATS in the subtropics (< 35°N). However, this bias is not reflected in the zonal average of the seasonal biases (Figure 5c). Similarly, the RMSE is also larger at Irminger station (14 μmol/kg) compared to HOT and BATS ($\sim$9 μmol kg^{-1}) in the subtropics). Despite these differences in the top-down error, the same amount of variability is represented by the OCEAN-SODA DIC for all three stations ($\sim$0.76) owing to the larger seasonal cycle at Irminger station.

The pH comparison with the GLODAP pH measurements shows that OceanSODA-ETHZ has a negligible bias (0.001). As with DIC, regional biases are larger than the global average, with the coastal and high latitude oceans contributing significantly to the regional biases. The GLODAP comparison of pH RMSE is also low (0.024), but is outperformed by the RMSE of pH calculated with LIARv2 TA and FFNNv2 pCO_2 (made available in the FFNNv2 data set, 0.023) but given the uncertainty of GLODAP pH (0.005), the difference is negligible.

For the "bottom-up" estimate, we propagate the total error of pCO_2 and only the measurement and prediction errors for TA. We do this to avoid including the representation error twice in the bottom-up estimate, as we hypothesize that the representation error of TA is largely accounted for by the representation error of pCO_2 and vice versa. We choose to use the representation error of pCO_2 rather than TA as the larger number of samples gives us greater confidence in the estimate. Moreover, we feel that the assumptions we make in the estimate of the TA representation error are larger than those for pCO_2, thus further justifying our choice in using only the representation error of pCO_2.

The comparison of top-down vs bottom-up error estimates for open and coastal oceans is shown in Figure 7 with values of the total errors also shown. The top-down and bottom-up error estimates for DIC are relatively accurate with the estimates being within 5% of each other in the open ocean and 9% in the coastal ocean. However, the estimates are not as coherent for pH where there the bottom-up error is 23% smaller top-down error than the in the open ocean. The difference is even bigger in the coastal ocean where the bottom-up is 31% smaller than the top-down error.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

4 Results

4.1 Comparison with other climatologies

A spatial comparison between OceanSODA-ETHZ and existing products might reveal potential biases in our product if the bias is present in all comparisons. We compare TA against LIARv2 and NNGv2 (Carter et al., 2018; Broullón et al., 2018) by first taking the difference for each month and then calculating the average for these differences. The same approach is used to compare pCO_2 with SOMFFN and FFNNv2 (Landschützer et al., 2016; Denvil-Sommer et al., 2019).

The differences in TA between OceanSODA-ETHZ and NNGv2 and LIARv2 are on the same order of magnitude in the open ocean as the prediction error (13 μmol/kg) but are slightly larger on average for NNGv2 than for LIARv2 (Figure 8a,b). There is also some agreement in the spatial distribution of the distribution, particularly in data sparse parts of the Pacific and Indian Oceans. The larger differences may stem from the fact that both LIARv2 and NNGv2 are more constrained by spatial coordinates than GRaCER. NNGv2 uses latitude and longitude as a predictor while the LIARv2 approach interpolates linear regression coefficients for every 5° grid cell (Broullón et al., 2018; Carter et al., 2018). The divergence in data sparse regions is thus not surprising. Though, it must be emphasized that this comparison serves more as a "sanity check" than as a ground truthing exercise.

The differences between OceanSODA-pCO_2 and FFNNv2 and SOMFFN are smaller and not as spatially coherent than those for TA. The differences are marginally larger for the OceanSODA - FFNNv2 than for OceanSODA - SOMFFN (Figure 8c,d). This is consistent with the data in Table 4, where the metrics for the SOMFFN are very similar to OceanSODA-ETHZ. The smaller difference between the latter should not come as a surprise as the GRaCER approach is built on the two-step cluster-regression approach of the SOMFFN, while the FFNNv2 approach includes spatial coordinates (Landschützer et al., 2016; Denvil-Sommer et al., 2019). In general, the dissimilarity between the differences is encouraging as it indicates that OceanSODA-ETHZ is not consistently biased relative to SOMFFN and FFNNv2.

4.2 Seasonal Climatologies

The climatological mean spatial distribution of TA, pCO_2, DIC, pH and Ω, obtained by averaging the estimates from 1985 through 2018, reveal a very rich and diverse pattern of variability with commonalities and differences (Figure 9). The climatological maps are accompanied by Hovmoeller diagrams that show the zonal average of the seasonal cycle for each of the variables (Figure 10). We also show climatological time series for each of the variables at high (55°N, 170°W), mid (30°N, 170°W) and low (10°N, 170°W) latitude locations.

Total alkalinity shows the largest differences between basins, with the mean alkalinity being much higher in the saltier Atlantic than in the Pacific and Indian basins (Boutin et al., 2018) (Figure 9a). The spatial variability of TA on a global scale exceeds the variability of the seasonal cycle (Figure 10a,e). Seasonal variability of TA is on the order of 20 μmol kg^{-1} at the chosen mid and high latitude locations, while the latitudinal gradient is as large as 150 μmol kg^{-1}. However, much of the TA seasonality is driven by seasonal changes in salinity due to precipitation and ice melt in the respective regions.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

Dissolved inorganic carbon is more homogeneous across the basins, but has a much larger meridional gradient than TA, amounting to more than 150 μmol kg^{-1} (Figure 9b). This meridional gradient is seasonally substantially more modified than is the case for TA, particularly in the high latitudes of the northern hemisphere where the seasonal cycle is as large as 100 μmol kg^{-1} 10b,g). The larger seasonal cycle in DIC is due to the carbon update by spring-time phytoplankton blooms and stratification during the warmer seasons (Siegel et al., 2002). The magnitude of the spring-time blooms is dampened by iron limitation in the Southern Ocean, visible by a smaller seasonal cycle amplitude (40 μmol kg^{-1}) (Watson et al., 2000; Tagliabue et al., 2017). However, the background DIC concentration in the Southern Ocean is much larger due to upwelling of DIC-rich circumpolar deep waters driven by the persistent westerlies south of 50°S (Marshall and Speer, 2012).

The spatial distribution and seasonal cycles of pCO$_2$ and pH (Figures 9c,d) and 10c,d,g,h) are strongly negatively correlated due to the inverse stoichiometric relationship between dissolved aqueous CO$_2$ and [H$^+$] (Dickson et al., 2007). The reduction in DIC is concomitant with the reduction in pCO$_2$ in the high latitudes due to the biological uptake of [CO2] (10c,g). However, this relationship does not hold true in the mid latitudes, where the slight decrease in DIC is contrasted by a relatively strong increase in pCO$_2$. This is due to the positive temperature dependence of pCO$_2$ (the opposite is true for pH) (Takahashi et al., 1993), which will be elaborated on in the discussion.

The spatial distribution of Ω (Figures 9e,f and 10e,j) strongly reflects the concentration of the carbonate ion, which can be well approximated by the difference between TA and DIC (Sarmiento and Gruber, 2006). Given that the seasonal cycle of TA is much weaker than DIC, the latter dominates the seasonal cycle of Ω_{arag} (Figures 10e,j). This would also be true for Ω_{calc} which only differs in magnitude and not in distribution or seasonality (the latter is not shown).

5 Discussion

5.1 Choosing the appropriate machine learning configuration

Here we consider here two notable decisions that have a large impact on the final estimates: 1) the use of the ensemble approach, and 2) the choice of regression algorithm. For details on the minor choices, see section A3 in the supplementary materials.

As previously motivated, we opt for the cluster-regression approach that is able to generalize estimates in sparse regions due to information sharing within a cluster. However, cluster boundaries are often semi-discrete, resulting in artifactual boundaries in estimates. This makes the output of cluster-regression approaches less suitable for studies where gradients over short time periods or distances are assessed, e.g. detection of extreme events. Our approach removes these boundaries and improves the robustness of the estimate by eliminating, to large extent, the sensitivity of the regression to the clustering algorithm.

The second major consideration is the choice of regression algorithm. Our choice of different algorithms for TA and pCO$_2$ may seem peculiar; however, this decision was informed by the nature of the problems. While regressing TA and pCO$_2$ are conceptually similar, the size of each data set and the distribution of training samples sets them apart. When gridded to a monthly $\times 1\,\mathrm{deg} \times 1\,\mathrm{deg}$ resolution the number of data are $\sim 300\,000$ for SOCAT pCO$_2$ and $\sim 16\,000$ for GLODAP TA; *i.e.*, nearly 20 times more data for the former. However, the strong linear correlation between TA and salinity ($\mathrm{r} = 0.96$) compensates

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

for the poor sampling distribution; that is, as long as the regression method is able to extrapolate - a criterion that support vector regression (SVR) meets.

pCO$_2$ is not highly correlated to any proxies, suggesting that the regression problem is more complex, thus requiring a method that is appropriately non-linear. Gradient-boosted decision trees (GBDT) and feed-forward neural-networks (FFNN) meet this criterion. But why not just use one of these approaches? Work by Gregor et al. (2019) found that an ensemble of methods (SVR, GBDT and FFNN) outperformed each individual member. And while SVR performed well in Gregor et al. (2019), the method does not scale to larger problems, which GBDT and FFNN are capable of leading to our choice of the latter two. One critique of GBDT in this application may be that, being a tree-based method, it is not able to extrapolate. However, we feel that the cluster-regression approach combined with the large number of training data for pCO$_2$ compensates for this shortcoming. GBDT provide also provide useful diagnostics such as feature importances that, when combined with the GRaCER approach, provides useful information about the spatial and seasonal importance of the proxies.

5.2 Why are pH uncertainties less well-constrained?

One of the novel contributions of this study is that we are able to assert the validity of our results by comparing the bottom-up (propagated) with the top-down errors (*in situ* comparisons). Using this approach, we show that the error estimates of DIC are remarkably well-constrained, with the top-down being within 5% of the bottom-up error estimate for the open ocean (7b). The same statistic for pH yields a 23% difference between the two budgets. The question is thus, why are pH bottom-up and top-down error estimates not so consistent?

To assess this problem from the bottom-up perspective, we need to consider the uncertainties of pCO$_2$ and TA. But given that the DIC error budgets are well constrained, we can, with some certainty, rule out the bottom-up error estimate as the source for the larger mismatch in pH.

The source of the mismatch must thus be driven primarily by uncertainties in the top-down perspective, where it may be that the representation error of pH is larger than for DIC. We can immediately rule out the measurement error as a contributor to the mismatch, as the bias of the measurements (provided accurate calibration to reference samples) should be normally distributed around zero. Thus, representation error is a more likely candidate, due to the temperature and pressure sensitive nature of pH compared to the conservative nature of DIC *w.r.t.* the same variables (Dickson et al., 2007). This is important given that our "surface" pH data from GLODAP can be as deep as 30 m, while pCO$_2$ is typically measured at a 10 m depth. A basic sensitivity study of the variability of pH in the surface layer shows that the median standard deviation of "surface" pH at a single station is 0.004 units (even when normalized to a standard temperature and depth). The same approach for DIC yields a standard deviation of 2 μmol kg^{-1}. This suggests that the vertical representation error of GLODAP alone already explains a good deal of the mismatch in the error budgeting. One might reduce this uncertainty by placing tighter constraints on the definition of "surface" pH, limiting the depths between 3 m and 15 m, for example, but at the loss of valuable test samples. The representation error from horizontal representation errors could further explain the remaining disparity in the pH error budget.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

5.3 Can we reduce the total error?

The last two decades have seen major advances in the reduction of the measurement uncertainty for both TA and $p\mathrm{CO}_2$. The introduction of certified reference materials for TA and a standardized approach for measuring $p\mathrm{CO}_2$ with reference gases means that the measurement uncertainties are low.

In contrast, the prediction uncertainty is the largest contributor to the total error for both DIC and TA suggesting that this could be a fruitful avenue to pursue. However, current literature suggests that this is unlikely. Gregor et al. (2019) showed that within a selection of six gap filling methods, all achieved similar accuracy scores when compared with independent data, upon which the authors suggested that we have hit a "wall".

This leaves the representation error, which contributes a moderate fraction to the total $p\mathrm{CO}_2$ and TA errors in the open ocean and even less in the coastal ocean. A back-of-the-envelope calculation shows that increasing the resolution of $p\mathrm{CO}_2$ fourfold (from monthly by 1°x1° to 8-daily by 0.25°x 0.25°) could decrease the representation error by 2.5 μatm (35%) for the open ocean and 3.2 μatm for the coastal ocean (20%). This is perhaps not as much as expected, but these small gains are larger than those that are currently being made in the prediction uncertainties (Gregor et al., 2019). This is not applicable for TA, where ungridded values are already used to train GRaCER and the predictor variables are more likely contributors.

Why are these gains smaller than hoped? It may be that our gradient approach for calculating the representation error breaks down as the resolution increases due to the decreasing number of adjacent grid points. This is hardly surprising considering that 78% of grid cells in the SOCAT v2019 monthly gridded product are represented by sampling on a single day that falls within that period (Bakker et al., 2016). Another possibility is that decreasing both the spatial and temporal resolution exposes the sharp mesoscale gradients that are otherwise averaged over at larger resolutions (Resplandy et al., 2009; Monteiro et al., 2015). The gains to be made in the total uncertainty are thus small by increasing the resolution of the prediction data.

5.4 Regional sensitivity of $p\mathrm{CO}_2$ to driver variables

Here we demonstrate one of the possible ways in which the OceanSODA-ETHZ data can be used to gain further insight into the marine carbonate system.

We decompose and attribute the mean seasonal cycle variability of $p\mathrm{CO}_2$ to its drivers, namely TA, DIC, temperature and salinity. Past studies using observation based-products have been limited to a simpler thermal/non-thermal decomposition of $p\mathrm{CO}_2$ due to the lack of DIC and TA (Takahashi et al., 2002; Landschützer et al., 2015). This is thus the first time that a full decomposition of $p\mathrm{CO}_2$ has been applied to observation-based data for a global domain. The decomposition is performed with *pyCO2SYS* by keeping all but one of the drivers constant (to the average) and assess the influence on $p\mathrm{CO}_2$ (Humphreys et al., 2020). This is conceptually similar to the decomposition applied to ocean simulation output by Lovenduski et al. (2007). An important note to make, considering that we intend that OceanSODA-ETHZ is used for ocean acidification studies, is that the decomposition of pH would result in virtually the same contribution of the drivers (Dickson et al., 2007).

The seasonal amplitude of $p\mathrm{CO}_2$ is driven predominantly by changes in DIC in the high and equatorial latitudes, and by temperature in the mid latitudes (Figure 11d). At first glance, these results may seem similar to those that the simpler thermal

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

decomposition might result in, requiring only temperature. However, in regions where the seasonal amplitude of $p\mathrm{CO}_2$ is smaller, the importance of TA becomes more apparent. For example, at the mid-latitude station (Figure 11b), TA and DIC synergistically act to dampen the impact of temperature on $p\mathrm{CO}_2$. Conversely, at the equatorial station (Figure 11c) the effect of TA on $p\mathrm{CO}_2$ opposes that of the more dominant DIC. Further, there are regions in the tropics where TA is the dominant driver due to the weak seasonal cycle of both temperature and DIC.

6 Summary

Our approach for estimating TA and $p\mathrm{CO}_2$ is an evolution of the cluster-regression approach: we create an ensemble of estimates by repeating the cluster-regression step multiple times, each with a different variation of clustering. We call this approach the Geo-spatial Random Cluster Ensemble Regression (GRaCER). The result is an estimates that are more robust with better generalization and the output does not have the discrete cluster boundaries that single member cluster-regression approaches have.

We find that our estimates of TA are within the ballpark of of previous methods with a prediction error (root mean square error) of 13 μmol kg^{-1} for open ocean estimates, while biases are < 1 μmol kg^{-1}. Taking into consideration all sources of error (measurement and representation errors), the total error is 17 μmol kg^{-1} for TA. The prediction error for $p\mathrm{CO}_2$ in the open ocean is 12 μatm, also with a bias of < 1 μatm. Including the measurement an representation errors for $p\mathrm{CO}_2$ results in a total error of 14 μatm for the open ocean. We estimate the total error of DIC and pH to be 19 μmol kg^{-1} and 0.022 units when compared with independent GLODAPv2 data for the open ocean. Finally, we compare the aforementioned "top-down" error estimates of DIC and pH with the "bottom-up" error estimates that are calculated by propagating the TA and pCO2 total error estimates through the marine carbonate system. This budgeting approach shows that we have a good grasp on the uncertainties of DIC for both the open and coastal oceans. However, pH uncertainties are not as well resolved, most likely due to a mismatch in the representivity of measured pH.

Lastly, we demonstrate a use case of the OceanSODA-ETHZ data set in which we decompose the seasonal variability of $p\mathrm{CO}_2$ into four driver components of DIC, TA, temperature and salinity. We find that DIC is the dominant driver in the high and equatorial latitudes, while temperature contributes the majority of the signal in the subtropics. Importantly, DIC and temperature are antagonistic drivers of pCO2, while alkalinity always acts in opposition to the stronger of the two primary drivers.

Finally, OceanSODA-ETHZ will be maintained and updated for future work.

7 Code and data availability

Software for the GRaCER framework is available on GitHub (access provided on request). The OceanSODA-ETHZ dataset is available at https://doi.org/10.25921/m5wx-ja34 (Gregor and Gruber, 2020).

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

Appendix A: Supplement to the Methods

A1 pCO_2 outlier removal

The first outlier removal method requires the pCO_2 to be adjusted from the ship intake temperature to the satellite SST as described by (Goddijn-Murphy et al., 2015):

$$pCO_2^{SST} = pCO_2^{SOCAT} \times \exp(0.0433 \cdot (T^{SST} - T^{SOCAT})) \quad \text{(A1)}$$

where the ship intake depth varies due to inconsistent depth between vessels and the water column state (e.g. well stratified or mixed). Here T^{SST} is the foundation temperature given by the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) v2. The OSTIA product is matched to the ungridded pCO_2^{SOCAT} at daily by 0.25°×0.25° resolution (Donlon et al., 2012). The corrected pCO_2 is then binned to monthly by 1°×1° without weighting. Data are excluded where the absolute difference between pCO_2^{SST} and pCO_2^{SOCAT} is larger than 40 μatm.

Secondly, we exclude data that lie outside the expected ranges for the monthly climatology of pCO_2. The expected ranges are defined using the interquartile range outlier detection method for each pixel in a given month with the following equation:

$$IQR = Q_3 - Q_1 \quad \text{(A2)}$$

$$\text{lower limit} = Q_1 - IQR \cdot 1.5 \quad \text{(A3)}$$

$$\text{upper limit} = Q_3 + IQR \cdot 1.5 \quad \text{(A4)}$$

where Q_1 and Q_3 are the 25^{th} and 75^{th} percentiles, respectively. This approach is only applied where there are enough data present for a particular month of the year.

A2 Target variable: ΔpCO_2 vs pCO_2

In this study, one of the avenues that explored was to predict ΔpCO_2 instead of pCO_2. Motivation for predicting ΔpCO_2 is that it might allow new measurements from recent years to add new information about the seasonal cycle in regions where sampling was previously seasonally biased, *e.g.* the SOCCOM float data (Gray et al., 2018). Bushinsky et al. (2019) showed that including the new information about the seasonal cycle of pCO2 in machine learning estimates resulted in stronger winter outgassing, but their results could only show this for the period that SOCCOM float data are present. One of the reasons that machine learning approaches are not able to propagate this information back through time is that the larger pCO2 is "anchored" by the atmospheric CO2 concentrations that are used as a proxy. Atmospheric pCO_2 is required as a predictor variable to capture the interannual signal of pCO_2. Predicting ΔpCO_2 might thus allow one to remove atmospheric pCO_2 as a driver because the interannual term trend of pCO_2 is removed.

The results appeared promising, but on further investigation we found that the regressions that were trying to predict ΔpCO_2 were not able to represent the increasing strength of the sink. Further, we found that the interannual variability of pCO_2 was reduced compared to results that include atmospheric xCO_2 as a driver. Ultimately, we abandoned the approach.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020


Open Access Earth System Science Data Discussions

A3 Hyper-parameter selection for regression methods

A3.1 Total alkalinity: support vector regression

Hyper-parameters for the support vector regression (SVR) were chosen on a per-cluster basis using grid search cross validation, where unshuffled K-fold cross validation with five splits was used. The νSVR variety of the algorithm from the *scikit-learn* package in Python was used. The parameters C, γ and ν were selected.

A3.2 pCO_2: Gradient boosted decision trees

We used the *LightGBM* package to perform the gradient boosted regression with decision trees (GBDT). The GBDT algorithm was trained using early stopping, which determines the number of trees used in the model – typically one of the most important hyper-parameters. Every every fifth year from 1987 to 2019 was set aside as the validation data used in the early stopping. The total number of leaves per tree and the minimum number of training points per terminal leaf were both set to $N^{0.5}$, where N is the number of training points in a given cluster. The number of leaves per tree determines the size of the tree. The difference with *LightGBM* compared to other packages, like *XGBoost*, is that trees are grown on a leaf wise basis rather than a level-wise basis, where the depth of the tree would be a more important hyper-parameter. The minimum number of training points per terminal leaf determines how many points are aggregated in an estimate – a small number could thus result in over-fitting. The value $N^{0.5}$ was determined experimentally with a single ensemble member, where the optimal values were determined with K-Fold cross validation. The results were in the ball-park of $N^{0.5}$ showing relatively low sensitivity to changes in these hyper-parameters. Further, the learning rate was set to 0.2 and $L1$ and $L2$ regularization were both set to 20.

A3.3 pCO_2: Feed-forward neural network

Given that the problem of solving pCO_2 is not very complex (*i.e.* it is within the capability of a single layer neural network), the multi-layer perceptron regressor from the *scikit-learn* package was used. The size of the hidden layer for each cluster was determined by shuffled K-fold cross validation with five splits. The maximum number of weights in a hidden layer was set to $N^{0.55}$. Back propagation was performed using the Adam optimizer. The learning rate of the optimizer was selected in the cross validation process. Early stopping was used to speed up the training process and prevent over-fitting where one random third of the data were used in early stopping.

A3.4 Feature importances of pCO_2

Author contributions. LG & NG conceived of the study and developed the method. LG performed the analysis and testing of the method and wrote the paper with substantial input by NG.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Competing interests. Both authors declare that they don't have any competing interests.

675 *Acknowledgements.* We are deeply indebted to the scientists who sampled, analyzed, and contributed to the global data bases for ocean carbon data, namely the Surface Ocean CO2 Atlas (SOCAT) and the Global Ocean Analysis Project (GLODAP). We also thank the funding agencies that made these efforts possible. SOCAT and GLODAP are international efforts, endorsed by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS) and the Integrated Marine Biosphere Research (IMBeR) program. This work was financially supported by ESA's OceanSODA project (contract No. 4000112091/14/I-LG) and the European Com-
680 mission through Horizon 2020 research and innovation programme. We than Meike Vogt, Fabio Benedetti and Peter Land for their valuable discussions.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

References

Bakker, D. C., Hoppema, M., Schr, M., Geibert, W., and Baar, H. J. W. D.: A rapid transition from ice covered CO2-rich waters to a biologically mediated CO2 sink in the eastern Weddell Gyre, Biogeosciences, 5, 1373–1386, 2008.

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S.-i., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T. T., Tilbrook, B., Wada, C., Wanninkhof, R. H., 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., Ibánhez, 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., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT), Earth System Science Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, http://www.earth-syst-sci-data.net/8/383/2016/, 2016.

Bates, N. R. and Peters, A. J.: The contribution of atmospheric acid deposition to ocean acidification in the subtropical North Atlantic Ocean, Marine Chemistry, 107, 547–558, https://doi.org/10.1016/j.marchem.2007.08.002, 2007.

Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santana-Casiano, J. M.: A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic CO2 and ocean acidification, Oceanography, 27, 126–141, https://doi.org/10.5670/oceanog.2014.16, 2014.

Bednaršek, N., Feely, R. A., Howes, E. L., Hunt, B. P. V., Kessouri, F., León, P., Lischka, S., Maas, A. E., McLaughlin, K., Nezlin, N. P., Sutula, M., and Weisberg, S. B.: Systematic Review and Meta-Analysis Toward Synthesis of Thresholds of Ocean Acidification Impacts on Calcifying Pteropods and Interactions With Warming, Frontiers in Marine Science, 6, https://doi.org/10.3389/fmars.2019.00227, 2019.

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 CO2 variables and nutrient concentrations from T, S, and O2 data using Bayesian neural networks, Frontiers in Marine Science, 5, 1–29, https://doi.org/10.3389/fmars.2018.00328, 2018.

Bockmon, E. E. and Dickson, A. G.: An inter-laboratory comparison assessing the quality of seawater carbon dioxide measurements, Marine Chemistry, 171, 36–43, https://doi.org/10.1016/j.marchem.2015.02.002, http://dx.doi.org/10.1016/j.marchem.2015.02.002, 2015.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P. R., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models, Biogeosciences, 10, 6225–6245, https://doi.org/10.5194/bg-10-6225-2013, 2013.

Boutin, J., Vergely, J. L., Marchand, S., D'Amico, F., Hasson, A., Kolodziejczyk, N., Reul, N., Reverdin, G., and Vialard, J.: New SMOS Sea Surface Salinity with reduced systematic errors and improved variability, Remote Sensing of Environment, 214, 115–134, https://doi.org/10.1016/j.rse.2018.05.022, https://doi.org/10.1016/j.rse.2018.05.022, 2018.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Garcia, H. E., Johnson, D. R., Mishonov, A. V., O'Brien, T. D., Seidov, D., Smolyar, I. I., Zweng, M. M., Paver, C. R., Locarnini, R. A., Reagan, J. R., Forgy, C. C., Grodsky, A., and Levitus, S.: World ocean database 2013, Tech. rep., National Oceanographic Data Center (U.S.), Ocean Climate Laboratory,, https://doi.org/10.7289/V5NZ85MT, 2013.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Broecker, W. S. and Peng, T.-h. T.-H.: Gas exchange rates between air and sea, Tellus, 26, 21–35, https://doi.org/10.3402/tellusa.v26i1-2.9733, http://tellusa.net/index.php/tellusa/article/view/9733, 1974.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T. T., Key, R. M., González-Dávila, M., Tanhua, T., Jeansson, E., Kozyr, A., and van Heuven, S. M. A. C.: A global monthly climatology of total alkalinity: a neural network approach, Earth System Science Data Discussions, 2013, 1–31, https://doi.org/10.5194/essd-2018-111, 2018.

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., Resplandy, L., Johnson, K. S., and Sarmiento, J. L.: Reassessing Southern Ocean air-sea CO2 flux estimates with the addition of biogeochemical float observations, Global Biogeochemical Cycles, 33, 1370–1388, https://doi.org/10.1029/2019GB006176, 2019.

Carter, B. R., Williams, N. L., Gray, A. R., and Feely, R. A.: Locally interpolated alkalinity regression for global alkalinity estimation, Limnology and Oceanography: Methods, 14, 268–277, https://doi.org/10.1002/lom3.10087, 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, Limnology and Oceanography: Methods, 16, 119–131, https://doi.org/10.1002/lom3.10232, http://doi.wiley.com/10.1002/lom3.10232, 2018.

Carton, J. A., Chepurin, G. A., and Chen, L.: SODA3: A new ocean climate reanalysis, Journal of Climate, 31, 6967–6983, https://doi.org/10.1175/JCLI-D-17-0149.1, 2018.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the Global Ocean with Biogeochemical-Argo, Annual Review of Marine Science, 12, 23–48, https://doi.org/10.1146/annurev-marine-010419-010956, 2020.

Cooley, S. R. and Doney, S. C.: Anticipating ocean acidification's economic consequences for commercial fisheries, Environmental Research Letters, 4, https://doi.org/10.1088/1748-9326/4/2/024007, 2009.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, https://cds.climate.copernicus.eu/cdsapp{#}!/home, 2017.

Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean p CO 2 over the global ocean, Geoscientific Model Development, 12, 2091–2105, https://doi.org/10.5194/gmd-12-2091-2019, https://www.geosci-model-dev.net/12/2091/2019/, 2019.

Dickson, A. G., Wesolowski, D. J., Palmer, D. A., and Mesmer, R. E.: Dissociation constant of bisulfate ion in aqueous sodium chloride solutions to 250°C, Journal of Physical Chemistry, 94, 7978–7985, https://doi.org/10.1021/j100383a042, 1990.

Dickson, A. G., Sabine, C. L., and Christian, J. R., eds.: Guide to best practices for ocean CO2 measurements, North Pacific Marine Science Organization, Sidney, British Columbia, 3 edn., https://www.oceanbestpractices.net/handle/11329/249, 2007.

Dlugokencky, E., Thoning, K., Lang, P., and Tans, P.: NOAA Greenhouse Gas Reference from Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, ftp://aftp.cmdl.noaa.gov/data/trace{_}gases/co2/flask/surface/, 2019.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO2 Problem, Annual Review of Marine Science, 1, 169–192, https://doi.org/10.1146/annurev.marine.010908.163834, 2009.

Doney, S. C., Busch, D. S., Cooley, S. R., and Kroeker, K. J.: The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities, Annual Review of Environment and Resources, 45, 1–30, https://doi.org/10.1146/annurev-environ-012320-083019, 2020.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sensing of Environment, 116, 140–158, https://doi.org/10.1016/j.rse.2010.10.017, http://dx.doi.org/10.1016/j.rse.2010.10.017, 2012.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

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, Proceedings of the National Academy of Sciences, 106, 12 235–12 240, https://doi.org/10.1073/pnas.0906044106, http://www.pnas.org/cgi/doi/10.1073/pnas.0906044106, 2009.

Fabry, V. J., Seibel, B. A., Feely, R. A., and Orr, J. C.: Impacts of ocean acidification on marine fauna and ecosystem processes, ICES Journal of Marine Science, 65, 414–432, https://doi.org/10.1093/icesjms/fsn048, https://academic.oup.com/icesjms/article/65/3/414/789605, 2008.

Fabry, V. J., McClintock, J. B., Mathis, J. T., and Grebmeier, J. M.: Ocean acidification at high latitudes: The Bellwether, Oceanography, 22, 160–171, https://doi.org/10.5670/oceanog.2009.105, 2009.

Feely, R. A., Doney, S., and Cooley, S.: Ocean Acidification: Present Conditions and Future Changes in a High-CO2 World, Oceanography, 22, 36–47, https://doi.org/10.5670/oceanog.2009.95, https://tos.org/oceanography/article/ocean-acidification-present-conditions-and-future-changes-in-a-high-co2-wor, 2009.

Franco, A. C., Gruber, N., Frölicher, T. L., and Kropuenske Artman, L.: Contrasting Impact of Future CO2 Emission Scenarios on the Extent of CaCO3 Mineral Undersaturation in the Humboldt Current System, Journal of Geophysical Research: Oceans, 123, 2018–2036, https://doi.org/10.1002/2018JC013857, 2018.

Gattuso, J. P., Frankignoulle, M., Bourge, I., Romaine, S., and Buddemeier, R. W.: Effect of calcium carbonate saturation of seawater on coral calcification, Global and Planetary Change, 18, 37–46, https://doi.org/10.1016/S0921-8181(98)00035-6, 1998.

Goddijn-Murphy, L., Woolf, D. K., Land, P. E., Shutler, J. D., and Donlon, C. J.: The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO2 fugacity in support of air-sea gas flux studies, Ocean Science, 11, 519–541, https://doi.org/10.5194/os-11-519-2015, 2015.

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

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., Wanninkhof, R. H., Williams, N. L., and Sarmiento, J. L.: Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean, Geophysical Research Letters, 45, 9049–9057, https://doi.org/10.1029/2018GL078013, http://doi.wiley.com/10.1029/2018GL078013, 2018.

Gregor, L. and Gruber, N.: OceanSODA-ETHZ: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification (NCEI Accession 0220059), Tech. rep., NOAA National Centers for Environmental Information, https://doi.org/https://doi.org/10.25921/m5wx-ja34, 2020.

Gregor, L., Kok, S., and Monteiro, P. M. S.: Interannual drivers of the seasonal cycle of CO2 in the Southern Ocean, Biogeosciences, 15, 2361–2378, https://doi.org/10.5194/bg-15-2361-2018, https://www.biogeosciences.net/15/2361/2018/, 2018.

Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M.: A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) - have we hit the wall?, Geoscientific Model Development, 12, 5113–5136, https://doi.org/10.5194/gmd-12-5113-2019, https://www.geosci-model-dev.net/12/5113/2019/, 2019.

Gruber, N., Sarmiento, J. L., and Stocker, T. F.: An improved method for detecting anthropogenic CO2 in the oceans, Global Biogeochemical Cycles, 10, 809–837, https://doi.org/10.1029/96GB01608, http://doi.wiley.com/10.1029/96GB01608, 1996.

Gruber, N., Hauri, C., Lachkar, Z., Loher, D., Frölicher, T. L., and Plattner, G.-K.: Rapid Progression of Ocean Acidification in the California Current System, Science, https://doi.org/10.1126/science.1216773, 2012.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

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., Lo Monaco, C., Mathis, J. T., Murata, A., Olsen, A., Perez, F. F., Sabine, C. L., Tanhua, T., and Wanninkhof, R. H.: The oceanic sink for anthropogenic CO 2 from 1994 to 2007, Science, 363, 1193–1199, https://doi.org/10.1126/science.aau5153, http://www.sciencemag.org/lookup/doi/10.1126/science.aau5153, 2019.

Hauri, C., Friedrich, T., and Timmermann, A.: Abrupt onset and prolongation of aragonite undersaturation events in the Southern Ocean, Nature Climate Change, 6, 172–176, https://doi.org/10.1038/nclimate2844, http://www.nature.com/doifinder/10.1038/nclimate2844{%}5Cnpapers3://publication/doi/10.1038/nclimate2844, 2016.

Holte, J., Talley, L. D., Gilson, J., and Roemmich, D.: An Argo mixed layer climatology and database, Geophysical Research Letters, 44, 5618–5626, https://doi.org/10.1002/2017GL073426, http://doi.wiley.com/10.1002/2017GL073426, 2017.

Humphreys, M. P., Gregor, L., Pierrot, D., van Heuven, S., Lewis, E. R., and Wallace, D. W. R.: PyCO2SYS: marine carbonate system calculations in Python, https://doi.org/10.5281/ZENODO.3967359, https://doi.org/10.5281/zenodo.3967359{#}.XzJOEdRZo2g.mendeley, 2020.

Iida, Y., Kojima, A., Takatani, Y., Nakano, T., Sugimoto, H., Midorikawa, T., and Ishii, M.: Trends in pCO2 and sea–air CO2 flux over the global open oceans for the last two decades, Journal of Oceanography, 71, 637–661, https://doi.org/10.1007/s10872-015-0306-4, 2015.

Jiang, L.-q., Carter, B. R., Feely, R. A., Lauvset, S. K., and Olsen, A.: Surface ocean pH and buffer capacity: past, present and future, Scientific Reports, 9, 18 624, https://doi.org/10.1038/s41598-019-55039-4, http://dx.doi.org/10.1038/s41598-019-55039-4http://www.nature.com/articles/s41598-019-55039-4, 2019.

Johnson, K. S., Jannasch, H. W., Coletti, L. J., Elrod, V. A., Martz, T. R., Takeshita, Y., Carlson, R. J., and Connery, J. G.: Deep-Sea DuraFET: A Pressure Tolerant pH Sensor Designed for Global Sensor Networks, Analytical Chemistry, 88, 3249–3256, https://doi.org/10.1021/acs.analchem.5b04653, 2016.

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array, Journal of Geophysical Research: Oceans, TBD, TBD, https://doi.org/10.1002/2017JC012838, http://doi.wiley.com/10.1002/2017JC012838, 2017.

Keppler, L., Landschützer, P., Gruber, N., Lauvset, S. K., and Stemmler, I.: Seasonal carbon dynamics in the global ocean based on a neural network mapping of observations, Global Biogeochemical Cycles, in review, 2020.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J., Lenton, A., Li, H., Lovenduski, N., Orr, J., Palmieri, J., Schwinger, J., Séférian, R., Stock, C., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper ocean nutrient decline from CMIP6 model projections, Biogeosciences Discussions, pp. 1–43, https://doi.org/10.5194/bg-2020-16, 2020.

Land, P. E., Findlay, H. S., Shutler, J. D., Ashton, I. G., Holding, T., Grouazel, A., Girard-Ardhuin, F., Reul, N., Piolle, J. F., Chapron, B., Quilfen, Y., Bellerby, R. G., Bhadury, P., Salisbury, J., Vandemark, D., and Sabia, R.: Optimum satellite remote sensing of the marine carbonate system using empirical algorithms in the global ocean, the Greater Caribbean, the Amazon Plume and the Bay of Bengal, Remote Sensing of Environment, 235, 111 469, https://doi.org/10.1016/j.rse.2019.111469, https://doi.org/10.1016/j.rse.2019.111469, 2019.

Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J.: A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, Biogeosciences, 10, 7793–7815, https://doi.org/10.5194/bg-10-7793-2013, 2013.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, Global and Planetary Change, pp. 927–949, https://doi.org/10.1002/2014GB004853.Received, http://onlinelibrary.wiley.com/doi/10.1002/2014GB004853/full, 2014.

Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., Van Heuven, S. M., Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T. T., Tilbrook, B., and Wanninkhof, R. H.: The reinvigoration of the Southern Ocean carbon sink, Science, 349, 1221–1224, https://doi.org/10.1126/science.aab2620, http://science.sciencemag.org/content/349/6253/1221.abstracthttp://www.sciencemag.org/cgi/doi/10.1126/science.aab2620, 2015.

Landschützer, P., Gruber, N., and Bakker, D. C.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochemical Cycles, 30, 1396–1417, https://doi.org/10.1002/2015GB005359, http://doi.wiley.com/10.1002/2015GB005359, 2016.

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

Lauvset, S. K., Gruber, N., Landschützer, P., Olsen, A., and Tjiputra, J.: Trends and drivers in global surface ocean pH over the past 3 decades, Biogeosciences, 12, 1285–1298, https://doi.org/10.5194/bg-12-1285-2015, 2015.

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 System Science Data, 8, 325–340, https://doi.org/10.5194/essd-8-325-2016, www.earth-syst-sci-data.net/8/325/2016/https://www.earth-syst-sci-data.net/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 Biogeochemical Cycles, 14, 979–994, https://doi.org/10.1029/1998GB001087, 2000.

Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., Park, G.-H., Wanninkhof, R. H., Feely, R. A., and Key, R. M.: Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans, Geophysical Research Letters, 33, L19 605, https://doi.org/10.1029/2006GL027207, http://doi.wiley.com/10.1029/2006GL027207, 2006.

Lefèvre, N. and Taylor, A.: Estimating pCO2 from sea surface temperatures in the Atlantic gyres, Deep-Sea Research Part I: Oceanographic Research Papers, 49, 539–554, https://doi.org/10.1016/S0967-0637(01)00064-4, 2002.

Lewis, E., Wallace, D., and Allison, L. J.: Program developed for CO2 system calculations, Tech. rep., Brookhaven National Lab., Dept. of Applied Science, Upton, NY, 1998.

Lovenduski, N. S., Gruber, N., Doney, S. C., and Lima, I. D.: Enhanced CO2 outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, Global Biogeochemical Cycles, 21, 1–14, https://doi.org/10.1029/2006GB002900, 2007.

Maritorena, S., Fanton D'andon, O. H., Mangin, A., Siegel, D. A., D'Andon, O. H. F., Mangin, A., and Siegel, D. A.: Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues, Remote Sensing of Environment, 114, 1791–1804, https://doi.org/10.1016/j.rse.2010.04.002, http://dx.doi.org/10.1016/j.rse.2010.04.002http://www.ioccg.org/, 2010.

Marshall, J. C. and Speer, K.: Closure of the meridional overturning circulation through Southern Ocean upwelling, Nature Geoscience, 5, 171–180, https://doi.org/10.1038/ngeo1391, http://www.nature.com/doifinder/10.1038/ngeo1391http://dx.doi.org/10.1038/ngeo1391, 2012.

Mehrbach, C., Culberson, C. H., Hawley, J. E., and Pytkowicx, R. M.: Measurement of the Apparent Dissociation Constants of Carbonic Acid in Seawater At Atmospheric Pressure, Limnology and Oceanography, 18, 897–907, https://doi.org/10.4319/lo.1973.18.6.0897, 1973.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., Good, S. A., Mittaz, J., Rayner, N. A., Berry, D., Eastwood, S., Taylor, M., Tsushima, Y., Waterfall, A., Wilson, R., and Donlon, C. J.: Satellite-based time-series of sea-surface temperature since 1981 for climate applications, Scientific data, 6, 223, https://doi.org/10.1038/s41597-019-0236-x, http://dx.doi.org/10.1038/s41597-019-0236-x, 2019.

Millero, F. J., Lee, K., and Roche, M.: Distribution of alkalinity in the surface waters of the major oceans, Marine Chemistry, 60, 111–130, https://doi.org/10.1016/S0304-4203(97)00084-4, http://www.sciencedirect.com/science/article/pii/S0304420397000844, 1998.

Monteiro, P. M. S., Gregor, L., Lévy, M., Maenner, S., Sabine, C. L., and Swart, S.: Intraseasonal variability linked to sampling alias in air-sea CO2 fluxes in the Southern Ocean, Geophysical Research Letters, 42, 8507–8514, https://doi.org/10.1002/2015GL066009, 2015.

Negrete-García, G., Lovenduski, N. S., Hauri, C., Krumhardt, K. M., and Lauvset, S. K.: Sudden emergence of a shallow aragonite saturation horizon in the Southern Ocean, Nature Climate Change, 9, https://doi.org/10.1038/s41558-019-0418-8, http://dx.doi.org/10.1038/s41558-019-0418-8, 2019.

Olafsson, J., Olafsdottir, S. R., Benoit-Cattin, A., and Takahashi, T. T.: The irminger sea and the Iceland sea time series measurements of sea water carbon and nutrient chemistry 1983-2008, Earth System Science Data, 2, 99–104, https://doi.org/10.5194/essd-2-99-2010, 2010.

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 System Science Data, 8, 297–323, https://doi.org/10.5194/essd-8-297-2016, https://www.earth-syst-sci-data.net/8/297/2016/, 2016.

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. J., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F. F., Yamanaka, Y., Yool, A., Doney, S. C., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F. F., Yamanaka, Y., Yool, A., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R. J., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F. 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, http://www.ncbi.nlm.nih.gov/pubmed/16193043, 2005.

Orr, J. C., Epitalon, J. M., and Gattuso, J.-P.: Comparison of ten packages that compute ocean carbonate chemistry, Biogeosciences, 12, 1483–1510, https://doi.org/10.5194/bg-12-1483-2015, 2015.

Orr, J. C., Epitalon, J.-M., Dickson, A. G., and Gattuso, J.-P.: Routine uncertainty propagation for the marine carbon dioxide system, Marine Chemistry, pp. 1–24, https://doi.org/10.1016/j.marchem.2018.10.006, https://linkinghub.elsevier.com/retrieve/pii/S030442031830149X, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, C., Thirion, B., Grisel, O., Blondel, M., Prettenhoffer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D.: Scikit-learn: Machine learning in Python, Journal of Machine Learning Research, 12, 2825–2830, https://doi.org/10.1007/s13398-014-0173-7.2, http://dl.acm.org/citation.cfm?id=2078195, 2011.

Pierrot, D., Neill, C., Sullivan, K. F., Castle, R. D., Wanninkhof, R. H., Lüger, H., Johannessen, T., Olsen, A., Feely, R. A., and Cosca, C. E.: Recommendations for autonomous underway pCO2 measuring systems and data-reduction routines, Deep-Sea Research Part II: Topical Studies in Oceanography, 56, 512–522, https://doi.org/10.1016/j.dsr2.2008.12.005, 2009.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Pörtner, H.-O. and Farrell, A. P.: Physiology and Climate Change, Science, 322, 690–692, https://doi.org/10.1126/science.1163156, http://epic.awi.de/epic/Main?puid=32305{&}lang=enhttp://www.sciencemag.org/cgi/doi/10.1126/science.1163156, 2008.

Raimondi, L., Matthews, J. B. R., Atamanchuk, D., Azetsu-Scott, K., and Wallace, D. W.: The internal consistency of the marine carbon dioxide system for high latitude shipboard and in situ monitoring, Marine Chemistry, 213, 49–70, https://doi.org/10.1016/j.marchem.2019.03.001, https://doi.org/10.1016/j.marchem.2019.03.001, 2019.

Resplandy, L., Lévy, M., D'Ovidio, F., and Merlivat, L.: Impact of submesoscale variability in estimating the air-sea CO2 exchange: Results from a model study of the POMME experiment, Global Biogeochemical Cycles, 23, 1–19, https://doi.org/10.1029/2008GB003239, 2009.

Rödenbeck, C., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C. L., Cassar, N., Reum, F., Keeling, R. F., and Heimann, M.: Interannual sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Biogeosciences, 11, 4599–4613, https://doi.org/10.5194/bg-11-4599-2014, https://www.biogeosciences.net/11/4599/2014/, 2014.

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

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T.-H., Kozyr, A., Ono1, T., and Rios, A. F.: The Oceanic Sink for Anthropogenic CO2, Science, 305, 367–371, https://doi.org/10.1126/science.1097403, http://www.sciencemag.org/content/305/5682/367.abstracthttp://www.sciencemag.org/cgi/doi/10.1126/science.1097403https://www.sciencemag.org/lookup/doi/10.1126/science.1097403, 2004.

Sarmiento, J. L. and Gruber, N.: Ocean Biogeochemical Dynamics, Princeton University Press, Princeton, NJ, 2006.

Sasse, T. P., McNeil, B. I., and Abramowitz, G.: A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks, Biogeosciences, 10, 4319–4340, https://doi.org/10.5194/bg-10-4319-2013, 2013.

Siegel, D. A., Doney, S. C., and Yoder, J. A.: The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis, Science, 296, 730–733, https://doi.org/10.1126/science.1069174, 2002.

Steinacher, M., Joos, F., Frölicher, T. L., Plattner, G.-K., and Doney, S. C.: Imminent ocean acidification projected with the NCAR global coupled carbon cycle-climate model, Biogeosciences Discussions, 5, 4353–4393, https://doi.org/10.5194/bgd-5-4353-2008, https://bg.copernicus.org/preprints/5/4353/2008/, 2009.

Tagliabue, A., Bowie, A. R., Boyd, P. W., Buck, K. N., Johnson, K. S., Saito, M. A., Philip, W., Buck, K. N., Johnson, K. S., Saito, M. A., Boyd, P. W., Buck, K. N., Johnson, K. S., and Saito, M. A.: The integral role of iron in ocean biogeochemistry, Nature, 543, 51–59, https://doi.org/10.1038/nature21058, http://dx.doi.org/10.1038/nature21058, 2017.

Takahashi, T. and Sutherland, C. S.: Climatological mean distribution of pH and carbonate ion concentration in global ocean surface waters in the unified pH scale and mean rate of their changes in selected areas, The National Science Foundation, OCE 10-388, 33, 2013.

Takahashi, T., Sutherland, S., and Kozyr, A.: Global Ocean Surface Water Partial Pressure of CO2 Database: Measurements Performed During 1957–2018 (LDEO Database Version 2018) (NCEI Accession 0160492), 2019.

Takahashi, T. T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S. C.: Seasonal variation of CO 2 and nutrients in the high-latitude surface oceans: A comparative study, Global Biogeochemical Cycles, 7, 843–878, https://doi.org/10.1029/93GB02263, http://doi.wiley.com/10.1029/93GB02263, 1993.

Takahashi, T. T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N. R., Wanninkhof, R. H., Feely, R. A., Sabine, C. L., Olafsson, J., and Nojiri, Y.: Global sea-air CO2 flux based on climatological surface ocean pCO2, and seasonal biological

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

and temperature effects, Deep-Sea Research Part II: Topical Studies in Oceanography, 49, 1601–1622, https://doi.org/10.1016/S0967-0645(02)00003-6, 2002.

Takahashi, T. T., Sutherland, S. C., Chipman, D. W., Goddard, J. G., and Ho, C.: Climatological distributions of pH, pCO2, total CO2, alkalinity, and CaCO3 saturation in the global surface ocean, and temporal changes at selected locations, Marine Chemistry, 164, 95–125, https://doi.org/10.1016/j.marchem.2014.06.004, http://dx.doi.org/10.1016/j.marchem.2014.06.004, 2014.

Tilbrook, B., Jewett, E. B., DeGrandpre, M. D., Hernandez-Ayon, J. M., Feely, R. A., Gledhill, D. K., Hansson, L., Isensee, K., Kurz, M. L., Newton, J. A., Siedlecki, S. A., Chai, F., Dupont, S., Graco, M., Calvo, E., Greeley, D., Kapsenberg, L., Lebrec, M., Pelejero, C., Schoo, K. L., and Telszewski, M.: An enhanced ocean acidification observing network: From people to technology to data synthesis and information exchange, Frontiers in Marine Science, 6, 1–21, https://doi.org/10.3389/fmars.2019.00337, 2019.

Turk, D., Dowd, M., Lauvset, S. K., Koelling, J., Alonso-Pérez, F., and Pérez, F. F.: Can empirical algorithms successfully estimate aragonite saturation state in the subpolar North Atlantic?, Frontiers in Marine Science, 4, https://doi.org/10.3389/fmars.2017.00385, 2017.

Uppström, L. R.: The boron/chlorinity ratio of deep-sea water from the Pacific Ocean, Deep-Sea Research and Oceanographic Abstracts, 21, 161–162, https://doi.org/10.1016/0011-7471(74)90074-6, 1974.

Watson, A. J., Bakker, D. C., Ridgwell, A. J., Boyd, P. W., and Law, C. S.: Effect of iron supply on Southern Ocean CO2 uptake and implications for glacial atmospheric CO2, Nature, 407, 730–733, https://doi.org/10.1038/35037561, 2000.

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., Dickson, A. G., Gray, A. R., Wanninkhof, R. H., Russell, J. L., Riser, S. C., and Takeshita, Y.: Calculating surface ocean pCO 2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis, Global Biogeochemical Cycles, 31, 591–604, https://doi.org/10.1002/2016GB005541, http://doi.wiley.com/10.1002/2016GB005541, 2017.

Wolf-Gladrow, D. A., Zeebe, R. E., Klaas, C., Körtzinger, A., and Dickson, A. G.: Total alkalinity: The explicit conservative expression and its application to biogeochemical processes, Marine Chemistry, 106, 287–300, https://doi.org/10.1016/j.marchem.2007.01.006, http://linkinghub.elsevier.com/retrieve/pii/S0304420307000047, 2007.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S.-i.: A global surface ocean fCO2 climatology based on a feed-forward neural network, Journal of Atmospheric and Oceanic Technology, 31, 1838–1849, https://doi.org/10.1175/JTECH-D-13-00137.1, http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-13-00137.1, 2014.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

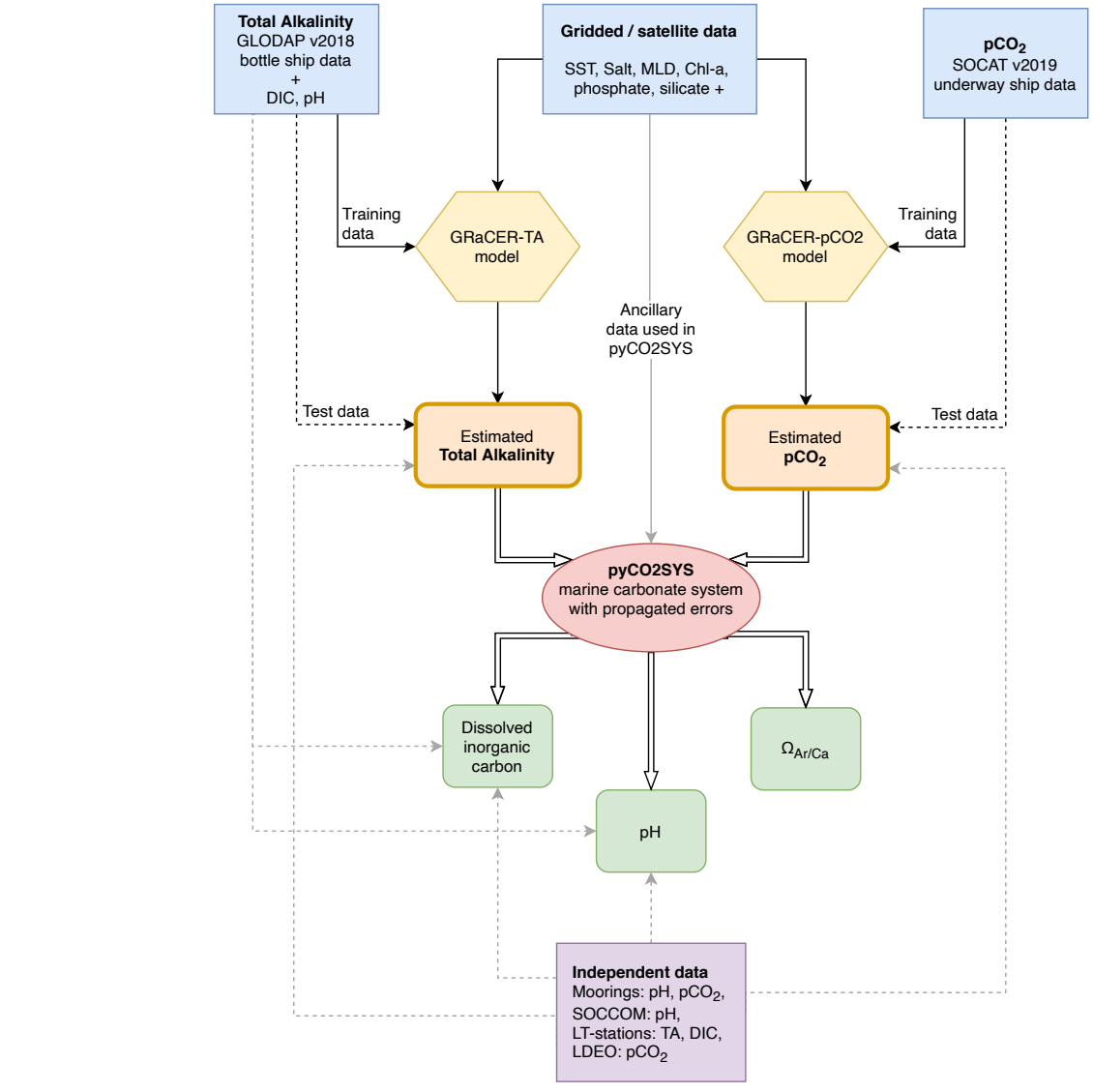
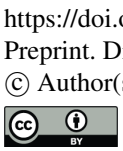


Figure 1. Schematic flow diagram showing the three steps required to reconstruct the the surface ocean carbonate system. In the first step, the GRaCER (Geospatial Random Cluster Ensemble regression) method is used to develop statistical models for the observed TA (left) and pCO_2 (right) fields. In the second step, these statistical models are used to extrapolate these two parameters over time and space using ancillary observations, primarily stemming from satellite observations. In the third step, the inter- and extrapolated TA and pCO_2 fields are then used to compute the remaining parameters of the surface ocean carbonate system, namely DIC, pH, and the saturation state of seawater with regard to mineral $CaCO_3$, Ω. The output of steps two and three is the OceanSODA-ETHZ product. Also shown are the various data sets and data flows used in this study. The different lines indicate whether data is used for training (solid lines) or testing (dashed lines) or output with an estimate of uncertainty, where independent test data are shown with gray dashed lines. The gridded/satellite data are summarized in Table 1. Independent test data are shown by the purple box. *pyCO2SYS* is the software used to solve the marine carbonate system and propagate uncertainties (Humphreys et al., 2020).

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020



Open Access Earth System Science Data Discussions

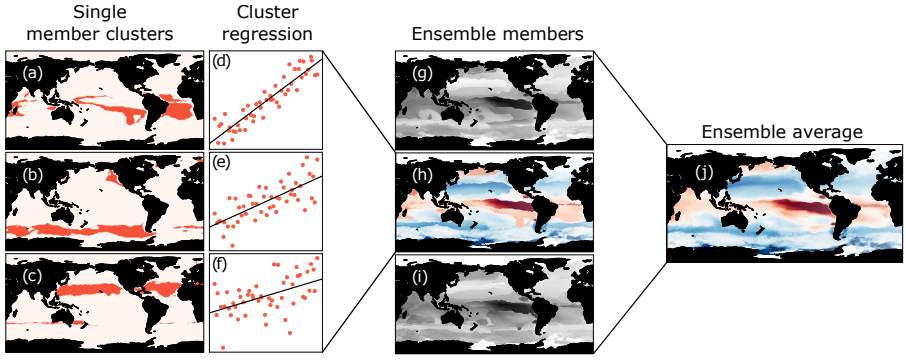


Figure 2. A schematic showing the steps used in the GRaCE-R method for a single month. (a-c) show a subset of the clusters of a single ensemble member (h), with the adjacent scatter plots (d-f) showing the training data for each cluster and the linear regression models for that cluster (with toy data). (g-i) show the ensemble member estimates for a subset of three members. (j) shows the ensemble mean for all ensemble members, which includes ensemble members not shown in (g-i).

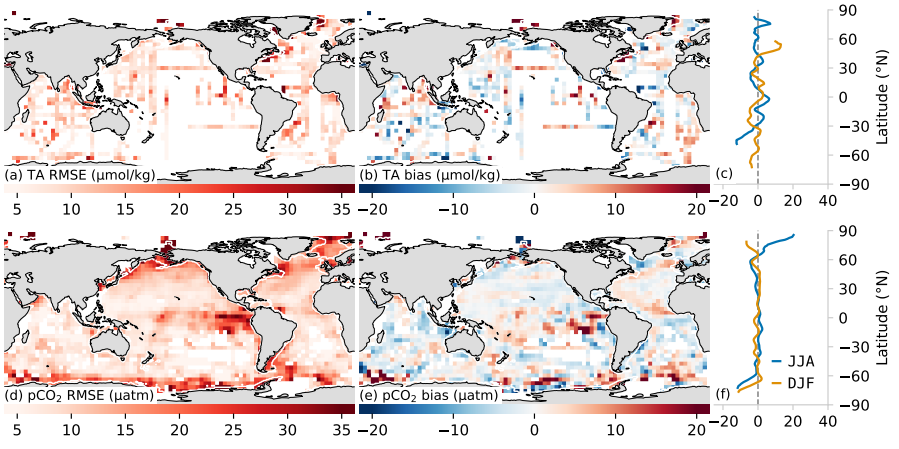
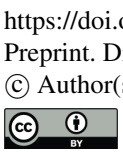


Figure 3. Test metrics for total alkalinity (a,b) and pCO_2 (c,d). The left-hand column (a, c) shows the root mean squared error (RMSE) compared with the target data. Similarly, the right-hand column (b, d) shows bias compared to the respective training data sets (GLODAP v2.2019 and SOCAT v2019). A 2D spatial convolution was first applied to the $1^\circ \times 1^\circ$ pixels to make regional patterns in the biases and RMSE clearer and data were then aggregated into $4^\circ \times 4^\circ$ pixels for clearer visualization.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020



Open Access Earth System Science Data Discussions

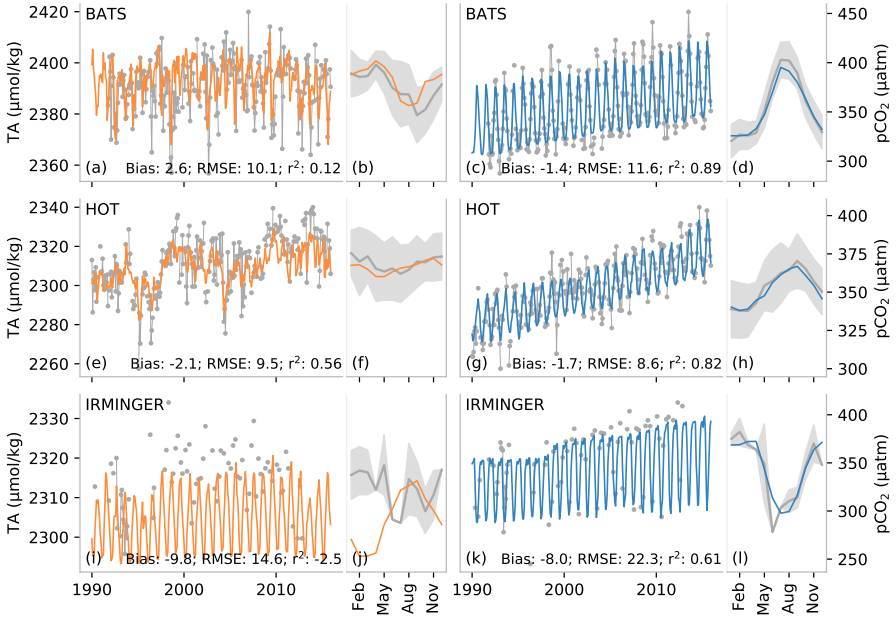


Figure 4. A comparison of a subset of measurements from long term observation stations (gray) with predicted total alkalinity (TA) (left: *a,b,e,f,i,j*) and partial pressure of CO_2 (pCO_2) (right: *c,d,g,h,k,l*). The top row (a-d) shows data for the Bermuda Ocean Time Series (BATS)Hawaii Ocean Time-series (HOT), the middle row (e-h) for the and the bottom row (i, l) shows the Irminger station. The narrow panels show the average of the seasonal climatology for the time series. The gray shading shows the standard deviation of the observations for the period 1990 to 2018, while the orange/blue lines show the average estimate. TA for the Irminger station is calculated from pCO_2 and DIC, and pCO_2 is calculated for BATS and HOT using DIC and TA, as described in section 2.4.

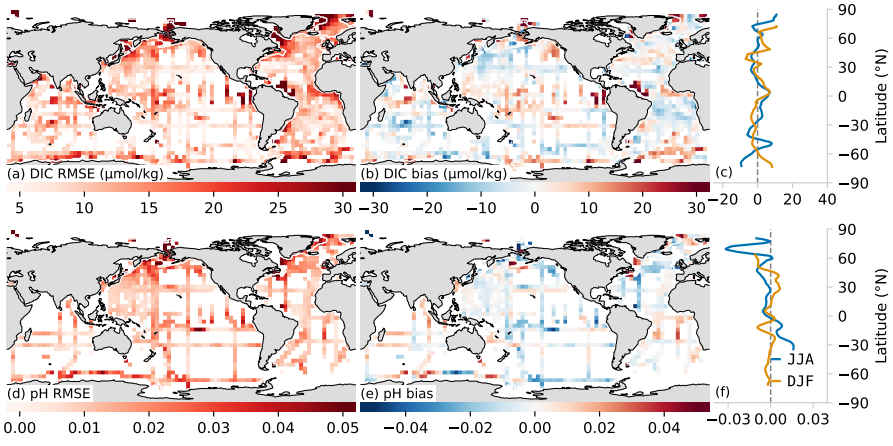


Figure 5. Root mean squared error (RMSE - a,d) and biases (b,c,e,f) for: dissolved inorganic carbon (DIC, top) and pH (bottom) compared with *in-situ* GLODAP v2.2019 data. The two subplots in the right most column compare the zonally averaged bias for JJA and DJF. Data were processed for plotting as described in 4.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

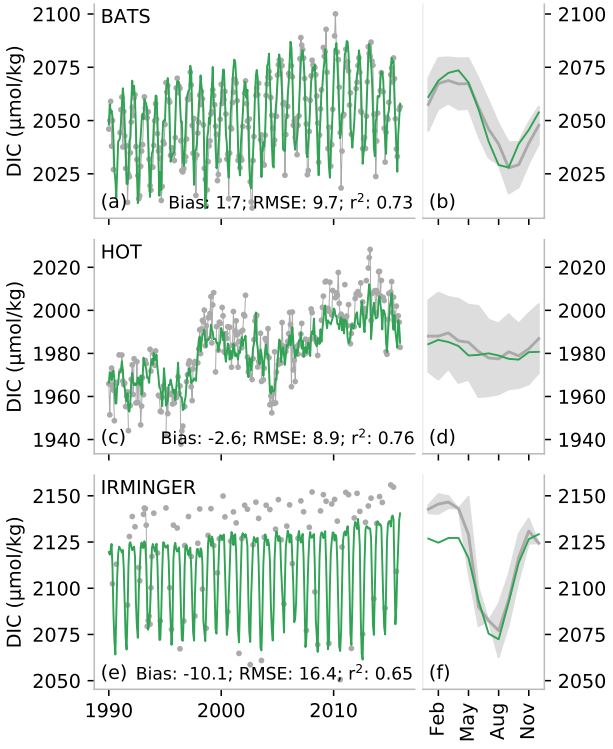


Figure 6. A comparison of observations from long term observation stations (gray) with predicted dissolved inorganic carbon (DIC). The top row (a,b) shows data for the Bermuda Ocean Time Series (BATS), the middle row (c,d) for the Hawaii Ocean Time-series (HOT) and the bottom row (e,f) shows the Irminger station. The narrow panels on the right show the average of the seasonal climatology for the time series, where the gray shading shows the standard deviation of the observations for the period 1990 to 2018, while the green line shows the average estimate.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

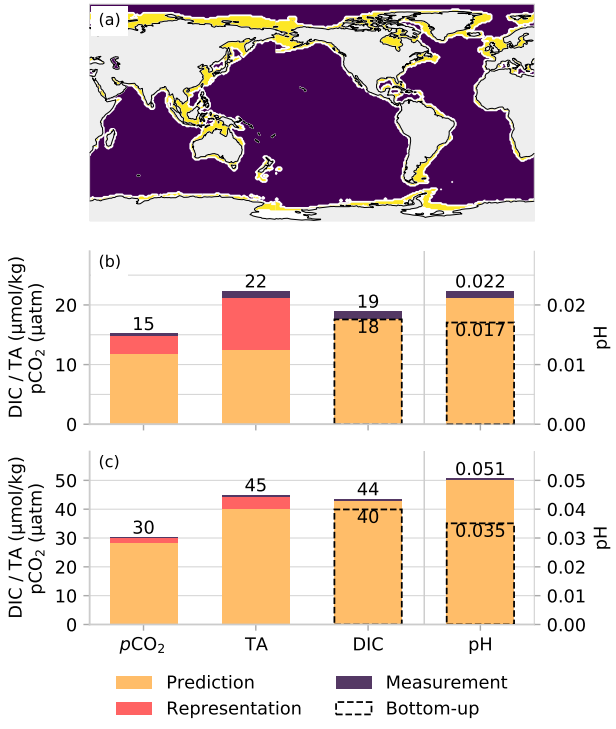
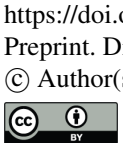


Figure 7. A comparison of propagated uncertainties with independent errors as an assertion of the validity of error estimates. The map (a) shows the separation between coastal and open ocean, (b) shows the error contributions in the open ocean, and (c) in the coastal ocean. The total error has been broken into the three different components. Note that the values represented by the bar plots are not equivalent to values in Table 3 as the latter shows pCO_2 and TA total errors for test data only, while the bar charts show total errors for all data; further the breakdown of the errors is proportional to the contribution of the sum of the squares.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020



Open Access Earth System Science Data Discussions

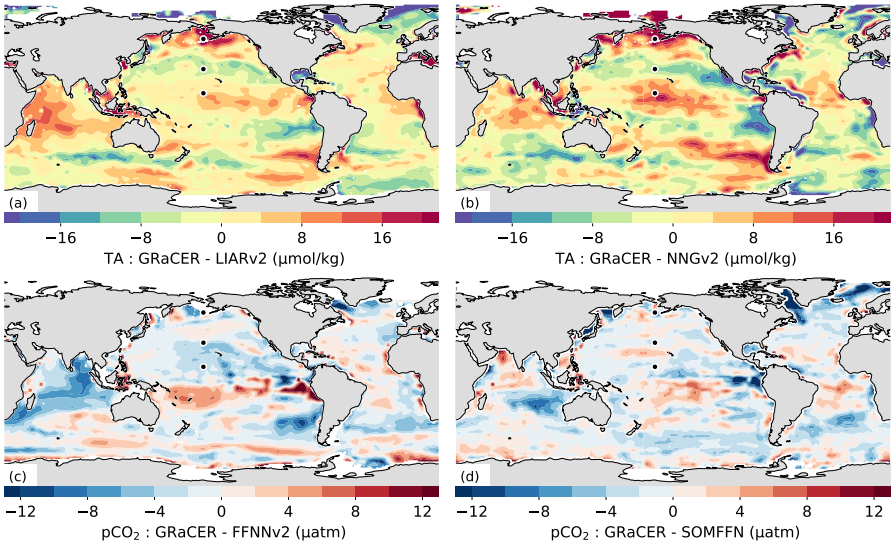


Figure 8. A comparison of the mean differences between TA (top) and pCO_2 (bottom) for OceanSODA-ETHZ and other methods in the literature: (a) LIARv2, (b) NNGv2, (c) LSCE-FFNNv2, and (d) MPI-SOMFFN. The markers in the North Pacific show the locations used in the comparison of the climatology in Figure 10.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

Open Access Earth System Science Data Discussions

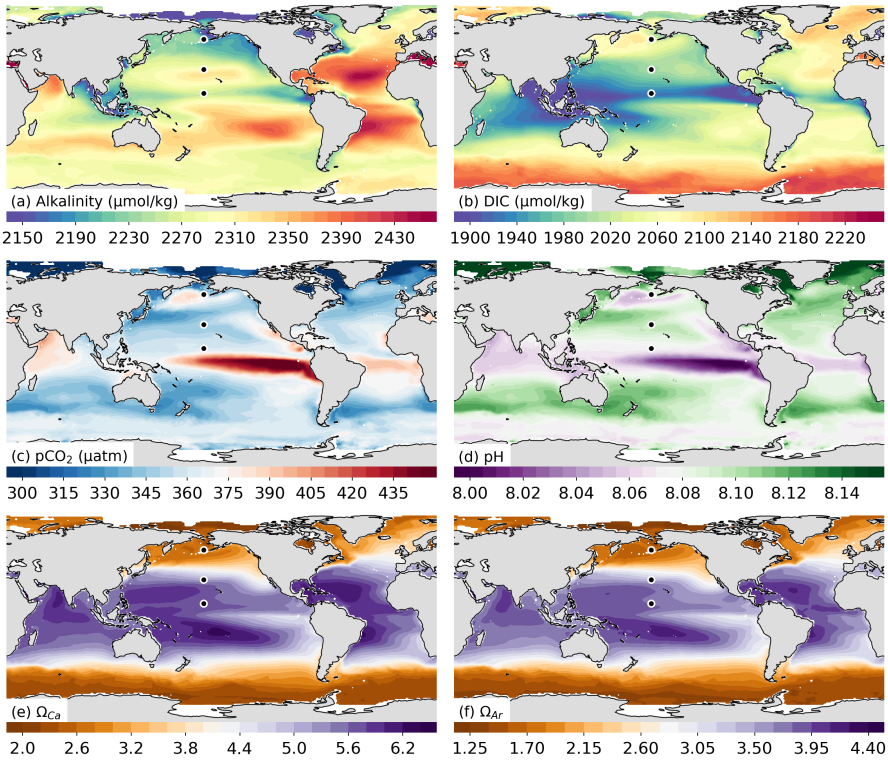


Figure 9. Climatological mean maps of the GRaCER-based estimates of (a) total alkalinity and (c) pCO_2, as well as those of the computed variables, (b) dissolved inorganic carbon, (d) pH, (e) Ω_{calc} (saturation state with regard to calcite) and (f) Ω_{arag} (saturation state with regard to aragonite). The three black markers in each plot show the locations chosen for the seasonal analysis in Figure 10*f-j*.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

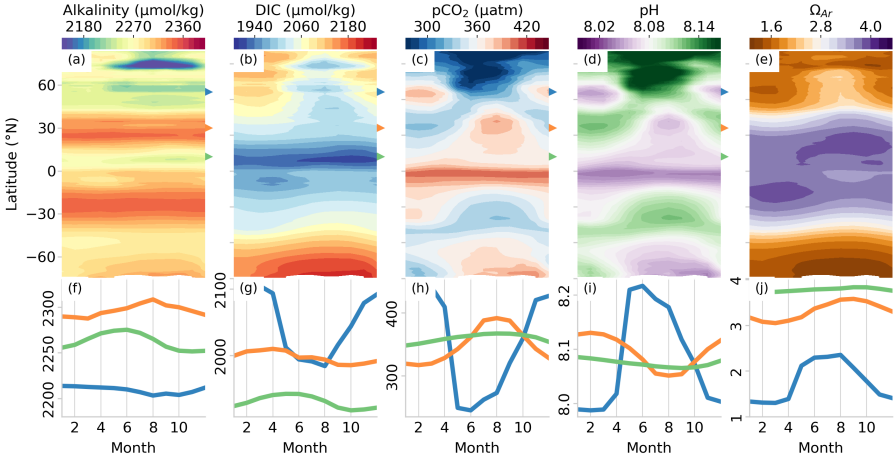


Figure 10. Hovmoeller plots (*a-d*) showing the zonally averaged seasonal climatology for (a) total alkalinity, (b) dissolved inorganic carbon, (c) pCO_2, (d) pH, and (e) aragonite saturation state (Ω_{Ar}). The second row of figures (*f-j*) show the corresponding variables for a high (55°N, 180°E), mid (30°N, 180°E) and low-latitude (10°N, 180°E) location. The units for (*f-j*) correspond with the units in (*a-e*).

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020

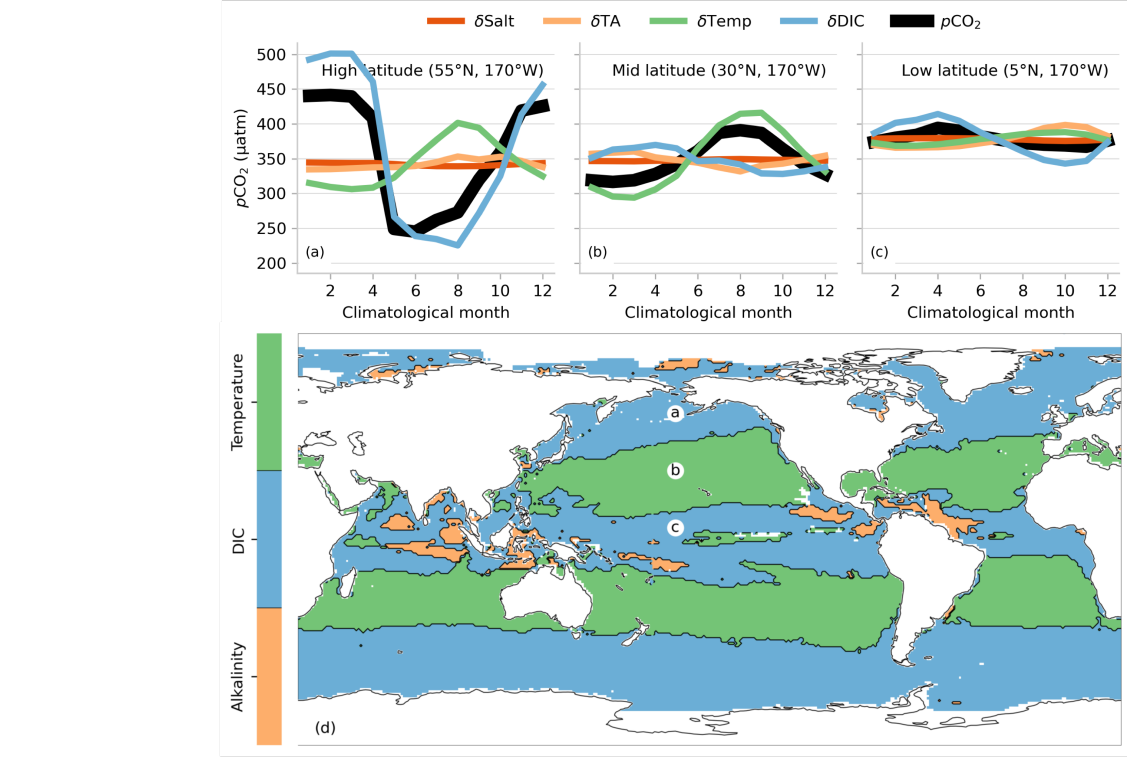
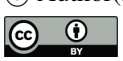


Figure 11. Attribution of DIC, TA and temperature to the seasonal cycle of pCO_2. The top two figures show the seasonal cycle of pCO_2 and the drivers thereof for (a) high latitudes, (b) mid latitudes, and (c) low latitudes. These locations are shown with the markers in the Pacific ocean in (c). The map (c) shows the dominant driver of the seasonal cycle for each region calculated as the value with the maximum seasonal amplitude.

https://doi.org/10.5194/essd-2020-300
Preprint. Discussion started: 22 October 2020



Open Access Earth System Science Data Discussions

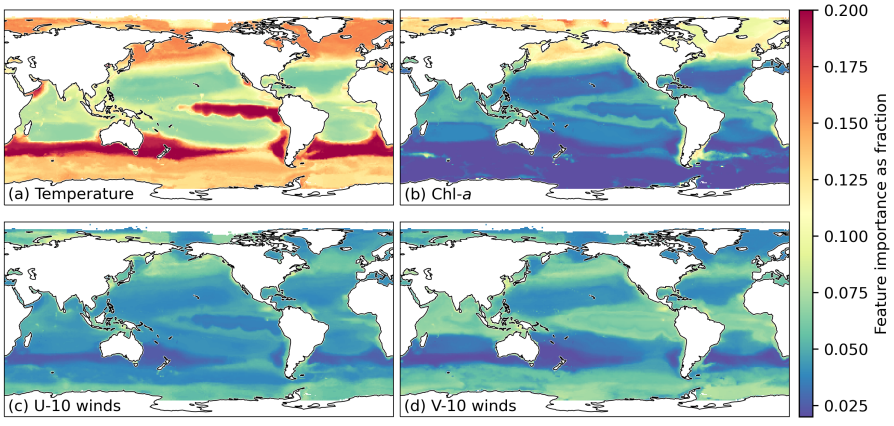


Figure A1. Feature importances determined by Gradient Boosted Decision Trees for pCO_2 predictions. A subset of four proxies are shown for the months of June, July and August. The feature importances allow one to make informed decisions about the inclusion or exclusion of proxy variables. Here, temperature (a) is one of the more important features, Chl-a is most important in the high northern latitudes, the U and V components of the winds are important along the coastal regions, particularly the eastern boundary upwelling systems.