Global Biogeochemical Cycles

Supporting Information for

Seasonal Carbon Dynamics in the Global Ocean

L.Keppler,[1,2] P. Landschützer,[1] N. Gruber[3], S.K. Lauvset[4], I. Stemmler[1]

1 Max-Planck-Institute for Meteorology (MPI-M), Hamburg, Germany

2 International Max Planck Research School on Earth System Modelling (IMPRS-ESM), Hamburg, Germany

3 Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Switzerland

4 NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

Contents of this file

Text S1. Background information for Self-Organizing Maps (SOMs) and Feed-Forward Networks (FFNs)

The purpose of the SOM-FFN method is to map sparse data, filling data gaps with the aid of better-constrained predictor data. First, we separate the ocean into clusters of similar biogeochemical and physical properties using SOMs, and second, we run an FFN in each of the clusters to approximate the non-linear best-fit relationship between the available observations of the target data (here: DIC) and a set of physical and biogeochemical predictor data. These predictor data exist as mapped (gap-filled) data at global scale, hence the approximated relationship between the target data and the predictor data can be applied where no target data exist to fill these observational gaps (Landschützer et al., 2013). The SOM-step is conducted, because the statistical relationships between the predictor and target data differ around the globe, while they should be similar within each SOM-cluster.

SOMs are a form of unsupervised machine learning that is commonly used to cluster data (Kohonen, 2001, 1989). In this clustering method, we first arrange each normalized multi-dimensional input variable (SST, SSS, climatological DIC; see Main Text and Table S1) as a 1D vector. The order of the 1D vector is less important as long as all multidimensional arrays are arranged in the same way. Next, we chose a number of neurons corresponding to the number of clusters we want to have. The network randomly places these neurons in a grid space, where each input vector represents one dimension. The network then identifies the Euclidean distance between the input data to these neurons. Next, the neurons are iteratively moved around in the grid space until the network has identified a set-up where the sum of the Euclidean distances between the input data and the neurons is minimal. Once this set-up is found, the input data is assigned the number of the neuron it is closest to, resulting in a 1D vector with the same length as our input variables. We then transfer this vector back to a multidimensional array (latitude, longitude, depth, and month) so that the clusters can be displayed on our multidimensional grid.

The choice of the number of neurons (and therefore the number of clusters) of a SOM is somewhat subjective. Too many clusters will result in only a few observations in each cluster, while too few will create large regions with a wide range of varying properties. As the surface ocean is less uniform than the intermediate and deep ocean, we chose six clusters for the surface slab (2.5 m–500 m), and four each for the intermediate (600 m–1500 m) and deep slabs (1600 m–1975 m; Fig. S1a-d, Table S1). Although the SOMs are computed for each climatological month, the clusters do not considerably change shape from one month to the next. Most clusters remain the same throughout the year, but near the cluster boundary, there is a small amount of variation in the top 200 m (Fig. S1e-f). The clusters are seasonally relatively static by design due to our weighting of the climatological DIC as a predictor variable.

FFNs are a form of supervised machine learning; they can approximate nearly any continuous function and are commonly used in Earth System Science (Hornik et al., 1989). In this step, we run an FFN in each cluster separately. We first co-locate the predictor data with the target data. During the FFN training, the network establishes the statistical relationship between the target data and the co-located predictor data (see Fig. S2 for our set-up). To do so, the predictors are connected by a transfer function to a set of neurons with random initial weights and biases at each connection. Next, these neurons are connected to the target data with a second transfer function, again with initial random weights and biases. The output of this initial set-up is a first guess estimation of the target data at the location of the observations. This output is then compared with the actual observations and the mean squared error (MSE) is calculated. This step is iteratively repeated using the Levenberg-Marquardt Algorithm that adjusts the neuron weights and biases until the MSE between the Output and the observations reaches a minimum. Next, this approximated relationship between the predictor and target data is applied to map the target data at all grid points where we have predictor data.

The input array consists of the predictor data described in the Main Text and summarized in Table S1 and Fig. S2. In our set-up, we use two layers, where the first layer (in the literature referred to as the hidden layer) uses 16 neurons, which are connected to a second layer via a sigmoid transfer function. The second layer, consisting of a single neuron, uses a linear transfer function to linearly extract the hidden layer output to produce the final DIC estimate (Fig. S2). This two-layer setup enables the network to represent both linear and non-linear relationships between predictor and target data (Broullón et al., 2019; Hagan et al., 2014). The number of neurons chosen in the set-up of the FFN is related to the complexity of the data sets (Gardner and Dorling, 1998). While too few neurons result in the network not learning complex relations, too many neurons may overfit the problem (e.g., Broullón et al., 2019). We tested several set-ups and found that 16 neurons lead to the best representation of the observations.

For each iteration in the training process, we use only a randomly chosen subset of the input data to train the network (the training set; here: 80% of the data), and we use the remaining data for internal validation (the validation set; here: 20% of the data). The validation set is used to stop the iterative training once the adjustment of the network weights does not improve the MSE towards the validation set. This process is often referred to as an "early-stopping approach" and ensures that the network can generalize and prevent the network from overfitting.

Text S2. Smoothing and uncertainty within our method

The internal validation of the SOM-FFN method is based on a randomly chosen subsample of the available observations by the network (the validation set for the early-stopping approach). Therefore, the resulting DIC fields vary slightly each time we run the network and could be biased depending on which data was chosen as training and validation data. To account for potential biases in the separation between training and validation data, we use a bootstrapping approach and run the SOM-FFN method ten times and take the mean of this ensemble, resulting in a smoother end product than a single ensemble member. We define the generalization uncertainty within the method as the standard deviation across this ensemble. We further smooth the mapped ensemble mean fields at each depth level with a filter that calculates the mean of the neighboring three grid cells in each horizontal direction (latitude and longitude). We then apply a non-linear least squares harmonic fit at each grid cell, at each depth level to smoothen the seasonal cycle.

Our final monthly climatology of the Mapped Observation-Based Oceanic DIC (MOBO-DIC) is hereafter called DIC_{MOBO}. Note that our mapped estimate is not scaled to a specific year, because it is based on only 14 years of data (2004 through 2017). As our estimate represents the monthly means of these 14 years, we consider it centered around the years 2010 and 2011.

Text S3. Discussion on including information on the time or location as predictors in FFNs

Some studies include a time-variable, such as the month of the year as a predictor in FFNs (e.g., Bittig et al., 2018; Sauzède et al., 2017). To represent the periodicity of the year, the cosine and/or sine of the time-variable is usually used (see Eq. S1 and S2 for the computation of the cosine and sine of the month of the year respectively). The same procedure is commonly used to represent the periodicity of longitude (e.g., Broullón et al., 2019).

$$\cos_{month} = \cos \frac{\pi}{n/2} \text{ month} \qquad \text{(Eq. S1)}$$

$$\sin_{month} = \sin \frac{\pi}{n/2} \text{ month} \qquad \text{(Eq. S2)}$$

where *n* is the number of months there are in a year (12).

However, a problem arises: both the cosine and sine curve cross the x-axis twice in one cycle (Fig. S3). Hence, months that are climatologically different, are assigned the same value. For example, in the cosine curve, the 3rd and 9th month have the save value (0). Hence, in this case, March would learn from October and vice versa, although they have different values in the real world. Similarly, in the sine curve, the 6th and the 12th month have the same value (0), and so June and December would learn from each other, which is not in line with our knowledge of the seasonal cycle of carbon.

During the set-up of our FFN, we analyzed what would happen if we did include the cosine and/or sine of the month of the year as predictors. Our results were considerably noisier in those set-ups and we could not reproduce the seasonal cycles. Presumably, the same problem would arise when using the cosine and/or sine of the day of the year as a predictor. Instead, the network obtains the seasonal information from the predictor (especially temperature and salinity) and can produce a seasonal cycle of DIC without being provided information about the time. Similarly, we expect the same problem to occur when using the cosine and/or sine of longitude. Our method overcomes this problem through the clustering with the SOMs before the FFN is run and so does not need explicit information on the location. Other studies have overcome this problem by feeding information on the location into the neural network using n-vector transformations of latitude and longitude (Gregor et al., 2017; Sasse et al., 2013).

Text S4. Validation tests, comparing DIC_{MOBO} with independent data

Compared to mapping the surface pCO_2, some additional challenges emerge when mapping the interior DIC. First, interior DIC measurements are even sparser than pCO_2 measurements at the surface, thus, larger spatio-temporal gaps need to be filled. Second, more potential predictors are available near the surface, for example from satellite data, than at depth. Therefore, substantial testing is required to check whether the method can be applied to map time-varying DIC fields. We test our method by comparing DIC_{MOBO} with various independent data that were not used to train the network, both observational and synthetic, as described in the following Subsections.

S4.1 Global mapped annual mean climatology (Lauvset et al., 2016)

We compare the annual mean of DIC_{MOBO} to the annual climatology by Lauvset et al. (2016). That product is on a 1°x1° grid and is normalized to the year 2002. To compare the two estimates, we linearly interpolate the Lauvset climatology onto the same 33 depth levels as our product (hereafter $DIC_{LAUVSET}$) and compute the annual mean of DIC_{MOBO}.

Generally, the two estimates agree on the distribution, and the RMSE between DIC_{MOBO} and $DIC_{LAUVSET}$ is 19.9 µmol/kg^{-1} and small bias of -1.5 µmol/kg^{-1} (negative bias indicates that our estimate is on average lower than the validation data). The isopycnals depicted in Fig. S4a,d,g demonstrate that the mean DIC profile largely follows the profile of the water masses. DIC_{MOBO} tends to have higher concentrations near the surface and lower concentrations in the interior than $DIC_{LAUVSET}$ (Fig. S4). The former can be linked to the difference in reference year: $DIC_{LAUVSET}$ is scaled to the year 2002, and DIC_{MOBO} is based on data after 2004, centered around the years 2010/2011. Hence, we expect that DIC_{MOBO} has more DIC near the surface than $DIC_{LAUVSET}$ due to the accumulation of anthropogenic carbon. The expected increase in surface ocean DIC due to the atmospheric perturbation is ~1.1 µmol kg^{-1} yr^{-1} or ~11 µmol kg^{-1} between 2002 and 2011 (following Sarmiento and Gruber, 2006). The positive differences near the surface approximately match the expected increase over one full decade: DIC_{MOBO} in the top 200 m is approximately 13 µmol kg^{-1} higher than $DIC_{LAUVSET}$, indicating most of the difference between the two estimates stems from the difference in time period and the anthropogenic perturbation.

In addition to this offset near the surface, our estimate in the interior (below ~200 m) is, on average, ~10 µmol kg^{-1} lower than $DIC_{LAUVSET}$, which cannot simply be explained by the difference in reference years. Furthermore, there is a striking difference between the two estimates in the Atlantic sector between ~100 m and 1000 m, where the time-average of DIC_{MOBO} is lower by ~50 µmol kg^{-1} than $DIC_{LAUVSET}$. This region of high DIC in the Lauvset product may be explained by data availability. All of the available information here stems from a single cruise (33MW19930704) as well as a few calculated DIC values (based on observed total alkalinity and pH) from cruise 74DI19980423. The DIVA mapping used by Lauvset et al. (2016) draws no other information apart from the observations directly, the correlation length scale, and the signal-to-noise ratio. The latter two are subjectively chosen, and for $DIC_{LAUVSET}$, the signal-to-noise ratio is such, that the observations are considered climatologically representative, and therefore, closely fit. Our method, however, takes the high DIC values in the Atlantic in combination with the additional information from the predictor data, and thus, DIC_{MOBO} might be more representative of the true climatological state. In addition, the differences in the ocean interior could be due to the difference in the timespan. While our approach only considers measurements between 2004 and 2017, the approach by Lauvset et al. (2016) also includes measurements from earlier campaigns.

S4.2 Validation with synthetic data

To test how accurately our method reconstructs time-varying fields at global scale, we can turn to synthetic data. We take the model field from the biogeochemical component of the Ocean General Circulation Model HAMOCC (Ilyina et al., 2013; Mauritsen et al., 2019), which was run on a 1.5°x1.5° grid in hindcast mode with historic atmospheric forcing for the Global Carbon Budget 2018 (Le Quéré et al., 2018). We first re-grid the HAMOCC output onto the same grid and format as the observational predictor and target data (monthly means between 2004 through 2017, 33 depth levels between 2.5 m and 1975 m, 1°x1° grid, from 65°N to 65°S). We call the full model field of DIC in HAMOCC hereafter DIC_{HAMOCC}.

To test how well our method reconstructs the full model field, we subsample DIC_{HAMOCC} at the month and location where we have DIC observations in GLODAPv2.2019. We then use the same SOM-FFN set-up and run the method using the same predictors, but from HAMOCC, to reconstruct the DIC in HAMOCC (hereafter $DIC_{MOBO.HAMOCC}$). Finally, we compare $DIC_{MOBO.HAMOCC}$ with DIC_{HAMOCC}.

We are aware that the use of models to validate empirical methods has its limitations; for example, because the model field is considerably smoother than data from measurements, and because here, "synthetic observations" are the monthly mean value of the model output, and not a snap-shot measurement. Nonetheless, the test with synthetic data provides us with a way to qualitatively test our method at each grid cell, overcoming the problem of the paucity of independent in-situ validation data.

Run with synthetic data, the SOM-FFN method is capable of reconstructing the mean DIC_{HAMOCC} distribution, as illustrated in Fig. S5. The differences between DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$ remain within 10 µmol kg^{-1} for the majority of the ocean and the overall RMSE between the two DIC fields is 13.8 µmol kg^{-1} and a bias of +1.4 µmol kg^{-1}, strengthening our trust in the reconstructed DIC field. However, a few exceptions are visible where differences reach up to 50 µmol kg^{-1} in the deep Indian and Pacific Ocean, where fewer observations exist. The Indian Ocean is a region where, due to data sparsity, the uncertainty of our method is largest globally, as illustrated by the ensemble spread (Fig. 2 in the Main Text). We thus link the difference here to the data sparsity and substantial spatial extrapolation in this region. The differences in the deep Pacific Ocean, however, cannot be attributed to the ensemble spread. Here, the ensemble spread is smaller than in most shallow regions and so the large differences between DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$ in this basin are likely linked to processes not represented in our predictor data. This illustrates again that regional uncertainties can be considerably large in our global approach.

The surface seasonal cycle of $DIC_{MOBO.HAMOCC}$ in large scale regions remains close to the seasonal cycle of DIC_{HAMOCC} (Fig. S6), with the maximum difference between DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$ of 11 µmol kg^{-1} in the northern temperate band, where the full model field is a bit jagged, and so $DIC_{MOBO.HAMOCC}$ is lower in boreal spring and higher in boreal summer. In the northern subtropics, $DIC_{MOBO.HAMOCC}$ is lower than DIC_{HAMOCC} by up to 9 µmol kg^{-1}, especially in boreal autumn and winter, while in the southern subtropics, $DIC_{MOBO.HAMOCC}$ is lower by up to 10 µmol kg^{-1} in austral winter. In the tropics, $DIC_{MOBO.HAMOCC}$ agrees best with DIC_{HAMOCC}, and this is likely linked to the lack of strong variations. Overall, this test demonstrates that our method, as well as the number of available observations, are well suited to reconstruct the climatological DIC distribution, and in particular, the seasonal representation of DIC_{HAMOCC}, adding confidence to our method. The RMSE between DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$ at the surface is 13.0 µmol kg^{-1}.

S4.3 The seasonal cycle at time-series stations HOT and BATS

We further compare our estimate with data from independent time-series sites that were not used to train the network and have a long enough record to extract the mean seasonality. Although there are many time-series stations across the globe (Bates et al., 2014; see also https://www.nodc.noaa.gov/ocads/oceans/time_series_moorings.html), only a few stations measured DIC in the upper ocean from 2004 through 2017 and at locations that are not excluded in our product (i.e., coastal regions and latitudes poleward of 65°). The time-series stations that fall within our temporal and spatial domains are the Hawaii Ocean Time-Series (HOT, Dore et al., 2009) and the Bermuda Atlantic Time Series Study (BATS, Bates et al., 2014).

The HOT (http://www.soest.hawaii.edu/HOT_WOCE/ftp.html) and BATS (http://batsftp.bios.edu/BATS/bottle/A_README_BOTTLE.txt) databases consist of physical and biogeochemical ship data. The DIC measurements that form a part of these time-series datasets were taken from bottled sea-water samples. The HOT time-series extends from 1988 through 2017 for the upper ocean at 22°45’N, 158°00’W, north of the Hawaiian island chain, while the BATS time series extends from 1988 through 2016 at 31°40’N, 64°10’W, near Bermuda in the northwestern Sargasso Sea (marked in Fig. 1a in the Main Text).

For the validation, we compile all DIC measurements from the HOT and BATS databases and only keep the data that overlap with the period from our study (2004 through 2017). At BATS, while conducting our analysis, data from 2017 were not available, so here, the dataset ends in December 2016. We then compute a monthly climatology by taking the mean monthly values (hereafter DIC_{HOT} and DIC_{BATS}). While the HOT data extend to 1000 m, at BATS, only a few observations exist below 600 m, so here we only use the top 600 m for our validation. We test DIC_{MOBO} at the 1°x1° grid point closest to the HOT location (hereafter $DIC_{MOBO.HOT}$) and compare it to DIC_{HOT}. We also test how DIC_{HAMOCC} at the grid point closest to HOT (hereafter $DIC_{HAMOCC.HOT}$) compares to our estimate thereof (hereafter $DIC_{MOBO.HAMOCC.HOT}$). We do the same test at BATS: we compare $DIC_{MOBO.BATS}$ to DIC_{BATS} and $DIC_{MOBO.HAMOCC.BATS}$ to $DIC_{HAMOCC.BATS}$.

Although $DIC_{MOBO.HOT}$ represents the DIC phase and amplitude at station HOT well, it tends to underestimate DIC_{HOT} at most depths, except at 500 m depth (Fig. S7). Most of the concentrations illustrated in Fig. S7b are based on only a few observations. Therefore, these differences might be subject to internal variability at HOT that is not represented in our mean climatology. Both $DIC_{MOBO.HOT}$ and DIC_{HOT} illustrate the weak seasonal cycle of surface DIC in the subtropics (Fig. S7d). The signal to noise ratio in DIC_{HOT} is high, and hence, no strong seasonal cycle can be observed here, whereas $DIC_{MOBO.HOT}$ demonstrates a slightly stronger seasonal cycle. Nonetheless, given the locality of the measurements compared to the global reconstruction, the mean surface values between DIC_{HOT} and $DIC_{MOBO.HOT}$ compare remarkably well (1983 and 1974 µmol kg^{-1}, respectively at 10 m) and the overall RMSE between DIC_{HOT} and $DIC_{MOBO.HOT}$ is 14.2 µmol kg^{-1} and the bias is -7.9 µmol kg^{-1}.

$DIC_{HAMOCC.HOT}$ is considerably lower than DIC_{HOT} (by ~80 µmol kg^{-1}, Fig. S7d). Nonetheless, our method reproduces the seasonal cycle of $DIC_{MOBO.HAMOCC.HOT}$ relatively well in terms of the mean and phase, with the highest DIC concentration in May. However, $DIC_{MOBO.HAMOCC.HOT}$, as observed before for the large scale regions, overestimates the amplitude of the seasonal cycle compared to $DIC_{HAMOCC.HOT}$ (~9 µmol kg^{-1} compared to ~4 µmol kg^{-1}). The RMSE between $DIC_{HAMOCC.HOT}$ and $DIC_{HAMOCC.HOT}$ is 8.1 µmol kg^{-1}.

$DIC_{MOBO.BATS}$ demonstrates a much more pronounced seasonal DIC cycle compared to the one observed at HOT. Overall, the concentrations are higher by ~5 µmol kg^{-1} than DIC_{BATS} in the top 100 m, while between 100 m and 600 m our estimate is lower by up to 18 µmol kg^{-1} (Fig. S8a-c). Again, given the locality of the time-series station, we find an encouraging agreement regarding the phase and amplitude of the seasonal cycle in $DIC_{MOBO.BATS}$ at the surface (Fig. S8d). The surface seasonal cycle of DIC_{BATS} has approximately the same mean concentration as $DIC_{MOBO.BATS}$ (2061 and 2067 µmol kg^{-1}, respectively), as well as a matching phase of the seasonal cycle (largest value in March). However, $DIC_{MOBO.BATS}$ underestimates the observed DIC concentrations in the winter months (up to 13 µmol kg^{-1}) and the overall RMSE between DIC_{BATS} and $DIC_{MOBO.BATS}$ is 26.6 µmol kg^{-1} and the bias is -15.2 µmol kg^{-1}.

We find that $DIC_{HAMOCC.BATS}$ is considerably lower than the DIC_{BATS} by ~90 µmol kg^{-1}. Our method reproduces the amplitude of $DIC_{HAMOCC.BATS}$ quite accurately ($DIC_{MOBO.HAMOCC.BATS}$), but there is a 2-month phase shift (Fig. S8d). The RMSE between $DIC_{HAMOCC.BATS}$ and $DIC_{HAMOCC.BATS}$ is 5.9 µmol kg^{-1}.

S4.4 Argo floats with biogeochemical sensors (SOCCOM floats)

To test our method in the southern hemisphere, we use data from biogeochemical Argo floats that take measurements as part of the Southern Ocean Carbon and Climate Observations and Modelling project (SOCCOM, https://soccom.princeton.edu/). We compare the monthly mean DIC concentration calculated from the SOCCOM floats to DIC_{MOBO} at the month and location of the float measurements ($DIC_{MOBO.SOCCOM}$). The DIC from the SOCCOM floats is calculated using a combination of pH measurements, total alkalinity estimated using the commonly used LIAR algorithm (Carter et al., 2018), and the CO_2SYS analysis tool (van Heuven et al., 2011). As the SOCCOM float data is only available after 2014, we take the monthly mean values of DIC from 2014 through 2017. We then interpolate all SOCCOM float DIC measurements onto a 1°x1° grid and linearly interpolate the result onto our 33 depth levels (hereafter DIC_{SOCCOM}). We then compute the mean monthly fields regardless of the float location within the Southern Ocean. In the domain until 65°S, there are, on average, 160 grid cells that contain at least one SOCCOM float in each month of the year (see Fig. 1b in the Main Text). The data density of the SOCCOM floats is relatively high, although the period of these observations only extends over four years (2014 through 2017).

We find that $DIC_{MOBO.SOCCOM}$ agrees well in phase with the DIC_{SOCCOM}, but DIC_{SOCCOM} is, on average, 16 µmol kg^{-1} higher than $DIC_{MOBO.SOCCOM}$ (Fig. S9). Comparatively higher carbon values measured by the SOCCOM floats have been reported in recent studies (Bushinsky et al., 2019; Gray et al., 2018; Williams et al., 2017), who found that SOCCOM floats demonstrated additional outgassing in austral winter months compared to estimates based on ship data. The mean surface seasonal cycle of $DIC_{MOBO.SOCCOM}$ has a lower amplitude by ~6 µmol kg^{-1} (Fig. S9d), owing to the disagreement in austral winter. The overall RMSE between $DIC_{MOBO.SOCCOM}$ and DIC_{SOCCOM} is 22.8 µmol kg^{-1} and the bias is -16.1 µmol kg^{-1}.

Comparing the mean seasonal cycle of DIC_{HAMOCC} with $DIC_{HAMOCC.SOCCOM}$, we find that the seasonal cycle in $DIC_{HAMOCC.SOCCOM}$ has a much larger amplitude (by ~19 µmol kg^{-1}) than DIC_{SOCCOM}, and the phase is shifted backward by ~2 months. However, $DIC_{MOBO.HAMOCC.SOCCOM}$ compares well with $DIC_{HAMOCC.SOCCOM}$, in phase, amplitude, and mean concentration, demonstrated by an RMSE of 7.4 µmol kg^{-1}.

S4.5 The surface seasonal cycle at Drake Passage time-series station

In addition to the time-series stations that measure DIC in the water column, here, we compare DIC_{MOBO} with a time-series station that contains surface measurements of DIC, the Drake Passage time-series station (Munro et al., 2015). The Drake Passage time-series is one of the most comprehensive datasets of carbon measurements in the Southern Ocean, including DIC data from bottled sea-water samples during multiple ship crossings per year from 2004 through 2017 (Munro et al., 2015, https://www.nodc.noaa.gov/archive/arc0118/0171470/2.2/data/0-data/). We use all DIC measurements from that time-series that are south of 54°S and east of 70°W, i.e. between the southern tip of Chile and the Antarctic Peninsula. Fig. 1b in the Main Text delimits the region of the ship cruises that we use from this time-series, and the ship tracks can also be found under https://data.nodc.noaa.gov/cgi-bin/gfx?id=gov.noaa.nodc:0171470. The exclusion of some cruises further away from the main routes is to ensure a relatively uniform dataset, enabling us to investigate the temporal variability in this region. We put the DIC measurements from this dataset onto a regular 1°x1° grid, and compute the monthly means from 2004 through 2017 (hereafter DIC_{DRAKE}). Next, we compare DIC_{DRAKE} to DIC_{MOBO} at the month and location at the grid points of the Drake time-series measurements ($DIC_{MOBO.DRAKE}$).

We find that the time-mean of $DIC_{MOBO.DRAKE}$ and DIC_{DRAKE} are mostly in agreement with each other (Fig. S10a-c). One exception is unusually high values in $DIC_{MOBO.DRAKE}$ in the north, which we expect are linked to internal variability and are not seasonally representative of this region. Overall, the RMSE between the two datasets is 29.6 µmol kg^{-1} and the bias is 3.0 µmol kg^{-1}, although most of the discrepancy between the two datasets stems from the high values in the north in DIC_{DRAKE}.

Comparing DIC_{HAMOCC} at the time and location of the Drake Passage measurements ($DIC_{HAMOCC.DRAKE}$) with $DIC_{MOBO.HAMOCC}$ at the same month and location (hereafter $DIC_{MOBO.HAMOCC.DRAKE}$), reveals broad agreement between the two estimates in terms of phase, mean, and amplitude, but $DIC_{MOBO.HAMOCC.DRAKE}$ is a lot smoother. The overall RMSE between these two datasets is 17.8 µmol kg^{-1}. As with the other validation tests with time-series, the HAMOCC model tends to be very different than the observational estimates, but our reconstruction thereof can adequately reproduce the model field.

In summary, given the assessments above, we demonstrate that our method can reconstruct the phase of the seasonal cycle at the sea surface well. The overall RMSE between our DIC estimates (DIC_{MOBO} and $DIC_{MOBO.HAMOCC}$) and the validation data is between 5.9 and 26.6 µmol kg^{-1} (see Fig. S11). As a large part of the discrepancies come from differences in time periods and internal variability rendering the observations not always seasonally representative, we argue that overall, our method adequately represents the monthly climatology of DIC. We demonstrate that DIC_{MOBO} is considerably closer to the independent test data that were not used to train the network (HOT, BATS, SOCCOM, Drake Passage) than the DIC_{HAMOCC} at those locations (Fig. S7d, S8d, and S9d), suggesting that our method may better capture the seasonal cycle of DIC than the HAMOCC model.

Text S5. Seasonal response function (statistical drivers)

To investigate how each of the predictors contributes to our estimate of the seasonal changes in DIC, we compute the seasonal response function for each of the predictors. We use an approach similar to the “profile method” described in Gevrey et al. (2003), which is commonly used in sensitivity analyses to determine how changes in the predictors affect the target data in a neural network. In the profile method, the network is trained as usual, and in the simulation step, each predictor is consecutively varied while holding the remaining predictors constant. As we are interested in the seasonal response in different regions, we adapt that method, only holding the time dimension constant (i.e., we use the time-mean of each grid-cell), while varying in space.

Our method works as follows: We first calculate DIC_{base} by training the network as usual and then apply the network while keeping all predictors constant in time (i.e., using the time-mean at each grid cell). Next, we simulate the network again consecutively for each predictor, while keeping all of the predictors except the predictor under evaluation constant in time. For example, we calculate $DIC_{temperature}$ by simulating the network with all of the predictors kept constant in time, except temperature. Lastly, for each predictor, we calculate DIC_{input} by subtracting the DIC_{input} of that predictor from the DIC_{base}; for example, for temperature: $\Delta DIC_{temperature} = DIC_{base} - DIC_{temperature}$. We repeat our bootstrapping approach by simulating these ten times to calculate the mean response over the ensemble.

Near the sea surface, i.e., where we observe the largest seasonal amplitude in the different climate regions (Fig. S12), we find that most of the seasonal changes of DIC_{MOBO} at the surface are linked to temperature as our main predictor. Temperature is inversely linked to DIC (Takahashi et al., 2002) and contributes to the seasonality two-fold. Colder waters are linked to higher solubility and increased vertical mixing, and both increase the surface DIC pool (Heinze et al., 2015; Sarmiento and Gruber, 2006). In the temperate regions, nitrate, representing nutrient input to the surface, is also a strong statistical driver of DIC_{MOBO}, thus affecting the seasonal cycle at the surface. The strong influence of nitrate highlights the importance of including upwelling and biology in reconstructing the seasonal cycle. Nutrient availability through vertical mixing or river input triggers biological production, lowering the DIC concentration at the surface (Sarmiento and Gruber, 2006; Takahashi et al., 2002). Hence, the effects of temperature and biology are competing in the temperate regions as statistical drivers of pCO2, and thus, DIC, and both need to be considered to reconstruct the seasonal DIC cycle faithfully. The remaining proxies, i.e. salinity, dissolved oxygen, and silicate play overall a smaller statistical role in our reconstruction.

Text S6. Interpretation of the nodal depth and validation of the nodal depth with synthetic data

To better interpret the distribution of the nodal depth, we presented the difference between the nodal depth and the mean depth of the euphotic zone, as well as the difference between the nodal depth and the mean winter mixed layer depth (MLD) in Fig. 8 in the Main Text. Fig. S13 presents the mean winter MLD (a) and the mean depth of the euphotic zone (b).

To test our estimate of DIC nodal depth, we return to the synthetic data from the HAMOCC model (Ilyina et al., 2013; Mauritsen et al., 2019). We compute the nodal depth the same way as described in the Main Text, but this time, we compute it first using DIC_{HAMOCC} and second using $DIC_{MOBO.HAMOCC}$ (Fig. S14). The seasonal cycle of inorganic carbon is not very well captured in HAMOCC (e.g., Mongwe et al., 2018), rendering this comparison challenging to interpret. There are many areas, where our algorithm to determine the nodal depth does not pick up a nodal depth (see white patches in Fig. S12a-b). Nonetheless, this comparison provides us with an idea of the error of the nodal depth in our reconstruction of DIC.

Comparing the nodal depth estimate using $DIC_{MOBO.HAMOCC}$ and DIC_{HAMOCC}, we find that our reconstruction overestimates the DIC nodal depth in many places, and there are various patches of very deep nodal depths in $DIC_{MOBO.HAMOCC}$ (Fig. S14a-c). However, the general distribution of the pattern is very similar in the two estimates and the RMSE between the nodal depth computed with $DIC_{MOBO.HAMOCC}$ and $DIC_{MOBO.HAMOCC}$ is 59 m. Fig. S14d depicts the DIC nodal depth using DIC_{MOBO} (adapted from Fig. 6b in the Main Text). Here, we also find patches of deeper nodal depths. Based on our test with synthetic data, we argue that the patchiness is likely a result of the data extrapolation and the sensitivity of the analysis towards uncertainties in the amplitude that can be significant.

Text S7. Validation of the summer net community production (NCP) with synthetic data

We test our estimate of the summer NCP, using the HAMOCC model (Ilyina et al., 2013; Mauritsen et al., 2019) to test how well the seasonal draw-down of DIC in our reconstruction of the model ($DIC_{MOBO.HAMOCC}$) represents the seasonal draw-down of DIC in the model (DIC_{HAMOCC}). To do so, we first compute the summer NCP the same way as described in the Main Text, but with the variables from HAMOCC (hereafter Summer NCP_{HAMOCC}). We then compute the summer NCP again with all HAMOCC variables and $DIC_{MOBO.HAMOCC}$ to derive Summer $NCP_{NN.HAMOCC}$.

We find that Summer $NCP_{NN.HAMOCC}$ compares well with Summer NCP_{HAMOCC} in terms of the distribution pattern, such as the large production in the Southern Ocean (Fig. S15a-c). However, there are some quantitative discrepancies, and the integrated Summer $NCP_{NN.HAMOCC}$ over the extra-tropics is 2.0 PgC summer^{-1}, while the summer NCP is 1.5 PgC summer^{-1} when computing it with DIC_{HAMOCC}. Upscaling the mean NCP onto the global ocean, we find a global summer NCP of 3.5 Pg summer^{-1} using DIC_{HAMOCC}, and 4.7 PgC summer^{-1} using $DIC_{MOBO.HAMOCC}$. The NCP estimate in the HAMOCC model is considerably lower than our estimate based on DIC_{MOBO}, and some regions show slightly negative values for the NCP, in both HAMOCC-based NCP estimates. We suspect that this is due to a less well-represented seasonality in HAMOCC, as well as the missing horizontal divergence, that we have to neglect in the calculation of the NCP (see Eq. 1 in the Main Text). Other sources of error in our NCP estimate are discussed in the Main Text (Eq. 5).

As an additional qualitative test for our summer NCP estimation, we show the carbon export at 100 m in HAMOCC (an output variable from the HAMOCC model that describes the sinking mole flux of particulate organic matter expressed as carbon in sea-water). Although the carbon export is not exactly the same as the NCP, as the latter does not account for the export of dissolved organic matter, and the production of biomass, it allows us to qualitatively compare it to Summer NCP_{HAMOCC}. Our method does capture the main features seen in the carbon export, such as the pronounced export in the Southern Ocean and the North Pacific, adding confidence in our method of calculating the Summer NCP from the seasonal draw-down of DIC. The summer export in the extra-tropics is 4.1 PgC summer^{-1}, which is considerably more than Summer NCP_{HAMOCC}, likely linked to some negative values in Summer NCP_{HAMOCC}, the missing horizontal divergence in Summer NCP_{HAMOCC}, and the fact that the export is not exactly the same as the NCP, as the export accounts for the export of dissolved organic matter, and the production of biomass, while the NCP does not.

Figures and Tables

Table S1. Input variables for the SOM and FFN for the three different depth slabs (2.5 to 500 m, 600 to 1500 m, 1600 to 1975 m). The depth levels are expressed where 75:25:150 means from 75 m to 150 m in steps of 25 m. For the SOM input variables, clim. DIC refers to the mean annual climatology by Lauvset et al. (2016).

Depth	Depth levels (m)	Number of SOM clusters	SOM input variables	FFN input variables (predictor data)
2.5–500m	2.5:2.5:10 20:10:50; 75:25:150; 200:50:500; (18 depth levels)	6	temperature, salinity, clim. DIC	temperature, salinity, dissolved oxygen, silicate, nitrate
600–1500m	600:100:1500 (10 depth levels)	4	temperature, salinity, clim. DIC	temperature, salinity, dissolved oxygen
1600–1975m	1600:100:1900; 1975 (5 depth levels)	4	temperature, salinity, clim. DIC	temperature, salinity

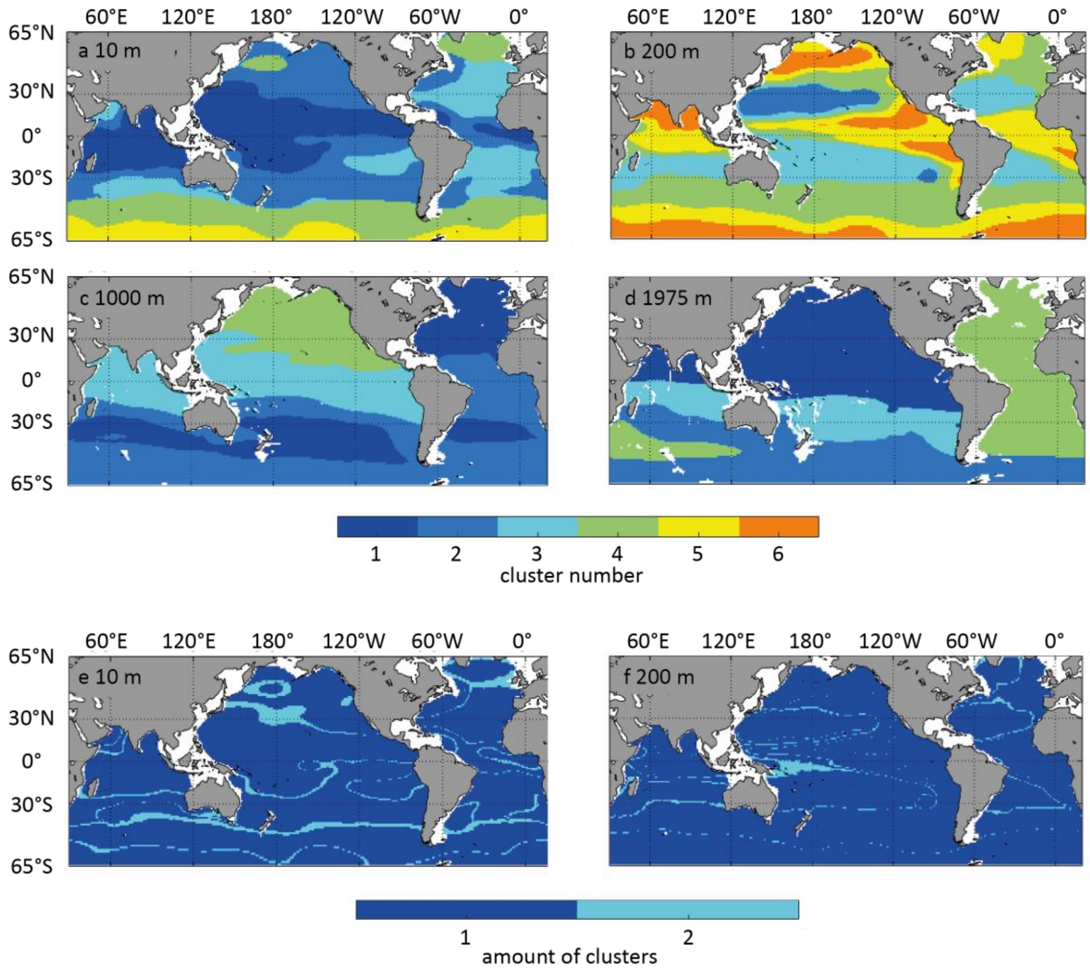

Figure S1. Location and variability of SOM clusters. Spatial distribution of the SOM clusters in January for 4 depth levels (a: 10 m, b: 200 m; c: 1000 m; d: 1975 m) and the number of different clusters throughout the monthly climatology at two depth levels (e: 10 m, f: 200 m).

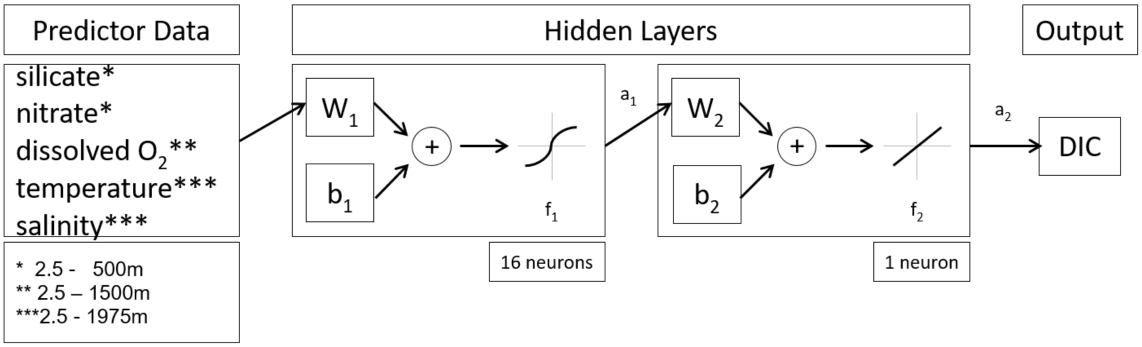

Figure S2. Schematic of our FFN configuration. Predictor data: silicate and nitrate until 500 m, dissolved oxygen until 1500 m, temperature and salinity until 1975 m; W: weight matrices; b: bias matrices, +: sum; f: transfer function; a: output matrices; subscripts indicate the number of the layer; boxes below the hidden layers indicate the number of neurons used. Modified from Hagan et al. (2014).

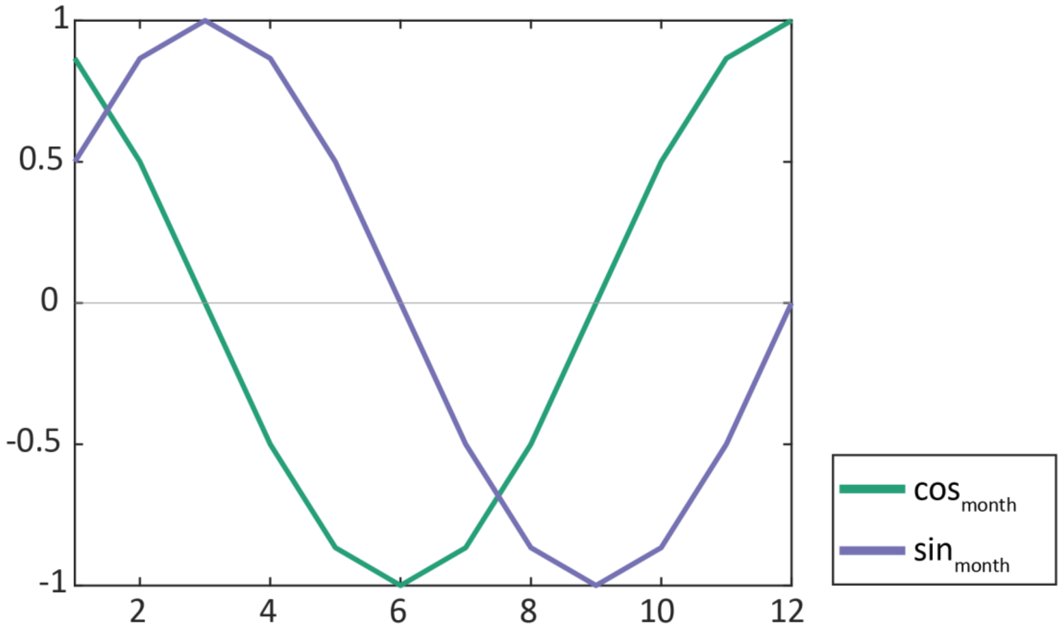

Figure S3. The curves of the cosine and sine of the month of the year.

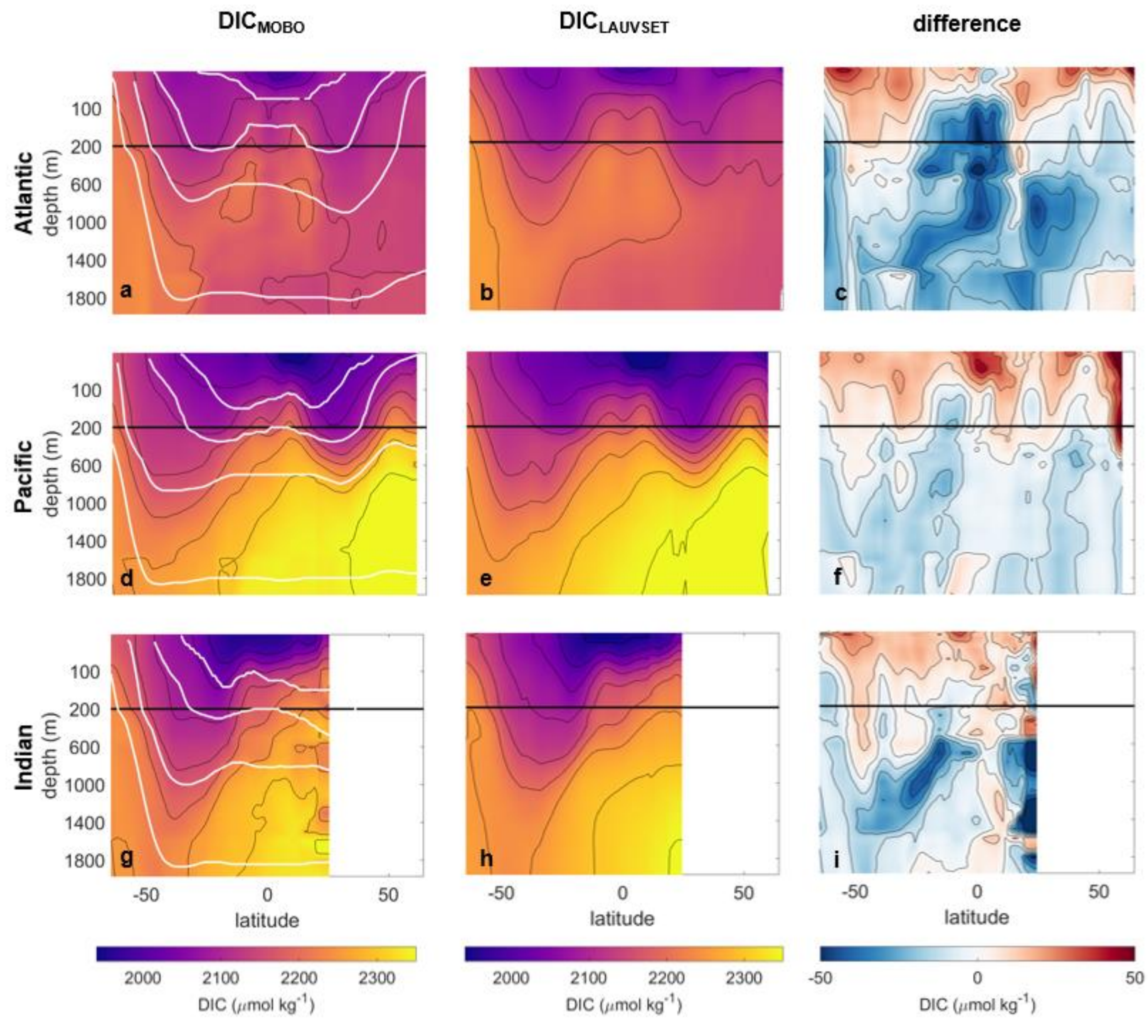

Figure S4. Comparison between DIC$_{LAUVSET}$ and DIC$_{MOBO}$. Zonal mean of the annual mean DIC$_{MOBO}$ (a,d,g), DIC$_{LAUVSET}$ (b,e,h), and the difference between the two (DIC$_{MOBO}$ - DIC$_{LAUVSET}$ (c,f,j). For each of the three sectors: Atlantic (a-c), Pacific (d-f); Indian (g-i). Zoomed into the top 200 m (delimited in black). Some isopycnals are illustrated as white lines in a,d,g (from top to bottom: 24.5, 26.2, 27.6, and 28.4 kg m^{-3}).

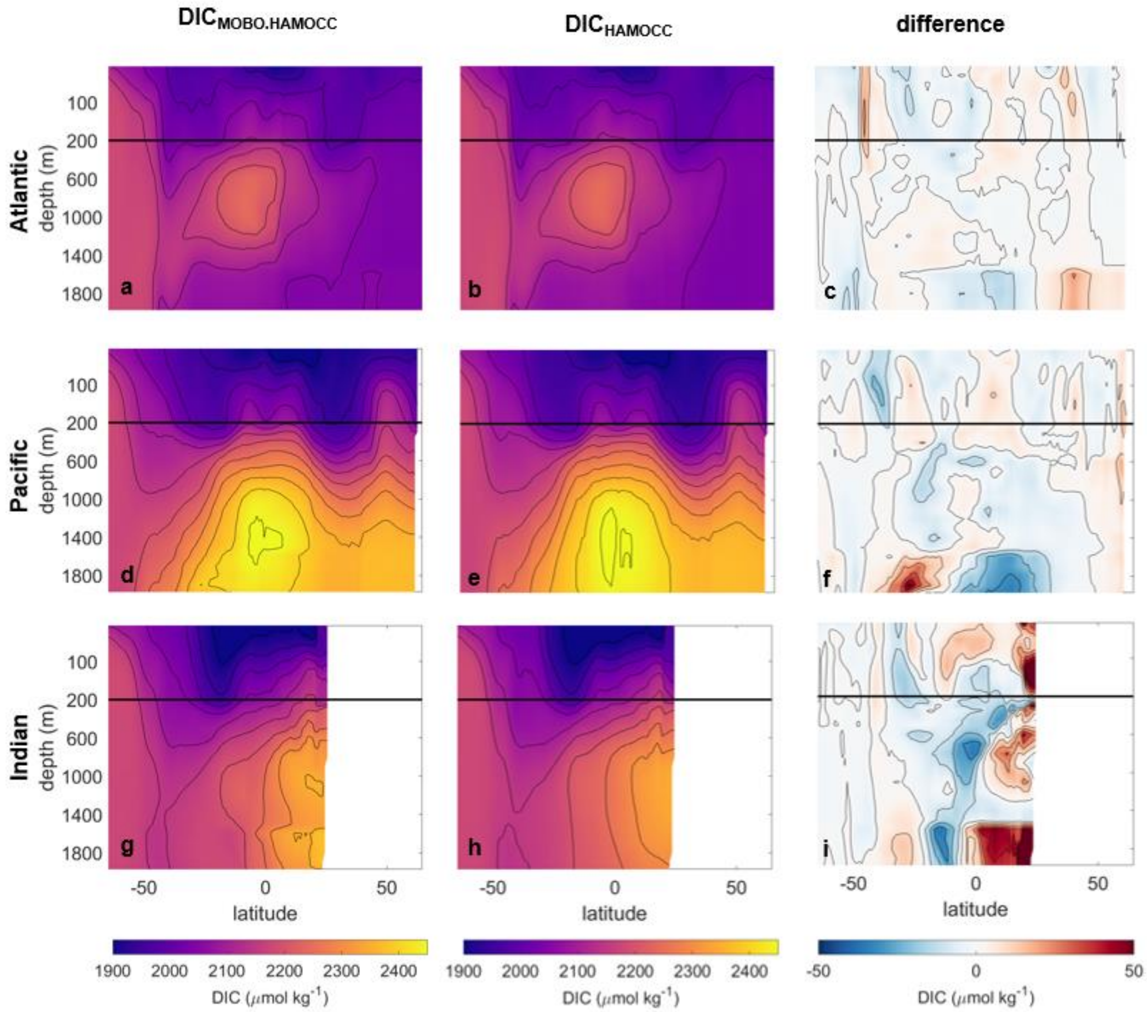

Figure S5. Comparison between the DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$. Zonal mean of the $DIC_{MOBO.HAMOCC}$ (a,d,g), DIC_{HAMOCC} (b,e,h), and the difference between the two ($DIC_{MOBO.HAMOCC}$ - DIC_{HAMOCC} (c,f,j). For each of the three sectors: Atlantic (a-c), Pacific (d-f); Indian (g-i). Zoomed into the top 200 m (delimited in black).

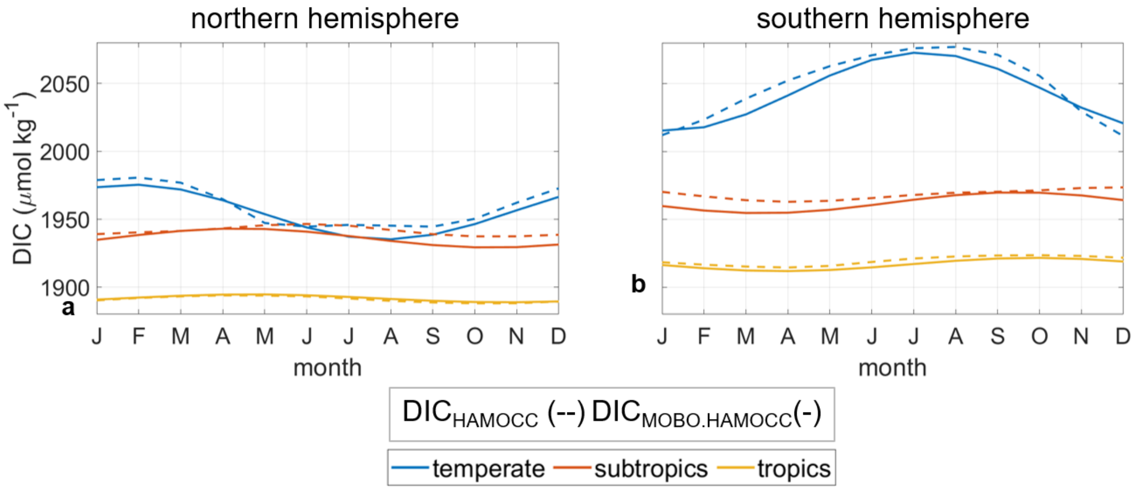

Figure S6. Seasonal cycle of DIC_{HAMOCC} and $DIC_{MOBO.HAMOCC}$ at 10 m in different climate regions. DIC_{HAMOCC} (dashed line) and $DIC_{MOBO.HAMOCC}$ (solid line): Temperate (35° to 65°, blue), subtropical (23° to 35°, orange), and tropical (0° to 23°, yellow) for the northern (a) and southern (b) hemispheres.

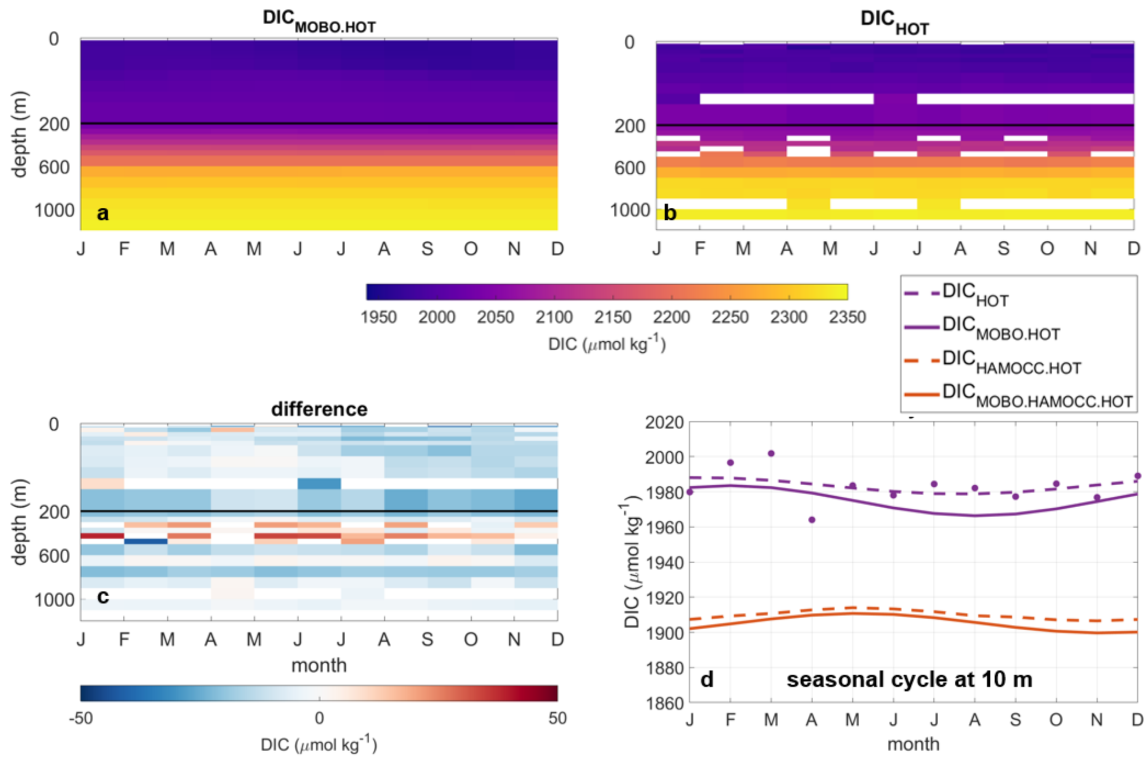

Figure S7. Comparison between the DIC_{HOT} and $DIC_{MOBO.HOT}$. a) $DIC_{MOBO.HOT}$; b) DIC_{HOT} c) the difference between the two ($DIC_{MOBO.HOT} - DIC_{HOT}$). d) Seasonal cycle at 10 m from DIC_{HOT} (purple dashed), $DIC_{MOBO.HOT}$ (purple) solid, $DIC_{HAMOCC.HOT}$ (orange dashed), $DIC_{MOBO.HAMOCC.HOT}$ (orange solid), illustrating the calculated value (filled circles) and the least-squares fit (solid lines); a-c are zoomed into the top 200 m.

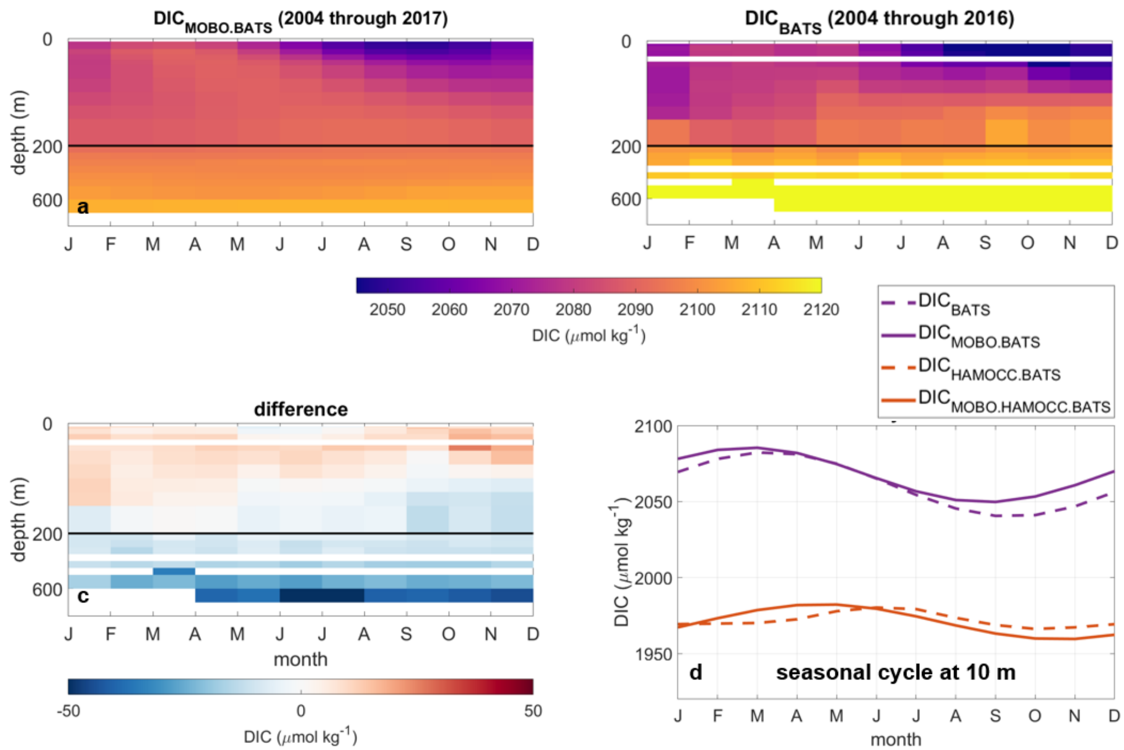

Figure S8. Comparison between the DIC$_{BATS}$ and DIC$_{MOBO.BATS}$. a) DIC$_{MOBO.BATS}$; b) DIC$_{BATS}$ c) the difference between the two (DIC$_{MOBO.BATS}$ – DIC$_{BATS}$). d) Seasonal cycle at 10 m from DIC$_{BATS}$ (purple dashed), DIC$_{MOBO.BATS}$ (purple solid), DIC$_{HAMOCC.BATS}$ (orange dashed), DIC$_{MOBO.HAMOCC.BATS}$ (orange solid); a-c are zoomed into the top 200 m.

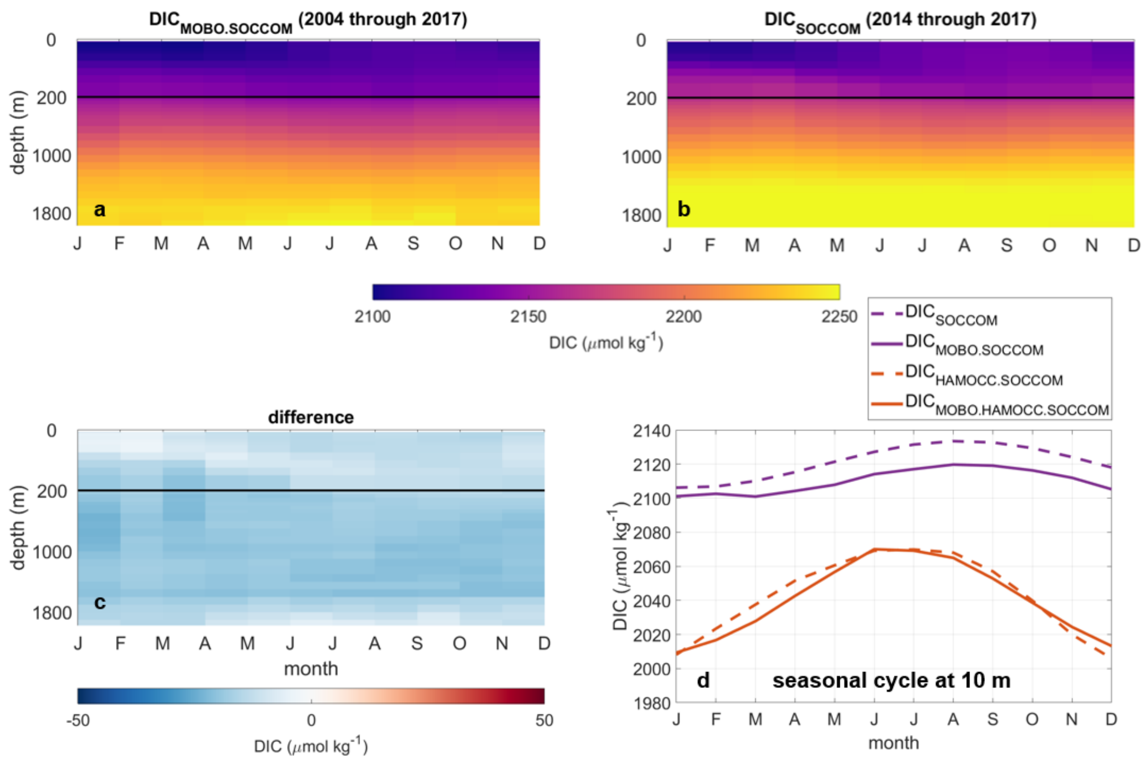

Figure S9. Comparison between the $\mathrm{DIC}_{\mathrm{SOCCOM}}$ and $\mathrm{DIC}_{\mathrm{MOBO.SOCCOM}}$. a) $\mathrm{DIC}_{\mathrm{MOBO.SOCCOM}}$; b) $\mathrm{DIC}_{\mathrm{SOCCOM}}$ c) the difference between the two ($\mathrm{DIC}_{\mathrm{MOBO.SOCCOM}}$ – $\mathrm{DIC}_{\mathrm{SOCCOM}}$). d) Seasonal cycle at 10 m from $\mathrm{DIC}_{\mathrm{SOCCOM}}$ (purple dashed), $\mathrm{DIC}_{\mathrm{MOBO.SOCCOM}}$ (purple solid), $\mathrm{DIC}_{\mathrm{HAMOCC.SOCCOM}}$ (orange dashed), $\mathrm{DIC}_{\mathrm{MOBO.HAMOCC.SOCCOM}}$ (orange solid); a-c are zoomed into the top 200 m.

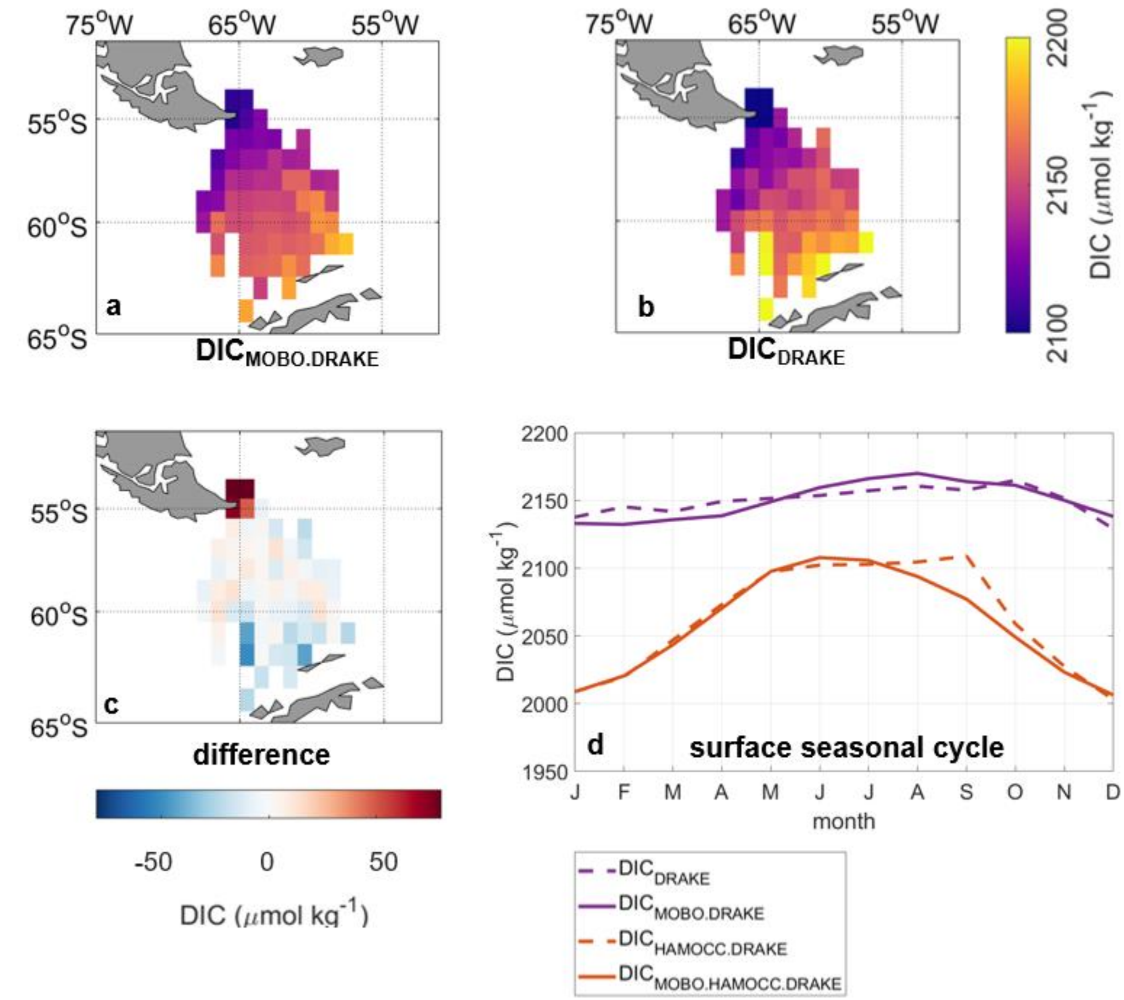

Figure S10. Comparison between the DIC_{DRAKE} and $DIC_{MOBO.DRAKE}$. a) $DIC_{MOBO.DRAKE}$; b) DIC_{DRAKE} c) the difference between the two ($DIC_{MOBO.DRAKE}$ – DIC_{DRAKE}). d) Surface seasonal cycle from DIC_{DRAKE} (purple dashed), $DIC_{MOBO.DRAKE}$ (purple solid), $DIC_{HAMOCC.DRAKE}$ (orange dashed), $DIC_{MOBO.HAMOCC.DRAKE}$ (orange solid).

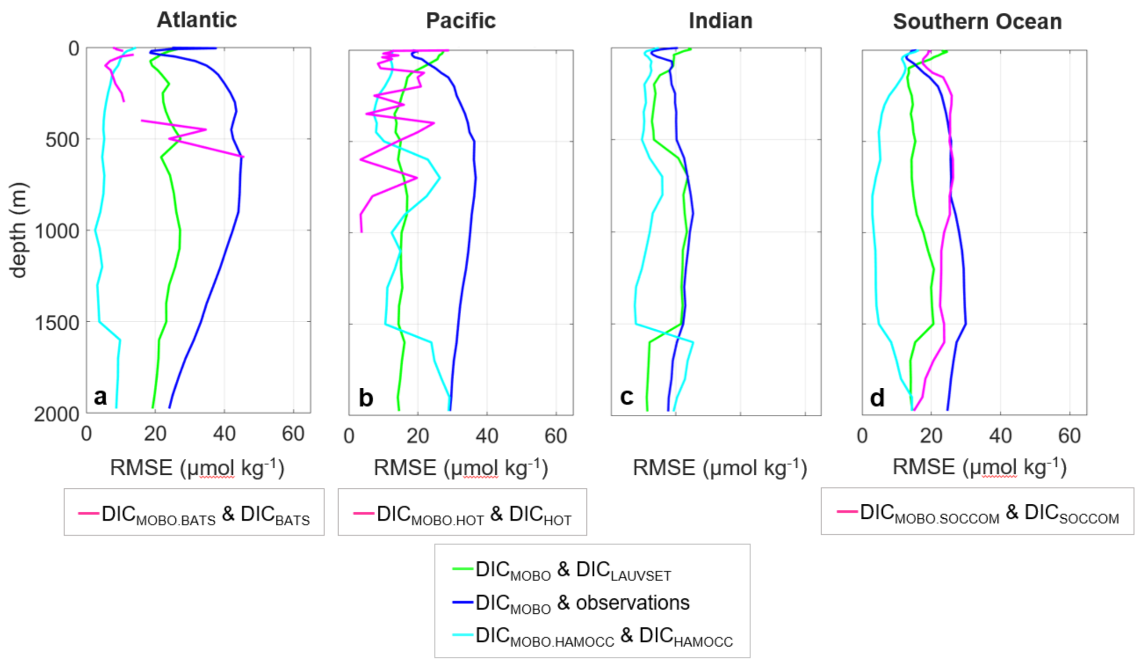

Figure S11. Summary of validation tests. RMSE as a function of depth for the Atlantic (a), Pacific (b), Indian (c), and Southern (d) Ocean. Showing the difference between DIC_{MOBO} and $DIC_{LAUVSET}$ (green). The residuals of DIC_{MOBO} from the observations (dark blue), and the difference between the $DIC_{MOBO.HAMOCC}$ and DIC_{HAMOCC} (light blue). The basins with independent observational data also show the difference between that (i.e. DIC_{BATS} (a), DIC_{HOT} (b), and DIC_{SOCCOM} (c)) and DIC_{MOBO} (magenta). As the Drake Passage time-series only covers the sea-surface, the RMSE is not included here.

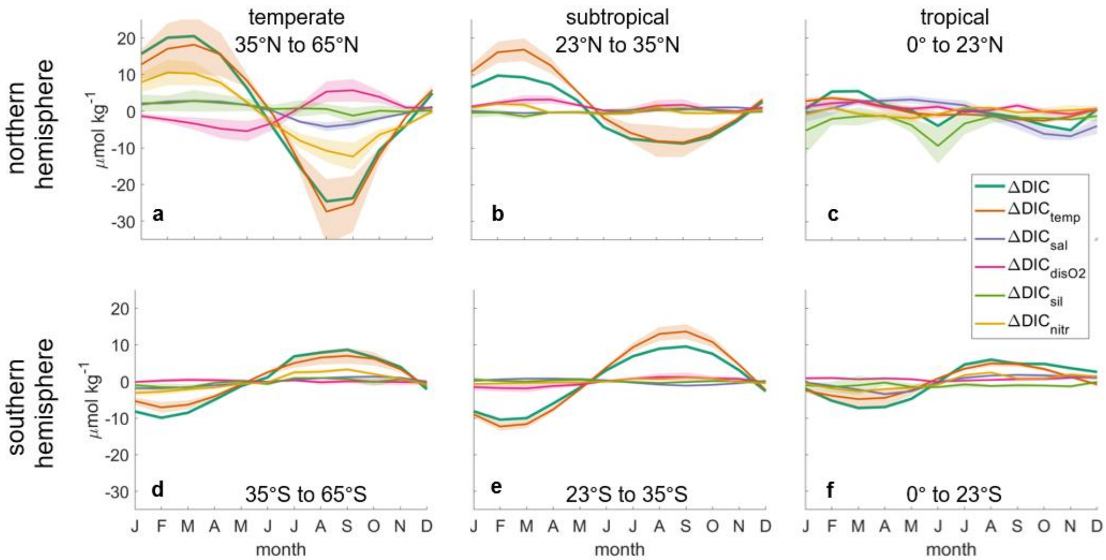

Figure S12. The seasonal response function at 2.5 m in different climate regions. Temperate (a,d; 35° to 65°), subtropical (b,e; 23° to 35°), and tropical (c,f; 0° to 23°) for the northern (a-c) and southern (d-f) hemisphere, $\Delta DIC_{temperature}$ (orange), $\Delta DIC_{salinity}$ (purple), $\Delta DIC_{dissolved.oxygen}$ (magenta), $\Delta DIC_{silicate}$ (light green), $\Delta DIC_{nitrate}$ (yellow). The mean of the 10-member ensemble is illustrated as solid line, and one standard deviation around the mean in shading. ΔDIC (dark green) is the mean seasonal anomaly at 10 m from our data estimate.

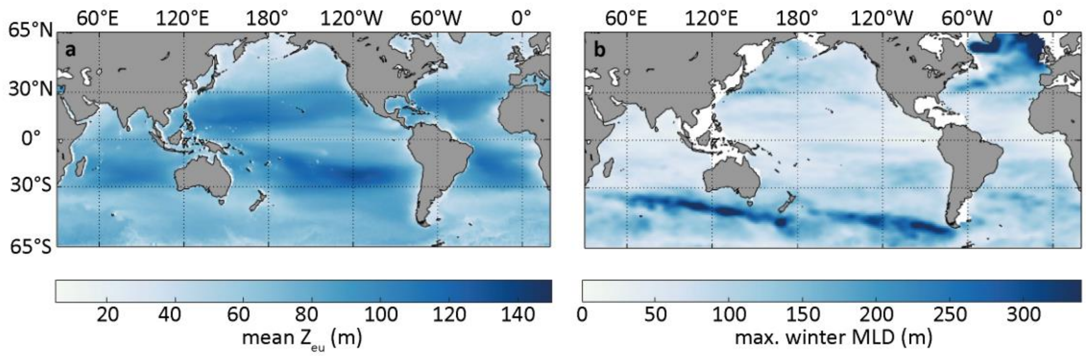

Figure S13. Additional plots for the analysis of the nodal depth. (a) Temporal mean depth of the 1% euphotic zone (Z_{eu}). (b) Maximum winter MLD. Note the different color scales in (a) and (b).

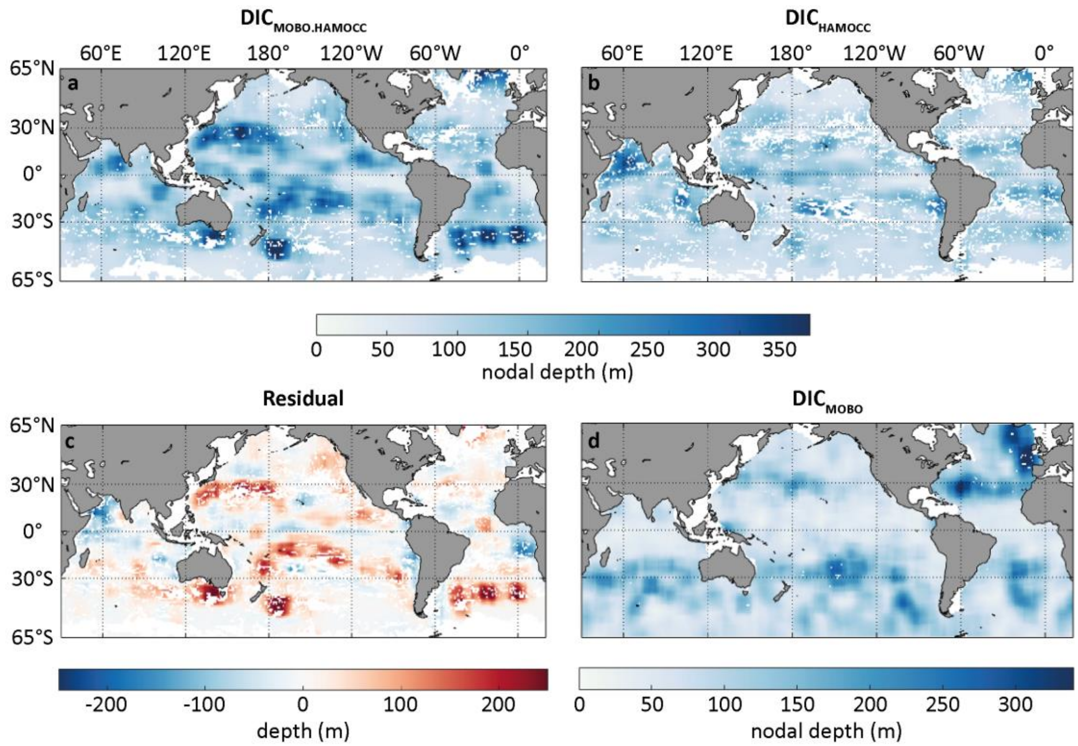

Figure S14. Test of the DIC nodal depth with synthetic data. a) Nodal depth calculated with $DIC_{MOBO.HAMOCC}$ b) Nodal depth calculated with DIC_{HAMOCC} c) Residual (Fig. S14a – Fig S14b). d) Nodal depth calculated with DIC_{MOBO} (modified from Fig. 6 in the Main Text).

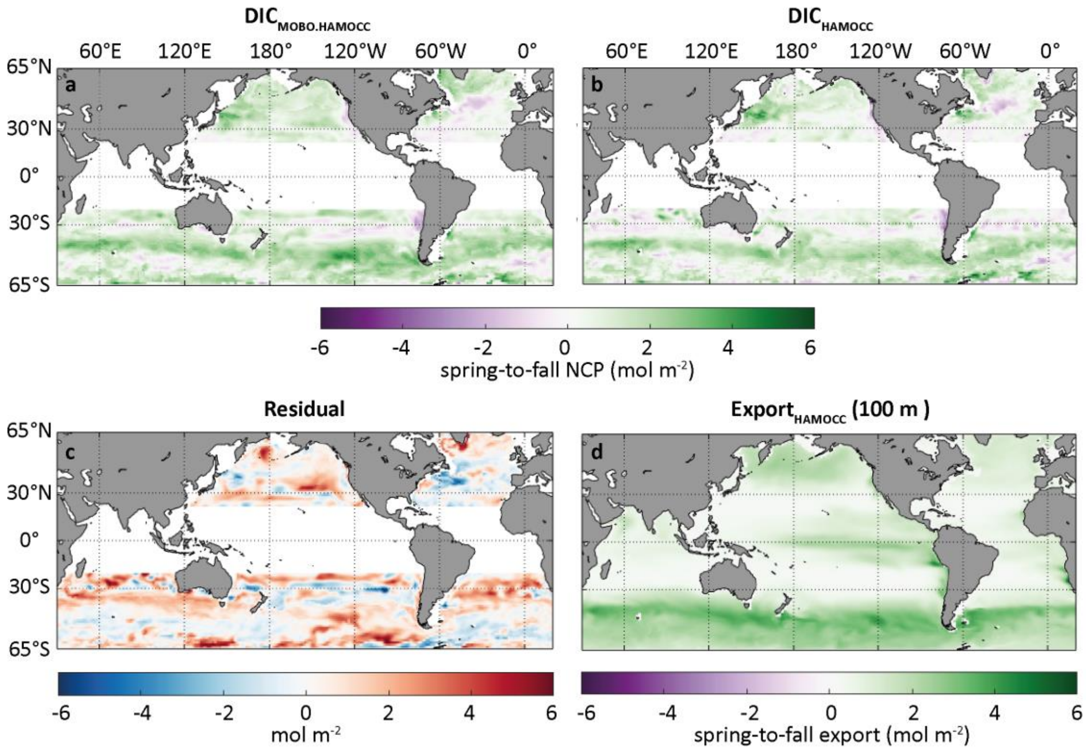

Figure S15. Test of Summer NCP with synthetic data. a) Summer NCP calculated with $DIC_{MOBO.HAMOCC}$ and variables from HAMOCC b) Summer NCP calculated with DIC_{HAMOCC} and variables from HAMOCC c) Residual (Fig. S15a – Fig S15b). d) Carbon export over hemispheric summer in HAMOCC (sinking mole flux of particulate organic).