Journal Of Marine Systems
June 2016, Volume 158 Pages 106-119
https://doi.org/10.1016/j.jmarsys.2016.02.011
https://archimer.ifremer.fr/doc/00383/49457/

Impacts of data assimilation on the global ocean carbonate system

Visinelli L. [1], Masina S. [1, 2, *], Vichi M. [3, 4], Storto A. [1], Lovato T. [1]

[1] Fdn CMCC Ctr Euro Mediterraneo Cambiamenti Climat, Bologna, Italy.
[2] INGV, Bologna, Italy.
[3] Univ Cape Town, Dept Oceanog, ZA-7700 Rondebosch, South Africa.
[4] Univ Cape Town, Nansen Tutu Ctr Marine Environm Res, ZA-7700 Rondebosch, South Africa.

* Corresponding author : S. Masina, email address : simona.masina@ingv.it

Abstract :

In an ocean reanalysis, historical observations are combined with ocean and biogeochemical general circulation models to produce a reconstruction of the oceanic properties in past decades. This is one possible method to better constrain the role of the ocean carbon cycle in the determination of the air–sea CO2 flux. In this work, we investigate how the assimilation of physical variables and subsequently the combined assimilation of physical data and inorganic carbon variables – namely dissolved inorganic carbon (DIC) and alkalinity – affect the modelling of the marine carbonate system and the related air–sea CO2 fluxes. The performance of the two assimilation exercises are quantitatively assessed against the assimilated DIC and alkalinity data and the independent ocean surface pCO2 observations from global datasets. We obtain that the assimilation of physical observations has contrasting effects in different ocean basins when compared with the DIC and alkalinity data: it reduces the root-mean square error against the observed pCO2 in the Atlantic and Southern oceans, while increases the model error in the North Pacific and Indian Oceans. In both cases the corrected evaporation rates are the major factor determining the changes in concentrations. The assimilation of inorganic carbon variables on top of the physical data gives a generalized improvement in the model error of inorganic carbon variables, also improving the annual mean and spatial distribution of air–sea fluxes in agreement with other published estimates. These results indicate that data assimilation of physical and inorganic carbon data does not guarantee the improvement of the simulated pCO2 in all the oceanic regions; nevertheless, errors in pCO2 are reduced by a factor corresponding to those associated with the air–sea flux formulations.

Highlights

► Impacts of physical–biogeochemical data assimilation on marine carbonate system are addressed. ► A coupled model (NEMO, LIM2, BFM) for the global ocean is combined with 3DVAR system. ► Physical-only data assimilation has contrasting effects on the regional carbonate system dynamic. ► Combined assimilation of physical and inorganic carbon data improves global estimate of DIC and ALK.

Keywords : Ocean biogeochemical dynamics, Data assimilation, Carbon cycle

1. Introduction

Both the land and the ocean act as sinks capable of absorbing fractions of atmospheric CO_2 (Le Quéré *et al.* 2015). Since the terrestrial sink turns out to be one of the most uncertain terms, it is usually derived as a difference between the atmospheric growth rate and ocean uptake (Canadell *et al.* 2007; Le Quéré *et al.* 2012). The reconstruction of the air-sea CO_2 flux is thus crucial in closing the global carbon budget. In addition, an adequate estimation of current ocean-atmosphere fluxes is required by the concern that the capability of the ocean in absorbing atmospheric CO_2 is likely to diminish in the future because of the saturation in the natural sinks due to surface ocean warming and reduced uptake efficiency (Sarmiento and Le Quéré 1996; Matear and Hirst 1999; Joos *et al.* 1999; Le Quéré *et al.* 2007, 2010; Ballantyne *et al.* 2012).

Recent studies aimed at assessing the value of the global and regional air-sea CO_2 flux, using inorganic carbon data from publicly available global ocean databases, account for ocean inversion methods (Gloor *et al.* 2001, 2003; Gurney *et al.* 2004; Patra *et al.* 2005; Jacobson *et al.* 2007; Mikaloff-Fletcher *et al.* 2007; Gruber *et al.* 2009; Maksyutov *et al.* 2013), interpolation procedures (Takahashi and Sutherland 2007; Takahashi *et al.* 2009; Jones *et al.* 2012; Park *et al.* 2010; Chen *et al.* 2011; Deng and Chen 2011; Gerber and Joos 2010), neural networks (Lefèvre *et al.* 2005; Telszewski *et al.* 2009; Landschützer *et al.* 2013), and prognostic Ocean Biogeochemical General Circulation Models (OBGCM) (Watson and Orr 2003; Matsumoto *et al.* 2004; Le Quéré *et al.* 2010).

In particular, the application of OBGCMs represents an alternative to the ocean and atmosphere inversion methods (Wanninkhof *et al.* 2013). In such a framework, a biogeochemical and physical oceanic models are coupled to reconstruct both the physical state and the biogeochemical properties of the ocean. The advantage offered by OBGCMs over statistical methods stems in the fact that the underlying models rely on diagnostic and prognostic equations, which in turn tests our knowledge of the main mechanisms involved. In a forward OBGCM, ocean physical dynamics are simulated with discretized primitive equations whose major uncertainties are mostly related to coarse spatial resolutions and sub-grid scale parameterizations. In particular, different realizations of the surface forcing or the model architecture used in the ocean dynamics have been demonstrated to give substantial differences in the resulting fields of inorganic carbon variables even when a rather simplified biogeochemical model is used (Doney *et al.* 2004; Sitch *et al.* 2015). A more complex alternative relies on the combination of an OBGCM with an atmospheric model to realize an Earth System model (Crueger *et al.* 2008; Vichi *et al.* 2011). Another source of uncertainty is represented by the parameterization of the air-sea CO_2 flux that is usually based on empirical estimates of the exchange rate at the interface (Wanninkhof 1992).

One approach at constraining the air-sea CO_2 fluxes consists in the realistic simulation of the space and time evolution of surface pCO_2, which is linked to the physical and biogeochemical dynamics of the two main carbonate system variables dissolved inorganic carbon (DIC) and total alkalinity (ALK). However, the large number of different biogeochemical models used in OBGCMs is an indication that there are few evidence-based constraints on biological processes, whose knowledge is derived heuristically from laboratory experiments and in situ measurements with necessarily limited spatial and temporal extents.

Only recently the scientific literature reported on the assimilation of these inorganic carbon variables or of the CO_2 partial pressure into an OBGCM (Ridgwell *et al.* 2007; Dowd *et al.* 2014; Valsala and Maksyutov 2010; While *et al.* 2012; Gregg *et al.* 2014). Ridgwell et al. (2007) used an ensemble Kalman filter method to assimilate alkalinity and phosphates into the global Grid ENabled Integrated Earth oceanic model, coupled to a model that resolves the biogeochemistry. Ensemble Kalman Filter was also used to estimate the parameters of the biological processes related to carbon cycling (Dowd *et al.* 2014). Valsala and Maksyutov (2010) modelled the ocean carbon cycle by coupling a biogeochemical model to an offline transport model for physical circulation, assimilating pCO_2 data with a variational method, but do not focus on the benefits of the assimilation of physical data. While et al. (2012) modified the FOAM data assimilation system to allow for the possibility of assimilating pCO_2 data, using the NEMO ocean model coupled to the HadOCC biogeochemical model.

At the best of our knowledge, the impact of physical data assimilation alone on the simulation of the carbonate system was not specifically addressed in previous works and only a limited literature deals with the effects on other biogeochemical variables. The first pioneering paper by Anderson et al. (2000) indicated the creation of spurious biogeochemical fluxes when physics and biology were not assimilated together and a joint assimilation process was suggested. Berline et al. (2007) reported some slight improvements of the assimilation of physics alone, mostly due to changes in the mixed layer depth in the North Atlantic. The impact on the ecosystem features was however deemed small and not necessarily positive. Following the previous work, Ourmieres et al. (2009) reported that the assimilation of physical data has a rather weak the impact on the ecosystem and, in some situations, it can even worsen the ecosystem response for areas where the prior nutrient distribution is significantly incorrect. They come to the conclusion that the combined assimilation of physical and nutrient data has a positive impact on the phytoplankton patterns, by claiming the urgent needed of more intensive in situ measurements of biogeochemical nutrients to overcome these issues. More recently, Raghukumar et al. (2015) used a physical-biogeochemical model of the California Current System with an incremental 4-dimensional variational method for physical data

assimilation. They found that the method improves correlation with observations, although artificially enhancing the phytoplankton standing stock that leads to a large bias particularly in regions of low mean concentration.

In this work, we investigate the benefits and drawbacks of using an assimilation system for physical-only observations (temperature and salinity) and the subsequent inclusion of inorganic carbon data (DIC and ALK) to simulate the evolution of the carbonate system and the related air-sea CO_2 fluxes in a forward OBGCM.

The approach based on the assimilation of physical quantities is worth exploring since an established and well-maintained monitoring network for the physical state of the global ocean exists (see http://www.argo.ucsd.edu). In view of a similar organized effort in collecting carbonate system observations through a global ocean network, we also aim at assessing the possible improvements emerging from the combination of data assimilation and inorganic carbon observations.

We used the Nucleus for European Modelling of the Ocean (NEMO, Madec 2008) general circulation model coupled online to the Biogeochemical Flux Model (BFM, Vichi *et al.* 2007a,b). The assimilation of the physical and biogeochemical components is performed with a three-dimensional variational ocean data assimilation system (Storto *et al.* 2011).

In this study, we run three different experiments under current climate conditions, which differ by the inclusion in the assimilation system of in-situ physical data, both physical and inorganic carbon data (DIC and ALK), compared against a control run. Beside the evaluation of the assimilated state variables, the overall assessment of the performed simulations focuses on the independent comparison with observations of the sea surface pCO_2 and with literature estimates of the air-sea CO_2 fluxes.

The manuscript is organized as follows. Section 2 describes the OBGCM, the data assimilation system, and observational data considered for the assimilation and the validation. In Section 3 we highlight the assessment of the assimilation for both the physical and inorganic carbon variables, and we present the results for the simulation of the pCO_2 and the air-sea CO_2 flux. In Section 4 we discuss the results obtained, drawing conclusions in Section 5.

2. Methods

2.1. Oceanic Biogeochemical General Circulation Model

The OBGCM used in the present work is composed by the NEMO general circulation model (Madec, 2008; see also http://www.nemo-ocean.eu, version 3.4), coupled with the Louvain-La-Neuve sea-ice model (Fichefet and Maqueda, 1997; version 2) and with the Biogeochemical Flux Model (Vichi *et al.* 2007a,b, Vichi and Masina, 2009; see http://bfm-community.eu, version 5).

148 The model is based on an ORCA grid with a horizontal grid resolution of 2° × 2°, except in the
149 20°N – 20°S belt where the meridional grid spacing reduces to 0.5°. The grid is irregular and
150 features three poles, two of which are located over the land regions in the northern hemisphere and
151 the third over Antarctica. The number of ocean vertical levels is 30, 20 of which are located in the
152 top 500 m.
153 The net freshwater flux is corrected by means of the relaxation towards the World Ocean Atlas
154 2009 (WOA09, http://www.nodc.noaa.gov/) monthly climatology of sea surface salinity, with a
155 relaxation timescale corresponding to 300 days for a 50 m deep mixed layer. An additional three-
156 dimensional relaxation is applied northward of 60°N and southward of 60°S in order to avoid high-
157 latitude model drifts. At each time step, the freshwater flux is adjusted according to the
158 climatological flux computed on the previous year. This adjustment directly modifies the Sea
159 Surface Salinity (SSS). At the same time, we do not enforce the relaxation in the Sea Surface
160 Temperature (SST).
161 The Biogeochemical Flux Model (BFM) describes the dynamics of major biogeochemical processes
162 occurring in global marine systems including the carbonate system. The model is based on a set of
163 differential equations describing the fluxes of matter and energy between inorganic pools and living
164 functional groups. The BFM describes through a continuum biomass representation the lower
165 trophic levels dynamics of the marine ecosystem. The model implements a set of biomass-based
166 differential equations that solve the fluxes of nutrients (carbon, nitrogen, phosphorus, silicate and
167 iron) among selected biological functional groups (namely, 1 bacterial, 3 phytoplanktonic and 3
168 zooplanktonic groups) representing the major components of the ocean ecosystem (Vichi *et al.*
169 2015).
170 It was here included the parameterization of calcite formation and dissolution proposed by Aumont
171 and Bopp (2006), with the reference phytoplankton content of particulate inorganic carbon (PIC) as
172 estimated by Gehlen et al. (2007). Calcite is produced by nanoflagellates and released as a
173 consequence of grazing by micro- and meso-zooplankton and loss processes involving particulate
174 matter originate by cells death. The sinking velocity of PIC is set to a constant value of 30 m/d and
175 changes in the calcite pool lead to a stoichiometric adjustment in DIC and ALK concentrations. A
176 scheme of the state variables and resolved physiological and ecological processes is available on the
177 model web page (http://bfm-community.eu), where it is also possible to download the code and
178 access the full documentation. Additional details for the parameterization of the advection and
179 diffusion schemes, the forcing, and the river runoff used in the experiments are provided in Tab. 1.
180
181
182

2.2. Ocean Data Assimilation system

The OBGCM is coupled with the global implementation of a three-dimensional variational data assimilation system, here OceanVar (Dobricic and Pinardi, 2008; Storto *et al.* 2011). The model assimilates data over the whole oceanic region with no depth exclusion over a fixed length time window. The data assimilation step consists in minimizing a cost function $J(\boldsymbol{x})$ with respect to the state vector $\boldsymbol{x}$, containing both physical and biogeochemical parameters (T, S, DIC, ALK) in the model three-dimensional grid, of the form (Courtier et al., 1994)

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^b)^T\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^b) + \frac{1}{2}[\boldsymbol{d}-H(\boldsymbol{x}-\boldsymbol{x}^b)]^T\boldsymbol{R}^{-1}[\boldsymbol{d}-H(\boldsymbol{x}-\boldsymbol{x}^b)], \quad (1)$$

where $\boldsymbol{x}^b$ is the background model state, $\boldsymbol{B}$ is the background-error covariance matrix, $\boldsymbol{d}$ is the vector of misfits, $\boldsymbol{H}$ is the observation operator that interpolates $\boldsymbol{x}$ over $\boldsymbol{d}$, and $\boldsymbol{R}$ is the observational error covariance matrix. In OceanVar, the background error covariances of the model state $\boldsymbol{B}$ are further split into a sequence of operators that account separately for the horizontal and vertical error covariances of the assimilated fields (Dobricic and Pinardi, 2008). The vector of misfits $\boldsymbol{d}$ is computed through the First Guess at Appropriate time (FGAT) method, namely observations are compared to the model equivalents closer in time to observations within 3-hourly time slots.

Due to the structure of the background error covariance matrix, vertical corrections are spread over the physical and biogeochemical variables by using Empirical Orthogonal Functions (EOFs). In order to derive the set of EOFs used in the experiments, we first run a non-assimilative experiment, from which we obtain an initial set of EOFs for both the physical and inorganic carbon variables. Then, we set to zero all cross-correlations between any physical and biogeochemical variable, and derive a new set of EOFs. These cross-correlations have been set to zero in order to ensure that the assimilation of biogeochemical quantities does not affect the physical reanalysis, as unrealistic correlations may arise when the number of biogeochemical observations is remarkably smaller than the physical ones. The EOFs thus computed retain all cross-correlations between temperature and salinity, as well as those between DIC and ALK. Since non-zero correlation between temperature and salinity exists, when only one of the two physical quantities is assimilated, vertical corrections apply to the other as well. Similarly, the assimilation of one specific biogeochemical variable affects the other assimilated biogeochemical property (DIC or ALK) through the specific cross-correlation term. For the assimilation, we use ten EOF modes for each vertical profile, which explained variance averaged over the global oceanic region is 98.9%.

In order to model horizontal correlations, a 4-iteration first-order recursive filter is used (Purser *et al.* 2003a,b), with a uniform horizontal correlation length-scale equal to 300 km for all assimilated variables. In the OceanVar system, cyclic conditions during the application of the recursive filter are approximated by imposing an extension of the domain with duplicated observations on the symmetric extension zones (Storto *et al.* 2011).

The assimilation system assumes the observational error covariance matrix to be diagonal, namely errors between observations are mutually uncorrelated, and the observation error variance is given by the sum of instrumental and representativeness errors variances.

Observational errors for the physical variables are derived from the profiles of instrumental errors provided by Ingleby and Huddleston (2007), which are subsequently multiplied by a coefficient that depends on the spatial variability at each point (Oke and Sakov 2008; Storto *et al.* 2014) and accounts for large representativeness errors in areas of strong variability. Since the corresponding information for the observational errors of inorganic carbon variables is not available within the GLODAP dataset, we have used the method proposed in Eq. (3) of Desroziers *et al.* (2005) to reconstruct the biogeochemical observation error. This method relates the error variance with the expectation value of the product between the observation minus background ($\boldsymbol{d}$) and the observation minus analysis, i.e. $\boldsymbol{y} - H(\boldsymbol{x}^a)$, with $\boldsymbol{x}^a$ representing the analysis.

The assimilation system performs several data quality controls on both physical and biogeochemical variables, among which a check against the climatology and one against background fields that rejects observations with a too large departure from the model fields. In detail, observations are rejected if the square of the errors $\boldsymbol{d}^2$ between the data and the model outcome is

$$\boldsymbol{d}^2 > \alpha_{\mathrm{INS}} \left(\sigma_b^2 + \sigma_o^2\right), \qquad (2)$$

where σ_b is a fixed parameter describing the error associated with the background and σ_0 is the error associated with the observation, and α_{INS} is a threshold factor. The value of α_{INS} was estimated to ensure that only a few data outliers are rejected and the use of observational information is maximized (see Storto et al. 2011). In particular, the assimilation system was repeatedly applied with different threshold values and α_{INS} was selected when both the magnitude of misfit range and rejection rates were in the +/- 2% range with respect to the initial standard statistics. The threshold factors α_{INS} are set as in Tab. 1.

An assimilation time window of 10 days was here adopted to balance between the frequency of available observations and the assumptions implied by the OceanVar scheme. In fact, a shorter time window would prevent the assimilation system from using a fairly homogenous observing network, while a longer one would lead to infrequent corrections with a detrimental impact on the skill scores.

2.3. Observations and statistical assessment

Global temperature and salinity data from the EN-ACT Quality Checked dataset (EN3, Ingleby and Huddleston 2007) were used for the assimilation. Each measurement has been quality controlled by using a set of objective tests, with data available from 1950 to present day. This collection comprises several independent data from different observing platforms, like e.g., moored and expendable bathythermographs, conductivity-temperature-depth profiles, moored buoys, and floats such as the instrumentation used in the ARGO project (see http://www.argo.ucsd.edu). More specifically, the moored buoys considered in EN3 are the PIRATA floats in the Atlantic basin, the TOGA-TAO floats in the Pacific basin, and the RAMA floats in the Indian Ocean (McPhaden *et al.* 1998).

Global biogeochemical data are obtained from the following datasets: GLODAP (Key *et al.* 2004), CARINA (Velo *et al.* 2009; Olsen 2009a,b), PACIFICA (Suzuki *et al.* 2013), the Bermuda Atlantic Time Series (BATS, Michaels and Knap 1996), and the Hawaii Ocean Time series (HOT, Lukas and Karl 1999). The available number of data considered in the assimilation system over the whole period of the simulations is shown in Fig. 1. For each basin, a different amount of data is available, clustered over specific years. For example, data in the Pacific Ocean were mostly collected during the period 1992-1994, while data in the Indian Ocean were mostly collected on 1995. Due to the shortage of available information, we have retained all data prior assimilation, which are subsequently filtered by the variational assimilation model under the condition described by Eq. (2).

To assess the overall efficiency of the OceanVar assimilation system, we computed the Root-Mean Square Error (RMSE) for temperature, salinity, DIC and ALK three-dimensional fields with respect to the datasets above over marine regions selected on the basis of the recent literature (see Schüster *et al.* 2013; Sarma *et al.* 2013; Ishii *et al.* 2014; Lenton *et al.* 2013; regions listed in Tab. 2). We only considered data collected in the entire water column, excluding the outliers larger than 3 standard deviations from the mean of the dataset. The latter condition selects data depending on the statistical properties of the data only, and it is independent from the filtering performed by the OceanVar scheme in Eq. (2). Although RMSE is a metric potentially affected by the sampling and the representativeness of the verifying observations, our choice stems from its widespread use as

intuitive and summarizing skill score and ease of comparison among the experiments and with other published works.

Given the multivariate nature of the experiments, it is useful to have a graphical representation of the RMSE metric for the different regions that enables for the comparison of changes due to the sequential assimilation of physical and biogeochemical data. This was done using the non-metric multidimensional scaling (MDS) that is an ordination method to refit the original data into a low dimensional space (here 2D). A symmetrical matrix for all pairwise distances among the original data is computed using a suitable distance metric, here assumed to be the Manhattan or cityblock distance (Cox and Cox 2000). An iterative procedure is then used to test the goodness of fit between the ordination-based distance matrix (computed for different refits of data) and the distance matrix of the original data. The Kruskal's Stress function is minimized through iteration and a value lower than 0.2 was here adopted to select the optimal refit of the original distances in the lower dimensional space.

In the case of physical data the validation step aims mostly to demonstrate the efficiency of the assimilation system in this specific simulation. However, it does have a merit in the assessment of the carbonate system variables because it allows evaluating the improvements due to the assimilation of physical variables (when DIC and ALK are not assimilated), as well as providing a reference for the combined assimilation.

The simulated surface pCO_2 fields were assessed against data from the Surface Ocean CO_2 Atlas (SOCAT2, Sabine *et al.* 2013; Bakker *et al.* 2013) a global pCO_2 dataset that reports measurements of pCO_2 at surface waters collected from 1968 to 2012.

The estimated air-sea flux of CO_2 comes from Takahashi *et al.* (2012) that provides data referred to the nominal year 2000 over a $4° \times 5°$ regular grid. Other global and regional estimates of CO_2 fluxes have been taken from Rödenbeck *et al.* (2014), Landschützer *et al.* (2014), Park *et al.* (2010), Peters *et al.* (2010), Peylin *et al.* (2013), Jacobson *et al.* (2007), Le Quéré *et al*. (2015) and Wanninkhof *et al.* (2013).

Both the SOCAT2 and Takahashi *et al.* (2012) datasets are not included in the assimilation system to enable an independent assessment of the pCO_2 fields and the air-sea CO_2 flux.

2.4. OBGCM setup and numerical simulations

We perform three different numerical simulations covering the period 1988-2010:

1. A control run (CTRL) without any data assimilation;
2. A physical reanalysis (TSRE) where we assimilate in-situ temperature and salinity data;

3. A reanalysis (here REAN) where we assimilate in-situ temperature, salinity, alkalinity and dissolved inorganic carbon.

In all experiments, we used the same parameterizations for the ocean and the biogeochemical models. The oceanic component has been spun up by repeating the year 1988 twenty-five times. Initial conditions for the temperature and salinity fields were obtained from the WOA09 dataset, whereas the zonal and the meridional components of the velocity fields start from rest. Initial conditions, external forcing, and parameterizations used are summarized in Tab. 1.

3. Results

3.1. Assessment of the assimilated variables

The RMSE errors computed against the assimilated datasets for the 3 simulations over a selected number of ocean regions and over the entire ocean depth (see Sec. 2.3) in the period 1988-2010 are presented in Tab. 2.

On average, the assimilation of physical data reduces the RMSE with temperature and salinity observations by about 20-30%, reducing the temperature error below 1°C in almost every region considered but the Indian Ocean. One major correction for temperature occurs in the Northern and Tropical Atlantic Ocean and for salinity in the South Pacific Ocean where in this case the RMSE is reduced to 40% with respect to the CTRL simulation.

The impact of the ocean physics improvements on DIC and alkalinity is however limited to certain regions, and not always the same ones where the physical model error is reduced the most. For instance, DIC benefits from the assimilation of temperature and salinity observations only in the Subpolar Atlantic region, and ALK only in the North and Tropical Atlantic regions. In all other areas, the simulations of DIC and ALK in TSRE show larger RMSE with observations with respect to the CTRL. The impact of the combined assimilation of physical and inorganic carbon variables is, on the other hand, positive in all the regions, reducing the RMSE with observations with respect to both the CTRL and TSRE experiments everywhere with peaks of about 40% reduction.

The values of the RMSE indicators for temperature, salinity, DIC and ALK in every region and every experiment are combined to give a measure of the distance from a perfect reference simulation without errors (all RMSE equal to 0) using the MDS method (see Sec. 2.3). This operation returns the plot shown in Fig. 2, where the axes are arbitrary and distances are representative of the quality of the simulation in that region.

It is evident that the combined assimilation in REAN reduces the multivariate RMSEs shifting almost all region points toward the reference and closer to each other. The figure shows for instance that Northern subtropical Atlantic (indicated with the code ATL-NST) is first shifted closer to the

353 reference because of the assimilation of physical data and further more by the carbonate variable
354 data. The use of inorganic carbon data also improves some of the regions that showed a worsening
355 with the physical assimilation. For instance, the discrepancies in the reconstruction of the inorganic
356 carbon variables obtained in TSRE for the Indian and Pacific oceans (see for instance the tropical
357 Pacific region, PAC-T) are corrected with the assimilation of the inorganic carbon variables in
358 REAN.
359 In Fig. 3 and 4 we show the time series of SST, SSS, surface DIC, and surface ALK for the three
360 experiments, compared with the sustained observations at the two stations at BATS and HOT. For
361 both stations, observed data present a positive trend in surface DIC because of the atmospheric
362 increase in the CO_2 concentration, which was generally reconstructed in all experiments.

363 BATS (Fig. 3) belong to the Subtropical North Atlantic region that shows a marked improvement
364 by physical data assimilation only. In fact, CTRL has lower salinity and warmer temperature than
365 observed; both facts contribute to high evaporation rates (Fig. 5a) and to a more stratified water
366 column that determine an unrealistic trend in the surface concentrations, especially for alkalinity.
367 The assimilation of temperature and salinity in TSRE determines a strong reduction of the
368 evaporation process, which in turn leads to a strong reduction of the positive trend in ALK and an
369 improved seasonal variability. The further assimilation of inorganic carbon data drives both DIC
370 and ALK toward a closer agreement with the observations, even if the latter is better constrained
371 only after the year 1992 when more data are available.

372 The CTRL simulation at HOT (Fig. 4) shows a remarkable overestimation of surface salinity, while
373 the surface temperature is satisfactorily reproducing the observations. Conversely to what found at
374 BATS, the assimilation of physical data is heavily correcting the salinity bias, namely RMSE
375 reduces from 0.36 to 0.19, and leads to a significant increase in the net freshwater flux (Fig. 5b).
376 Such an enhanced evaporation determines both the bias and positive trend obtained in TSRE for the
377 simulated DIC and ALK. However, these unrealistic changes in the inorganic carbon variables are
378 successfully corrected in the REAN simulation.

379 The results of the REAN simulation at BATS and HOT show the advantage of having a relatively
380 high abundance of DIC and ALK data: the assimilation system corrects the discrepancy moving the
381 model toward the observations. This is a demonstration of the efficiency of the adopted assimilation
382 system that is expected to improve the carbonate system equilibrium also in oceanic areas where
383 data are less abundant compared to these stations. This will be analysed in the next section
384 comparing the results over the various regions against the independent surface pCO_2 data.

3.2. Regional analysis of surface pCO_2 field

In Fig. 6 we assess the simulated surface pCO_2 against the Surface Ocean CO_2 Atlas (see Sect. 2.3), by computing the RMSE between the monthly pCO_2 data of the dataset and the corresponding values from each simulation for the period 1988-2010. Since we do not assimilate pCO_2 data into the OBGCM, this assessment represents a totally independent validation of the assimilation system. In the whole Atlantic Ocean, the RMSE against data decreases with the subsequent assimilation of physical (TSRE) and both physical and inorganic carbon (REAN) data. In general, the assimilation of physical data worsens surface pCO_2 in the Arctic, Pacific and Indian oceans, with respect to those computed for CTRL. The assimilation of DIC and ALK counters this effect and improves pCO_2 in the Tropical Indian ocean, also lowering the RMSE obtained in the South Pacific and South Indian oceans, even if with values still larger than those obtained in the CTRL simulation. However, despite the assimilation of DIC and ALK data that has led to an overall improvement of the main carbonate system variables (see Tab. 2 and Fig. 2), the RMSE of surface pCO_2 in the Arctic and North Pacific is larger in REAN than in the control run without any data assimilation. Finally, the Southern Ocean is the unique region where the physical-only reanalysis slightly improves the simulation of pCO_2, while the additional assimilation of inorganic carbon data leads to an RMSE that is larger than CTRL. All these cases will be compared and further discussed in Sec. 4.

The North Pacific Ocean is the most notable region where physical data assimilation increases the error of surface pCO_2 and, whereas improvements in DIC and ALK are obtained with their direct assimilation, its error becomes even larger. By looking at the differences between the three experiments (Fig. 7), the physical data assimilation leads to colder but less saline surface waters similarly to what shown in the results at HOT (Sec. 3.1 and Fig. 4). Despite the data-driven cooling of surface waters, the salinity corrections in REAN have a major effect on the evaporation rate in the North Pacific Ocean that has a positive bias in comparison to CTRL (Fig. 7c). This modified freshwater flux is physically consistent with the correction, but it also impacts the concentrations of DIC and ALK whose surface concentrations steadily increase in the REAN simulation (Fig. 7d,e). The assimilation of inorganic carbon data is not sufficient to correct this imbalance in the North Pacific as a whole, since observations in this area are rather scarce and concentrated over specific years (supplementary Fig. S1). The available data allow to partly reduce the overall RMSE (see Tab. 2), but after year 2000, the REAN experiment is very little constrained by observations and the

lack of combined assimilation of physical and carbon data thus leads to a spurious departure in the REAN pCO_2 (Fig. 7f).

3.3. Assessment of air-sea CO_2 fluxes

The statistical indicators for the globally integrated air-sea CO_2 flux for the three simulations are summarized in Tab. 3, where we have computed the average, decadal trend, InterAnnual Variability (IAV) and Seasonal Variability (SV) over the period 1992-2010 for the three simulations. The air-sea CO_2 fluxes indicate that the global ocean is always acting as a sink (negative values) and the sequential assimilation of physical and inorganic carbon data leads to a significant reduction, namely from -4.7 PgC/yr in the CTRL simulation to -2.4 PgC/yr in the REAN one. Similarly, the interannual and seasonal metrics are characterized by decreasing values, but IAV shows a major change in the TSRE run and SV mainly reduces with the additional assimilation of DIC and ALK. The decadal trends obtained with a least-squared fit for each simulations have quite different values, which are not distinguishable from zero since the uncertainty on the trend is on the same order of the trend itself.

The impacts of assimilation on the regional distribution of mean annual fluxes are illustrated in Figure 8, which shows the air-sea CO_2 flux climatological field obtained from the dataset by Takahashi *et al.* (2012) and compare it with the three experiments for the period 1992-2010. The Atlantic Ocean stands out as the ocean with the largest differences from the climatological estimates; the large sink in the Northern Atlantic region is substantially reduced in TSREAN and further improved in REAN. A similar consideration can be done for the Northern Indian Ocean, where the large sink bias of the CTRL is appropriately reduced toward a source by the assimilation of physics. The assimilation of carbon data is then enhancing this source, possibly exceeding the estimates because of the presence of coastal data with high alkalinity. The inorganic carbon source in the tropical Pacific is evidently increased as estimated in the Takahashi *et al.* (2012) thanks to data assimilation, although the region of source is still too confined within the tropics (Vichi *et al.* 2011). The tropical Atlantic is also not well represented in both CTRL and TSRE, while the additional assimilation of inorganic carbon data determines a reduction of the CO_2 sinking flux.

Estimates of the climatological air-sea CO_2 flux integrated over the selected marine regions for the three experiments are presented in Tab. 4. Overall, the assimilation system leads to a reduction of the CO_2 fluxes in the regions located at mid latitudes, in the Tropical Indian and Pacific oceans, and to a lower extent in the Tropical Atlantic. In particular, the major sinks in the REAN simulation occur in ATL-NST, PAC-N, and PAC-S regions. Conversely, the Arctic and Subpolar Atlantic

454 regions have nearly unchanged values within the different simulations. The only positive value of
455 the CO_2 flux was obtained for the Southern Ocean region in the REAN simulation.
456
457
458
459 **4. Discussion**
460
461 The set of experiments presented in this work combines the assimilation of physical data and the
462 assimilation of inorganic carbon data with the final aim to better constrain the dynamics of the
463 carbonate system and the related air-sea CO_2 flux.
464 The key question addressed in the previous sections is whether the assimilation of physical and
465 carbon data is likely to improve the surface carbon fluxes by improving the representation of pCO_2.
466 A summary of all the findings presented in Sec. 3 is given in Fig. 9, where the change in pCO_2
467 RMSE for each region and experiment is plotted against the corresponding change in the
468 multivariate combined RMSE (shown Fig. 2). The latter is computed as the change of the Euclidean
469 distance between the region points and the reference in Fig. 2 for each assimilation run against the
470 CTRL one. Figure 9 clearly shows that the assimilation of physics is a necessary first step to
471 improve pCO_2 simulation in the Northern and Tropical Atlantic (bottom-left quadrant), whereas an
472 opposite effect was obtained, i.e., for the North Pacific (top-right quadrant). The bottom-line is that
473 the assimilation of temperature and salinity leads to significant changes in surface physical
474 processes, which have a twofold impact on the inorganic carbon variables. Beside the improved
475 solution of biogeochemical processes under more realistic physical conditions, the modification of
476 evaporation rates will directly impact other conservative oceanic properties, like e.g. alkalinity.
477 Such a feedback of the physical assimilation was clearly identified at BATS and HOT stations (see
478 Sec. 3.1), as the changes in surface concentrations of DIC and ALK were driven by those of the
479 evaporation rates (Fig. 5). This condition occurs very likely for the whole North Pacific and North
480 Atlantic regions, thus explaining the opposite effect produced by the physical assimilation and
481 indirectly pointing to an imbalance due to whether the atmospheric forcing fields or the bulk
482 formulations. In all other regions, an improved representation of physical variables - as
483 demonstrated by the reduced RMSE for temperature and salinity in Tab. 2 – does not lead to
484 significantly better inorganic carbon in the ocean, although there may still be an improvement of
485 pCO_2 due to better constraints on temperature control.
486 On the other hand, once DIC and ALK are assimilated, pCO_2 improves almost everywhere, which
487 is expected to better constrain the air-sea CO_2 fluxes. It is not a substantial improvement but
488 certainly in the desired direction. An average error reduction of 3-5 μmol is comparable to the error

attributed to the parameterization of gas exchange piston velocity (Takahashi et al., 2009), which implies that data assimilation may help to reduce the overall uncertainty.

How is it possible to reconcile the regions where pCO_2 worsens due to the assimilation of inorganic carbon data, even if the improvement against those variables is large (Fig. 9 bottom-right quadrant)? The typical example is the Arctic, where the assimilation of a limited number of observations markedly reduces the RMSE for DIC and ALK (see Tab. 2). However, these data are just enough to reduce the model bias but not sufficient to capture the pCO_2 variability found in the SOCAT data and the RMSE increases (Fig. 6). A similar consideration can be done for the Southern Ocean. Both the Arctic and Southern Ocean have been sampled for pCO_2 only for one month out of the year (Sabine *et al.* 2013, most likely during summer) and the eddy-driven spatial variability, which is estimated to account for an error of about 3 μmol (Lenton et al., 2006), is likely to become the dominant factor. Given the coarse resolution of this model, the corrections applied to the simulated DIC and ALK are actually worsening the performance of REAN against TSRE. The assimilations of the more numerous physical data in the Southern Ocean lead to a representation of the pCO_2 field that is better than assimilating sparse observations of inorganic carbon (see Fig. S2 for the map of the assimilated data per model grid point).

The case of the North Pacific has been explored in detail in Sec. 3.3. The results of this work indicate that both physical and carbon variables should be assimilated to reach an improved representation of carbonate system variables. Supplemental figures S1 and S2 show that the North Pacific carbon data are unevenly distributed in time and have lower data density with respect to the northern Atlantic Ocean (Landschützer et al., 2013). This implies that there must be more adequate temporal and spatial collection of carbon data to improve pCO_2 in the North Pacific as it occurs for the Atlantic.

The final question is whether data assimilation actually improves the simulation of the carbon flux between ocean and atmosphere. This issue is necessarily related to the quality of the specific model being analysed, and in this particular case the model presented a substantial overestimation of the annual mean ocean uptake that can be reduced by means of the assimilation of physical and inorganic carbon data (Sec. 3.3). As presented in the introduction, global carbon models have been demonstrated to agree on the annual global means and to some extent on some regional means. Table 3 compare the global air-sea CO_2 flux simulated in the three experiments with the results obtained in previous literature works using different forward ocean models and atmospheric inversion models.

In Le Quéré *et al*. (2015), the ocean flux is computed over the period 1959-2013 by using an ensemble containing seven OBGCMs. For each model, the reconstructed air-sea CO_2 flux has been

normalized to observations by dividing it by the observed average over the period 1990-1999 (Keeling *et al*. 2011), then multiplying the result by 2.2 Pg C/yr, obtaining a multi-model mean of -1.9 Pg C/yr with standard deviation 0.5 Pg C/yr.

Wanninkhof *et al.* (2013) also obtained −1.9 ± 0.3 PgC/yr with a trend −0.14 Pg C/yr/decade, by using an ensemble mean consisting of six OBGCMs, and −2.1 ± 0.3 PgC/yr with trend −0.13 PgC/yr/decade, by using eleven Ocean Inversion Models, for the period 1990-2009. As the trend for the considered period is mostly due to human activities, our findings are closely comparable to those of Wanninkhof *et al.* (2013). The air-sea flux simulated by CTRL in the present OBGCM is clearly higher if compared to the results of the other global models. The assimilation of the inorganic carbon variables allows reducing the bias.

In general, OBGCMs simulate a higher air-sea CO_2 flux with respect to the available ocean inversion models (Rödenbeck *et al.* 2014, Landschützer *et al.* 2014, Park *et al.* 2010, Peters *et al.* 2010, Peylin *et al.* 2013), although the values obtained by the inversions in Jacobson *et al.* (2007) and Peters et al. (2010) lie within the inter-annual variability (IAV) obtained in our simulations. Also the seasonal variability decreases with data assimilation towards values that are close to what obtained in other works. The IAV of the air-sea CO_2 flux obtained in the experiments varies from 0.22 to 0.40, and the estimates over the period 1990-2009 in the literature give a similar range. The errors associated to this quantity are however large, although all of the results from our experiments fall within the range given in the literature.

These comparisons indicate the positive role played by data assimilation, although without an independent measure of the carbon fluxes in the various regions it is not possible to assess the overall quality. The comparison of model results with some recent assessment of the regional air-sea CO_2 flux presented in Tab. 4 indicates an improvement in some key regions. The main differences with the results reported in the literature are found in the tropical regions, which however have been demonstrated to improve against the independent pCO_2 SOCAT data (Fig. 6). These discrepancies are probably linked more to methodological differences and to the scarcity of data rather than to substantial problems.

5. Conclusions

This work has shown that the data-driven correction of the factors regulating the concentration of carbonate system variables do not guarantee that pCO_2 is closer to the observations in a global ocean carbon model having the spatial resolution used in the last round of CMIP simulations. However, some important findings have emerged. The assimilation of physical data only has shown

557 to improve pCO_2 in the North Atlantic Ocean and in the Southern Ocean, the latter being a region
558 where the extensive collection of carbon data is much more difficult. The correction of physical
559 model errors has a direct effect on evaporation that helps to constrain alkalinity biases, although this
560 does not occur in all the regions and especially in the North Pacific. The concurrent assimilation of
561 dissolved inorganic carbon and alkalinity may help to reduce the errors in some regions such as the
562 tropics, also leading to improved fluxes. In general, errors in pCO_2 are reduced of a factor
563 corresponding to those introduced by to the air-sea flux formulations. However, the spatial and
564 temporal distribution of the available data appears to be an important constraint to the effective
565 improvements.
566 Since a global-scale network for collecting inorganic carbon data is still under development, the
567 current monitoring network for the global physical ocean is likely to be the most readily available
568 resource to increase the confidence on air-sea carbon fluxes especially in remote regions like the
569 Southern Ocean.
570

Figures

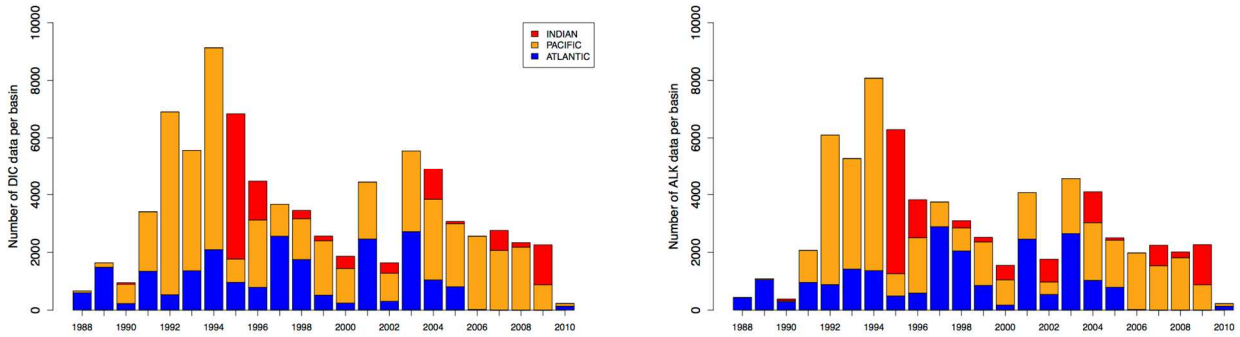

Figure 1. Number of DIC (left) and ALK (right) data used in the assimilation for the Atlantic (blue), Pacific (Orange) and Indian (red) oceans, for each year of the simulation. A description of this composite inorganic carbon dataset is given in Sec. 2.3.

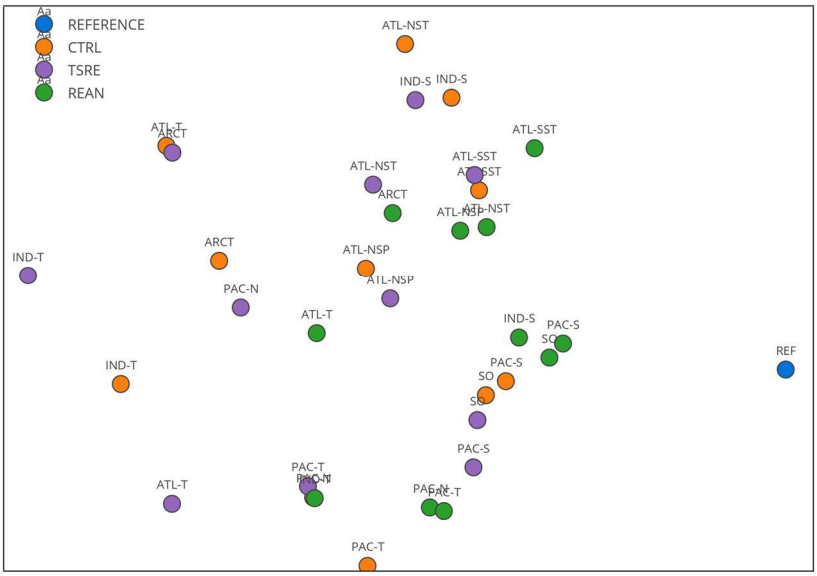

Figure 2. Two-dimensional ordination for the multivariate RMSE indicators (temperature, salinity, DIC and ALK), as obtained with the MDS method (see Sec. 2.3), in CTRL (orange), TSRE (purple), and REAN (green) experiments over the different marine regions listed in Tab. 2. The blue dot is a reference perfect experiment when all RMSE values are zero.

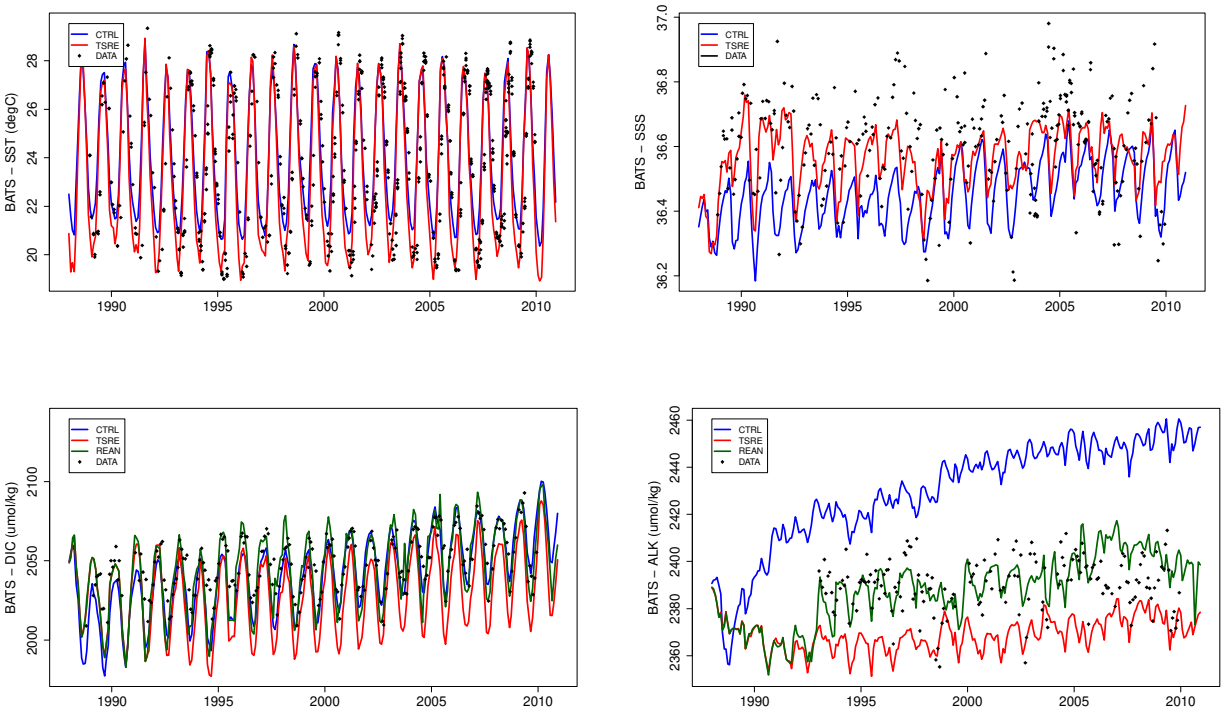

Figure 3. Time series of ocean variables at surface at BATS, obtained from each experiment considered: CTRL (blue), TSRE (red), REAN (dark green). The black dots show the data collected over the first 10 m depth at each station.

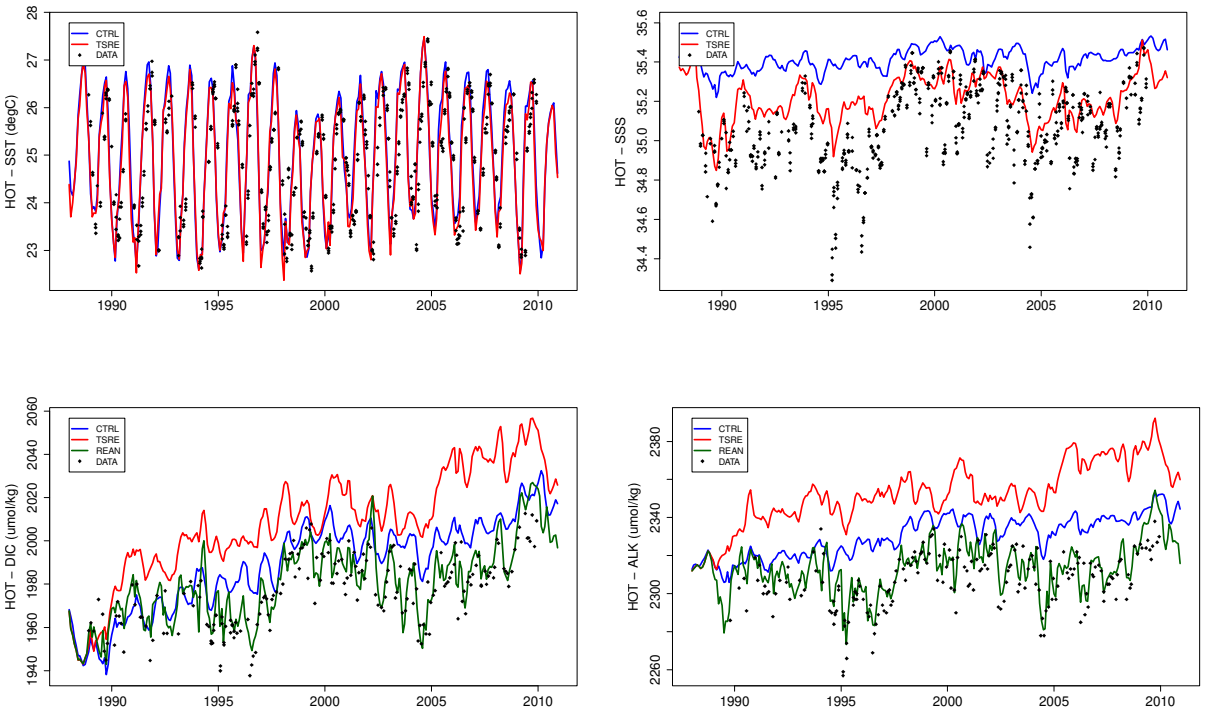

Figure 4. Time series of ocean variables at surface at HOT, obtained from each experiment considered: CTRL (blue), TSRE (red), REAN (dark green). The black dots show the data collected over the first 10 m depth at each station.

578

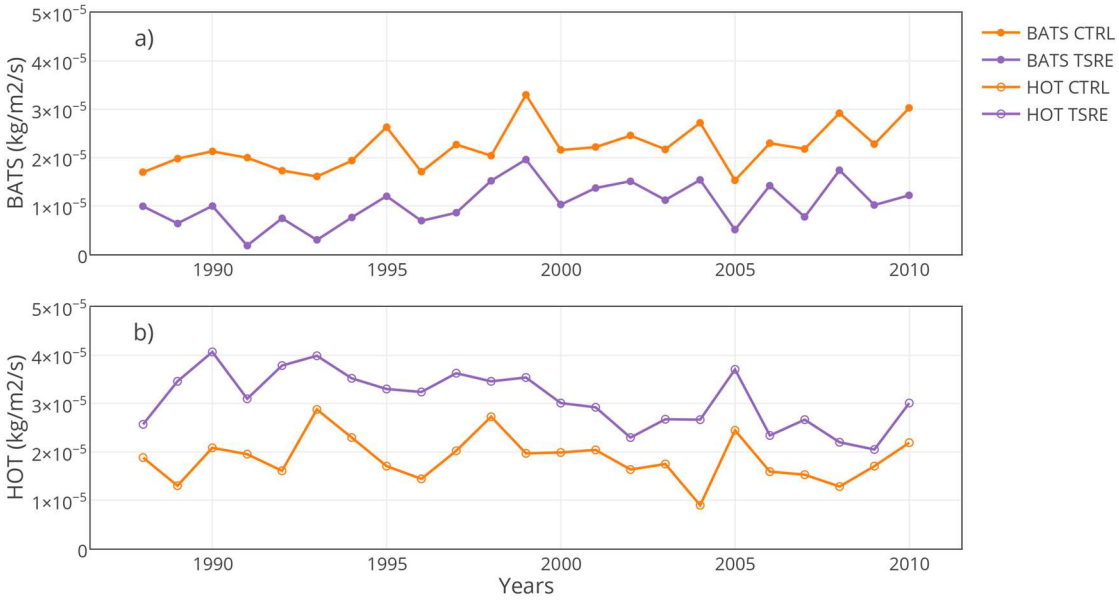

Figure 5. Time series of mean annual net freshwater flux at (a) BATS and (b) HOT for the CTRL and TSRE simulations.

579
580

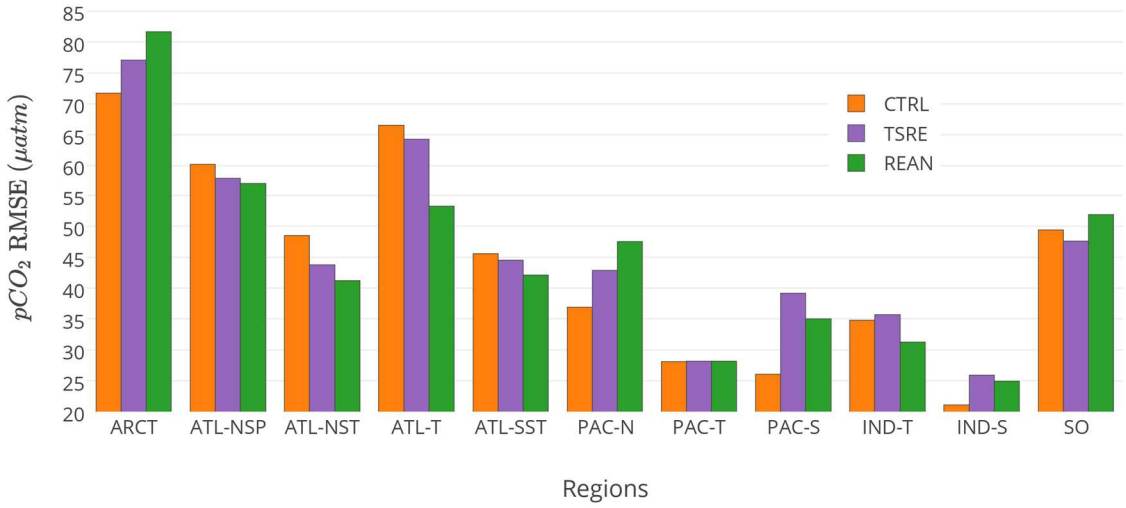

Figure 6. Root-Mean Square Error (RMSE) of surface pCO_2 (μatm) against the SOCAT2 dataset in the selected marine regions (see caption of Tab. 2 for a description).

581
582
583

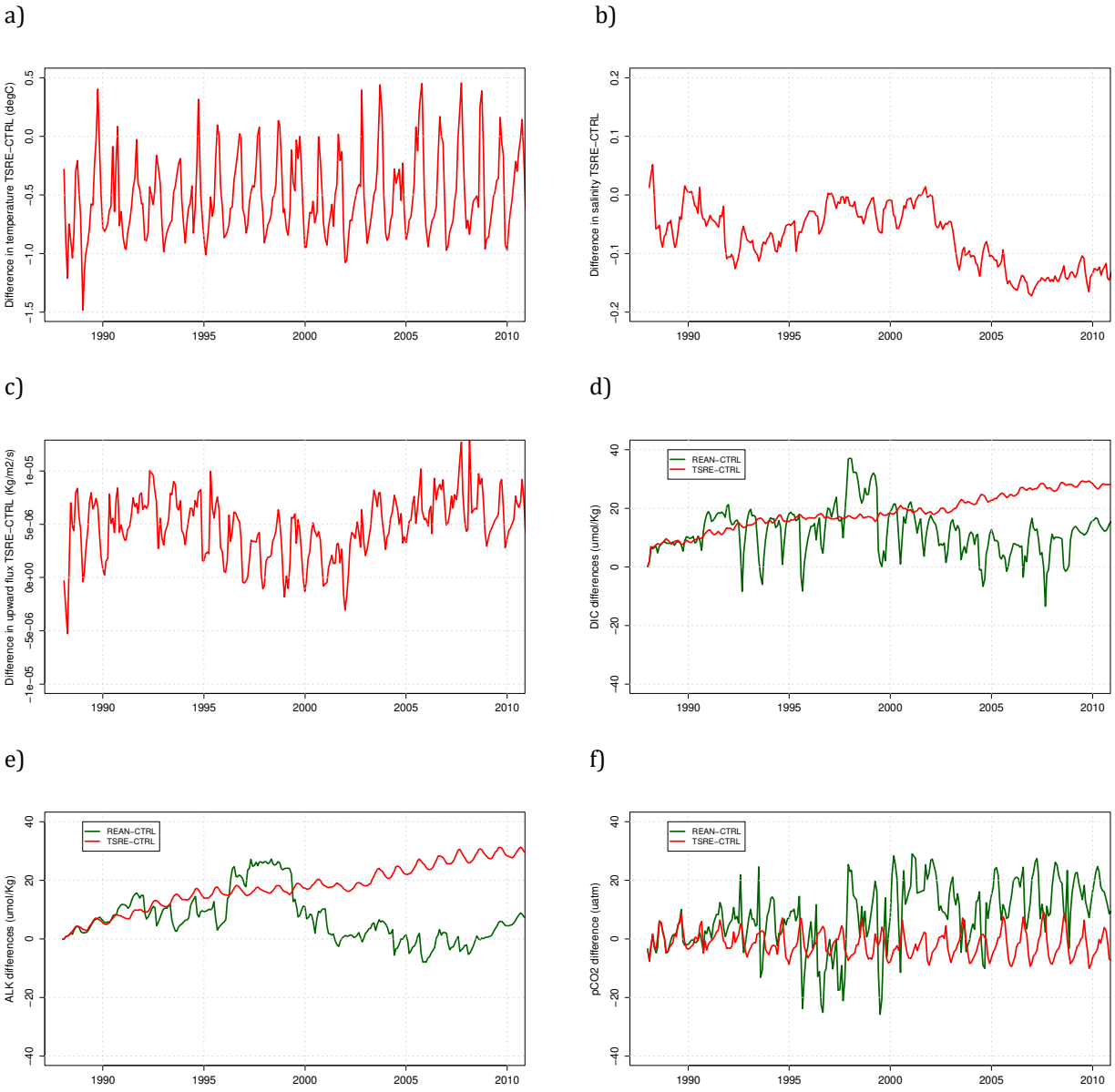

Figure 7. Difference in the mean SST, SSS, evaporation, surface DIC, surface ALK, and surface pCO_2, averaged over the North Pacific region. The difference is taken between TSRE and CTRL for temperature, salinity, and evaporation flux, and also between REAN and CTRL for the inorganic carbon variables.

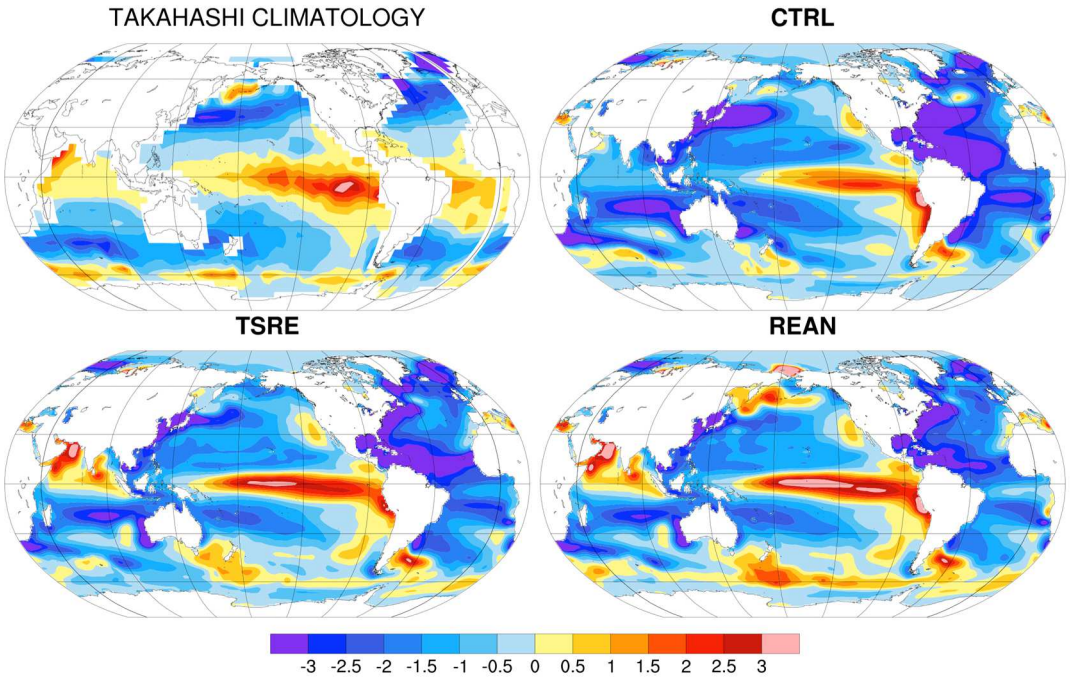

Figure 8. Map of the climatological mean air-sea CO_2 flux (in mol C/m^2/year) over the period 1992-2010, for the dataset from Takahashi and the three experiments.

587
588

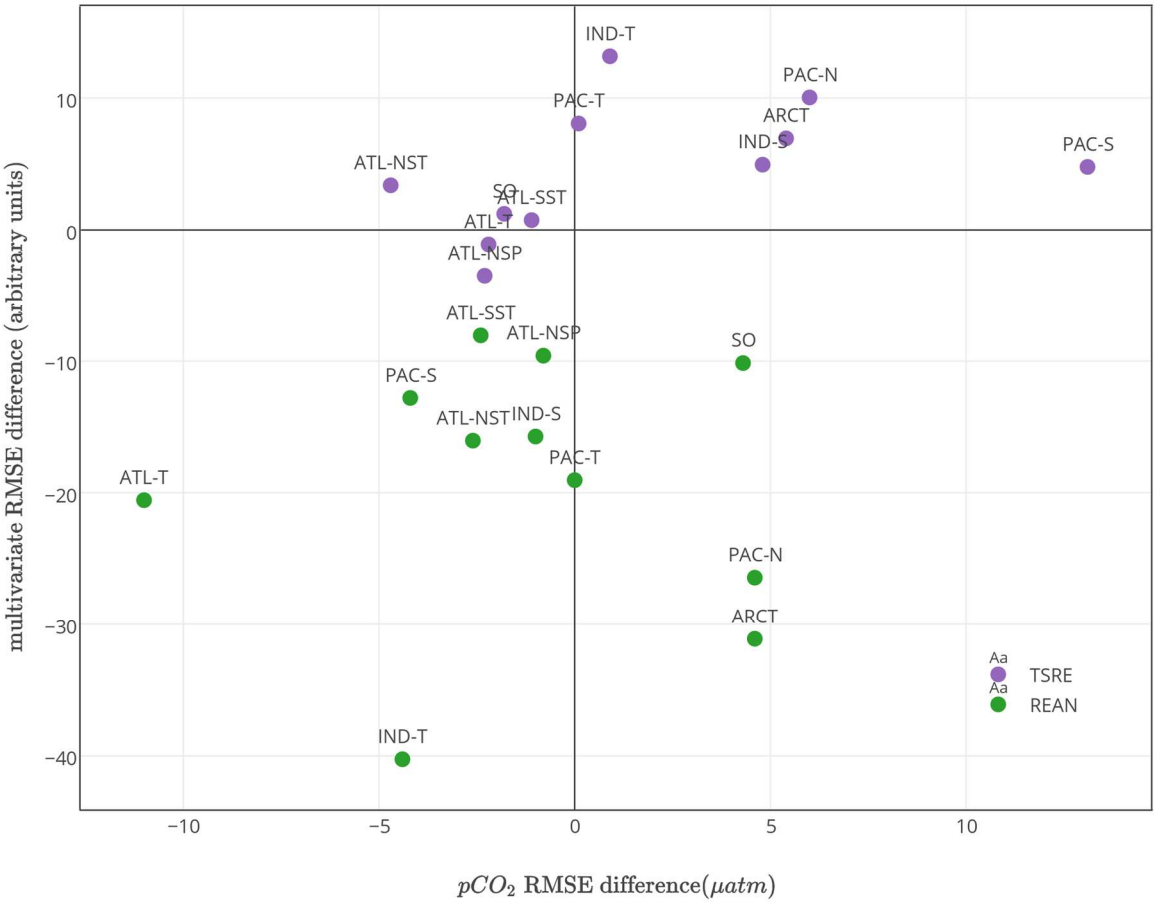

Figure 9. Comparison of improvements in pCO_2 RMSE and errors in physical and inorganic carbon variables for the two assimilation experiments. The multivariate RMSE difference is computed as the change of the Euclidean distance in Fig. 2 between a region point and the reference for each assimilation experiment against the CTRL one. Regions in the lower left quadrant are the ones where improvements in physics and/or carbonate variables lead to concurrent improvements in the pCO_2 RMSE.

Tables

Table 1. Description of the initial conditions, external forcing, and parameterizations used in the present OBGCM and shared by all numerical simulations.

Model	NEMO	BFM
Atmospheric forcing	CORE bulk formulae (Large and Yeager 2008) ERA-INTERIM atmospheric data (Dee *et al.* 2011)	Atmospheric CO_2 concentration of historical and RCP8.5 scenarios (Moss *et al.* 2010) Climatological iron deposition (Moore *et al.* 2002) based on Tegen and Fung (1994, 1995).
River forcing	Runoff (Dai and Trenberth 2002)	Inorganic nutrients (Cotrim da Cunha *et al.* 2007)
Experiment IC	25years spin-up from WOA09	DIC, ALK: GLODAP (Key *et al.* 2004) Oxygen, inorganic nutrients: WOA09 Iron: Vichi et al. (2007b)
Advection	TVD scheme (Harten 1997)	MUSCL scheme (Van Leer 1979)
Lateral diffusion	Laplacian operator	Laplacian operator
Vertical diffusion	Turbulent Kinetic Energy (Blanke and Delecluse 1993)	Same as T and S
Assimilation system	3DVar	3DVar
α_{INS},	17	17
EOFs	10	10
Data output	Ten days	Ten days

Table 2. Root-Mean Square Error (RMSE) of temperature (in °C), salinity, DIC (in µmol/kg), and ALK (in µmol/kg), computed against the datasets presented in Sect. 2.3 in the period 1988-2010. Region codes: ARCT: Arctic Sea (>80N); ATL-NSP: Subpolar North Atlantic (50÷80N); ATL-NST: Subtropical North Atlantic (14÷50N); ATL-T: Tropical Atlantic (15S÷14N); ATL-SST: Subtropical South Atlantic (40÷15S); PAC-N: North Pacific (18÷66N); PAC-T: Tropical Pacific (18S÷18N); PAC-S: South Pacific (44.5÷18S); IND-T: Tropical Indian (18S÷30N); IND-S: South Indian (18S÷44S); SO: Southern Ocean (44÷75S). The last column reports the total number of DIC and ALK data used in REAN (Sec. 2.3).

Code	**Temperature** (degC)		**Salinity**		**DIC** (µmol/kg)			**ALK** (µmol/kg)			**# Data**
	CTRL	*TSRE*	*CTRL*	*TSRE*	*CTRL*	*TSRE*	*REAN*	*CTRL*	*TSRE*	*REAN*	*DIC+ALK*
ARCT	0.89	0.79	0.55	0.33	47.6	49.9	30.1	30.8	35.5	24.4	31
ATL-NSP	1.23	0.92	0.30	0.21	33.8	32.0	24.6	24.3	22.8	19.9	8557
ATL-NST	1.22	0.95	0.30	0.26	26.2	31.4	22.4	27.8	26.0	18.2	20364
ATL-T	1.34	0.95	0.28	0.20	50.6	58.2	40.4	35.6	27.2	24.3	2579
ATL-SST	1.17	1.00	0.21	0.18	22.3	22.3	16.8	19.5	20.5	17.3	1785
PAC-N	1.37	1.02	0.21	0.16	43.7	47.0	33.0	20.9	28.4	15.2	34507
PAC-T	1.11	0.92	0.19	0.15	39.8	44.3	32.1	17.5	21.1	14.3	22285
PAC-S	1.18	0.94	0.22	0.13	24.2	28.7	18.9	13.2	13.7	11.0	4695
IND-T	1.36	1.23	0.26	0.18	59.8	66.8	43.6	32.4	38.6	20.8	1325
IND-S	1.43	1.12	0.23	0.16	22.8	26.1	22.4	23.5	25.8	13.6	2576
SO	1.16	0.82	0.23	0.16	26	27.3	20.9	14.3	14.4	11	9147

611 **Table 3.** Statistical properties (average, trend, inter-annual and seasonal variability, IAV and SV, respectively) of the
612 global average value of the air-sea CO_2 flux over the period 1992-2010. Also shown are the comparisons with results
613 from literature.
614

Air-sea CO_2 flux	**Average** (PgC/yr)	**Trend** (*0.01PgC/yr/ decade)	**IAV** (PgC/yr)	**SV** (Pg C/yr)	**Method**	**Notes**
CTRL	-4.7	0.9±1	0.40	0.97	OBGCM	1992-2010
TSRE	-3.3	0.1±0.9	0.28	0.80	OBGCM	1992-2010 T+S Assimilation
REAN	-2.4	0.7±0.9	0.22	0.69	OBGCM	1992-2010 T+S+DIC+ALK Assimilation
Wanninkhof *et al.* (2013)	-1.9±0.3	-0.14	0.16	0.38	Ensemble of 6 OBGCM	1990-2009 Anthropogenic,
Wanninkhof *et al.* (2013)	-2.1±0.3	-0.13	0.40	0.41	Ensemble of 11 OIM	1990-2009 Anthropogenic
Wanninkhof *et al.* (2013)	-2.0±0.6	-	-	-	Tier 1 (Canadell *et al.* 2011)	2000 Anthropogenic
Zeng *et al.* (2014)	-1.9 ÷ -2.3	-	-	-	NNM	1990-2011 Anthropogenic
Le Quéré *et al.* (2015)	-1.9 ± 0.5	-	-	-	Ensemble of 8 OBGCM	1959-2015
Rödenbeck *et al.* (2014)	-1.45	-0.64	0.29	-	OC V1.2	2001-2011
Landschützer *et al.* (2014)	-1.70	-1.13	0.08	-	CDIAC Global Carbon Project	2001-2010
Park *et al.* (2010)	-1.28	-0.23	0.09	-	Diagnostic model with empirical relationships	2001-2011
Jacobson *et al.* (2007)	-2.62	-0.62	0.03		CTE2014	2001-2011
CTE2014 (Peters *et al.* 2010)	-2.27 ± 0.77	-0.69			CTE2014	2001-2013

615
616

Table 4. Annual air-sea CO_2 flux (in Pg C/yr) averaged over the selected marine regions (see caption of Tab. 2 for a description) for the three experiments in the period 1992-2010, compared with the recent literature findings from Schüster *et al.* 2013 (a), Ishii *et al.* 2014 (b), Sarma *et al.* 2013 (c), and Lenton *et al.* 2013 (d).

Marine Region	CTRL	TSRE	REAN	Literature
ARCT	-0.01	-0.01	-0.01	-0.12±0.06[a]
ATL-NSP	-0.09	-0.10	-0.09	-0.21±0.06[a]
ATL-NST	-0.77	-0.61	-0.54	-0.26±0.06[a]
ATL-T	-0.58	-0.50	-0.35	0.12±0.04[a]
ATL-SST	-0.38	-0.32	-0.28	-0.14±0.04[a]
PAC-N	-0.47	-0.42	-0.31	-0.47±0.13[b]
PAC-T	-0.47	-0.16	-0.07	0.44±0.14[b]
PAC-S	-0.42	-0.28	-0.32	-0.37±0.08[b]
IND-T	-0.50	-0.15	-0.10	0.08±0.04[c]
IND-S	-0.50	-0.41	-0.41	-0.43±0.07[c]
SO	-0.26	-0.20	0.04	-0.42±0.07[d]

Acknowledgements

The authors wish to thank the Centro Euro-Mediterraneo sui Cambiamenti Climatici for its financial and scientific support of the activities presented in this work. The implementation and the following improvements of the global ocean assimilation system were carried out in the framework of the GEOCARBON and MYOCEAN projects. The research leading to these results has received funding from the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea under the GEMINA project. M.V. has been partly funded by the SANAP project TRAIN-SOPP (UID93089). The authors acknowledge the public availability of the BFM system (http://bfm-community.eu). The authors want to thank Dr. Simon Good (U.K. Met Office) for the support in the use of the EN3 dataset, and the TAO Project Office of NOAA/PMEL for the availability of the TAO/RAMA/PIRATA dataset. L.V. would like to thank I. van der Laan-Luijkx, P. Landschützer, and C. Rödenbeck for providing part of the useful data presented in Tab. 3.

References

L. A. Anderson, A. R. Robinson, C. J. Lozano, C. J. Physical and biological modeling in the Gulf Stream region:: I. Data assimilation methodology. Deep Sea Research Part I: Oceanographic Research Papers, 47(10), 1787-1827, 2000.

O. Aumont and L. Bopp. Globalizing results from ocean *in-situ* iron fertilization studies. *Global Biogeochemical Cycles*, 20, 2006.

D. C. E. Bakker, B. Pfeil, K. Smith, S. Hankin, A. Olsen, S. R. Alin, C. Cosca, S. Harasawa, A. Kozyr, Y. Nojiri, K. M. O'Brien, U. Schüster, M. Telszewski, B. Tilbrook, C. Wada, J. Akl, L. Barbero, N. Bates, J. Boutin, W.-J. Cai, R. D. Castle, F. P. Chavez, L. Chen, M. Chierici, K. Currie, H. J. W. de Baar, W. Evans, R. A. Feely, A. Fransson, Z. Gao, B. Hales, N. Hardman-Mountford, M. Hoppema, W.-J. Huang, C. W. Hunt, B. Huss, T. Ichikawa, T. Johannessen, E. M. Jones, S. D. Jones, S. Jutterström, V. Kitidis, A. Körtzinger, P. Landscützer, S. K. Lauvset, N. Lefèvre, A. B. Manke, J. T. Mathis, L. Merlivat, N. Metzl, A. Murata, T. Newberger, T. Ono, G.-H. Park, K. Paterson, D. Pierrot, A. F. Rıos, C. L. Sabine, S. Saito, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, R. Sieger, I. Skjelvan, T. Steinhoff, K. Sullivan, H. Sun, A. J. Sutton, T. Suzuki, C. Sweeney, T. Takahashi, J. Tjiputra, N. Tsurushima, S. M. A. C. van Heuven, D. Vandemark, P. Vlahos, D. W. R. Wal lace, R. Wanninkhof, and A. J. Watson. An update to the surface ocean CO_2 atlas (socat version 2). *Earth System Science Data Discussions*, 6(2):465–512, 2013. doi: 10.5194/essdd-6-465-2013. URL http://www.earth-syst-sci-data-discuss.net/6/465/2013/.

A. P. Ballantyne, C. B. Alden, J. B. Miller, P. P. Tans, and J. W. C. White. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature*, 488(7409):70–72, Aug 2012.

L. Berline, J. M. Brankart, P. Brasseur, Y. Ourmières, J. Verron. Improving the physics of a coupled physical–biogeochemical model of the North Atlantic through data assimilation: Impact on the ecosystem. Journal of Marine Systems, 64(1), 153-172, 2007.

B. Blanke and P. Delecluse. Variability of the tropical Atlantic ocean simulated by a general-circulation model with 2 different mixed-layer physics. 23:1363–1388, 1993.

J. G. Canadell, C. Le Quéré, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R A Houghton, and G. Marland. Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc Natl Acad Sci USA*, 104(47): 18866–18870, Nov 2007. ISSN 1091-6490 (Electronic).

J. G. Canadell, P. Ciais, K. Gurney, C. Le Quéré, S. Piao, M. R. Raupack, and C. L. Sabine. An international effort to

quantify regional carbon fluxes. *Eos Trans.*, 92:81–88, 2011.

L. Chen, S. Xu, Z. Gao, H. Chen, Y. Zhang, J. Zhan, and W. Li. Estimation of monthly air-sea CO_2 flux in the southern Atlantic and Indian ocean using in-situ and remotely sensed data. *Remote Sensing of Environment*, 115(8):1935-1941, 2011. ISSN 0034-4257. doi: http://dx.doi.org/10.1016/j.rse.2011. 03.016. URL http://www.sciencedirect.com/science/article/pii/S0034425711000939.

L. Cotrim da Cunha, E.T. Buitenhuis, C. Le Quéré, X. Giraud, and W. Ludwig. Potential impact of changes in river nutrient supply on global ocean biogeochemistry. page GB4007, 2007. doi: 10.1029/2006GB002718.

P. Courtier, J.-N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519): 1367–1387, 1994. ISSN 1477-870X. doi: 10.1002/qj.49712051912. URL http://dx.doi.org/10. 1002/qj.49712051912.

T. F. Cox and M. A. A. Cox. Multidimensional Scaling. CRC press. 2000.

T. Crueger, E. Roeckner, T. Raddatz, R. Schnur, and P. Wetzel. Ocean dynamics determine the response of oceanic CO_2 uptake to climate change. 31(2-3):151–168, Aug 2008. ISSN 0930-7575.

A. Dai and K.E. Trenberth. Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeor*, 3:660–687, 2002.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Bamaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hölm, L. Isaksen, P. Kållberg, M. Koehler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. The´paut, and F. Vitart. The era-interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, 137: 553–597, 2011.

F. Deng and J. M. Chen. Recent global CO_2 flux inferred from atmospheric CO_2 observations and its regional analyses. *Biogeosciences*, 8(11):3263–3281, 2011. doi: 10.5194/bg-8-3263-2011. URL http://www.biogeosciences.net/8/3263/2011/.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613): 3385–3396, 2005. ISSN 1477-870X. doi: 10.1256/qj.05.108. URL http://dx.doi.org/10.1256/ qj.05.108.

S. Dobricic and N. Pinardi. An oceanographic three-dimensional variational data assimilation scheme. *Ocean Modelling*, 22:89–105, 2008.

S. C. Doney, K. Lindsay, K. Caldeira, J. M. Campin, H. Drange, J. C. Dutay, M. Follows, Y. Gao, A. Gnanadesikan, N. Gruber, A. Ishida, F. Joos, G. Madec, E. Maier-Reimer, J. C. Marshall, R. J. , Matear, P. Monfray, A. Mouchet, R. Najjar, J. C. Orr, G. K. Plattner, J. Sarmiento, R. Schlitzer, R. Slater, I. J. Totterdell, M. F. Weirig, Y. Yamanaka, and A. Yool. Evaluating global ocean carbon models: The importance of realistic physics. *Global Biogeochemical Cycles*, 18(3):GB3017, 2004. doi: 10.1029/2003GB002150. URL http://dx.doi.org/10.1029/2003GB002150.

M. Dowd, E. Jones, and J. Parslow. A statistical overview and perspectives on data assimilation for marine biogeochemical models. *Environmetrics*, 25(4):203–213, 2014. ISSN 1099-095X. doi: 10. 1002/env.2264. URL http://dx.doi.org/10.1002/env.2264.

T. Fichefet and M. A. M. Maqueda. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J. Geophys. Res.*, 102(C6):12609–12646, 1997.

H. E. Garcia, R. A. Locarnini, T. P. Boyer, J. I. Antonov, M. M. Zweng, O. K. Baranova, and D. R. Johnson. *World Ocean Atlas 2009, Volume 4: Nutrients*, volume NOAA Atlas NESDIS 71. U.S. Government Printing Office, Washington D.C., 2010.

M. Gehlen, R. Gangstø, B. Schneider, L. Bopp, O. Aumont, and C. Ethe. The fate of pelagic $CaCO_3$ production in a high CO_2 ocean: a model study. *Biogeosciences*, 4(4):505–519, 2007. doi: 10.5194/ bg-4-505-2007. URL http://www.biogeosciences.net/4/505/2007/.

M. Gerber and F. Joos. Carbon sources and sinks from an ensemble Kalman filter ocean data assimilation. *Global Biogeochemical Cycles*, 24(GB3004), 2010.

M. Gloor, N. Gruber, T P Hughes, and J. L. Sarmiento. An inverse modelling method for estimation of net air-sea fluxes from bulk data: Methodology and application to the heat cycle. *Global Biogeochem. Cycles*, 15:767–782, 2001. doi: 10.1029/2000GB001301.

M. Gloor, N. Gruber, J. L. Sarmiento, C. L. Sabine, R. A. Feely, and C. Röedenbeck. A first estimate of present and pre-industrial air-sea CO_2 fluxes patterns based on ocean interior carbon measurements and models. *Geophys. Res.*

Lett., 30(1):1010, 2003. URL http://dx.doi.org/10.1029/ 2002GL015594.

W. W. Gregg, N. W. Casey, and C. S. Rousseaux. Sensitivity of simulated global ocean carbon flux estimates to forcing by reanalysis products. Ocean Modelling, 80(0):24-35, 2014. ISSN 1463- 5003. doi: http://dx.doi.org/10.1016/j.ocemod.2014.05.002.

URL http://www.sciencedirect. com/science/article/pii/S1463500314000651.

N. Gruber, M. Gloor, S. E. Mikaloff-Fletcher, S. C. Doney, S. Dutkiewicz, M. J. Follows, M. Gerber, A. R. Jacobson, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, S. A. Mueller, J. L. Sarmiento, and T. Takahashi. Oceanic sources, sinks, and transport of atmospheric CO_2. *Global Biogeochemical Cycles*, 23(1):GB1005, 02 2009. URL http://dx.doi.org/10.1029/2008GB003349.

K. R. Gurney, R. M. Law, A. S. Denning, B. C. Rayner, P. J .and Pak, D. Baker, P. Bousquet, L. Bruhwiler, Y.-H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, P. Peylin, M. Prather, and S. Taguchi. Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochem. Cycles*, 18(1), 2004. doi: 10.1029/2003GB002111.

A. Harten. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 135:260–278, Aug 1997. doi: 10.1006/jcph.1997.5713. URL http://adsabs.harvard.edu/abs/ 1997JCoPh.135..260H.

B. Ingleby and M. Huddleston. Quality control of ocean temperature and salinity profiles - historical and real-time data. *Journal of Marine Systems*, 65(1-4):158–175, 2007. ISSN 0924-7963. doi: http://dx.doi.org/10.1016/j.jmarsys.2005.11.019.

M. Ishii, R. A. Feely, K. B. Rodgers, G.-H. Park, R. Wanninkhof, D. Sasano, H. Sugimoto, C. E. Cosca, S. Nakaoka, M. Telszewski, Y. Nojiri, S. E. Mikaloff Fletcher, Y. Niwa, P. K. Patra, V. Valsala, H. Nakano, I. Lima, S. C. Doney, E. T. Buitenhuis, O. Aumont, J. P. Dunne, A. Lenton, and T. Takahashi. Air-sea CO_2 flux in the pacific ocean for the period 1990-2009. *Biogeosciences*, 11(3):709–734, 2014. doi: 10.5194/bg-11-709-2014.

A. R. Jacobson, S. E. Mikaloff-Fletcher, N. Gruber, J. L. Sarmiento, and M. Gloor. A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Global Biogeochem. Cycles*, 21(1), 2007. doi: 10.1029/2005GB002556.

A. R. Jacobson, S. E. Mikaloff-Fletcher, N. Gruber, J. L. Sarmiento, and M. Gloor. A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, *Global Biogeochem. Cycles* 21 (2007), doi:10.1029/2006GB002703.

S. D. Jones, C. Le Quéré, and C. Rödenbeck. Autocorrelation characteristics of surface ocean pCO_2 and air-sea CO_2 fluxes. *Global Biogeochem. Cycles*, 26(GB2042), 2012. doi: 10.1029/2010GB004017.

F. Joos, G.-K. Plattner, T. F. Stocker, O. Marchal, and A. Schmittner. Global warming and marine carbon cycle feedbacks on future atmospheric pCO_2. *Science*, 284:464-467, 1999.

R. F. Keeling, A. C. Manning, and M. K. Dubey. The atmospheric signature of carbon capture and storage, 2011. URL http://rsta.royalsocietypublishing.org/roypta/369/1943/2113.full. pdf.

R. M. Key, A. Kozyr, C. L. Sabine, K. Lee, R. Wanninkhof, J. L. Bullister, R. A. Feely, F. J. Millero, C. Mordy, and T. H. Peng. A global ocean carbon climatology: Results from global data analysis project (GLODAP). 18(4):GB4031, Dec 2004. ISSN 0886-6236.

P. Landschützer, N. Gruber, D. C. E. Bakker, U. Schüster, S. Nakaoka, M. R. Payne, T. P. Sasse, and J. Zeng. A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic ocean carbon sink. *Biogeosciences*, 10(11):7793–7815, 2013. doi: 10.5194/bg-10-7793-2013

P. Landschützer, N. Gruber, D. C. E. Bakker, and U. Schüster. Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, 28(9):927–949, 2014. ISSN 1944-9224. doi: 10.1002/ 2014GB004853. URL http://dx.doi.org/10.1002/2014GB004853.

P. Landschützer, N. Gruber, D. C. E. Bakker, and U. Schüster. An observation-based global monthly gridded sea surface pCO_2 product from 1998 through 2011 and its monthly climatology (2014).

doi:10.3334/CDIAC/OTG.SPCO2_1998_2011_ETH_SOM-FFN

url: http://cdiac.ornl.gov/ftp/oceans/spco2_1998_2011_ETH_SOM-FFN

W.G. Large and S.G. Yeager. The global climatology of an interannually varying air-sea flux data set. *Climate Dynamics*, 33:341–364, 2008.

C. Le Quéré, Christian Rödenbeck, Erik T Buitenhuis, Thomas J Conway, Ray Langenfelds, Antony Gomez, Casper

Labuschagne, Michel Ramonet, Takakiyo Nakazawa, Nicolas Metzl, Nathan Gillett, and Martin Heimann. Saturation of the Southern Ocean CO_2 sink due to recent climate change. *Science*, 316(5832):1735–1738, Jun 2007. ISSN 1095-9203 (Electronic).

C. Le Quéré, Taro Takahashi, Erik T. Buitenhuis, Christian Röedenbeck, and Stewart C. Sutherland. Impact of climate change and variability on the global oceanic sink of CO_2. *Global Biogeochemical Cycles*, 24(4):1–10, 10 2010. ISSN 1944-9224. doi: 10.1029/2009GB003599. URL http://dx. doi.org/10.1029/2009GB003599.

C. Le Quéré, R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. van der Werf, A. Ahlstro¨m, R. M. Andrew, L. Bopp, J. G. Canadell, P. Ciais, S. C. Doney, C. Enright, P. Friedlingstein, C. Huntingford, A. K. Jain, C. Jourdain, E. Kato, R. F. Keeling, K. Klein Goldewijk, S. Levis, P. Levy, M. Lomas, B. Poulter, M. R. Raupach, J. Schwinger, S. Sitch, B. D. Stocker, N. Viovy, S. Zaehle, and N. Zeng. The global carbon budget 1959 - 2011. *Earth System Science Data Discussions*, (5):1107 – 1157, 2012. URL http://www.earth-syst-sci-data-discuss.net/5/ 1107/2012/.

C. Le Quéré, R. Moriarty, R. M. Andrew, G. P. Peters, P. Ciais, P. Friedlingstein, S. D. Jones, S. Sitch, P. Tans, A. Arneth, T. A. Boden, L. Bopp, Y. Bozec, J. G. Canadell, L. P. Chini, F. Chevallier, C. E. Cosca, I. Harris, M. Hoppema, R. A. Houghton, J. I. House, A. K. Jain, T. Johannessen, E. Kato, R. F. Keeling, V. Kitidis, K. Klein Goldewijk, C. Koven, C. S. Landa, P. Landschützer, A. Lenton, I. D. Lima, G. Marland, J. T. Mathis, N. Metzl, Y. Nojiri, A. Olsen, T. Ono, S. Peng, W. Peters, B. Pfeil, B. Poulter, M. R. Raupach, P. Regnier, C. Ro¨denbeck, S. Saito, J. E. Salisbury, U. Schüster, J. Schwinger, R. Seferian, J. Segschneider, T. Steinhoff, B. D. Stocker, A. J. Sutton, T. Takahashi, B. Tilbrook, G. R. van der Werf, N. Viovy, Y.-P. Wang, R. Wanninkhof, A. Wiltshire, and N. Zeng. Global carbon budget 2014. *Earth System Science Data*, 7(1):47–85, 2015. doi: 10.5194/essd-7-47-2015. URL http://www.earth-syst-sci-data.net/7/47/2015/.

N. Lefèvre, A. J. Watson, and A. R. Watson. A comparison of multiple regression and neural network techniques for mapping in situ pCO_2 data. *Tellus B*, 57(5):375–384, 2005. ISSN 1600-0889. doi: 10.1111/j.1600-0889.2005.00164.x. URL http://dx.doi.org/10.1111/j. 1600-0889.2005.00164.x.

A. Lenton, Matear RJ, Tilbrook B. 2006 Design of an observational strategy for quantifying the Southern Ocean uptake of CO_2. Glob. Biogeochem. Cycles 20, GB4010. doi:10.1029/2005GB002620

A. Lenton, B. Tilbrook, R. M. Law, D. Bakker, S. C. Doney, N. Gruber, M. Ishii, M. Hoppema, N. S. Lovenduski, R. J. Matear, B. I. McNeil, N. Metzl, S. E. Mikaloff Fletcher, P. M. S. Monteiro, C. Rödenbeck, C. Sweeney, and T. Takahashi. Sea-air CO_2 fluxes in the southern ocean for the period 1990-2009. *Biogeosciences*, 10(6):4037–4054, 2013. doi: 10.5194/bg-10-4037-2013. URL http://www.biogeosciences.net/10/4037/2013/.

R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, and D. R. Johnson. *World Ocean Atlas 2009, Volume 1: Temperature*, volume NOAA Atlas NESDIS 68. U.S. Government Printing Office, Washington D.C., 2010.

R. Lukas and D. Karl. Hawaii ocean time-series (hot), 1988–1998: A decade of interdisciplinary oceanography. CD-ROM 99-05, School of Ocean and Earth Science and Technology, University of Hawaii, 1999.

G. Madec and M. Imbard. A global ocean mesh to overcome the North Pole singularity. 12:381–388, 1996.

S. Maksyutov, H. Takagi, V. K. Valsala, M. Saito, T. Oda, T. Saeki, D. A. Belikov, R. Saito, A. Ito, Y. Yoshida, I. Morino, O. Uchino, R. J. Andres, and T. Yokota. Regional CO_2 flux estimates for 2009-2010 based on GOSAT and ground-based CO_2 observations. *Atmos. Chem. Phys.*, 13(18):9351– 9373, 2013. doi: 10.5194/acp-13-9351-2013. URL http://www.atmos-chem-phys.net/13/ 9351/2013/.

R.J. Matear and Andrew Hirst. Climate change feedback on the future oceanic CO_2 uptake. *Tellus B*, 51: 722-733, 1999.

K. Matsumoto, J. L. Sarmiento, R. M. Key, O. Aumont, J. L. Bullister, K. Caldeira, J.-M. Campin, S. C. Doney, H. Drange, J.-C. Dutay, M. Follows, Y. Gao, A. Gnanadesikan, N. Gruber, A. Ishida, F. Joos, K. Lindsay, E. Maier-Reimer, J. C. Marshall, R. J. Matear, P. Monfray, A. Mouchet, R. Najjar, G.-K. Plattner, R. Schlitzer, R. Slater, P. S. Swathi, I. J. Totterdell, M.-F. Weirig, Y. Yamanaka, A. Yool, and J. C. Orr. Evaluation of ocean carbon cycle models with data-based metrics. *Geophysical Research Letters*, 31(7):1–4, 2004. ISSN 1944-8007. doi: 10.1029/2003GL018970. URL http://dx.doi. org/10.1029/2003GL018970.

M. J. McPhaden, A. J. Busalacchi, R. Cheney, Jean-Ren Donguy, Kenneth S. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. T. Mitchum, P. P. Niiler, J. Picaut, R. W. Reynolds, N. Smith, and K. Takeuchi. The tropical ocean-global atmosphere observing system: A decade of progress. *J. Geophys. Res.*, 103 (C7):14169–14240, 1998. ISSN 2156-2202. doi: 10.1029/97JC02906. URL http://dx.doi.org/ 10.1029/97JC02906.

A.F. Michaels and A.H. Knap. Overview of the U.S. JGOFS Bermuda Atlantic Time Series study. 43: 157–198, 1996.

S. E. Mikaloff Fletcher, N. Gruber, A. R. Jacobson, M. Gloor, S. C. Doney, S. Dutkiewicz, M. Gerber, M. Follows, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, S. A. Mueller, and J. L. Sarmiento. Inverse estimates of the oceanic sources and sinks of natural CO_2 and the implied oceanic carbon transport. 21(1):GB1010, 2007. ISSN 1944-9224. doi: 10.1029/2006GB002751. URL http://dx. doi.org/10.1029/2006GB002751.

J. K. Moore, S. C. Doney, D. M. Glover, and I. Y. Fung. Iron cycling and nutrient-limitation patterns in surface waters of the World ocean. 49:463–507, 2002.

R.H. Moss, J.A. Edmonds, K.A. Hibbard, M.R. Manning, S.K. Rose, et al. The next generation of scenarios for climate change research and assessment. Nature 463, 747-756, 2010.

P. R. Oke and P. Sakov. Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol.*, (25):1004–1017, 2008.

A. Olsen. Nordic seas total dissolved inorganic carbon data in carina. *Earth System Science Data*, 1: 35–43, 2009a.

A. Olsen. Nordic seas total alkalinity data in carina. *Earth System Science Data*, 1:77–86, 2009b.

Y. Ourmières, P. Brasseur, M. Lévy, J. M. Brankart, J. Verron. On the key role of nutrient data to constrain a coupled physical–biogeochemical assimilative model of the North Atlantic Ocean. Journal of Marine Systems, 75(1), 100-115, 2009.

G.-H. Park, R. Wanninkhof, Scott C. Doney, T. Takahashi, Kitack Lee, Richard A. Feely, Christopher L. Sabine, J. Trinanes, and Ivan D. Lima. Variability of global net sea–air CO_2 fluxes over the last three decades using empirical relationships. *Tellus B*, 62(5):352–368, 2010. ISSN 1600-0889. doi: 10.1111/ j.1600-0889.2010.00498.x. URL http://dx.doi.org/10.1111/j.1600-0889.2010.00498.x.

S. K. Park and L. Xu, editors. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, volume 2. Springer-Verlag Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-35088-7.

P. K. Patra, M. Ishizawa, S. Maksyutov, T. Nakazawa, and G. Inoue. Role of biomass burning and climate anomalies for land-atmosphere carbon fluxes based on inverse modeling of atmospheric CO_2. *Global Biogeochem. Cycles*, 19(3), 2005. doi: 10.1029/2004GB002258. URL http://dx.doi.org/10. 1029/2004GB002258.

W. Peters, M. C. Krol, G. R. van der Werf, S. Houweling, C. D. Jones, J. Hughes, K. M. Schaefer, K. A. Masarie, A. R. Jacobson, J. B. Miller, C. H. Cho, M. Ramonet, M. Schmidt, L. Ciattaglia, F. Apadula, D. Heltai, F. Meinhardt, A. G. Di Sarra, S. Piacentino, D. Sferlazzo, T. Aalto, J. Hatakka, J. Strom, L. Haszpra, H. A. J. Meijer, S. van der Laan, R. E. M. Neubert, A. Jordan, X. Rodo, J.- A. Morgui, A. T. Vermeulen, E. Popa, K. Rozanski, M. Zimnoch, A. C. Manning, M. Leuenberger, C. Uglietti, A. J. Dolman, P. Ciais, M. Heimann, and P. P. Tans. Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations. *Global Change Biology*, 16(4):1317–1337, 2010. ISSN 1365-2486. doi: 10.1111/j.1365-2486.2009.02078.x. URL http://dx.doi.org/10.1111/j.1365-2486.2009.02078.x.

P. Peylin, R. M. Law, K. R. Gurney, F. Chevallier, A. R. Jacobson, T. Maki, Y. Niwa, P. K. Patra, W. Peters, P. J. Rayner, C. Rödenbeck, I. T. van der Laan-Luijkx, and X. Zhang. Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions. *Biogeosciences*, 10(10):6699– 6720, 2013. doi: 10.5194/bg-10-6699-2013. URL http://www.biogeosciences.net/10/6699/ 2013/.

R. James Purser, Wan-Shu Wu, David F. Parrish, and Nigel M. Roberts. Numerical aspects of the application of recursive filters to variational statistical analysis. Part i: Spatially homogeneous and isotropic Gaussian covariances. *Monthly Weather Review*, 131(8):1524–1535, 2014/10/24 2003a. doi: 10.1175//1520-0493

R. J. Purser, W.-S. Wu, D. F. Parrish, and N. M. Roberts. Numerical aspects of the application of recursive filters to variational statistical analysis. Part ii: Spatially inhomogeneous and anisotropic general covariances. *Monthly Weather Review*, 131(8):1536–1548, 2014/10/24 2003b. doi: 10.1175//2543.1.

K. Raghukumar, C. A. Edwards, N. L. Goebel, G. Broquet, M. Veneziani, A. M. Moore, J. P. Zehr. Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the California Current System. Progress in Oceanography, 138, 546-558, 2015.

A. Ridgwell, J. C. Hargreaves, N. R. Edwards, J. D. Annan, T. M. Lenton, R. Marsh, A. Yool, and A. Watson. Marine geochemical data assimilation in an efficient earth system model of global biogeochemical cycling. *Biogeosciences*, 4(1):87–104, 2007. ISSN 1726-4170. URL http: //www.biogeosciences.net/4/481/2007/.

C. Rödenbeck, S. Houweling, M. Gloor, and M. Heimann. CO_2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, 3(6):1919–1964, 2003. doi: 10.5194/acp-3-1919-2003. URL http://www.atmos-chem-phys.net/ 3/1919/2003/.

C. Rödenbeck, R. F. Keeling, D. Bakker, N. Metzl, A. Olsen, C. Sabine, and M. Heimann. Global surface-ocean pCO_2 and sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme, *Ocean Science*, 9(2),

193–216, 2013.

C. Rödenbeck, D. C. E. Bakker, N. Metzl, A. Olsen, C. Sabine, N. Cassar, F. Reum, R. F. Keeling, and M. Heimann. Interannual sea-air CO_2 flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences*, 11(17):4599–4613, 2014. doi: 10.5194/bg-11-4599-2014. URL http: //www.biogeosciences.net/11/4599/2014/.

C. L. Sabine, S. Hankin, H. Koyuk, D. C. E. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, J. Malczyk, J. Akl, S. R. Alin, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, R. A. Feely, M. Gonza´lez-Da´vila, C. Goyet, N. Hardman- Mountford, C. Heinze, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Rìos, J. Salisbury, J. M. Santana-Casiano, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schüster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, D. Vandemark, T. Veness, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue. Surface ocean CO_2 atlas (socat) gridded data products. *Earth System Science Data*, 5(1):145–153, 2013. URL http://www.earth-syst-sci-data.net/5/145/2013/.

V. V. S. S. Sarma, A. Lenton, R. M. Law, N. Metzl, P. K. Patra, S. Doney, I. D. Lima, E. Dlugokencky, M. Ramonet, and V. Valsala. Sea-air CO_2 fluxes in the Indian Ocean between 1990 and 2009. *Biogeosciences*, 10(11):7035–7052, 2013. doi: 10.5194/bg-10-7035-2013. URL http: //www.biogeosciences.net/10/7035/2013/.

J. L. Sarmiento and N. Gruber. *Ocean Biogeochemical Dynamics*. Princeton Univ. Press, Princeton, NJ, 2006.

J. L. Sarmiento and C. Le Quéré. Oceanic carbon dioxide uptake in a model of century-scale global warming. *Science*, 274:1346–1350, 1996.

U. Schüster, Galen A. McKinley, Nicholas R. Bates, F. Chevallier, S. C. Doney, A. R. Fay, M. Gonzalez- Davila, Nicolas Gruber, S. Jones, J. Krijnen, P. Landschützer, N. Lefèvre, N. Manizza, J. Mathis, Nicolas Metzl, A. Olsen, A. F. Rios, Christian Rödenbeck, J. M. Santana-Casiano, Taro Takahashi, Rik Wanninkhof, and Andrew J. Watson. An assessment of the Atlantic and Arctic sea–air CO_2 fluxes, 1990–2009. *Biogeosciences*, 10:607-627, 2013. doi: 10.5194/bg-10-607-2013.

D. Seidov T. P. Boyer R. A. Locarnini A. V. Mishonov Antonov, J. I. and H. E. Garcia. *World Ocean Atlas 2009 Volume 2: Salinity*, volume NOAA Atlas NESDIS 69. U.S. Government Printing Office, Washington D.C., 2010.

S. Sitch, P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, A. Ahlstro¨m, S. C. Doney, H. Graven, C. Heinze, C. Huntingford, S. Levis, P. E. Levy, M. Lomas, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, A. Arneth, G. Bonan, L. Bopp, J. G. Canadell, F. Chevallier, P. Ciais, R. Ellis, M. Gloor, P. Peylin, S. L. Piao, C. Le Quéré, B. Smith, Z. Zhu, and R. Myneni. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, 12(3):653–679, 2015. doi: 10.5194/ bg-12-653-2015. URL http://www.biogeosciences.net/12/653/2015/.

A. Storto, S. Dobricic, S. Masina, and P. Di Pietro. Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system. *Mon. Weather Rev.*, 139:738–754, 2011.

A. Storto, S. Masina, and S. Dobricic. Estimation and impact of non-uniform horizontal correlation length-scales for global ocean physical analyses. *J. Atmos. Oceanic Technol.*, (31):2330–2349, 2014.

T. M. Suzuki, M. Ishii, J. R. Aoyama, K. Christian, T. Enyo, R. M. Kawano, N. Key, A. Kosugi, L. A. Kozyr, A. Miller, T. Murata, T. Nakano, T. Ono, K. Saino, D. Sasaki, Y. Sasano, M. Takatani, Wakita, and C. L. Sabine. Pacifica data synthesis project. *ORNL/CDIAC-159*, (NDP-092), 2013. doi: 10.3334/ CDIAC/OTG.PACIFICA NDP092.

T. Takahashi and S. C. Sutherland. Global ocean surface water partial pressure of CO_2 database: Measurements performed during 1968–2007. ORNL/CDIAC-152 NDP-088a, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, 2007.

T. Takahashi, S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. P. Chavez, C. L. Sabine, A. Watson, D. C. E. Bakker, U. Schüster, N. Metzl, H. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, A. Körtzinger, T. Steinhoff, M. Hoppema, J. Olafsson, T. S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. G. J. Bellerby, C. S. Wong, B. Delille, N. R. Bates, and H. J. W. de Baar. Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(8–10):554 – 577, 2009. ISSN 0967-0645. doi: http://dx.doi.org/10.1016/j.dsr2.2008.12.009. URL http://www.sciencedirect.com/science/article/pii/S0967064508004311. Surface Ocean CO_2 Variability and Vulnerabilities.

T. Takahashi, S. C. Sutherland, and A. Kozyr. Global ocean surface water partial pressure of CO_2 database: Measurements performed during 1968–2012. ORNL/CDIAC-160 NDP-088(V2012), Carbon Dioxide Information

Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee (USA), 2012.

I. Tegen and I. Fung. Modeling of mineral dust in the atmosphere - sources, transport, and optical- thickness. 99:22897–22914, 1994.

I. Tegen and I. Fung. Contribution to the atmospheric mineral aerosol load from land-surface modification. 100:18707–18726, 1995.

M. Telszewski, A. Chazottes, U. Schüster, A. J. Watson, C. Moulin, D. C. E. Bakker, M. Gonzalez-Davila, T. Johannessen, A. Körtzinger, H. Lüger, A. Olsen, A. Omar, X. A. Padin, A. F. Rıos, T. Steinhoff, M. Santana-Casiano, D. W. R. Wallace, and R. Wanninkhof. Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network. *Biogeosciences*, 6(8):1405–1421, 2009. doi: 10.5194/bg-6-1405-2009. URL http://www.biogeosciences.net/6/1405/2009/.

V. K. Valsala and S. Maksyutov. Simulation and assimilation of global ocean pCO_2 and air–sea CO_2 fluxes using ship observations of surface ocean pCO_2 in a simplified biogeochemical offline model. *Tellus B*, (62):821–840, 2010.

B. Van Leer. Towards the ultimate conservative difference scheme, v. a second order sequel to Godunov's method. *J. Com. Phys.*, 32:101–136, 1979.

A. Velo, F. F. Perez, P. Brown, T. Tanhua, U. Schüster, and R. M. Key. Carina alkalinity data in the Atlantic ocean. *Earth System Science Data*, 1:45–61, 2009.

M. Vichi and S. Masina. Skill assessment of the PELAGOS global ocean biogeochemistry model over the period 1980-2000. *Biogeosciences*, 6(11):2333–2353, 11 2009. URL http://www. biogeosciences.net/6/2333/2009/.

M. Vichi, S. Masina, and A. Navarra. A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: numerical simulations. *Journal of Marine Systems*, 64:110–134, 2007a.

M. Vichi, N. Pinardi, and S. Masina. A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: theory. *Journal of Marine Systems*, 64:89–109, 2007b.

M. Vichi, E. Manzini, P. Fogli, A. Alessandri, L. Patara, E. Scoccimarro, S. Masina, and A. Navarra. Global and regional ocean carbon uptake and climate change: sensitivity to a substantial mitigation scenario. 37(9):1929–1947, 11 2011. URL http://dx.doi.org/10.1007/s00382-011-1079-0.

Vichi, M., T. Lovato, P. Lazzari, G. Cossarini, E. Gutierrez Mlot, G. Mattia, W. McKiver, S. Masina, N. Pinardi, C. Solidoro and M. Zavatarelli (2015), The Biogeochemical Flux Model (BFM): Equation Description and User Manual. BFM version 5.1 (BFM-V5). Release 1.1, BFM Report Series 1, Bologna, Italy.

R. Wanninkhof. Relationship between windspeed and gas exchange over the ocean. 97:7373–7382, 1992.

R. Wanninkhof, G. H. Park, T. Takahashi, C. Sweeney, R. Feely, Y. Nojiri, N. Gruber, S. C. Doney, G. A. McKinley, A. Lenton, C. Le Quéré, C. Heinze, J. Schwinger, H. Graven, and S. Khatiwala. Global ocean carbon uptake: magnitude, variability and trends. *Biogeosciences*, 10(3):1983–2000, 2013. doi: 10.5194/bg-10-1983-2013.

A. J. Watson and J. C. Orr. *Ocean Biogeochemistry*. Global Change — The IGBP Series (closed). Springer Berlin Heidelberg, 2003.

J. While, I. Totterdell, and M. Martin. Assimilation of pCO_2 data into a global coupled physical- biogeochemical ocean model. *Journal of Geophysical Research: Oceans*, 117(C3):1–12, 2012. ISSN 2156-2202. doi: 10.1029/2010JC006815. URL http://dx.doi.org/10.1029/2010JC006815.

J. Zeng, Y. Nojiri, P. Landschützer, M. Telszewski, and S. Nakaoka. A global surface ocean fCO_2 climatology based on a feed-forward neural network. *Journal of Atmospheric and Oceanic Technology*, 31(8):1838–1849, 2015/06/09 2014. doi: 10.1175/JTECH-D-13-00137.1. URL http: //dx.doi.org/10.1175/JTECH-D-13-00137.1.

640
641

Supplemental material

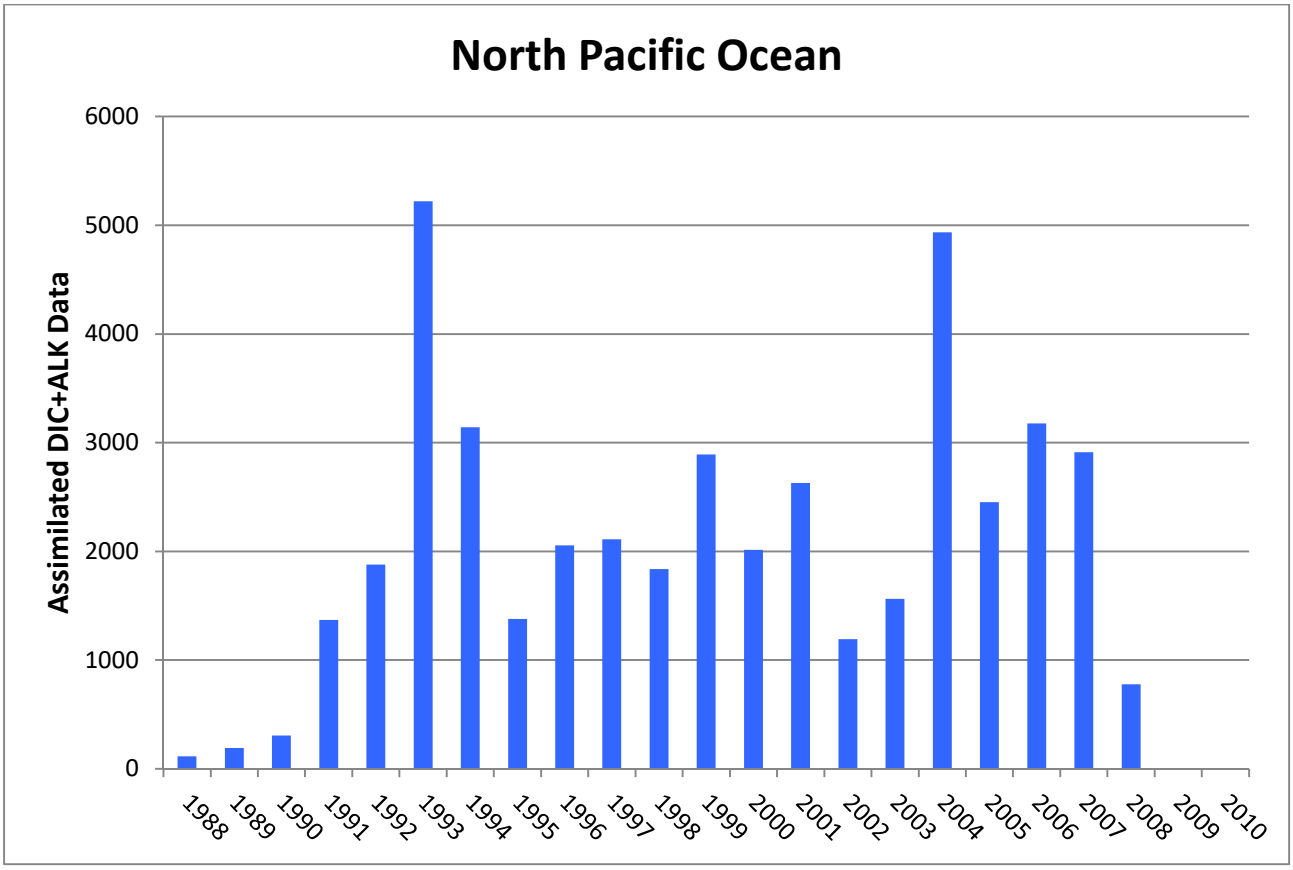

Figure S1. Number of combined DIC and ALK data used in the assimilation for the North Pacific for each year of the simulation. A description of this composite inorganic carbon dataset is given in Sec. 2.3.

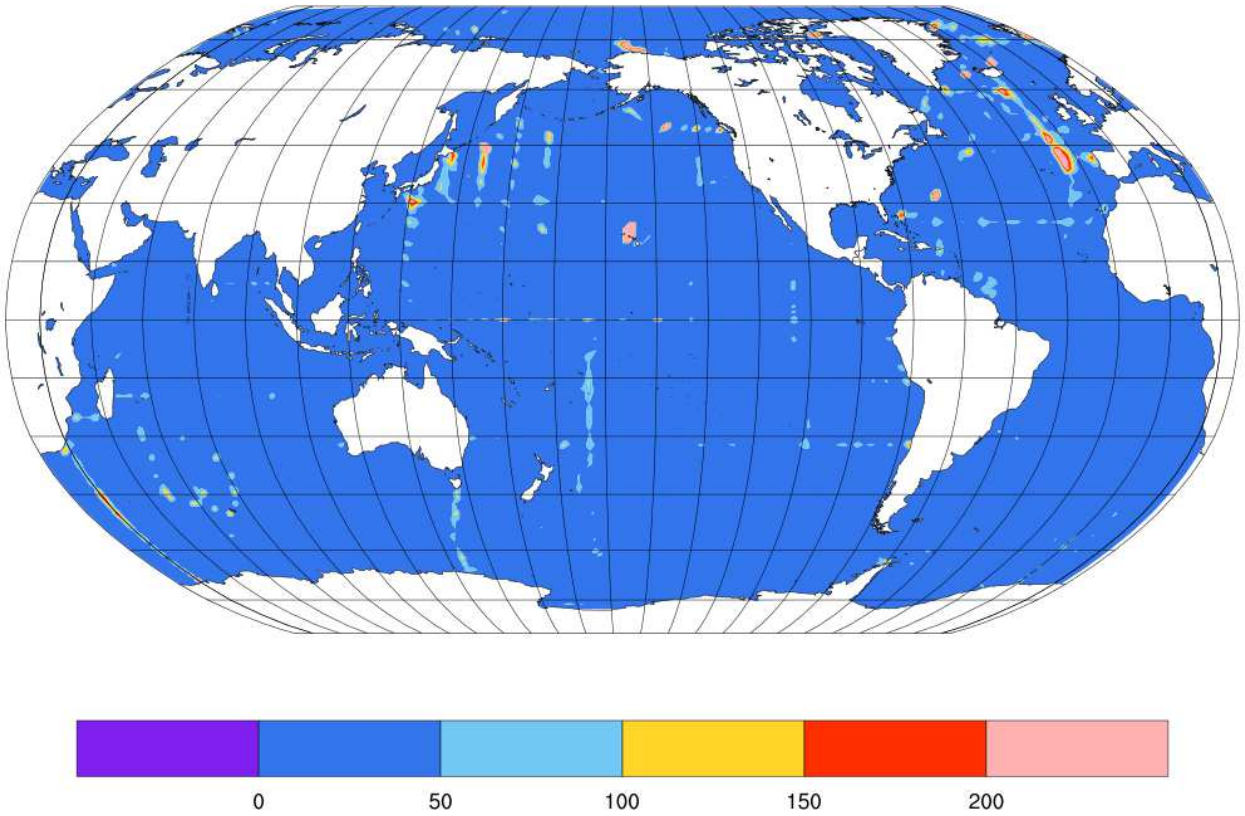

Figure S2. Number of assimilated DIC and ALK data (paired) computed for every model grid point over the entire study period.