Biogeosciences Discuss., 9, 14589–14638, 2012
www.biogeosciences-discuss.net/9/14589/2012/
doi:10.5194/bgd-9-14589-2012

Biogeosciences Discussions

Role of regression model selection and station distribution on the estimation of oceanic anthropogenic carbon change by eMLR

Y. Plancherel[1,*], **K. B. Rodgers**[2], **R. M. Key**[2], **A. R. Jacobson**[3], **and J. L. Sarmiento**[2]

[1]Department of Geosciences, Princeton University, Princeton, New Jersey, USA
[2]AOS Program, Princeton University, Princeton, New Jersey, USA
[3]NOAA/ESRL/GMD, Boulder, Colorado, USA
[*]now at: Department of Earth Sciences and Oxford Martin School, University of Oxford, Oxford, OX1 3AN, UK

Received: 2 September 2012 – Accepted: 4 October 2012 – Published: 19 October 2012

Correspondence to: Y. Plancherel (yvesp@earth.ox.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Differencing predictions of linear regression models generated from hydrographic data collected at different times (the eMLR method) was proposed as a means of quantifying the dominant patterns of change in oceanic anthropogenic carbon in the context of sparse data sets subject to natural variability. The ability of eMLR to recover the anthropogenic carbon signal in the North Atlantic was tested using a global circulation and biogeochemistry model. Basin-scale applications of eMLR on horizontal layers can estimate the change in anthropogenic carbon inventory with an accuracy typically better than 10 %. Regression model selection influences the distribution of the recovered anthropogenic carbon change signal. The systematic use of statistically optimum regression formulae does not produce the best estimates of anthropogenic carbon change if the distribution of the station locations emphasizes hydrographic features differently in time. Additional factors, such as a balanced station distribution and vertical continuity of the regression formulae should be considered to guide model selection. Accurate results are obtained when multiple formulae are used throughout the water column. Different formulae can yield results of similar quality. The fact that good results are obtained in the hydrographically complex North Atlantic suggests that eMLR can produce accurate estimates in other basins.

1 Introduction

Since publication of the global cumulative mid-1990s anthropogenic carbon inventory estimate (Sabine et al., 2004), a measure of the time-integrated anthropogenic signal, attention has turned toward methodologies capable of monitoring the carbon uptake directly. Owing to the size of the oceanic carbon stores and the role of the ocean as a long-term sink of excess carbon dioxide, perturbations, progressive saturation or a decrease of the oceanic uptake rate (Schuster and Watson, 2007; Corbiere et al., 2007; Le Quéré et al., 2007; Khatiwala et al., 2009) can have large impacts on the

BGD
9, 14589–14638, 2012

atmospheric concentrations. Accurate knowledge of the uptake rate and its interannual variability (McKinley et al., 2011) has thus important policy implications for carbon mitigation.

Independant assessements using atmospheric and oceanic carbon observations for the period 1995–2000 constrain the mean oceanic uptake rate of anthropogenic carbon to $2.2 \pm 0.3\,\mathrm{Pg\,C\,yr^{-1}}$ (Gruber et al., 2009). While estimates of the uptake rate tend to converge (Wetzel et al., 2005; Takahashi et al., 2002, 2009; Mikaloff-Fletcher et al., 2006; Khatiwala et al., 2009), assessments diverge on a regional level, showing different uptake and storage patterns (Sabine et al., 2004; Waugh et al., 2006), especially in the Southern Ocean (Caldeira and Duffy, 2000; Lo Monaco et al., 2005a,b; Le Quéré et al., 2007). These differences have important mechanistic implications for the understanding and prediction of the marine carbon cycle and argue for improved observational estimates.

Quantification of the oceanic anthropogenic carbon concentration and the characterization of its time rate of change are challenging. First, anthropogenic carbon is the difference between the comtemporary dissolved inorganic carbon (DIC), i.e. the measured DIC, and an estimate of the natural DIC; that is, the DIC field thought to have existed in the absence of human activity (Gruber et al., 2009). This implies assumptions regarding the cylcing of natural carbon. The anthropogenic carbon fraction is small relative to the background DIC concentration (of order $\leq 5\,\%$ of the DIC in the upper ocean). Even if the current analytical precision is sufficient to detect DIC changes on interannual to decadal time-scales (Brewer et al., 1997; Winn et al., 1998; Bates, 2001), natural variability confounds efforts to quantify the dynamics of the marine anthropogenic carbon sink on these scales (Keeling, 2005; Sabine et al., 2008; McKinley et al., 2011).

Additional difficutlies are associated with monitoring and sampling strategies. Basin or global scale databases represent assemblages of data collected by individual cruises over many years. Owing to logistical limitations and since each cruise has its own scientific objectives, the large-scale spatio-temporal distribution of the data is not

BGD
9, 14589–14638, 2012

ideal. While new samples are often collected close to older stations, this is not always the case. As such, direct point-by-point data comparison in time is only possible for specific cruises and cannot necessarily be used to infer temporal changes on the basin scale. While a point-by-point analysis allows for a good control of the time difference between repeat samples locally (Levine et al., 2008; Sabine et al., 2008; Wanninkhof et al., 2010), this approach would only be applicable to a subset of the data for which repeat measurements exist. A strict section-by-section or station-by-station strategy would thus not be able to exploit the many samples for which no repeat exists. A form of extrapolation, which considers data in entire regions instead of constrained along sections, is thus desireable to exploit all available data.

Wallace (1995) and Friis et al. (2005) proposed to compare empirical regression model representations of the measurements instead of directly comparing time-separated measurements to maximize data use, filter out the natural spatio-temporal variability and to generate spatial prediction; this is known as the extended Multiple Linear Regression (eMLR) approach. A few studies have described various aspects and limitations of the eMLR methodology either in models or applied to data (Friis et al., 2005; Tanhua et al., 2007; Levine et al., 2008; Wanninkhof et al., 2010; Goodkin et al., 2011). In preparation to an application of the eMLR approach to global data sets, we add to these previous efforts by addressing two points not thoroughly covered in the existing eMLR literature: the influence of regression model selection and the effect of variable observational sampling networks on eMLR-derived estimates of the interannual to decadal change in the ocean interior carbon inventory.

The eMLR procedure, under the constraint of the observational sampling network available, is here evaluated using a circulation model that includes carbon and nutrient biogeochemistry in which the true anthropogenic signal is known exactly. The model framework used provides a means of estimating absolute errors in the presence of natural temporal and spatial variability patterns that are consistent with many observed climate processes on a variety of time and space scales.

BGD
9, 14589–14638, 2012

An accuracy target for the rate of change of anthropogenic carbon inventory of 0.1 $PgC\,yr^{-1}$ for each of the major ocean basins (3 PgC globally over 10 yr, 10 % of the expected anthropogenic input for that period) was suggested in the Large Scale CO_2 Observing Plan (LSCOP) report (Bender et al., 2002) as a target for the Repeat CO_2/Hydrography program. Given this criterion, our results show that, while many choices are possible with regards to the structure of the linear regression model used in eMLR caculations, temporal changes in the observational network exert a first order control on regression model selection and so on the eMLR-inferred changes in anthropogenic concentration. Although limited to the North Atlantic, our analysis outlines a general strategy, focusing on the vertical continuity of regression formulae, that can be used to guide model selection in the oceanographic context.

The principles of the eMLR theory are described first, using matrix notation to cast eMLR into the general framework of inverse problems. This is followed by a methodology section giving the details of the circulation and biogeochemistry model experiments used to generate the synthetic dataset on which the eMLR methodology is tested. The methodology section also includes a description of the regression calculations and the mapping scheme used. Result are presented in three parts. First, we describe the structure and variability of the anthropogenic carbon signal. Then, a summary of the regression results focusing on regression quality and formulae structure is given. The influence of the various regressions on the absolute error of the eMLR solution is presented in the last section. A discussion of potential errors, particularly the problem of inhomogenous data distribution in time and space, precedes the conclusions.

2 eMLR theory

By design, regression models separate the fraction of the variance that can be explained by the model and the part that is due to noise. Assuming suitable empirical regression models can be found for DIC and that the physical and biogeochemical processes underlying the model are stationary and not affected by the anthropogenic

BGD
9, 14589–14638, 2012

perturbation, the noise can be filtered out and the anthropogenic signal revealed as the difference between model predictions of DIC at different times (Friis et al., 2005). Conceptually,

$$\Delta\mathrm{DIC}^{\mathrm{anth}} = \mathcal{G}_2(D_2) - \mathcal{G}_1(D_2) \tag{1}$$

where $\mathcal{G}_t$ are empirical model fits at times t derived from the respective data sets D_t. Note that, since it is the difference between predictions from two models derived from two different data sets that is used, a set of DIC predictions obtained from a model obtained from one data set but applied to the other data set is necessary (in this example, $\mathcal{G}_1(D_2)$), so that DIC predictions exist for all samples in the data set D_2. If the models $\mathcal{G}_t$ are linear, the expression on the r.h.s. of Eq. (1) can be reduced to a difference between regression coefficients (Friis et al., 2005).

Regarding models $\mathcal{G}_t$, Tarantola (2005) gives the following expressions (his equations 3.37 and 3.38) as possible forms of the least squares estimator of the regression coefficients $\widetilde{c}$ and the associated posterior covariance matrix $\widetilde{\mathbf{C}_c}$:

$$\widetilde{c} = \left(\mathbf{Z}^T\mathbf{C}_Y^{-1}\mathbf{Z} + \mathbf{C}_c^{-1}\right)^{-1}\left(\mathbf{Z}^T\mathbf{C}_Y^{-1}\boldsymbol{Y} + \mathbf{C}_c^{-1}\boldsymbol{c}_{\mathrm{prior}}\right) \tag{2}$$

$$\widetilde{\mathbf{C}_c} = \left(\mathbf{Z}^T\mathbf{C}_Y^{-1}\mathbf{Z} + \mathbf{C}_c^{-1}\right)^{-1}. \tag{3}$$

$\mathbf{C}_Y$ is the data covariance matrix and $\mathbf{C}_c$ is the prior covariance matrix of the estimator with mean prior densities given in the vector $\boldsymbol{c}_{\mathrm{prior}}$. Although this study uses noiseless synthetic data, consideration of these covariance matrices will be key for the application of eMLR on real data. $\mathbf{Z}$ is any design matrix containing the variables used as predictors and $\boldsymbol{Y}$ is a vector containing the DIC observations.

As indicated by Eq. (1), the eMLR estimate of anthropogenic carbon change is obtained by using the different sets of regression coefficients with one of the reference design matrices, either forward or backward in time. Using Eq. (2), the eMLR quantity

BGD
9, 14589–14638, 2012

predicted with the set of samples taken at time t_2 is given by

$$\widetilde{\Delta C}^{\mathrm{eMLR}}_{\mathrm{anth}|t_2} = \widetilde{\boldsymbol{Y}_{t_2}} - \widetilde{\boldsymbol{Y}_{t_1|t_2}}$$
$$= \mathbf{Z}_{t_2} \cdot \left[\left(\mathbf{Z}^T \mathbf{C}_{\mathrm{Y}}^{-1} \mathbf{Z} \right)^{-1} \left(\mathbf{Z}^T \mathbf{C}_{\mathrm{Y}}^{-1} \boldsymbol{Y} \right) \right]_{t_2}$$
$$- \mathbf{Z}_{t_2|t_1} \cdot \left[\left(\mathbf{Z}^T \mathbf{C}_{\mathrm{Y}}^{-1} \mathbf{Z} \right)^{-1} \left(\mathbf{Z}^T \mathbf{C}_{\mathrm{Y}}^{-1} \boldsymbol{Y} \right) \right]_{t_1} \quad (4)$$

in the limit of no available prior information ($\mathbf{C_c}^{-1} \rightarrow 0$) and with the "tilde" indicating prediction estimates ($\widetilde{\boldsymbol{Y}} = \mathbf{Z}\widetilde{\boldsymbol{c}}$). The subscripts t_2 and t_1 associated with the square brackets apply to every term in the brackets. $\mathbf{Z}_{t_2|t_1}$ is the design matrix built from data at time t_2, although adjusted to utilize the variables included in the regression model derived from time t_1. The notation $t_2 \,|\, t_1$ is introduced to allow for different sets of predictor variables (i.e. different regression formulae) to be used in the derivation of the regression coefficients at either t_1 or t_2, a generalization of original eMLR (Friis et al., 2005).

Ideally, since the physical and biogeochemical processes are assumed constant in time, the structure of the regression formulae should also be constant in time for a given region of the ocean. In reality, as sampling intensity in different regions changes, the formulae of the regression models that minimize residuals in a region may change in time. Equation (4) explicitly accounts for this possibility.

Equation (4) shows that predicted changes in the carbon concentration can occur as expected from differences in the vectors $\boldsymbol{Y}_{t_1}$ and $\boldsymbol{Y}_{t_2}$, but also from differences in the matrices $\mathbf{Z}_{t_1}$ and $\mathbf{Z}_{t_2}$ and from differences in the covariance matrices associated with variable $\mathbf{Y}$, $\mathbf{C}_{\mathrm{Y},t_1}$ and $\mathbf{C}_{\mathrm{Y},t_2}$. The measurement accuracy of DIC and alkalinity (Alk) has improved after the introduction of the certified reference material such that, for most samples taken during and after the World Ocean Circulation Experiment (WOCE), $\mathbf{C}_{\mathrm{Y}_{t_1}} \approx \mathbf{C}_{\mathrm{Y}_{t_2}}$. The measurement accuracy for DIC between cruises would vary by a factor of 2–5 prior to the introduction of reference material, such that these terms

BGD
9, 14589–14638, 2012

can significantly contaminate the eMLR signal when using older data sets, as shown experimentally by Matear and McNeil (2003) and Tanhua et al. (2007).

For uncorrelated errors, an estimate of the errors associated with $\widetilde{\Delta C}^{\mathrm{eMLR}}_{\mathrm{anth}|t_2}$ can be obtained by linear propagation of the individual projected uncertainties; generally, the error around a function $D(x_1, \ldots, x_n)$ can be estimated to first-order by $\delta D = \sqrt{\sum_{i=1}^{n} \left(\frac{\mathrm{d}D}{\mathrm{d}x_i} \delta x_i \right)^2}$. Given Eq. (4),

$$\delta \left(\widetilde{\Delta C}^{\mathrm{eMLR}}_{\mathrm{anth}|t_2} \right) \approx \sqrt{\widetilde{\mathbf{C}}^2_{\mathrm{Y}t_2} + \widetilde{\mathbf{C}}^2_{\mathrm{Y}t_1|t_2}} \tag{5}$$

since the Gaussian posterior probability density of the response variable at each time (t_1 and t_2) is constrained by expresions $\widetilde{\boldsymbol{Y}} = \mathbf{Z}\widetilde{\boldsymbol{c}}$ and $\widetilde{\mathbf{C}}_{\mathrm{Y}} = \mathbf{Z}\widetilde{\mathbf{C}}_{\mathrm{c}}\mathbf{Z}^T$ (Tarantola, 2005). This form of error propagation would be appropriate even if non-linear regression models were considered: $\widetilde{\Delta C}^{\mathrm{eMLR}}_{\mathrm{anth}|t_2}$ is expressed as a difference between two terms, which is a linear operation. One *caveat* is that Eqs. (3) and (5) do not consider errors associated with the predictor variables in **Z**. Only errors associated with variable **Y** are included in these expressions. One way to propagate the errors from the elements of **Z** would be to use a Monte-Carlo approach, although more direct methods exist to approximate errors from the predictor and the response variable in some cases (Tarantola, 2005).

3 Methodology

3.1 Synthetic data set and description of the model

A synthetic data set with known anthropogenic carbon concentrations is used as a testbed. The synthetic data set is constructed by sampling a global circulation-biogeochemistry model (output provided by J. Dunne, Geophysical Fluid Dynamic Laboratory, NOAA, Princeton, NJ, USA) at the station coordinates given by the GLODAP

BGD
9, 14589–14638, 2012

(Key et al., 2004) and CLIVAR (defined operationally as data collected after GLODAP) data sets (Fig. 1) to reproduce the observed sampling grid. To isolate the effect of regression model selection from other sources of error, the synthetic data are assumed free of measurement errors throughout this work.

The analyses focus on 1995 and 2005. These years were chosen as they are representative of the modal sample density available for the GLODAP and CLIVAR data sets. Similarly, emphasis is given to July 1995 and July 2005 as July mimics the summer bias inherent in the data sets.

Our choice of the North Atlantic for this study is motivated by a number of factors. First, it is clearly a region of consequence for carbon uptake by the ocean. Second, the complexity of the hydrographic conditions and water mass structures for the North Atlantic can be expected to pose particular challenges for empirically-based detection methods. Third, the relatively large number of measurements in this region suggests that it is an appropriate context within which to deconvolve uncertainties associated with the eMLR approach itself and uncertainties associated with mapping errors.

The simulator is composed of the NOAA/GFDL z-level coordinate Modular Ocean Model MOM4 general circulation model (Griffies et al., 2004, 2005; Gnanadesikan et al., 2006) and the Tracers in the Ocean with Allometric Zooplankton (TOPAZ) lower-trophic level biogeochemistry model (Dunne et al., 2005, 2007, 2008, 2010). Sea-ice dynamics are modeled by the GFDL Sea Ice Simulator (SIS) (Winton, 2000).

The ocean model has 50 vertical layers and is resolved on a tripolar grid with an approximate resolution of 1°, improved to 1/3° meridionally near the equator. Synthetic profiles isolated from each station are not further sub-sampled in the vertical to mimic the observations, however. This results in a slight overestimation of the vertical sampling relative to the resolution of the data but the ocean is sufficiently well-sampled in the vertical. Horizontal interpolation errors are, for this problem, larger than vertical ones.

The TOPAZ biogeochemistry module is fully prognostic and includes all major nutrients (NO_3, PO_4, O_2, Si, DIC, Alk), labile and semi-labile dissolved organic matter pools,

BGD
9, 14589–14638, 2012

an iron cycle, ballasting of sinking particles, nutrient and light co-limitation, a microbial loop, three classes of phytoplankton and zooplankton. Details about the model formulation and performance are available in Dunne et al. (2010), Sarmiento et al. (2010) and Henson et al. (2009, 2010).

3.2 Simulation configurations and definition of anthropogenic carbon in the model

The model was initialized with World Ocean Atlas (2001) temperature, salinity and nutrients, GLODAP carbon and forced with the NCEP-derived CORE representation of atmospheric fields and fluxes (Large and Yeager, 2004, 2009; Griffies et al., 2009) over the period 1958–2006. Surface salinity was restored to observation with a relaxation time of 60 days.

The strategy used to isolate the anthropogenic carbon concentration from the model is described by Rodgers et al. (2009). Briefly, the model was spun up for two repeating CORE cycles with fixed pre-industrial atmospheric CO_2 concentration after initialization. At this point, parallel integrations were performed: one with a prescribed atmospheric carbon dioxide transient boundary condition, yielding the contemporary carbon signal and one without, giving an estimate of what the evolution of the natural carbon would have been had the atmosphere remained stable at pre-industrial pCO_2 levels. These parallel simulations were repeated for 5 additional CORE cycles with the atmospheric CO_2 concentration increasing monotonically throughout the 5 cycles as prescribed by the known evolution of historical atmospheric pCO_2. The last cycle is used as a model surrogate for years 1958–2006 and provides the basis for this work. Since both branches of integration were forced with exactly the same forcing fields, the physical fields are identical and the only difference between the two runs are the concentrations of carbon dioxide in the oceanic and atmospheric reservoirs. The anthropogenic carbon concentration is operationally defined by difference between the two runs.

BGD
9, 14589–14638, 2012

3.3 Regressions and statistics

First-order additive linear models were fitted to the synthetic DIC data sets extracted from the monthly mean fields of the MOM4/TOPAZ simulations in 1995 and 2005 sampled at both GLODAP and CLIVAR station locations. All 255 possible models, from single-term to 8-term models, were considered, using the following set of oceanographic variables (salinity, potential temperature, nitrate, phosphate, silicate, apparent oxygen utilization, oxygen, salinity): $Z \subseteq \{S, \theta, NO_3, PO_4, Si, AOU, O_2, Alk\}$. An offset term (i.e. y-intercept) is implicitly included in each fit but this term is not included in the following discussion for simplicity. Given the set of variables available in Z, the 255 linear models are given by $a + bS$, $a + b\theta$,..., $a + bS + c\theta$, $a + bS + cNO_3$, etc.

The best regression models chosen from all possible first-order models were identified for each size class (1 to 8 term models) and across all size classes and for each horizontal layer and each month from January to December for the nominal years 1995 and 2005 to investigate the effect of temporal, physical and biological variability on the ability of simple linear regression models to fit oceanographic data. The minimum Akaike Information Criterion (AIC) was used as a guide for model selection across the complexity spectrum. AIC addresses the bias-variance tradeoff problem when comparing models of different complexity and minimizes the risk of over-fitting. AIC is defined as $AIC = -2\ln(L) + 2k$,where L is the log-likelihood and k is the number of parameters in the model. AIC is a measure of residual sum of squares misfit (L) with a penalty added ($2k$) that is a function of the number of terms in the model (Burnham and Anderson, 1998).

3.4 Mapping

Mapping, that is the horizontal extrapolation of point samples to a basin-scale grid, is a necessary step in calculating column inventories from GLODAP and CLIVAR sampling networks. Mapping was performed using a fixed exponential covariance function with a longitudinal correlation scale of 15.5° and a latitudinal scale of 7.4° above

BGD
9, 14589–14638, 2012

3500 m, or 7.4° for both scales below that depth. Analysis of the semi-variograms, experimentation with the length-scales and other kriging control parameters showed these scales to be appropriate. This scheme was chosen to mimic the objective mapping process used by Key et al. (2004) who used typical length scales of 1550 and 740 km above 3500 m and 740 km in both direction below that depth, and to ease the computational burden. In light of the thousands of maps that were produced, a fully adaptable kriging scheme for each map was not practical. Inventories were calculated from fields mapped to a 1° × 1° grid.

4 Changes in DIC distribution

A description of the target signal (change in anthropogenic carbon) and its components (change in natural and contemporary carbon) is provided first, before the eMLR results, to provide context for the signal in relation to the variability captured by the model and the sampling network.

4.1 The "true" target signal

Figure 1a shows the modeled change in column inventory of anthropogenic carbon between July 1995 and July 2005. This panel shows the target signal that eMLR aims to recover. Figure 1a is calculated directly from the transient (contemporary carbon) and control (natural carbon) components on the original model grid. Figure 1a shows that regions with large inventory changes associate closely with water mass formation regions that are also high uptake regions, notably the Labrador Sea water and the North Atlantic subtropical mode water formation regions, but also reflect water mass reorganization, gyre wobble and frontal shifts in the control simulation.

Both the GLODAP and the CLIVAR observational networks are overlain on Fig. 1a, showing how some notable high-change regions are entirely missed by the sampling. One such high-change feature, with column inventory differences above 20 mol m^{-2}

BGD
9, 14589–14638, 2012

and centered around 35° W–35° N, is missed entirely by both the GLODAP or the CLIVAR stations. Another localized high-change feature is situated near 60° W–38° N and is similarly omitted in the respective data sets. The Labrador Sea is currently only represented by the GLODAP stations in our data compilation but data in this region will soon become available.

Figure 1b shows the same quantity as Fig. 1a (i.e. the change in vertical anthropogenic carbon inventory), although panel b is obtained from horizontally interpolated values sampled at GLODAP stations on each model layer and thus includes the vertically integrated mapping error. Mapped fields, using either the GLODAP or CLIVAR station distribution, tend towards overestimation in the subtropics and towards underestimation in the subtropical/subpolar transition and in the Labrador Sea. Local differences are only important in small restricted regions (Fig. 1c, d). Absolute errors due to mapping are small and result in maximum vertically integrated column inventory biases of order $\pm 10\,\mathrm{mol\,m^{-2}}$ (Fig. 1c, d). The kriging uncertainty (uncertainty around the central kriging estimator) is of similar magnitude as the absolute error (difference between the true value and the central kriging estimator) on given depth layers. Propagation of the mapping uncertainty is not considered in the following discussion, where only the central kriging estimator is used as a diagnostic.

4.2 Changes in simulated contemporary and natural carbon distributions

Figure 1e, f show the vertically integrated changes in the transient and control simulations (the two components used to calculate the anthropogenic signal), interpolated from the set of samples taken at GLODAP locations. The July 2005–1995 difference in the transient simulation (Fig. 1e, $\Delta C_{\mathrm{contemporary}}$) reveals substantial carbon accumulation in the Subpolar gyre, the European Basin and at the southern edge of the subtropical gyre ($\approx 15°$ N) but little change in carbon inventory in the subtropical mode water formation region, south of the Gulf Stream. This later point is unexpected as the mode waters are expected to take up anthropogenic carbon (Bates et al., 1996; Lee et al., 2003). These features are in fact compensated in the control simulation (Fig. 1f,

BGD
9, 14589–14638, 2012

$\Delta C_{natural}$). The Greenland Current region, the Eastern Atlantic and the southern edge of the subtropical gyre all show increases in vertical carbon inventories in the control run (Fig. 1f). In contrast, the Western Subtropical Atlantic shows a drastic decrease between 1995 and 2005, which, when added to the transient run, results in substantial carbon uptake in that region, confirming expectations (Fig. 1a, b).

Inspection of horizontal maps in the ocean interior of DIC change in the control simulation between 1995 and 2005 suggest that the systematic negative change in vertical inventory (Fig. 1f) in the North American Basin is caused primarily by a decrease in the DIC concentrations ($>$ 5–10 µmol kg^{-1}) in the deep model ocean ($>$ 2200 m). These deep DIC changes are accompanied by a decrease in the concentration of the other nutrients, an increase in oxygen and a slight warming. The Labrador Sea and subpolar basin show large increases in carbon and in nutrient concentrations, a decrease in oxygen concentrations and strong increases in salinity and potential temperature. These changes are topographically constrained to the west of the Mid-Atlantic ridge below 3000 m, but the changes between 2200 and 3000 m suffice to explain the drop in column inventory visible in the Northeastern Atlantic (25° W, 50° N, Fig. 1f). These patterns indicate that variability in the convective activity and export of the Labrador Sea and downstream adjustments of the western boundary current properties are responsible for the large-scale column inventory changes in the Northern and Western Atlantic (Fig. 1f). The increase in column inventory simulated by the control run at the southern edge of the subtropical gyre and Eastern Atlantic are due to gyre dynamics. Increases in the DIC field are observed in this region between 150 and 700 m, along with changes in other tracers. These changes are consistent with a northward contraction of the subtropical gyre.

Differences in the simulated annual mean sea surface height (SSH) between 1995 and 2005 agree with the interpretation given above. The patterns of change in SSH do not reflect the North Atlantic Basin drop in carbon inventory seen in Fig. 1f, suggesting the source of that feature is in the deep ocean. SSH varies consistently with the signal observed at the eastern and southern edge of the gyre (Fig. 1f). The regions with

BGD
9, 14589–14638, 2012

positive carbon uptake in the control simulation (Fig. 1f) coincide with the regions of highest interannual variability identified by Cromwell (2006) from an analysis of satellite SSH data in the North Atlantic. The source of the positive deviation of the carbon inventory in the subequatorial and Eastern North Atlantic is in the upper few hundred meters. This pattern likely reflects a real mode of interannual variability captured by the model.

The subtropical region with strong negative change in the column carbon inventory (Fig. 1f) is identified as a low SSH variability region by Cromwell (2006). This is further evidence that the strong and coherent signal of Fig. 1f is not due to interannual variability in the upper thermocline. This signal is clearly associated with the Labrador Sea water. Curry et al. (1998) have reported on the export of deep subpolar perturbations caused by variable convection in the Labrador Sea to the subtropics from an analysis of historical hydrographic data. These patterns of DIC inventory changes are qualitatively consistent with the known dominant patterns of SSH variability over the North Atlantic that also affect real data.

5 Model selection and variability of regression performance

Model selection for eMLR has relied on two approaches. On one hand, models are assumed a priori based on knowledge of the physical and biogeochemical processes or data availability. On the other hand, model selection is purely statistical, relying on stepwise linear regression. In this section, we explore the ability of regression formulae to explain the data as a function of depth and time and explore the spatio-temporal continuity of statistically selected models. We show that, for the most part, there is convergence of statistically selected model formulae across multiple depth intervals. This is consistent with the fact that water mass differences are responsible for most of the variance along horizontal layers. Continuity of the regression formulae should thus be used as a model selection criterion in addition to standard measures of statistical misfit.

BGD
9, 14589–14638, 2012

Figures 2 and 3 show the vertical continuity of the model structures selected by minimum AIC for the July 1995 GLODAP or the July 2005 CLIVAR data sets. In these figures, the horizontal axis is regression model number. This represents the full suite of possible permutations for eight predictor variables, beginning with a model with only one term (model 1) to the model containing all eight terms (model 255). The color strip at the top that maches the figure background summarizes information about model complexity. Panel c in these figures indicates the models that are statistically best in each size class (black vertical segments) or overall (white vertical segments), as a function of depth. Panel a and b summarize the frequency with which particular models are selected as optimal throughout the water column, plotted either as a number frequency (panel a) or weighted as a function of the thickness that a particular model layer represents (panel b).

Parallel analyses (not shown) for the complementary July 2005 GLODAP and July 1995 CLIVAR cases indicate that interannual variability exerts only a small influence on the selection of regression formulae. Furthermore, the few observed temporal changes in formula obtained for a constant observational network typically only involved one of the terms in the formula. These term swaps are consistent with the vertical patterns of changes in standard deviation between data sets constructed from the 1995 and 2005 sampling of the model fields (Appendix A).

Different variance patterns imposed by changing sampling networks is a more important factor influencing model selection than interannual variability. A regression model derived with data taken from a particular depth range or from a particular sampling network does not necessarily represent the best fit when applied to a different depth or a different set of stations since the processes governing the distribution of that tracer may be different at other depths and the regression model was optimized for a different variance pattern.

As Figs. 2 and 3 show, the set of regression models selected between the GLODAP or CLIVAR observational networks differ in each complexity class. These observational networks put slightly different emphasis on hydrographic structures owing to

BGD
9, 14589–14638, 2012

the presence, absence, and density of sampling stations in certain areas. CLIVAR emphasizes the Eastern Atlantic and the subtropical gyre. GLODAP samples the North Atlantic more homogeneously, yielding a more balanced contribution of the subsurface Northern Atlantic region, which contains larger nutrient levels and higher oxygen concentrations than the subtropical region. GLODAP also includes coverage in the Irminger Sea, the Iceland Basin and the Labrador Sea: regions characterized by low temperature and low salinities. As a result of the differences in sampling, regressions using the North Atlantic GLODAP data are more representative on the basin-scale.

Network differences are reflected in the formulae of the models identified by minimum AIC. A quantitative analysis of the terms in the selected formulae in Figs. 2 and 3 as a function of depth highlighted the importance of salinity in the top 300 m as an explanatory variable in the regressions applied to the GLODAP data set. Temperature and oxygen replaced salinity in many of the formulae produced by the CLIVAR synthetic data in this depth range. This is a consequence of the fact that the dominant source of variance in the CLIVAR set is the subpolar to subtropical contrast and not the East/West Greenland Current and the Labrador Sea as in the GLODAP case. Salinity took a relatively more important role in CLIVAR between 400 to 1200 m. This reflects the influence of the Mediterranean Sea overflow water in the Eastern Atlantic, which is relatively more heavily sampled during CLIVAR. Silicate was more frequently present in the formulae in that depth range given the GLODAP set of samples. Common features also exist, however, between the formulae structures generated by the two grids. For instance, the role of phosphate at intermediate depths (200–1500 m) was clear for both networks. Similarly, alkalinity was recurrently selected in the deep ocean (below 2000 m). Overall, nitrate and AOU were the least often selected variables in the formulae.

Regression quality varies also with depth. Overall, the quality of the best fits is lower in the top 1000 m than below 3000 m (Fig. 4a, c). There is a thin layer centered around 1500 m where fit quality is better for many models than for the layers just below it (2000 m). This layer corresponds to the position of the Mediterranean overflow water

BGD
9, 14589–14638, 2012

in the circulation model. Given that the vertical profile of the range of AIC values on each layer (Fig. 4b) show a maximum between 800 and 1500 m, model selection can make a significant difference in the Mediterranean overflow layer. On the other hand, the vertical profile of the AIC ranges shows a minimum between 2000 and 2400 m suggesting this depth layer is less sensitive to the form of the particular regression formula. The difference between the maximum and minimum AIC value is lowest below 4000 m where the fits are also best, suggesting that many models can be used to properly represent the DIC field in that range.

While Figs. 2 and 3 indicate that there is some volatility in terms of the models identified by minimum AIC as a function of depth, quite a few models have AIC values within 10 % of the depth-specific AIC range from the minimum AIC in each layer (Fig. 4a). Differences in AIC values can be relatively small between many of the regression models. Often, these closely fitting formulae fall in related groups, e.g. the nitrate term replaces the phosphate term, oxygen and AOU swap. While a strict identification of the minimum AIC values can result in model formulae with different structures, the DIC data can be fitted to similar degrees of precision using a variety of different models. Although this work did not consider measurement uncertainty, this additional source of noise will further contribute to blur boundaries between regression formulae such that multiple regression models will be statistically indistinguishable from one another in terms of fit quality. Investigators will be able to use closely related models for convenience, for example to maximize data coverage in cases when measurements for particular tracers are missing.

6 Recovery of the change in anthropogenic carbon signal by eMLR

While the question of variable station coverage and associated data set variance is not an issue when dealing with exactly repeated data sets, it is a dominant consideration in the present basin-scale eMLR application. Previous applications of eMLR have required the structure of the regression formula to be constant as a function of time and

BGD
9, 14589–14638, 2012

derived the anthropogenic signal by difference between the regression coefficients. However, direct subtraction of the regression coefficients is only possible because the models are linear. The equivalent signal can also be obtained by subtracting the predicted DIC values obtained after parallel application of the regression equations to the data from one of the time points (Eq. 4). This second approach opens the conceptual possibility of using separate regression models, possibly non-linear models, derived independently at each time point.

The main argument for using a constant model structure in time is that the physical and biogeochemical processes maintaining the DIC field are relatively constant and should thus be constrained by the same empirical models. In practice, there is no guarantee that empirical formulae represent these physical and biogeochemical processes accurately. In addition, if the observational network varies, the variance in the data will change and regression formulae will match these different patterns of variance, such that the concept of "best" formula becomes a function of depth and of the particular geographic distribution of the samples.

This section investigates the overall performance of eMLR and contrasts results obtained by the two conceptual approaches described above, namely: strategy (1) use of a composite of statistically optimized formulae with sets of explanatory variables that are allowed to vary in time and as a function of depth, and strategy (2) use of regression formulae with a constant set of explanatory variables at all depths and times but with regression coefficients optimized independently at each depth and time. Results for the basin-scale inventory changes are presented first, followed by column inventory changes and layer inventory changes.

6.1 Basin-scale inventories

6.1.1 "Best AIC" strategy

The basin-scale inventory changes inferred from the composite best-AIC eMLR methodology (strategy 1) slightly underestimate the true inventory changes (Fig. 5).

BGD
9, 14589–14638, 2012

The relative error of eMLR estimates varies seasonally from about −3% in November to −8% in February for the estimate projected onto the CLIVAR stations (green lines). Due to the ≈2.5% overestimation introduced by the mapping process in GLODAP, the underestimation is less severe in the GLODAP case due to compensating errors; the relative error in the inventory change estimate is offset by about 2.5% and varies seasonally from about −1% in November to −6% in April.

The simulated (true) change in North Atlantic carbon inventory between 1995 to 2005 (difference taken month-by-month: January 2005–January 1995, etc.) is 4.43 Pg C (Fig. 5) and does not vary significantly through the year indicating that seasonality is fairly constant between these two years. The error in the "best AIC" eMLR carbon uptake estimate is smallest in fall, early winter and largest in the spring. Since the mapping error is nearly constant, the exaggerated seasonal cycle of the eMLR estimates mainly originate from the seasonally varying ability of linear models to fit the data. Deep convection, shoaling of the mixed layer and initiation of blooms in late winter and spring all contribute to the presence of sharp property gradients that are difficult to properly represent in simple linear models empirically defined over broad geographic scales.

Analysis of regression statistics indicates that it is more difficult to fit first-order linear models to summer and fall data than to winter data, irrespective of the sampling grid. Diagnostics or bulk regression quality are usually very good, however. Aside from a few summer and fall months in the top 100 m where the standard error of the residuals hovers near 8 µmol kg^{-1}, the standard errors of the residuals are typically smaller than the typical modern measurement uncertainty for DIC (≈ 4 µmol kg^{-1}) and consistently smaller than 2 µmol kg^{-1} below 500 m. Corresponding seasonal profiles of R^2 values produced by the "best AIC" models are typically better than 0.995 below 200 m. Summer and fall values are slightly lower in the top 80 m and between 80 m and 180 m throughout the year ($R^2 \approx 0.98$). All regressions are significant at $p \ll 0.001$.

BGD
9, 14589–14638, 2012

6.1.2 Constant formula strategy

The relative and absolute errors in the determination of the change in North Atlantic carbon inventory resulting from the use of eMLR with fixed regression structures (strategy 2) for both the GLODAP and CLIVAR data sets, projected either backward or forward in time onto the corresponding stations, are shown on Fig. 6 for all 255 first-order models. Most calculated inventory changes are within 20 % of the true value, with a large fraction of the models resulting in an underestimation. The mean relative error for eMLR across all models is −6 % (±2 %) and is significantly different from 0 (two-tailed t-test, $p < 0.001$). The GLODAP and CLIVAR estimates are well-correlated ($\rho = 0.92$, Pearson correlation, $p < 0.001$), confirming that mapping errors are small.

The across-model average underestimation of 6 % over 10 yr (≈ 0.2 Pg C) obtained in this study for the North Atlantic easily meets the LSCOP criterion (Bender et al., 2002): 75 % of the models tested produce carbon uptake estimates within 1 Pg C of the true value. Considering that about one third of the global carbon inventory is in the North Atlantic (Steinfeldt et al., 2009), 1 Pg C is proposed as a North Atlantic target over 10 yr in that region. Half of the regression formulae yield results within 0.5 Pg C of the true estimate.

From these results, it is clear that most models will produce estimates of the decadal basin-scale inventory change that meet desired accuracy limits. For example, all models with 7 or more terms and the composite best-AIC solutions (strategy 1) for every month (Fig. 5) produce uptake estimates that are better than the 0.5 Pg C error limit and thus exceed the success criterion proposed in the LSCOP report (Bender et al., 2002). Simpler models, such as models Z_{140} and Z_{150}, which stood out particularly in Figs. 2 and 3, also fall within the 0.5 Pg C accuracy limit. Owing to cancelation of errors, however, intergral measures may not be particularly sensitive tests of quality. Furthermore, as simulated changes indicate (Fig. 1a, e, f), the evolution in the distribution of anthropogenic carbon in the basin is far from uniform. Do all regression models produce the same distribution or are some models better able to reproduce local features

BGD
9, 14589–14638, 2012

than others? This question is addressed in the next section, looking at the column and layer inventory changes.

6.2 Column inventories

Changes in carbon column inventories resulting from the use of fixed model structures at all depth on both the GLODAP and CLIVAR data sets (strategy 2) are shown in Fig. 7a–h. Only small differences exist between the case when the regressions are projected backwards in time onto the GLODAP data (Fig. 7a, c, e, g) or forward in time onto the CLIVAR data (Fig. 7b, d, f, h). Forward projection onto the CLIVAR stations results in improved results in the East Greenland Current region, where the overestimation relative to the true values is less in the CLIVAR case than in the GLODAP case. One systematic feature is the slight worsening of the overestimation near Gibraltar in the CLIVAR case in spite of the availability of a few samples in that region. This is typical of most regression models tested in this study. Overall, backward or forward projection of the results does not make a big difference. In basins where mapping errors may be a problem owing to data sparsity, however, mapping of the combined backward and forward projected results could be performed to improve spatial coverage.

Results from the 8-term model (Z_{255}, 7g, h) produce results similar to the "best AIC" compilation of models (Fig. 7i). Models with the lowest overall AIC values tend to be the more complex ones (Figs. 2 and 3) and the model with the maximum number of terms is the most frequently selected statistically for each depth and each sampling network.

The error pattern of Fig. 7g–i look similar to the column inventory change pattern resulting from the natural carbon run (Fig. 1f), although with absolute errors of overall smaller magnitudes than the vertically integrated natural carbon change. This pattern similarity indicates that, while eMLR accounts for some of the natural variability, large-scale natural variability patterns are not fully corrected for, even when the statistically best models are used systematically.

The "best AIC" estimate (Fig. 7i), which represents the best statistical solution from the perspective of minimizing fit residuals, is able to account for many of the dynamical

BGD
9, 14589–14638, 2012

changes but not all. Strategy 1 is most sensitive to changes in the sampling network. Preliminary inspection of the station distribution elsewhere indicates that this effect will be stronger in other basins. Selecting regression models for eMLR solely based on their ability to fit data in a statistical sense is not sufficient. Other factors, such as a balanced basin-wide representativeness of the station distribution and consistency with the basic oceanographic structure (spatial continuity of the models) should be considered.

In comparison to the "best AIC" case, absolute error patterns of the vertical inventory change produced by simpler (4-term) constant model structures (strategy 1) are shown in Fig. 7a, b (model $Z_{100} = \{S, \theta, PO_4, O_2\}$) Fig. 7c, d ($Z_{140} = \{\theta, PO_4, Si, Alk\}$), and Fig. 7e, f (model $Z_{150} = \{NO_3, PO_4, Si, Alk\}$). Models Z_{100}, Z_{140} and Z_{150} were selected based on analyzes of Figs. 2 and 3.

The use of model Z_{140} produces an error pattern that is most similar to the "best-AIC" and Z_{255} cases. The main differences are the amplitude of the overestimation in the Eastern Atlantic, at the southward edge of the subtropical gyre and the extrapolation to the Caribbean. The use of model Z_{100}, which in contrast to Z_{140} does not use silicate or alkalinity but rather salinity and oxygen in addition to phosphate and temperature, produces smaller overestimates than model Z_{140} in the subequatorial and Eastern Atlantic. However, it also results in inflated and extended underestimations over the North American Basin and the Labrador Sea. Model Z_{150}, which uses neither temperature, salinity nor oxygen as parameters, produces yet other absolute error patterns, typified by large overestimations over the Eastern Atlantic and large underestimations over the Northwest Atlantic. Clearly, various regression models and implementation strategies result in different column inventory error patterns, even if they produce similar basin-scale inventory change estimates.

Since some variables are clearly more appropriate in certain layers of the water column than others (Appendix A), the next section looks at the vertical distribution of absolute errors. A strategy is proposed that is intermediate to strategies 1 and 2. This compromise uses a composite of models in the vertical, selected based on both

BGD
9, 14589–14638, 2012

statistical fitness and considers other factors, such as vertical continuity and oceanographic relevance.

6.3 Layer inventories and the role of vertical continuity

Figure 6 shows that good inventory change estimates can be obtained from different regression formulae. Yet, different regression formulae produce different horizontal distributions of the column inventory of the anthropogenic carbon change (Fig. 7). Results in Fig. 6 can give a false sense of confidence in the ability of eMLR to recover the true signal and in the value of selecting appropriate regression models. Additional diagnostics, such as the vertical continuity of the regression formulae and an investigation of the geographical pattern of the residuals (not shown), are necessary to help evaluate the quality of anthropogenic carbon estimates produced by eMLR.

Absolute errors of the 1995 to 2005 inventory changes calculated layer-by-layer for all first-order models (strategy 2) are shown in Fig. 8a. Vertical patterns of the absolute errors in Fig. 8 are consistent with patterns resulting from the AIC analysis in Fig. 4. One notable similarity is the band of relative AIC highs centered around 2000 m (Fig. 4a, c) which corresponds to a region of systematic underestimation of the true layer inventories in the water column (Fig. 8a). The region between 1500 to 3500 m generates most of the error (underestimation, Fig. 6) in the basin-scale inventory change estimates. The vertical distribution of the change in layer inventory associated with the natural carbon simulation (green line in Fig. 8b, c) clearly shows the effect of the Labrador Sea Water variability and water mass reorganization.

Illustrative layer-specific inventory change profiles are shown in Fig. 8b, together with the "best AIC" estimate. Panel c shows the corresponding cumulative layer-by-layer vertical inventory change, integrated from the bottom to the surface. The composite "best AIC" eMLR case (yellow line in Fig. 8b, c) does not produce the most accurate profile. Based on Fig. 8b, c, a simpler eMLR solution relying on models Z_{100} in the upper 1500 m and Z_{140} below that depth would produce the true uptake inventory profile almost exactly.

BGD
9, 14589–14638, 2012

Absolute errors cannot be used as guides for model selection when working with real data. Instead, vertical continuity of statistically selected models, in conjunction with a general oceanographic assessment of the regression residuals, are proposed as model selection criteria. These criteria can be used to isolate model formulae that outperform a purely statistical application of eMLR. This approach mitigates the influence of changing the observational network and reduces the risk of using empirial models that are biased towards particular features.

As indicated in Fig. 2, the family of 4-term models shows strong vertical continuity between layers, with essentially four formulae able to cover all depths from 100 to 4000 m. These 4-term models (numbers 140, 99, 100 and 150, ordered by relative frequencies) are $Z_{140} = \{\theta, PO_4, Si, Alk\}$, $Z_{99} = \{S, \theta, PO_4, AOU\}$, $Z_{100} = \{S, \theta, PO_4, O_2\}$, $Z_{150} = \{NO_3, PO_4, Si, Alk\}$.

Model Z_{140} is the model structure used by Friis et al. (2005) for their North Atlantic analysis and Z_{100} is the model used by Levine et al. (2008). Interestingly, while Levine et al. (2008) applied this structure for the model fields between 200 and 2000 m, Fig. 2 suggest this structure is more appropriate in the upper 200 m. Models Z_{99} and Z_{100} are nearly identical, the only difference between the two being the use of O_2 or AOU. Model Z_{150}, which fits the data well in the range 2000 to 4000 m is interesting in that it does not include either θ or S in its formula, whose dynamic range in that depth range is small compared to other tracers (Appendix A). This is qualitatively consistent with the classic studies of Broecker (1974) and Broecker et al. (1985) who used nutrient-based composite tracers ("NO", "PO") to characterize the flow path of deep waters in the Atlantic.

While some of the particular models identified from the GLODAP analysis (Fig. 2) are also present in the CLIVAR equivalent (Fig. 3), their vertical stacking can differ. This is the case for models Z_{99}, Z_{100} and Z_{140}. Given the CLIVAR setup, Z_{140}, the model of Friis et al. (2005), takes a prominent role in the top 200 m while models Z_{99} and Z_{100}, the model of Levine et al. (2008), occupy the space between 300 and 500 m. Models $Z_{97} = \{S, \theta, NO_3, Alk\}$ and $Z_{98} = \{S, \theta, PO_4, SiO_4\}$ belong to the same model group as Z_{99}

BGD
9, 14589–14638, 2012

and Z_{100}, which features salinity, temperature and phosphate (or nitrate) as dominant variables. Z_{97}, Z_{98} extend the influence of this model group down to about 3000 m, although the continuity is not as clear as with models Z_{140} or Z_{150} in the GLODAP case.

The differences between Friis et al. (2005) and Levine et al. (2008) can be explained by the results of our analysis. Friis et al. (2005) used data located in the Subpolar North Atlantic (North of 40° N, South of Iceland) and many of the data used in the Friis et al. (2005) are the same data that partly constitute GLODAP (Key et al., 2004) in that region. It is then reassuring that both Friis et al. (2005) and our results converge towards the same model (Z_{140}). Similarly, the data set used by Levine et al. (2008) was most heavily influenced by the subtropical regions, more like CLIVAR, and it is again reassuring that model Z_{100}, or related models, be most representative in that case. Tanhua et al. (2007) used model $Z_{205} = \{\theta, NO_3, SiO_4, AOU, Alk\}$. Since the data set used by Tanhua et al. (2007) was oriented along East-West sections in the Subtropical North Atlantic, model Z_{205} is unrepresentative of the basin-scale data set used here and Figs. 2 and 3 show this formula is never selected.

Given that this analysis uses a physical and biogeochemistry model as a source of data, that Levine et al. (2008) used another circulation/biogeochemistry model and that Friis et al. (2005) used real data, it is encouraging to note how well the regression formulae proposed by each study converge when presented in the context of their associated sampling grids. Whether a simple combination of the regression formlae $Z_{100} = \{S, \theta, PO_4, O_2\}$ and $Z_{140} = \{\theta, PO_4, Si, Alk\}$ is appropriate for application of eMLR to the real data set remains to be seen. Based on the analysis presented above, the fact that the TOPAZ model is a state-of-the-art biogeochemistry model and the robust correspondence with other studies, it would appear, however, that these are good candidate formulae in the North Atlantic.

BGD
9, 14589–14638, 2012

7 Discussion

The focus of this study was to evaluate the impact of regression model selection on the eMLR results and to make a step towards basin-scale eMLR implementation with real data by looking at the effects of realistic spatial station distribution on the results. Although the context of noiseless synthetic data was practical and appropriate for our purpose, many simplifying assumptions were made. Some of these assumptions and other remaining issues for the applications of basin-scale eMLR to real data are discussed below.

7.1 Unresolved temporal variability

The analysis performed here relied on snapshots of the ocean state taken either in 1995 or 2005, a situation which is overly idealistic since hydrographic sampling programs are never instantaneous. July was chosen to approximate the summer bias that exists in the real data sets, and the years 1995 and 2005 were selected as they represent peaks in sampling intensity. This section dicusses the possible effects of unresolved variability on the eMLR results.

7.1.1 Seasonality

These analyses have shown that the ability of regression models to fit the DIC data varies through the seasonal cycle. The summer to winter contrast in the standard error of the regressions for either the GLODAP or CLIVAR sampling is about 5 $\mu mol\,kg^{-1}$. This effect is restricted to the upper water column, however. The GLODAP 1995 to CLIVAR 2005 differences in the standard errors of the residuals are typically smaller than 2 $\mu mol\,kg^{-1}$, but are mostly caused by differences in the sampling grid and not temporal changes, as comparisons with corresponding GLODAP 2005 and CLIVAR 1995 cases show. Even if the upper ocean contains large anthropogenic carbon concentrations, the volume is relatively small and changes in seasonality only result in a small signal

BGD
9, 14589–14638, 2012

(4 % amplitude, Fig. 5), small relative to changing the selection strategy of the empirical models, on the basin-scale estimate of the inventory change.

Seasonal inventory changes estimates (Fig. 5) were derived using month-by-month comparisons, however, where January 1995 is directly compared to January 2005, etc. In practice, real data sets are composed of samples taken from different seasons. Unless the seasonal biases in sample distribution contained in real data sets were to change drastically (e.g. all winter versus all summer values), the seasonal bias inherent in the data is not expected to vary greatly between sampling campaigns. Given the available sample distribution, differences in representativeness of the sampling grids have a larger effect than changes in the seasonal distribution of the samples. Furthermore, since regression misfits are largest in the summer and early fall, addition of winter and spring data should result in an overall improvement of the fit quality, a consequence of reduced biogeochemical gradients during the winter and spring seasons due to more intense mixing. While seasonal effects can produce local extrema in residuals at particular near-surface stations, seasonal variability tends to be filtered out and is not expected to bias the change in carbon inventory estimates obtained by eMLR on the space and time-scales relevant to the repeat hydrography program.

7.1.2 Sub-monthly variability

The synthetic data sets were generated from monthly mean fields such that sub-monthly variability is filtered out by design. The magnitude of seasonal variability outweighs sub-monthly variability. Since seasonal variability does not introduce large errors in the eMLR inventory results, and since eMLR is a statistical method that relies on a large number of data points, sub-monthly variability is not expected to play a role as long as spatial covariances typical of these temporal scales are small.

There exists an implicit relationship between the spatial scales of the system under study and the temporal scales that are smoothed out by regression: any perturbation must affect a substantial fraction of the data for it to have a noticeable effect on the regression statistics and the eMLR results. Considering data on the basin-scale for

BGD
9, 14589–14638, 2012

the regression analysis is then equivalent to filtering out temporal variability that is uncharacteristic of that scale and is averaged out. Since about a decade separates WOCE and CLIVAR, consideration of large domains consistent with spatial patterns of interannual variability limits aliasing of shorter term variability.

7.2 Temporal sampling density

The target North Atlantic average uptake rate in the simulation is $0.443\,\mathrm{Pg\,C\,yr^{-1}}$. This number is of course obtained from the knowledge that exactly 10 yr separate the measurements. Allowing for uncertainty in the timing of ±2 yr, the uptake rate would vary from 0.55 to $0.37\,\mathrm{Pg\,C\,yr^{-1}}$ were one to spread the true signal over 8 or 12 yr. These values are close to the accuracy limits on the uptake rate implied by the LSCOP criterion ($0.443 \pm 0.1\,\mathrm{Pg\,C\,yr^{-1}}$). Since most of the model formulae produce North Atlantic uptake estimates that underestimate the true value (by 6 % on average, Fig. 6), assuming smaller time intervals (by 1 or 2 yr) between data sets would compensate this.

The degree to which the use of a nominal time interval between sampling campaigns, as done here, biases the estimated uptake rate is not clear. This depends on the spatial distribution of the data and on how the time interval is distributed spatially, i.e. how much each station influences the regressions. Based on the noise-free ideal calculations performed here, it is suggested that if a true inventory change can be approximated precisely, basin-scale eMLR-estimated uptake rates will remain within the desired accuracy of the true value if the bias in the characteristic time interval is smaller than ±2 yr.

The important problem of temporally staggered samples is not yet resolved in the context of eMLR. Although interior DIC values can be adjusted to a nominal year, it is not possible to do the same with all tracers. This may result in possible inconsistencies as samples in different regions can be influenced by different phases of natural variability patterns. This effect has yet to be quantified.

BGD
9, 14589–14638, 2012

7.3 Spatial sampling density

This analysis has shown that one important aspect of eMLR implementation lies in the spatial representativeness of the GLODAP and CLIVAR data sets. Ideally, data sets should measure approximately the same hydrographic regions in the same amount, otherwise optimized empirical formulae may contain different explanatory variables. Emphasis of particular water masses or gradients may result not only from the presence or absence of data in the region, but also from the station density along hydrographic cruises.

Because of inhomogenous and non-random sampling of the ocean, an eMLR implementation based purely on statistical arguments (i.e. "best AIC", strategy 1) will not necessarily yield the most accurate answer. Analysis of the geographical distribution of the residuals indicate that misfit is not homogenous but that residuals form large spatially coherent patterns. These patterns tend to be analogous for similar regression formulae applied to the different sampling grids, however. As such, structures due to regression misfit partly cancel when subtracting predictions from similar formulae (strategy 2) as part of the computation of the anthropogenic signal.

Although model selection does not influence basin-scale inventory changes very much, model selection is very important locally. The concepts of a balanced station coverage and of vertical continuity were used, in addition to statistical measures of fit, as guides for model selection. Formally quantitative methodologies that account for these additional aspect as part of the eMLR calculation are desirable.

In a few regions, multiple repeated cruises are available (e.g. OVIDE section in the Northeast Atlantic, Lherminier et al., 2007). Using all of these sections in the analysis will bias the data set towards these particular regions. In such cases, it is best to use the one cruise track that is most representative of the nominal year used in the analysis (i.e. 2005). Because of the temporal data distribution, the eMLR estimate of the carbon uptake represents a weighted average over a few years. Repeat sections provide a rare and valuable opportunity to evaluate the sensitivity of the final eMLR estimates

BGD
9, 14589–14638, 2012

to temporal data inconsistencies by replacing the temporally most representative section with the others. These repeat cruises can also be used to estimate the detection limit of any eMLR implementation.

7.4 Additional recommendations

As global eMLR implementations are being attempted, we list here a few additional points not addressed in this paper but relevant to the application of eMLR to real data. First, this analysis was performed on depth layers for convenience (the model output is gridded to depth levels). An analysis performed on isoneutral surfaces instead of horizontal surfaces would likely perform better because property gradients are smaller on isoneutrals as water masses mostly mix along these surfaces. Secondly, while working in smaller geographical regions will improve the regression fits, the size of regions should not be so small as to be prone to strong aliasing by time scales shorter than the time scale inherent to the repeat hydrography program (about 10 yr). Third, solution of inverse problems, such as eMLR, are best when variables contained in the design matrix **Z** are independent. Oceanographic tracers tend to be highly correlated. Optimizing tracer orthogonality, perhaps using quasi-conservative (Si^*, N^*, C^*, PO, NO) or dynamic tracers (potential vorticity), will improve conditioning of the problem. Finally, the influence of measurement errors and possible biases between data from different cruises will have to be addressed.

8 Conclusions

The eMLR method was evaluated using output from a global circulation and biogeochemistry model with a known anthropogenic signal. eMLR has so far mostly been used on exact repeat hydrographic sections or small regions, but not over large scales. The model was sampled at observed station locations to create synthetic data sets that mimic the spatial structure of the observed data sets. Analysis of these synthetic

BGD
9, 14589–14638, 2012

data sets has shown that both the station distribution and the selection of regression models exert strong influences on the eMLR's ability to recover the true signal. The model selection process is not independent of the station distribution.

The basin-scale application of eMLR exhibits skill in detecting ocean carbon changes that fall within the threshold of acceptable uncertainty (10 %) proposed in the LSCOP report (Bender et al., 2002). The eMLR method thus presents an opportunity to evaluate the evolution of the ocean carbon sink and its rate of change independently from other estimates, such as Khatiwala et al. (2009), who assumed a steady circulation and thus did not account for natural variability explicitly.

The interior distribution of the change in anthropogenic carbon shows a complex structure. Interior layer inventories can err by as much as 100 % or more when the analysis is performed on horizontal surfaces and uses inappropriate regression models, even if the basin-scale inventory change is in agreement with the true value. The depth range between 1500 and 3500 m is the source of most of the difference between the true signal and the eMLR-inferred signals in the North Atlantic. This challenge is associated with variability in the formation and export of Labrador Sea water. The North Atlantic eMLR results are particularly sensitive to model selection in that depth range.

This analysis has shown that careful application of the eMLR technique is able to produce accurate estimates of the anthropogenic carbon change. The analysis was performed in the North Atlantic as this basin contains about one third of the global anthropogenic signal. The North Atlantic is also one of the most hydrographically complex and dynamically variable region. The fact that eMLR produces good results in that basin suggests that it will likely perform well in other hydrographically simpler and less variable basins.

BGD
9, 14589–14638, 2012

Appendix A

Spatio-temporal variance patterns in the synthetic data set

Given that regression analysis aims to explain the dominant variance patterns in a data set, changes in the spatial and temporal patterns of variance can affect eMLR results by influencing regressio model selection. Variance variability can arise either from temporal variability or by altering the sampling grid, which acts by weighting certain regions differently in the data set. The structure and quality of linear regressions vary depending on whether the analyses are performed on sections, on regions, or on isopycnals such that the regression models used in the eMLR context are ad hoc. This section contrasts the spatial variance patterns captured by the GLODAP and CLIVAR sampling networks and discusses the seasonal to interannual changes of these spatial patterns in the synthetic data set used in this study.

Figure A1 shows the seasonal evolution of vertical profiles of the standard deviation in the synthetic North Atlantic GLODAP data set for year 1995 for 8 variables. The standard deviation is calculated horizontally and independently for each month and each model layer. A parallel analysis using the CLIVAR sampling grid shows similar broad-scale patterns, although with slightly different magnitudes owing to the different emphasis put on the Labrador Sea and the Eastern Tropical Atlantic between the two sampling networks. The variables exhibit different zones of low or high variance (Fig. A1), indicating a priori the role each tracer will take in the regression models as a function of depth and highlighting the value of each variable as a tracer for each water masses.

The seasonal evolution of variance profiles reflects the mechanisms of water mass formation, gas exchange and ecological succession in the basin. The magnitude of the seasonal cycle of the standard deviation is typically 10 to 15 % in the upper 200 m for the nutrients (O_2, AOU, NO_3, PO_4, SiO_4), and 5 % for θ, S, Alk and DIC. Seasonality is small below 200 m (< 1–2 %). Nutrients show large variances in late summer and fall in

BGD
9, 14589–14638, 2012

the top 150 m and relatively smaller standard deviations in winter and spring (Fig. A1), consistent with the development of the North Atlantic Bloom (Henson et al., 2009). Temperature shows a maximum variance in spring and summer when the subtropical-subpolar gradients are strongest. The variance of salinity is small in summer and is large in winter, reflecting sea-ice dynamics in the northern subpolar region. Seasonality of the variance is associated to a seasonal cycle in the misfit error of linear regression models and in the eMLR results.

Relative to the basin-scale horizontal variance in the data set, 1995 to 2005 variance changes in the vertical profiles are small. These changes are typically less than 3 % above 500 m and less than 1 % below that depth. These changes reflect processes such as water mass reorganization, gyre wobble, thermocline oscillation, frontal shifts, etc. Although the level of variance on horizontal slices in the data are relatively constant, this is not to say that point-by-point differences in tracer values or concentrations do not routinely exceed the standard deviation calcuated over the whole layer. In fact, point-by-point differences between July 1995 and 2005 for the North Atlantic can be as high as 50–100 % in specific regions (East Greenland Current, Labrador Sea, across the North Atlantic Current, near the equatorial boundary of the subtropical gyre). The relative constancy of the data set variance in time sampled from a constant observational network suggests that the point-by-point changes are not a priori systematic enough as to greatly bias the large-scale representativeness of a given sampling grid: the GLODAP or CLIVAR sets of stations would measure features in the same approximate proportions in 1995 and in 2005.

The GLODAP and CLIVAR sampling grids emphasize hydrographic features differently because of their variable spatial sampling densities (Fig. 1). Calculated differences between the basin-scale variances of the GLODAP and CLIVAR data sets show typical standard deviation differences of order 10 % between the two observational networks. These differences also exhibit vertical patterns clearly different from changes induced by natural variability or either seasonal or interannual time-scales (Fig. A2). Interannual variability and variations in the sampling grid both alter the data set variance

BGD
9, 14589–14638, 2012

patterns and affect misfit error but interannual variability is secondary to the variance changes imposed by changing sampling network between GLODAP 1995 and CLIVAR 2005.

Acknowledgements. This study was supported by the National Science Foundation under Grant No. OCE-0727170 and OCE-0327189 and by BP and Ford Motor Company through the Carbon Mitigation Initiative at Princeton University. We thank J. Dunne for providing model output. The contributions of KBR, ARJ, and RMK were supported through NASA award NNX09AI13G. Additionally, KBR was supported through NOAA awards NA17RJ2612 and NA08OAR4320752, which includes support through the NOAA Office of Climate Observations (OCO). RMK acknowledges additional support from NOAA awards NA08OAR4310820 and NA08OAR4320752.

References

Bates, N. R.: Interannual variability of oceanic CO_2 and biogeochemical properties in the Western North Atlantic subtropical gyre, Deep-Sea Res. II, 48, 1507–1528, 2001. 14591

Bates, N., Michaels, A., and Knap, A.: Seasonal and interannual variability of oceanic carbon dioxide species at the US JGOFS Bermuda Atlantic Time-series Study (BATS) site, Deep-Sea Res. II, 43, 347–383, 1996. 14601

Bender, M., Doney, S., Feely, R. A., Fung, I., Gruber, N., Harrison, D. E., Keeling, R., Moore, J. K., Sarmiento, J., Sarachik, E., Stephens, B. Takahashi, T., Tans, P., and Wanninkhof, R.: A large-scale CO_2 observing plan: in situ oceans and atmosphere (LSCOP), OAR Special Report NTIS: PB2003–00377, NOAA/OAR/PMEL, Seattle, WA, 2002. 14593, 14609, 14620

Brewer, P. G., Goyet, C., and Friederich, G.: Direct observation of the oceanic CO_2 increase revisited, PNAS, 94, 8308–8313, 1997. 14591

Broecker, W. S.: "NO", a conservative water-mass tracer, Earth Planet. Sci. Lett., 23, 100–107, 1974. 14613

Broecker, W. S. and Takahashi, T.: Sources and flow patterns of deep-ocean waters as deduced from potential temperature, salinity, and initial phosphate concentration, J. Geophys. Res.-Oceans, 90, 6925–6939, 1985. 14613

Burnham, K. and Anderson, D.: Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach, 2nd Edn., Springer-Verlag, New York, 1998. 14599

BGD
9, 14589–14638, 2012

Caldeira, K. and Duffy, P. B.: The role of the Southern Ocean in uptake and storage of anthropogenic carbon dioxide, Science, 287, 620–622, 2000. 14591
Corbiere, A., Metzl, N., Reverdin, G., Brunet, C., and Takahashi., T.: Interannual and decadal variability of the oceanic carbon sink in the North Atlantic subtropical gyre, Tellus, 59B, 168–178, 2007 14590
Cromwell, D.: Temporal and spatial characteristics of sea surface height variability in the North Atlantic Ocean, Ocean Sci., 2, 147–159, doi:10.5194/os-2-147-2006, 2006. 14603
Curry, R., McCartney, M. S., and Joyce, T. M.: Oceanic transport of subpolar climate signals to mid-depth subtropical waters, Nature, 391, 575–577, 1998. 14603
Dunne, J. P., Armstrong, R. A., Gnanadesikan, A., and Sarmiento, J. L.: Empirical and mechanistic models for the particle export ratio, Global Biogeochem. Cy., 19, GB4026, doi:10.1029/2004GB002390, 2005. 14597
Dunne, J., Sarmiento, J., and Gnanadesikan, A.: A synthesis of global particle export fromthe surface ocean and cycling through the ocean interior and on the seafloor, Global Biogeochem. Cy., 21, GB4006, doi:10.1029/2006GB002907, 2007. 14597
Dunne, J., Gnanadesikan, A., and Sarmiento, J. L.: Coupling between the C, N, P, Fe, Si, and Ca, and lithogenic cycles in a global biogeochemical and ecological model, Ocean Sciences Meeting, Orlando, Abstract ID:1166, 2–7 March 2008. 14597
Dunne, J. P., Gnanadesikan, A., Sarmiento, J. L., and Slater, R. D.: John P. Technical description of the prototype version (v0) of Tracers Of Phytoplankton with Allometric Zooplankton (TOPAZ) ocean biogeochemical model as used in the Princeton IFMIP* model†, Version-4.1 10 November 2010, 22 pp., 15 September 2010. 14597, 14598
Friis, K., Kortzinger, A. J. P., and Wallace, D. W. R.: On the temporal increase of anthropogenic CO_2 in the subpolar North Atlantic, Deep-Sea Res. I, 52, 681–698, 2005. 14592, 14594, 14595, 14613, 14614
Gnanadesikan, A., Dixon, K. W., Griffies, S. M., Balaji, V., Barreiro, M., Beesley, J. A., Cooke, W., Delworth, T. L., Gerdes, R., Harrison, M. J., Held, I. M., Hurlin, W. J., Lee, H.-C., Liang, Z., Nong, G., Pacanowski, R. C., Rosati, A., Russell, J., Samuels, B. L., Song, Q., Spelman, M. J., Stouffer, R. J., Sweeney, C. O., Vecchi, G., Winton, M., Wittenberg, A. T., Zeng, F., Zhang, R., and Dunne, J. P.: GFDL's CM2 global coupled climate models, Part II: The baseline ocean simulation, J. Climate, 19, 675–697, 2006. 14597

BGD
9, 14589–14638, 2012

Goodkin, N., Levine, N., Doney, S., and Wanninkhof, R.: Impacts of temporal CO_2 and climate trends on the detection of ocean anthropogenic CO_2 accumulation, Global Biogeochem. Cy., 25, GB3023, doi:10.1029/2010GB004009, 2011. 14592

Griffies, S. M., Harrison, M., Pacanowski, R. C., and Rosati, A.: A technical guide to MOM4, Technical report, GFDL Ocean Group, Princeton, 2004. 14597

Griffies, S. M., Gnanadesikan, A., Dixon, K. W., Dunne, J. P., Gerdes, R., Harrison, M. J., Rosati, A., Russell, J. L., Samuels, B. L., Spelman, M. J., Winton, M., and Zhang, R.: Formulation of an ocean model for global climate simulations, Ocean Sci., 1, 45–79, doi:10.5194/os-1-45-2005, 2005. 14597

Griffies, S., Biastoch, A., Boening, C., Bryan, F., Danabasoglu, G., Chassignet, E. P., England, M. H., Gerdes, R., Haak, H., Hallberg, R. W., Hazeleger, W., Jungclaus, J., Large, W. G., Madec, G., Pirani, A., Samuels, B. L., Scheinert, M., Sen Gupta, A., Severijns, C. A., Simmons, H. L., Treguier, A. M., Winton, M., Yeager, S. and Yin, J.: Coordinated ocean-ice reference experiments (COREs), Ocean Model., 26, 1–46, 2009. 14598

Gruber, N., Gloor, M., Mikaloff Fletcher, S. E., Doney, S. C., Dutkiewicz, S., Follows, M. J., Germer, M. Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Mueller, S. A., Sarmiento, J. L., and Takahashi, T.: Oceanic sources, sinks, and transport of atmospheric CO_2, Global Biogeochem. Cy, 23, GB1005, doi:10.1029/2008GB003349, 2009. 14591

Hartigan, J. and Wong, M.: A K-means clustering algorithm, Appl. Stat., 28, 100–108, 1979.

Henson, S., Dunne, J., and Sarmiento, J.: Decadal variability in North Atlantic phytoplankton blooms, J. Geophys. Res., 114, C04013, doi:10.1029/2008JC005139, 2009. 14598, 14622

Henson, S. A., Sarmiento, J. L., Dunne, J. P., Bopp, L., Lima, I., Doney, S. C., John, J., and Beaulieu, C.: Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity, Biogeosciences, 7, 621–640, doi:10.5194/bg-7-621-2010, 2010. 14598

Keeling, R. F.: Comment on "The ocean sink for anthropogenic CO_2", Science, 308, 5729, doi:10.1126/science.1109620, 2005. 14591

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T.-H.: A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP), Global Biogeochem. Cy., 18, GB4031, doi:10.1029/2004GB002247, 2004. 14597, 14600, 14614

Khatiwala, S., Primeau, F., and Hall, T.: Reconstruction of the history of anthropogenic CO_2 concentrations in the ocean, Nature, 462, 346–350, 2009. 14590, 14591, 14620

BGD
9, 14589–14638, 2012

Large, W. G. and Yeager, S. G.: Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies, Tech. Rep. TN-460+STR, National Center for Atmospheric Research, Boulder, Colorado, 2004. 14598

Large, W. G. and Yeager, S. G.: The global climatology of an internanually varying air-sea flux data set, Clim. Dynam., 33, 341–364, 2009. 14598

Lee, K., Choi, S. D., Park, G. H., Wanninkhof R., Peng, T.-H., Key, R. M., Sabine, C. L., Feely, R. A., Bullister, J. L., Millero, F. J., and Kozyr, A.: An updated anthropogenic CO_2 inventory in the Atlantic ocean, Global Biogeochem. Cy., 17, 1116, doi:10.1029/2003GB002067, 2003. 14601

Le Quéré, C., Rodenbeck, C, Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramoney, M., Nakazawa, T., Metzl, N., Gillett, N., and Heinmann, M.: Saturation of the Southern Ocean CO_2 sink due to recent climate change, Science, 316, 1735–1738, 2007. 14590, 14591

Levine, N., Doney, S. C., Wanninkhof, R., Lindsay, K., and Fung, I.: Impact of ocean carbon system variability on the detection of temporal increases in anthropogenic CO_2, J. Geophys. Res., 113, C03019, doi:10.1029/2007JC004153, 2008. 14592, 14613, 14614

Lherminier, P., Mercier, H., Gourcuff, C., Alvarez, M., Bacon, S., and Kermabon, C: Transport across the 2002 Greeland-Portugal Ovide section and comparsion with 1997, J. Geophys. Res., 112, C07003, doi:10.1029/2006JC003716, 2007. 14618

Lo Monaco, C., Metzl, N., Poisson, A., Brunet, C., and Shauer, B.: Anthropogenic CO_2 in the Southern Ocean: distribution and inventory at the Indian-Atlantic boundary (world ocean circulation experiment line I6), J. Geophys. Res., 110, C06010, doi:10.1029/2004JC002643, 2005a. 14591

Lo Monaco, C., Goyet, C., Metzl, N., Poisson, A., and Touratier, F.: Distribution and inventory of anthropogenic CO_2 in the Southern Ocean: comparison of three data-based methods, J. Geophys. Res., 110, C09S02, doi:10.1029/2004JC002571, 2005b. 14591

Matear, R. and McNeil, B.: Decadal accumulation of anthropogenic CO_2 in the Southern Ocean: a comparison of CFC-age derived estimates to multiple-linear regression estimates, Global Biogeochem. Cy., 17, 1113, doi:10.1029/2003GB002089, 2003. 14596

McKinley, G., Fay, A. R., Takahashi, T., and Metzl, N.: Convergence of atmostpheric and North Atlantic carbon dioxide trends on multidecadal timescales, Nat. Geosci., 4, 606–610, 2011. 14591

BGD
9, 14589–14638, 2012

Mikaloff-Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Mueller, S. A., and Sarmiento, J. L.: Inverse estimates of anthropogenic CO_2 uptake, transport, and storage by the ocean, Global Biogeochem. Cy., 20, GB2002, doi:10.1029/2005GB002530, 2006. 14591

Rio, M.-H., Guinehut, S., and Larnicol, G.: New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry and in situ measurements, J. Geophys. Res., 116, C07018, doi:10.1029/2010JC006505, 2011.

Rodgers, K. B., Key, R., Gnanadesikan, A., Sarmiento, J. L., Aumont, O., Bopp, L., Doney, S. C., Dunne, J. P., Glover, D. M., Ishida, A., Ishii, M., Jacobson, A. R., Lo Monaco, C., Maier-Reimer, E., Mercier, H., Metzl, N., Perez, F. F., Rios A. F., Wawnninkhof R., Wetzel, P., Winn, C., and Yamanaka, Y.: Using altimetry to help explain patchy changes in hydrographic carbon measurments, J. Geophys. Res., 114, C09013, doi:10.1029/2008JC005183, 2009. 14598

Sabine, C., Feely, R., Millero, F., Dickson, A. G., Langdon, C., Mecking, S., and Greeley, D.: The oceanic sink for anthropogenic CO_2, Science, 305, 367–371, 2004. 14590, 14591

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T.-H., Kozyr, A., Ono, T., and Rios, A. F.: Decadal changes in Pacific carbon, J. Geophys. Res., 113, C07021, doi:10.1029/2007JC004577, 2008. 14591, 14592

Sarmiento, J. L., Slater, R. D., Dunne, J., Gnanadesikan, A., and Hiscock, M. R.: Efficiency of small scale carbon mitigation by patch iron fertilization, Biogeosciences, 7, 3593–3624, doi:10.5194/bg-7-3593-2010, 2010. 14598

Schuster, U. and Watson, A. J.: A variable and decreasing sink for atmospheric CO_2 in the North Atlantic, J. Geophys. Res., 112, C11006, doi:10.1029/2006JC003941, 2007. 14590

Steinfeldt, R., Rhein, M., Bullister, J., and Tanhua, T.: Inventory changes in anthropogenic carbon from 1997–2003 in the Atlantic Ocean between 20S and 65N, Global Biogeochem. Cy., 23, GB3010, doi:10.1029/2008GB003311, 2009. 14609

Takahashi, T, Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Reely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, Deep-Sea Res. II, 49, 1601–1622, 2002. 14591

Takahashi, T, Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Koertzinger, A., Steinhoff,

BGD
9, 14589–14638, 2012

T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans, Deep-Sea Res. II, 56, 554–577, 2009. 14591

Tanhua, T., Koertzinger, A., Friis, K., Waugh, D. W., and Wallace D. W. R.: An estimate of anthropogenic CO_2 inventory from decadal changes in oceanic carbon content, PNAS, 104, 3037–3042, 2007. 14592, 14596, 14614

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, SIAM, Philadelphia, PA, 2005. 14594, 14596

Wallace, D. W. R.: Monitoring Global Ocean Carbon Inventories, Tech. rep., Ocean Observing System Development Panel, Texas A+M University, College Station, TX, 1995. 14592

Wanninkhof, R., Doney, S. C., Bullister, J., Levine, N. M., Warner, M., and Gruber, N.: Detecting anthropogenic CO_2 changes in the interior Atlantic ocean between 1989–2005, J. Geophys. Res., 115, C11028, doi:10.1029/2010JC006251, 2010. 14592

Waugh, D. W. and Hall, T. M. and McNeil, B. I., Key, R. M., and Matear, R. J.: Anthropogenic CO_2 in the oceans estimated using transit time distributions, Tellus, 58B, 376–389, 2006. 14591

Wetzel, P., Winguth, A., and Maier-Reimer, E.: Sea-to-air CO_2 flux from 1948 to 2003: a model study, Global Biogeochem. Cy., 19, GB2005, doi:10.1029/2004GB002339, 2005. 14591

Winn, C. D., Li, Y. H., Mackenzie, F. T., and Karl, D. M.: Rising surface ocean dissolved inorganic carbon at the Hawaii ocean time-series site, Mar. Chem., 60, 33–47, 1998. 14591

Winton, M.: A reformulated three-layer sea ice model, J. Atmos. Ocean. Tech., 17, 525–531, 2000. 14597

BGD
9, 14589–14638, 2012

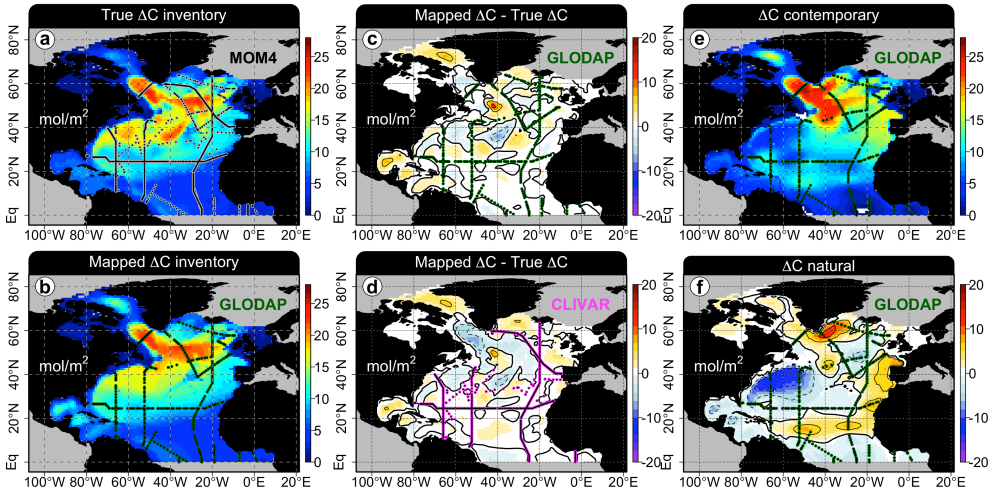

Fig. 1. (a) Change in anthropogenic carbon column inventory, in mol m^{-2}, between July 1995 and 2005 calculated on the original MOM4/TOPAZ grid. **(b)** Inventory change calcualting after mapping the true values sampled at GLODAP stations. **(c)** Mapping error, difference between panels **(b)** and **(a)** for GLODAP. **(d)** Mapping error for CLIVAR. **(e)** Changes in contemporary and **(f)** natural carbon column inventories between July 1995 and 2005 mapped from GLODAP stations. Station locations are show in green (GLODAP) magenta (CLIVAR). Both GLODAP and CLIVAR stations are plotted in **(a)**. In **(c)**, **(d)** and **(f)**, dashed (negative) and solid (positive) contour lines are drawn in increment of 5 mol m^{-2}. Thick contours mark 0 mol m^{-2}.

BGD
9, 14589–14638, 2012

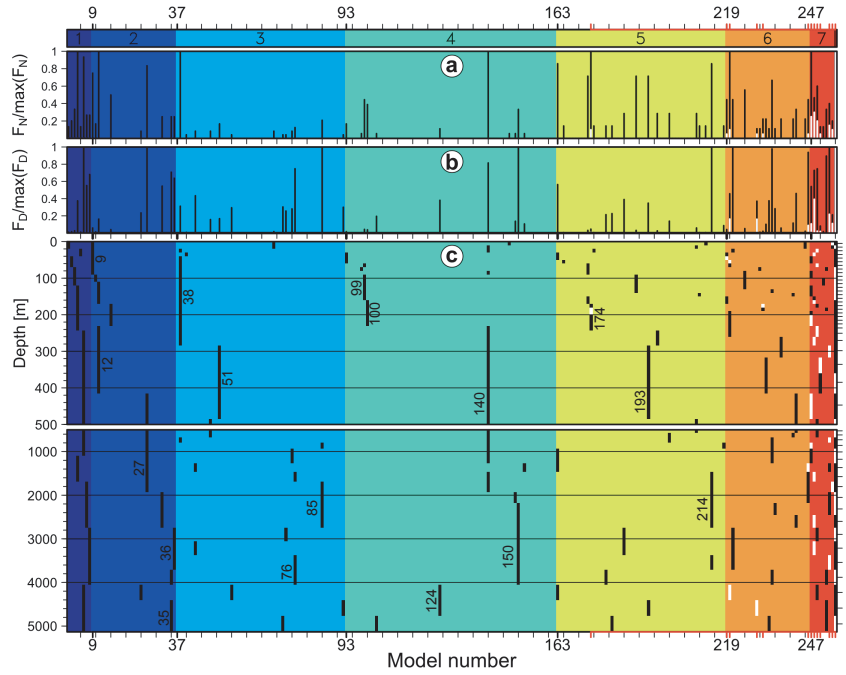

Fig. 2. Summary of the best fitting linear models for the July 1995 GLODAP synoptic synthetic data set. Background colors identify models size classes (1 to 8). **(a)** Relative frequency (F_N/max(F_N)) with which models are selected in each size class (minimum root-mean-square error, black bars) and overall (minimum AIC, white bars). Frequency is computed based on the number of model layers (F_N) normalized to the most frequently identified model (max(F_N)). **(b)** Same as **(a)** but for frequency weighted by the thickness of each layer (F_D/max(F_D)). **(c)** Models with with lowest AIC in each size class (black bars) and overall (white bars) and each depth layer. Tick marks on the right show boundary between model layers. Tick marks on top and bottom show model number (in steps of 5). The first model number of each size class is indicated, except for size classes 1 and 8 (number 1 and 255). Red ticks on the top and bottom identify models with minimum AIC.

BGD
9, 14589–14638, 2012

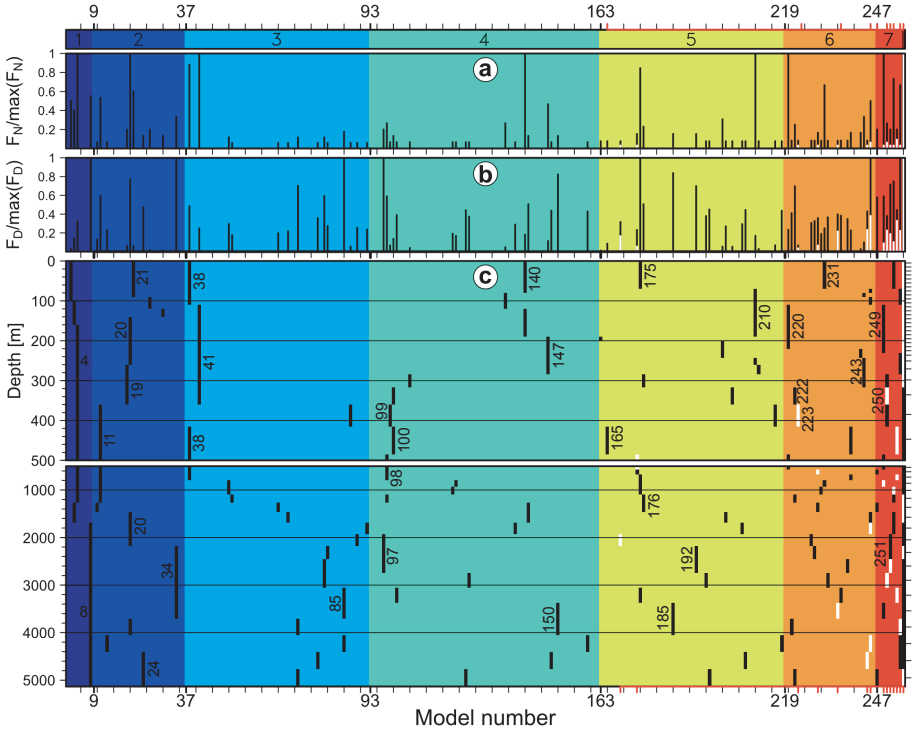

Fig. 3. Same as for Fig. 2 but using the July 2005 CLIVAR synoptic synthetic data set.

BGD
9, 14589–14638, 2012

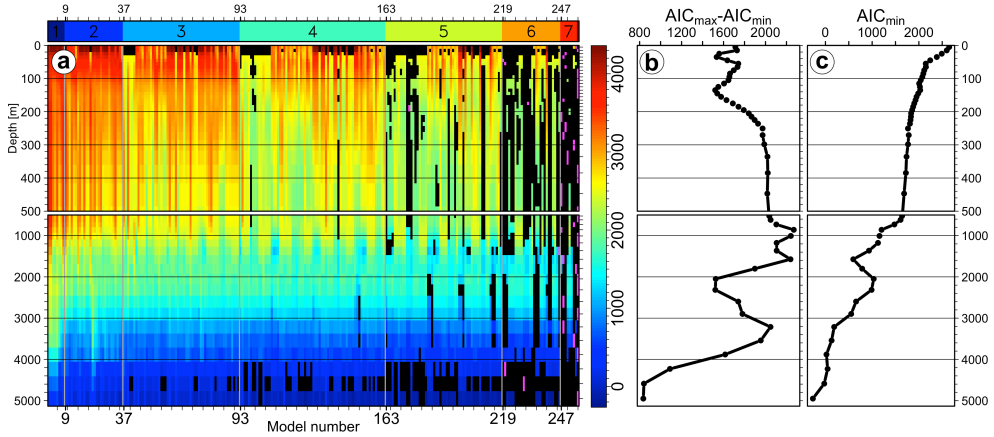

Fig. 4. (a) AIC values as a function of model number (strategy 2) and depth for the July GLODAP 1995 data set. All models with AIC values within 10 % of the depth-specific range in AIC of the minimum AIC value at each depth (highlighted in magenta) are highlighted in black. Tick marks on the right show the vertical location of model layers. Corresponding vertical profiles of **(b)** the depth-specific range in AIC and **(c)** the minimum AIC values.

BGD
9, 14589–14638, 2012

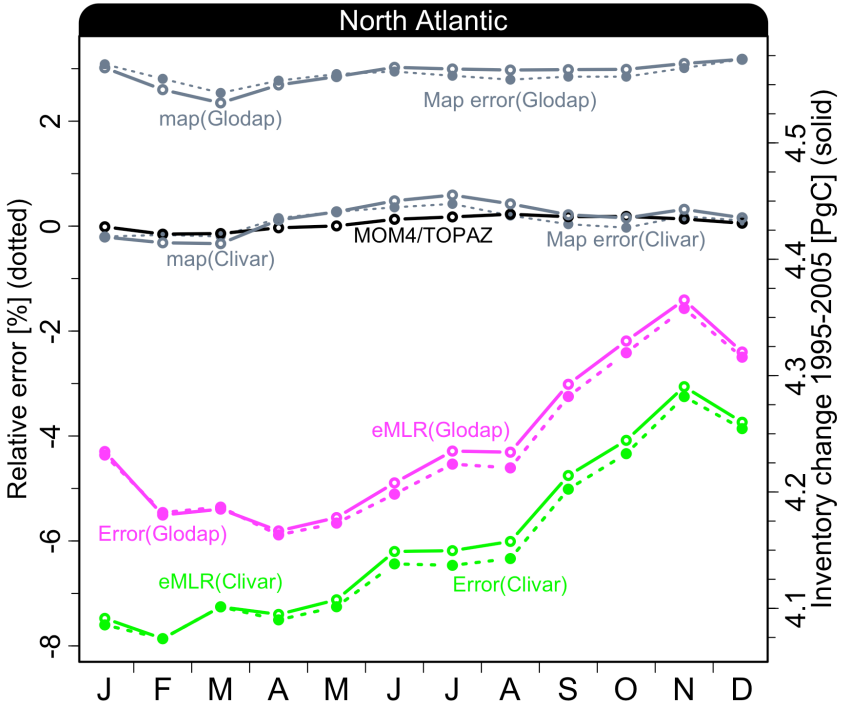

Fig. 5. Month-by-month evolution of the anthropogenic carbon inventory change (right y-axis, solid lines) and associated relative errors (left y-axis, dotted lines) between 1995 and 2005 computed on the original model grid (black), after mapping the "true" values sampled at the GLODAP or CLIVAR stations (grey) and after mapping the composite best-AIC eMLR solutions projected onto either the GLODAP (magenta) or CLIVAR (green) observational networks.

BGD
9, 14589–14638, 2012

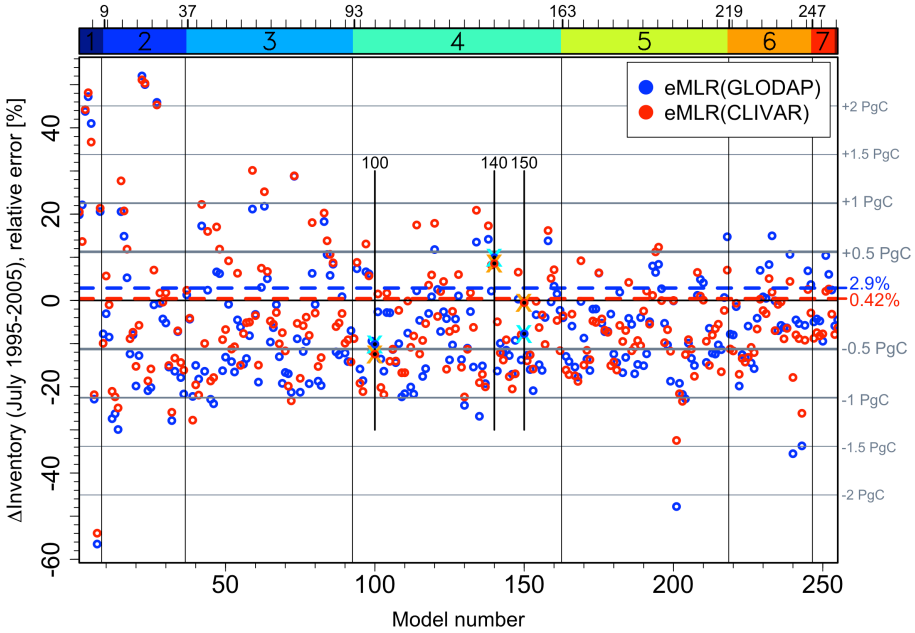

Fig. 6. Relative (left y-axis) and absolute (right y-axis) errors in the determination of the change in the North Atlantic anthropogenic carbon inventory, projected either onto the GLODAP (blue) or the CLIVAR (red) stations, by strategy 2 (constant model structure for all layers) between July 1995 and July 2005 for each of the 255 first order linear models. Model size is indicated by the color strip on top. Mapping errors calculated from "true" values are show as the horizontal dahsed blue (2.9 %, GLODAP) and red (0.42 %, CLIVAR) lines.

BGD
9, 14589–14638, 2012

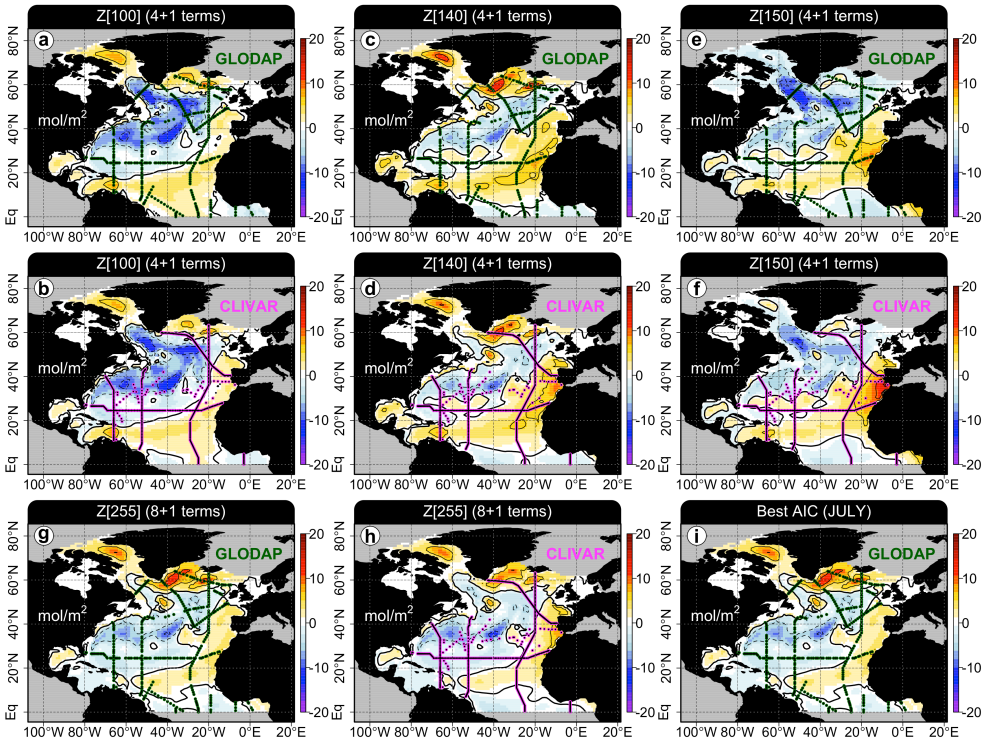

Fig. 7. Absolute errors in calculated eMLR anthropogenic carbon column inventory change between July 2005 and 1995 mapped from either the GLODAP (green) or CLIVAR (magenta) stations. Dashed (negative) and solid (positive) contours are drawn in increment of 5 mol m^{-2}. The thick lines marks the 0 contour. Results shown for models **(a, b)** $Z_{100} = \{S, \theta, PO_4, O_2\}$, **(c, d)**: $Z_{140} = \{\theta, PO_4, Si, Alk\}$, **(e, f)**: $Z_{150} = \{NO_3, PO_4, Si, Alk\}$, **(g, h)** Z_{255} (= all 8 variables), using the same model structure at all layers and both times but with coefficients that vary in time and between layers (strategy 2) and **(i)** for the "best AIC" models selected by minimum AIC for each layer and each time point (strategy 1).

BGD
9, 14589–14638, 2012

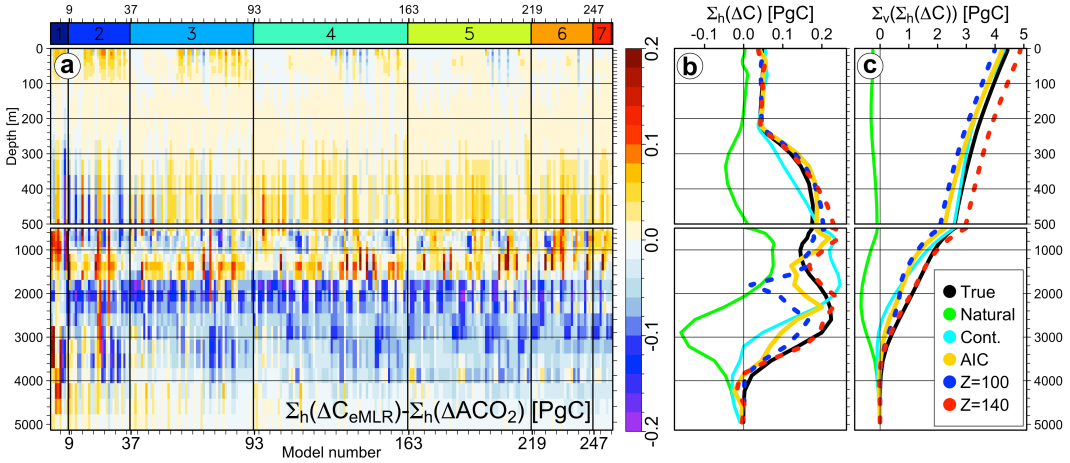

Fig. 8. (a) Absolute errors between the North Atlantic eMLR predicted inventory change, mapped from estimates at GLODAP stations, and the true inventory changes integrated on each horizontal model layer (Σ_h) and for each first order regression model (strategy 2). **(b)** Vertical profiles of the layer inventory changes and **(c)** vertically integrated layer inventory change (from the bottom to the surface, Σ_v). The true, natural and contemporary (Cont.) layer inventory changes between July 1995 and July 2005 are shown, together with the "best AIC" composite solution and results from models Z_{100} and Z_{140} (dotted).

BGD
9, 14589–14638, 2012

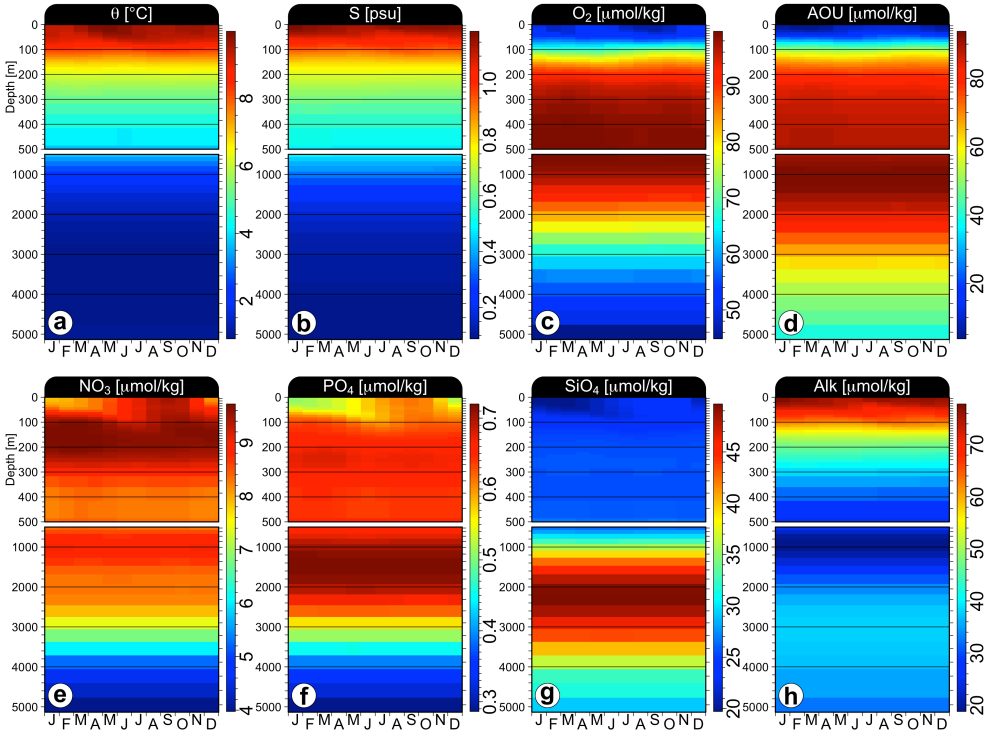

Fig. A1. Monthly vertical profiles of the horizontal spatial standard deviation, from January to December, expressed in **(a)** °C, **(b)** psu or **(c–h)** µmol kg^{-1}, for the hydrographic variables used in this study for the year 1995 as sampled on the GLODAP grid. Tick marks to the right of the main panels show the vertical position of the vertical layers in the circulation model.

BGD
9, 14589–14638, 2012

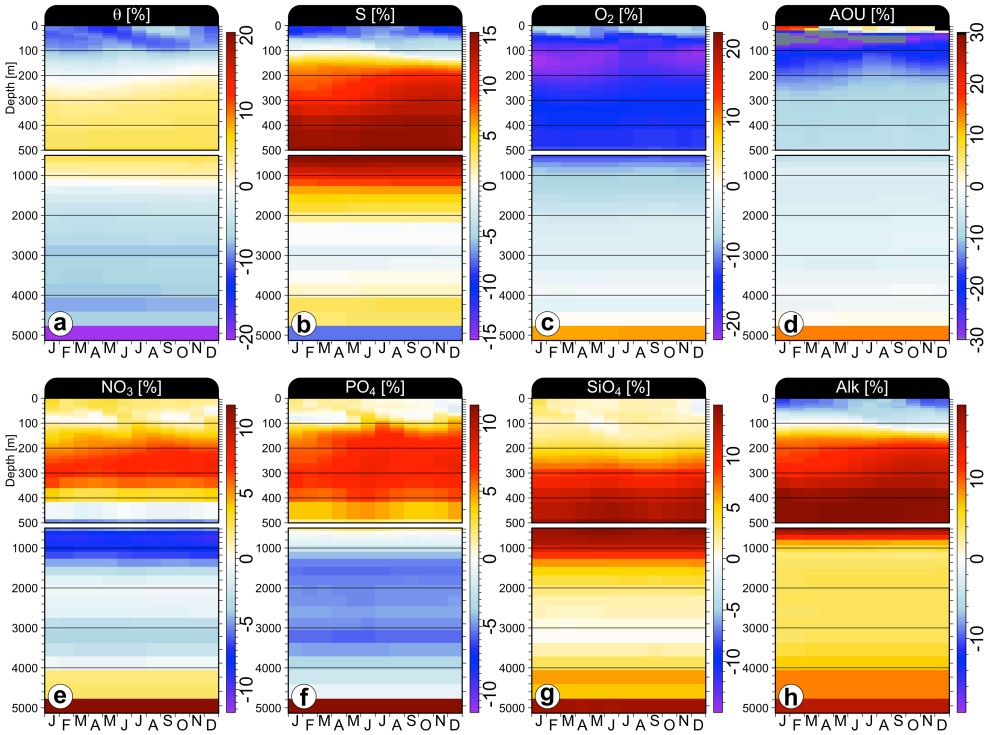

Fig. A2. Monthly vertical profile of the month-by-month relative changes of the horizontal spatial standard deviation, from January to December, expressed in percent relative to the 1995 values, between the synthetic 2005 CLIVAR data set and the 1995 GLODAP data set for the variables used in this study. Tick marks to the right of the main panels show the vertical position of the layers in the circulation model.

BGD
9, 14589–14638, 2012