

Total alkalinity estimation using MLR and neural network techniques

Velo, A.¹, Pérez, F.F.¹, Tanhua, T.², Gilcoto, M.¹, Ríos, A.F.¹, Key, R.M.³

¹ Instituto de Investigaciones Mariñas (CSIC). Eduardo Cabello 6, 36208 Vigo, Spain.

² GEOMAR Helmholtz Centre for Ocean research Kiel, Marine Biogeochemie, Kiel, Germany

³ Princeton University, Program in Atmospheric and Oceanic Science, Forrestal Campus/Sayre Hall, Princeton, NJ 08544, USA

Correspondence e-mail: avelo@iim.csic.es Tlf: +34986231930

Abstract

During the last decade, two important collections of carbon relevant hydrochemical data have become available: GLODAP and CARINA. These collections comprise a synthesis of bottle data for all ocean depths from many cruises collected over several decades. For a majority of the cruises at least two carbon parameters were measured. However, for a large number of stations, samples or even cruises, the carbonate system is under-determined (i.e., only one or no carbonate parameter was measured) resulting in data gaps for the carbonate system in these collections. A method for filling these gaps would be very useful, as it would help with estimations of the anthropogenic carbon (C_{ant}) content or quantification of oceanic acidification. The aim of this work is to apply and describe, a 3D moving window multilinear regression algorithm (MLR) to fill gaps in total alkalinity (A_T) of the CARINA and GLODAP data collections for the Atlantic. In addition to filling data gaps, the estimated A_T values derived from the MLR are useful in quality control of the measurements of the carbonate system, as they can aid in the identification of outliers. For comparison, a neural network algorithm able to perform non-linear predictions was also designed. The goal here was to design an alternative approach to accomplish the same task of filling A_T gaps. Both methods return internally consistent results, thereby giving confidence in our approach.

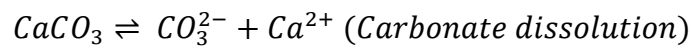
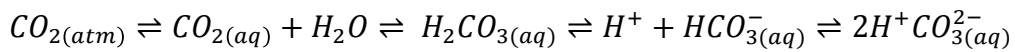
Highlights: > Estimation of alkalinity by multilinear regression (MLR) techniques. > Estimation of alkalinity by neural network techniques. > Intercomparison between alkalinity prediction techniques. > Use of alkalinity for carbon calculations. > Use of alkalinity for quality control of measurements.

Keywords: Seawater Alkalinity; Ocean carbonate system; MLR techniques; Neural Network techniques; Data Quality Control

1. Introduction

GLODAP (Key et al., 2004; Sabine et al., 2005) and CARINA (Tanhua et al., 2009; Key et al., 2010) are data collections of ocean bottle samples from the synthesis of many oceanic cruises. GLODAP provides reasonable global coverage except for the Arctic. CARINA focuses on the Atlantic with particular emphasis on the North Atlantic and Arctic Ocean. Both collections are the result of international efforts to summarise all available data into compiled products.

Total dissolved inorganic carbon (C_T), total alkalinity (A_T), total hydrogen ion concentration (pH) and fugacity of CO_2 are the four measured parameters of the carbonate system in seawater. A large fraction of the samples in these databases has at least one carbon parameter measured and their merging results in the distribution of measured parameters, as shown in Fig 1. The seawater carbonate system is governed by the thermodynamics of the carbonate system through the following equation (*¡Error! No se encuentra el origen de la referencia.*) (Millero, 1995; Zeebe and Wolf-Gladrow, 2001; Dickson, 2010):



The system is defined by having any two of the four possible parameters (Millero, 1995; Lewis et al., 1998; van Heuven et al., 2009).

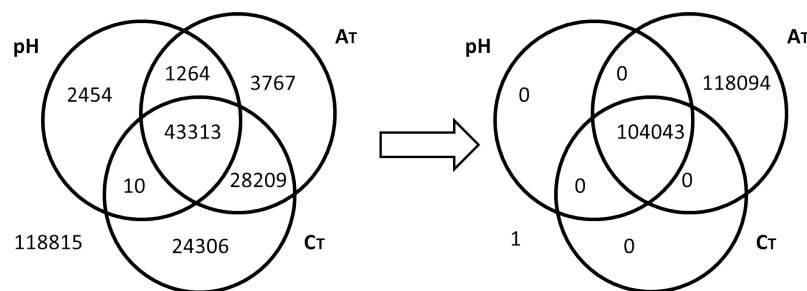


Fig 1: Venn diagram with carbon data availability before and after MLR calculations; the numbers within the different fields are the number of samples that have been assigned with a value for each of the parameters pH, A_T, C_T, a combination of them, or no carbonate system data (lower left). Left panel: original data from the CARINA/GLODAP Atlantic merge; right panel: after adding the additional data obtained in this study.

Among the carbon system parameters, A_T can be estimated most accurately due to its low variability in the ocean (Wolf-Gladrow et al., 2007; Zeebe and Wolf-Gladrow, 2001). A_T is supposed to be only marginally affected by increasing C_T concentration, at

least at the current decadal scale, because the carbonate system tends to buffer the alkalinity when dissolving CO₂ in contemporary seawater (Ilyina et al., 2009; Riebesell et al., 2010). Most of the A_T variability is known and is well correlated with salinity, silicate and even temperature (Wallace, 1995; Millero et al., 1998; Lee et al., 2006). This is advantageous for this work, as the full dataset of available A_T data from old cruises performed in the 1970s to the most recent can be used directly as a climatology for the multilinear regression (MLR) computations. There is no need for partitioning the dataset in a time series or for introducing the time parameter in estimations, which would result in fewer available data, or less precision for estimations. However, alkalinity measurements performed without using certified reference materials (CRMs) have proven to be less accurate, which is true for most data obtained prior to the mid-1990s.

GLODAP bottle cruise data for the Atlantic region were used together with CARINA bottle cruise data for the Atlantic and Southern Ocean regions. Both data collections (i.e., the data products with adjustments applied to certain parameters, see Sabine et al. (2005) for the GLODAP data and Tanhua et al. (2010) for the CARINA data) were merged and outlier filtering was performed to prepare the dataset. The existence of depth, temperature and salinity data for each sample was required and so, samples without these variables were removed. The absence or bad data of nitrate and phosphate samples were recovered by using the same multilinear regression presented here for alkalinity, once the algorithm was fine-tuned. Multiparametric spatial/WMP interpolation techniques (Velo et al., 2010) with the World Ocean Atlas 2009 (WOA09) (Garcia et al., 2009) were used to recover silicate samples where no data were available. The Results and Discussion section (2.1) provides details on these procedures.

Two very different techniques are being proposed here to infer the A_T values. Firstly, a customised 3D moving window MLR method (hereinafter named 3DwMLR), fine-tuned and focused for this purpose; and secondly, a customised neural network, also parameterised and fine-tuned for this task. Both alternatives provide fundamentally different approaches to the same problem and thus, a comparison of their results was used as a validation tool for the analysis.

1. Methodology

1.1. The 3D Moving Window MLR (3DwMLR) estimation

The classical multilinear regression (MLR, or LM) scheme allows modelling of the relationship between a response variable and some predictors by using a linear equation combining these predictors. The problem with using a classical MLR applied

to the full CARINA and GLODAP dataset is that not all A_T natural variability can be modelled this way and large errors may result. The A_T relationship with salinity and temperature could be assumed to be linear, but only when data are restricted to specific confined areas in basins and mainly for surface waters (Millero et al., 1998; Lee et al., 2006).

The objective of this work is to enhance the results of the classical MLR by combining the following techniques:

1. Use of a 3D moving window to define a data box around the node for which we are trying to calculate A_T and to perform the regression using only the data inside that box.
2. Use a robust regression algorithm instead of ordinary least squares regression.
3. Include other parameters in the regression in addition to salinity and temperature (Millero et al., 1998; Lee et al., 2006), which have influence, directly or indirectly, on the natural variability of A_T ; these can be silicate, nitrate, phosphate and oxygen.

1.1.1. 3D moving window

An algorithm was designed to extract a pool of data from a box around each bottle sample. This box is referred to as the 3D moving window as its central node changes for and is centred on each sample. All samples of the merged GLODAP-CARINA dataset were used and all of them have their specific data window in order to make the MLR calculations. The width and depth of the data window were chosen according to the previous procedures of Velo et al. (2010) and multiple comparisons of different parameterisations were done to improve the results.

Summarising them; variable-width and fixed-width moving windows have been tried, both with the requirement of a minimum number of samples inside the window to perform the MLR. The variable-width tests started with a radius of 2 latitudinal degrees around the node, allowing the algorithm to increase the size degree by degree until reaching 100 samples around the node. For depth, a direct depth-dependent equation of $height = 150 + 0.15 * [sample\ depth]$ around the node was used, allowing the algorithm to grow 5% in depth for each degree of latitudinal increase. Fixed-width tests used a wider area of 5, 10 or even 20 latitudinal degrees around the node and the same equation for depth.

Variable-width parameterisation more closely mimics the nearby sample values, but is affected more by the scarcity of data. However, fixed-width settings tend to smooth the natural variability or even add noise if the window is too wide.

A mostly-fixed moving window of 10° in latitude has proven to be the most consistent parameterisation for A_T calculations with the CARINA-GLODAP merge for the Atlantic Ocean, compared with observations. The term mostly-fixed means that the algorithm was allowed to increment the size of its moving window only when less than 100 neighbouring samples were found in the box, which only happened for 5,324 of a total 222,136 estimated A_T values (2.4% of samples). The increments were made in steps of 1° in latitude and 5% in depth, up to a maximum window of 20° (+63% in depth). By using this combination of settings, the A_T dataset gets the most consistent result among the presented options, allowing for the recovery of large amounts of data.

1.1.2. Robust regression

A robust regression algorithm (Filzmoser et al., 2007; Hubert, 2008) was used as the MLR technique. The robust regression combines an ordinary regression algorithm with outlier detection by performing iterations to re-weight predictors. This technique is also known as interactive reweighting least squares (IRLS) (Green, 1984). Re-weighting is performed using a bi-square weighting function over residuals. Thus, the inferred A_T is more consistent compared with observations than when an ordinary least-squares regression is used.

In ordinary least-squares (OLS) regression it is a critically important task to keep input values under strict control, as the OLS regression is very sensible to outliers. By using a robust regression technique, the weighting function is able to detect most of these outliers by reiterating and thus, discarding them or giving them a minimum impact on the regression. A bi-square function with a cut-off value of 4.685/residuals (Holland and Welsch, 1977) was used for weighting in each iteration step. Weights decay quickly as residuals deviate from zero and the points are discarded (weight 0) when residuals exceed the cut-off value. The robust regression technique remains statistically robust with up to 50% of outliers present in the dataset, whereas for the OLS regression, the presence of one single outlier influences the result.

1.1.3. Selection of predictor parameters

As predictor parameters pressure, potential temperature (θ), salinity (S), nitrate, phosphate, silicate and oxygen were used. The main objective of the addition of these parameters was to incorporate as much of the natural variability as possible in order to get better predictions for each regional A_T regression. These parameters have been chosen as they are related to alkalinity and their abundance is larger than the carbon parameters in the dataset. The relationship between A_T and both S and θ is well-known (Millero et al., 1998; Lee et al., 2006). The relationship of A_T with nutrients (nitrate, phosphate, silicate) is related to organic matter synthesis and re-

mineralisation that adds or removes these ions from dissolution and thus, affects A_T (Chen and Millero, 1979; Chen and Pytkowicz, 1979). Furthermore, it can be considered that the sum of the A_T , phosphate and nitrogen species remains constant during the re-mineralisation of organic matter (Brewer et al., 1975; Fraga and Álvarez-Salgado, 2005). Oxygen also influences A_T through its use in organic synthesis and re-mineralisation processes. The pressure relationship with A_T can also be related to carbonate formation and its dissolution at depth in the ocean. All these processes have been treated extensively in the literature (Zeebe and Wolf-Gladrow, 2001; Millero, 2007; Dickson, 2010).

The influence of these parameters on the A_T regression has also been analysed from a statistical point of view, to verify that the parameter selection is adequate and that all parameters provide information to the model. It should be noted that the ultimate purpose of this work is the prediction of A_T values with the most realistic fit, with less emphasis on the search for the relative significance of each predictor in A_T . The existence of correlations and colinearity among a number of predictors is well known by the authors, but from a prediction point of view, they should not be removed if their presence improves the prediction. Moreover, we are using a 3D moving window and the colinearities may change from window to window. In addition, the relative importance of the parameters may vary for each window fit resembling the oceanic variability. Therefore, we do not discard a priori any parameter; the MLR of each window will adapt the relative contribution for each parameter to the local conditions of the 3D window.

To assess the overall colinearity among predictors, we have performed two statistical analyses. First, variance inflation factors (VIF) were calculated for a multilinear regression with the full dataset, following the procedures of Juranek et al. (2009) and Kim et al. (2010). The results are presented in Table 1, where each row shows the statistics for each MLR model considering as predictors the parameter of that row and below. The VIF factor indicates colinearity for values greater than 5–10 (Kutner et al., 2004) and very high colinearity is found when nitrate and phosphate are used. It is well-known that statistically, nitrate correlates highly with phosphate (Redfield, 1934), but it is also known that the N:P relationship has biogeochemical sources of variability (Anderson and Sarmiento, 1994; Geider and La Roche, 2002). Thus, as both presented models make use of geographical information, the inclusion of both phosphate and nitrate can improve the predictions.

Second, Akaike Information Criterion (AIC) values were also calculated for a full model with all the indicated predictors and for MLR models excluding the highest VIF values. In Table 1, the AIC is presented as ΔAIC , i.e., the AIC for the MLR model of that row minus the AIC for the MLR with all (7) predictors (first table row). Lower ΔAIC values indicate better significance (Anderson and Burnham, 2002) of the model. R squared

values have also been calculated for each subset combination of predictors (Miller, 1984), indicating the best fit when all seven parameters are used.

Parameter	VIF(7)	VIF(6)	VIF(5)	VIF(4)	Nº par. used	ΔAIC	R^2
Nitrate	117.15				7	0.0	0.930
Phosphate	116.26	17.18			6	1539.1	0.929
Theta	14.41	10.88	5.54		5	9883.7	0.923
Oxygen	5.76	4.97	1.29	1.19	4	13249.6	0.921
Salinity	5.31	4.97	3.57	1.46	3	15282.8	0.919
Silicate	3.67	3.50	1.77	1.59	2	273075.5	0.039
Pressure	1.93	1.86	1.75	1.32	1	276923.6	0.002

Table 1 Statistical test on colinearity and relevance of predictors

Inside each 3D moving window, the distribution of predictor values around the node value was tested for each parameter. If this node value was statistically inside the data cloud, only samples with more than three times the standard deviation (STD) from the node value were discarded. Otherwise, if most of the cloud of samples around this node was deviated, the entire predictor was discarded for the MLR in that node, in order to avoid extrapolation biases in the predicted A_T . The algorithm was designed with freedom to remove up to two parameters, giving no predicted A_T if more than two parameters had to be discarded.

The relationship between the estimated A_T and the predictors has been assumed linear within each 3D window, i.e., the ranges of variation of A_T and predictors in the windows are small enough to allow linear approximations. However, the method to select predictors inside the 3D windows assures that the predictor node values, those used to calculate the interpolated A_T , are inside the ranges of the predictors in the 3D window.

1.2. Neural Network estimation

For comparison with the previous algorithm, a neural network was constructed and tested to predict A_T in the same area and with the same predictors, plus the geographical information. Two main steps were needed: 1) design, configure and implement an adequate network model and 2) tune the model parameters to optimise performance and results, i.e., what commonly is referred to as training. An overview of these procedures can be found in Sarle (1997).

1.2.1. Design

A feed-forward multilayer network with a hidden log-sigmoidal layer and a linear output was designed and configured (Gardner and Dorling, 1998). The advantage of such architecture is that the log-sigmoidal layer can learn the non-linear relationships between predictors and results, whereas the final linear output layer deals with linear variability (Bishop, 1995; Hagan et al., 1996). Each node in a neural network applies a transfer function from its inputs to its outputs by weights tuned in the training step. A schematic diagram is shown in Fig 2. The combination and superposition of results from all simple nodes and layers is the key for modelling very complex non-linear problems (Gardner and Dorling, 1998).

Latitude, longitude (in -180° to 180° scale) and depth were used as predictor parameters to inform the network about the sample position, while θ , S , phosphate, nitrate, silicate and oxygen informed about the seawater properties. Apart from sample position, this choice matches the parameters chosen for the 3DwMLR estimation.

There is no direct way to establish the optimal number of nodes in the hidden layer (Sarle, 1997); the typical evaluation procedure is to try different network sizes and evaluate their confidence with observations. Network models with a range from 2^1 to 2^9 hidden neurons (units) were tested. A network size of 64 hidden units proved to be the best compromise between underestimation and overfitting and thus, was deemed to be the optimal network size. The results for a number of different training techniques are presented in Fig 3.

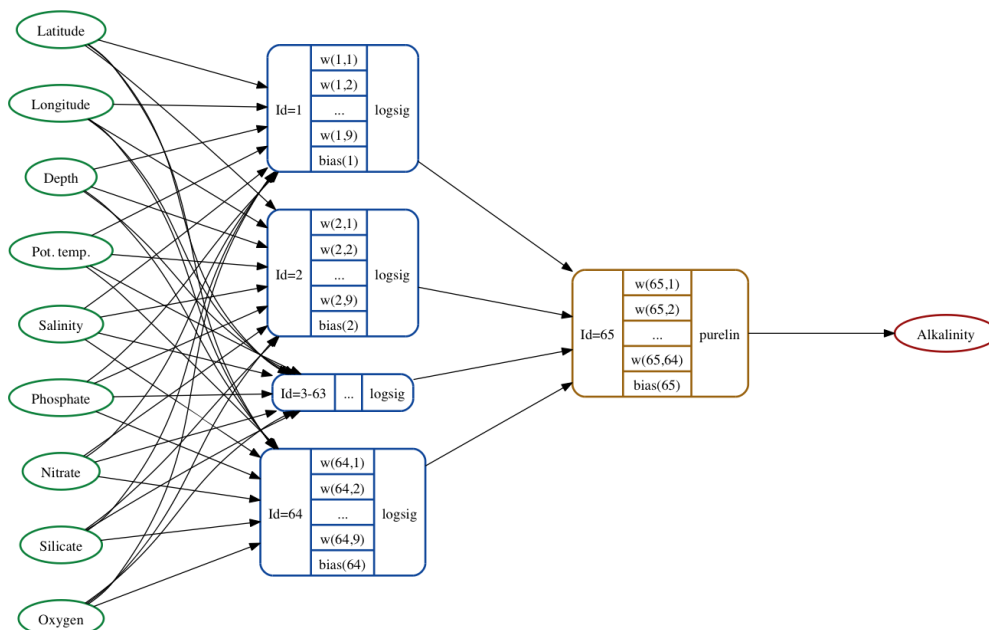


Fig 2: Schematic diagram of the feed-forward neural network designed for A_T prediction. Nodes with Id=1-64 represent the first log-sigmoidal hidden layer (logsig) and node with Id=65 represents the pure linear layer (purelin). Nodes 3-63 are represented by dots in the middle box. Values w_{i1-9} and w_{1-64} together with their bias are the weights established thorough training and that makes the network useful to predict.

1.2.2. Tuning

Artificial feed forward neural networks need training and they are commonly trained by back-propagation techniques (Hagan et al., 1996; Sarle, 1997). These techniques feed the network iteratively and in batch-mode with values of the training dataset and back-propagate the errors (deltas) of the results through the weights determined in the network. There are many training techniques in the literature, which differ on how the weights are modified by the back-propagation. Some of these training techniques were analysed and tested to tune the design of the network. Fig 3 shows the STD of the differences between the predicted A_T and the observations, plotted for the proposed different training techniques and against increasing network sizes (number of neurons in the hidden layer). As can be seen, the performance of techniques such as Resilient Backpropagation (trainRP), Scaled Conjugate Gradient (trainSCG) or Bayesian Regularisation (trainBR) quickly degrades as the size of the hidden layer exceeds a number of neurons (Hagan et al., 1996). The best training technique, according to observations, has proven to be the Levenberg-Marquardt (trainLM) method (Hagan and Menhaj, 1994), using a network size of 64 neurons in the hidden layer. Larger network sizes have revealed a tendency to overfit (Bishop, 1995; Hagan et al., 1996).

One of the main problems with neural networks is overfit. Neural networks are very powerful and by increasing the number of neurons in a non-linear layer, they can eventually model perfectly (producing null residuals). Therefore, care must be taken not to overfit the network or the model will also incorporate the measurement errors. This can be detected by testing or validating the network with data not used in the learning steps of the training process.

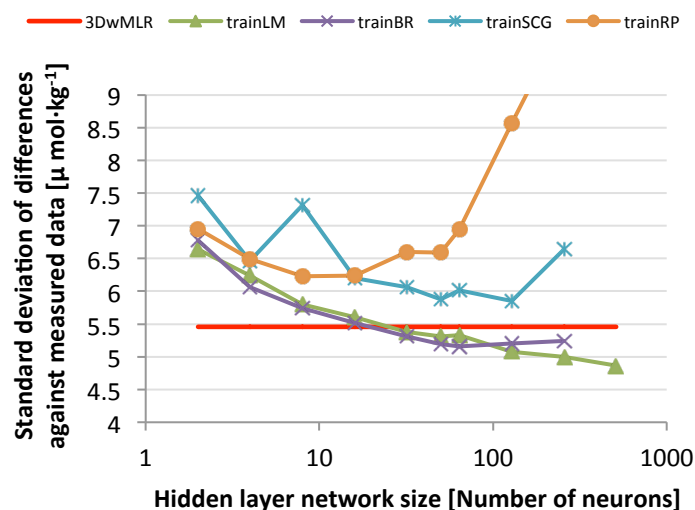


Fig 3: Comparison of some different neural network training techniques and MLR with the measured data

To train a neural network, some data are needed to validate the improvement in each learning iteration and other data for testing the results. Thus, the dataset has to be partitioned. The most consistent scheme consisted of using an arbitrary selection of data for the partitions and re-doing the arbitrary partitioning in each iteration. 80% of the dataset was used for learning procedures, 15% for validation and 5% for testing. However, by design, the Bayesian regularisation training technique always uses the whole dataset for learning.

Mean squared error was chosen as the convergence statistic in each iteration for all training methods, except Bayesian regularisation, that used the sum of squared errors. These analyses are designed to maximise the fit between learning and validation partitions by iteratively tuning the weights.

As the convergence approaches a minimum, the learning process stops and the neural network saves the weighting data information for use in predicting A_T from inputs (Fig 2).

2. Results and Discussion

2.1. Initial Quality Control of Input Data

As a starting point, the GLODAP and CARINA databases were merged, providing a total of 282,853 samples available for the Atlantic basin. From these, only 84,734 initially had A_T values (29.96%).

An initial check resulted in 18,063 samples being discarded as they lacked pressure or depth, θ or S data. Additionally, 32,910 samples without oxygen data were removed.

Measurements of nutrient parameters were missing for a large number of samples. In order to recover some of these data points we used the World Ocean Atlas (WOA09) dataset as follows:

Silicate: From 14,085 samples without silicate, 4,661 new values were incorporated using a combined Spatial/WMP interpolation (Velo et al., 2010) applied to the WOA09 gridded climatology. The climatology of silicate was interpolated to cruise sample nodes and θ , S , latitude, longitude and depth were used for the parametric distances in the interpolation algorithm. The remaining 9,424 samples without silicate were removed from the dataset.

Nitrate and phosphate: 3DwMLR technique was used to obtain nitrate or phosphate in samples where their counterpart phosphate or nitrate was available. For these MLR, the respective available phosphate or nitrate parameter was used as the predictor parameter and also oxygen, θ , S and distances were incorporated as input in the MLR. For samples without these nutrients, a 3DwMLR with only oxygen, θ , S and distances was used. As a result, 6,780 samples without nitrate and 15,220 samples without phosphate were filled and no data without nitrate or phosphate were left in the database.

Some further quality control was performed for the nitrate and phosphate dataset of measurements. The 3DwMLRs indicated above, were also applied to samples with measured data in order to compare predicted with measured data for these parameters. The difference between the measured data with the two MLR was taken as quality criterion for keeping or replacing the measured value. This check yielded 6,268 nitrate samples and 4,146 phosphate samples replaced by 3DwMLR values.

As a final quality control of input data, station profiles were manually supervised to identify outliers and were removed from the dataset for the A_T calculations. Because of this control, 4,425 sample values were removed.

2.2. Alkalinity estimation

After the initial checks and quality control, the dataset was reduced to 222,138 samples useful for calculations, with 76,553 samples having observations of A_T . Then, both the 3DwMLR and the neural network algorithm were used as stated in the Methodology section to predict A_T for the full dataset. Both algorithms were able to generate 222,136 estimated values of A_T , i.e., for all but two single samples.

2.3. Quality Control for observed A_T data

The estimated A_T data was used to perform a quality control of the observations, comparing them with the values predicted by the algorithm. The STD of the difference

between A_T observations and the predicted A_T by both methods was used as quality control criterion for the observations. The magnitude of this STD is a very good reference for the overall performance of the algorithms to predict the A_T data. An STD value of $5.4 \mu\text{mol kg}^{-1}$ was obtained for the 3DwMLR method and a value of $5.2 \mu\text{mol kg}^{-1}$ for the chosen neural network trained with the Levenberg-Marquardt method and a hidden layer of 64 units. This indicates good agreement between observations and predicted values. The value is slightly lower than the minimum adjustment applied to the CARINA dataset ($6 \mu\text{mol/kg}$) for the secondary quality control (QC2) procedures (Tanhua et al., 2010).

A boundary of $2 \times \text{STD}$ (i.e., $10.8 \mu\text{mol kg}^{-1}$ for 3DwMLR and $10.4 \mu\text{mol kg}^{-1}$ for the neural network) has been set as the confidence interval of observations to keep. This way 3,462 and 3,526 A_T observations (i.e., 4.5% and 4.6%), respectively fall outside this $2 \times \text{STD}$ limit and could then be flagged as questionable or alternatively, replaced by their 3DwMLR or neural network values.

2.4. Data results

A comparison of the data results of both methods has been graphically summarised in Fig 4 and Fig 5. The first column of both figures displays the difference between A_T obtained by neural network against the 3DwMLR technique. In this column, the full set of 222,136 predicted A_T values is compared. The scale is $\pm 15 \mu\text{mol kg}^{-1}$ and vertical dashed grey lines are placed at -12, -6, 0, 6 and 12 to allow an easy comparison with the ± 6 boundary of minimum adjustment established in CARINA. For each box in the boxplots, the height is proportional to the quantity of samples available between the boxes boundaries. Boxes have black lines but they were filled in red colour in order to emphasise the intervals of the plotted parameters (presented in left axis) where most of the data were located. The red colour filling is not even perceptible for intervals with less data.

2.4.1. Intercomparison between 3DwMLR and neural network

The top-left histogram of Fig 4 indicates a good level of agreement between both estimation techniques, giving a standard deviation of $2.36 \mu\text{mol}\cdot\text{kg}^{-1}$ for the residuals and a mean of $0.074 \mu\text{mol}\cdot\text{kg}^{-1}$. Thus, 95% of the residuals fall between $\pm 4.7 \mu\text{mol}\cdot\text{kg}^{-1}$ of A_T .

By analysing the box plots against some other parameters in the first column (Fig 4 and Fig 5), a good agreement can also be seen. Especially remarkable is the agreement in the estimations of both methods for the same cruises (Cruise ID subplot), indicating that both approaches give similar mean results for each cruise, even when both methods have some disagreements with the measured data, as can be seen in the

second and third columns of the CruiseID subplot in Fig 4. This can be interpreted as an indication of data bias for these cruises. Conversely, poor agreement can be identified as pertaining to samples with low salinity and high oxygen, corresponding to surface samples (shallower than about 200m) close to coastal areas affected by continental inputs, i.e., in the Labrador Current and Southern Ocean (Fig 5). Variability in the short time scale is high due to water mixing (Wootton et al., 2008; Rintoul, 2011), such that large residuals are expected for prediction methods.

2.4.2. Comparison of both methods against measured data

The second (the 3DwMLR method) and third (the neural network method) columns of both Fig 4 and Fig 5 compare estimated A_T vs. observations. Along these columns, only the subset of 76,553 samples with A_T observations can be compared.

After analysing both figures, a good overall agreement between observations and estimated A_T can be stated for both techniques, with some minor variations between them. The neural network technique seems to behave slightly more consistently with geographical parameters, as can be seen in subplots against depth and even with longitude.

Subplots against Cruise Year in Fig 4 show a slightly reduced scatter for more recent cruises, which is to be expected due to the improvement of analytical techniques and instrumentation. An abrupt step near 1992 was expected to be seen in these yearly subplots, due to the progressive introduction of CRM in successive years, but it is not appreciable. This is probably due to the use of the GLODAP and CARINA data products that are already adjusted in order to minimise offsets between cruises (Sabine et al., 2005; Tanhua et al., 2009).

As previously indicated, a significant bias against measured A_T data can be stated for some cruises when analysing Cruise ID subplots. Even when the agreement between both estimation techniques themselves is very good for these cruises, it is poor when the comparison is made against the observations of A_T . Cruises with a mean bias greater than $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ are listed in Table 2. All of them correspond to cruises with none or too few deep samples (>1500m) to qualify for secondary quality control and to be adjusted in the CARINA Product (Tanhua et al., 2009, 2010; Velo et al., 2010). As a result of this work, we propose adjusting these five cruises to compensate the bias indicated in Table 2.

Carina ID	Expocode	3DwMLR bias	Neural Net. bias	Proposed adjust
9	06MT19920509	-7.8±9.0	-7.6±8.6	-7.7
28	06MT20021013	-7.2±19.8	-3.8±15.8	Flag bad

51	29CS19771007	7.6±3.9	6.0±2.6	6.8
57	29GD19860904	-6.9±4.0	-5.1±4.5	-6.0
158	67SL19881117	-19.0±4.8	-14.1±6.0	-16.6

Table 2: Mean biases between 3DwMLR and neural technique against observed A_T data. Units for biases and adjusts are $\mu\text{mol}\cdot\text{kg}^{-1}$. Only values with mean bias greater than $\pm 6 \mu\text{mol}\cdot\text{kg}^{-1}$ are shown.

By visual inspection of the residuals against seawater properties (Fig 5) similar patterns are observed for both techniques. The neural network technique seemed to behave more consistently for most parameters, but these differences are not significant.

The highest scatter in Fig 5 corresponds to low salinity samples for all three columns. Conversely, high oxygen samples show good agreement against observations for both techniques, even though poor agreement between the two estimation methods is indicated in the first column.

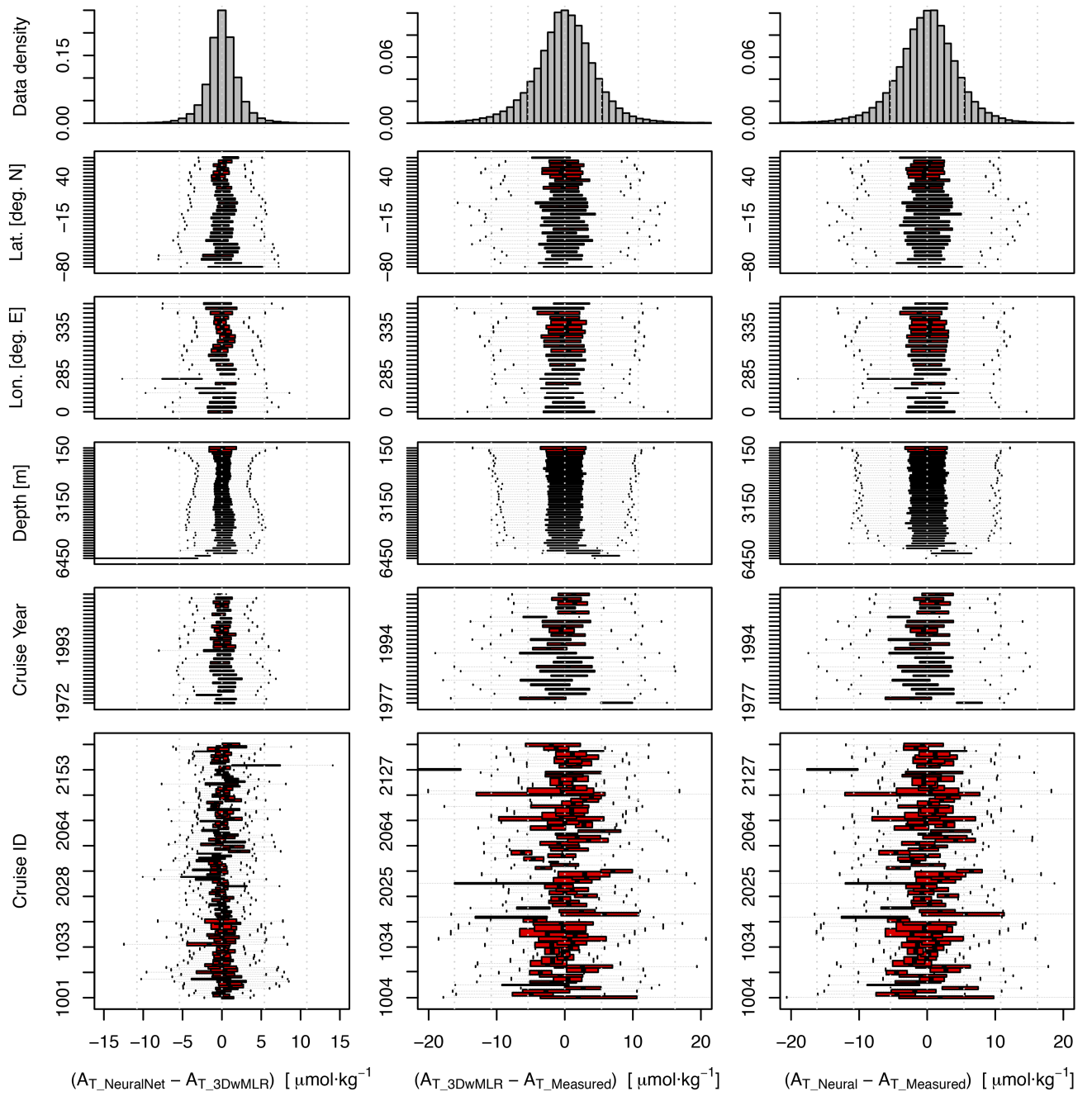


Fig 4: Histograms and box plots of A_T differences between, from left to right: Neural network minus 3DwMLR, 3DwMLR minus observations and neural network minus observations. From top to bottom, first row are histograms of data density, and box-whisker diagrams of the A_T difference against the parameter are indicated on the left: Latitude, Longitude, Depth, year of cruise, and a numerical identifier for each cruise.

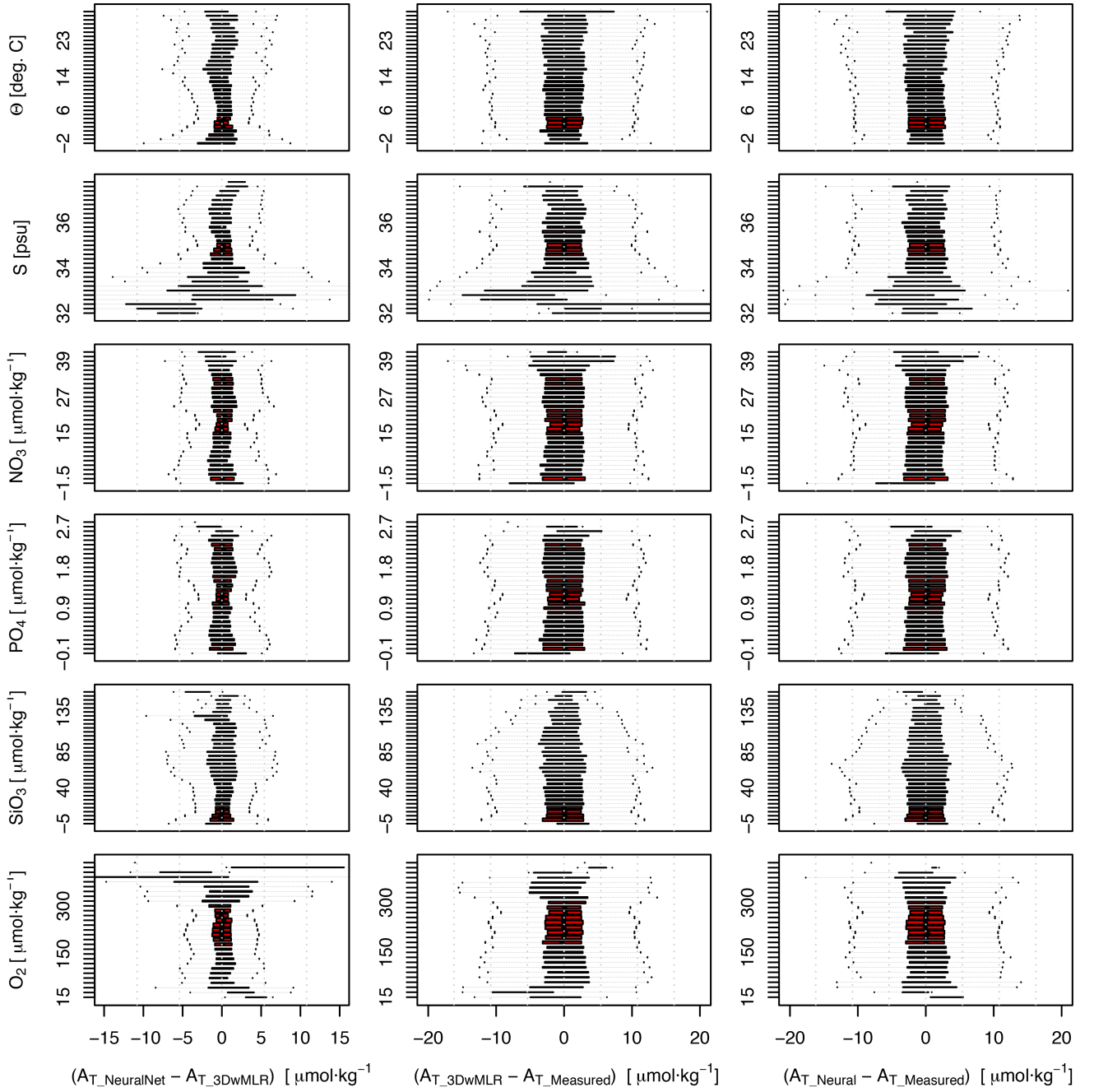


Fig 5: Box and whisker diagrams of A_T differences between, from left to right: Neural network minus 3DwMLR, 3DwMLR minus observations and neural network minus observations. From top to bottom, the parameter being compared against the A_T residuals (indicated on the left): θ , salinity, nitrate, phosphate, silicate and oxygen.

3. Conclusions

From an initial dataset with 72,796 samples useful for carbon calculations, i.e., with at least two carbon parameters measured, the algorithms presented here allowed for the extension of the dataset up to 104,043 useful samples. This implies a 42.9% increase in the abundance of carbonate system parameters in the dataset.

Moreover, both proposed methods of estimating alkalinity in the ocean give acceptable results, as the average error in the prediction is lower than $5 \mu\text{mol}\cdot\text{kg}^{-1}$, which can be considered an acceptable uncertainty for alkalinity measurements. Nevertheless, the two approaches to solve the problem are very different and take advantage of different considerations.

Considering only the performance of the two methods used here, measured on how well they can model the natural variability of the estimated parameter, the neural network performs better, because it is the best at predicting the natural variability (Fig 3). The results shown in Fig 3 indicate clearly that the proposed neural network with 64 neurons in the hidden layer and Levenberg-Marquadt tuning (Hagan and Menhaj, 1994) achieves better overall estimations than the 3DwMLR method.

Considering the scalability of the methodology used to obtain the results, the 3DwMLR is the best option, because it delivers a good agreement with the measured data and this methodology can be easily tuned if some misfit is found in some areas. In the 3DwMLR, all the steps performed for estimating every single sample can be debugged and improved. The neural network represents the very opposite approach. It is true that a tuned network can be retrained, but the weights and the way they are obtained are not transparent, to the point that the combination of the network with its tuning can be considered as a black box that works with the given dataset.

The advantage of using both techniques simultaneously for this purpose is that their results can be easily compared against each other and thus, their performance and agreement assessed.

4. Acknowledgements

We wish to acknowledge the hard work and dedication of analysts, investigators and crew that collect the data at sea; without them, this work would not be possible. We also gratefully acknowledge those who have contributed their data to the CARINA and GLODAP project. Thanks also to those people in the fisheries group of IIM-CSIC for their support in statistical concerns. This work and in particular T.T. was supported through EU FP7 project CARBOCHANGE “Changes in carbon uptake and emissions by oceans in a changing climate”, which received funding from the European Community’s Seventh Framework Programme under grant agreement no. 264879

5. Bibliography

- Anderson, D.R., Burnham, K.P., 2002. Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management* 66, 3, 912. doi: 10.2307/3803155
- Anderson, L.A., Sarmiento, J.L., 1994. Redfield Ratios of Remineralization Determined by Nutrient Data Analysis. *Global Biogeochemical Cycles* 8, 65–80. doi: 10.1029/93GB03318
- Bishop, C.M., 1995. *Neural networks for pattern recognition*. Clarendon Press, Oxford.
- Brewer, P.G., Wong, G.T.F., Bacon, M.P., Spencer, D.W., 1975. An oceanic calcium problem? *Earth and Planetary Science Letters* 26, 1, 81–87. doi: 10.1016/0012-821X(75)90179-X
- Chen, C.-T., Millero, F.J., 1979. Gradual increase of oceanic CO₂. *Nature* 277, 5693, 205–206. doi: 10.1038/277205a0
- Chen, C.-T., Pytkowicz, R.M., 1979. On the total CO₂-titration alkalinity-oxygen system in the Pacific Ocean. *Nature* 281, 5730, 362–365. doi: 10.1038/281362a0
- Dickson, A.G., 2010. The carbon dioxide system in seawater: equilibrium chemistry and measurements, in: *Guide to Best Practices for Ocean Acidification Research and Data Reporting*. Publications Office of the European Union, Luxembourg, pp. 17–52.
- Filzmoser, P., Serneels, S., Maronna, R., Van Espen, P.J., 2007. Robust multivariate methods in chemometrics. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Elsevier, Amsterdam 681–722.
- Fraga, F., Álvarez-Salgado, X.A., 2005. On the variation of alkalinity during phytoplankton photosynthesis. *Ciencias Marinas* 004, 627–639.
- Garcia, H.E., Locarnini, R.A., Boyer, T.P., Antonov, J.I., Zweng, M.M., Baranova, O.K., Johnson, D.R., 2009. *World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, and silicate)*. S. Levitus, Ed., NOAA Atlas NESDIS 71, U.S. Government Printing Office, Washington, D.C. 398.
- Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 14–15, 2627–2636. doi: 10.1016/S1352-2310(97)00447-0
- Geider, R., La Roche, J., 2002. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *European Journal of Phycology* 37, 1, 1–17. doi: 10.1017/S0967026201003456
- Green, P.J., 1984. Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *Journal of the Royal Statistical*

Society. Series B (Methodological) 46, 2, 149–192.

Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural network design. PWS Publishing, Boston, MA.

Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. Neural Networks, IEEE Transactions on 5, 6, 989–993.

Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. Communications in Statistics-Theory and Methods 6, 9, 813–827.

Hubert, M., 2008. High-Breakdown Robust Multivariate Methods. Statist. Sci. 23, 1, 92–119. doi: 10.1214/088342307000000087

Ilyina, T., Zeebe, R.E., Maier-Reimer, E., Heinze, C., 2009. Early detection of ocean acidification effects on marine calcification. Global Biogeochemical Cycles 23, GB1008, 11 PP. doi: 10.1029/2008GB003278

Juranek, L.W., Feely, R.A., Peterson, W.T., Alin, S.R., Hales, B., Lee, K., Sabine, C.L., Peterson, J., 2009. A novel method for determination of aragonite saturation state on the continental shelf of central Oregon using multi-parameter relationships with hydrographic data. Geophys. Res. Lett. 36, 6 PP. doi: 200910.1029/2009GL040778

Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J.L., Feely, R.A., Millero, F.J., Mordy, C., Peng, T.H., 2004. A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). Global Biogeochemical Cycles 18, GB4031.

Key, R.M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D.W.R., Mintrop, L., 2010. The CARINA data synthesis project: introduction and overview. Earth System Science Data 2, 1, 105–121. doi: 10.5194/essd-2-105-2010

Kim, T.-W., Lee, K., Feely, R.A., Sabine, C.L., Chen, C.-T.A., Jeong, H.J., Kim, K.Y., 2010. Prediction of Sea of Japan (East Sea) acidification over the past 40 years using a multiparameter regression model. Global Biogeochem. Cycles 24, 14 PP. doi: 201010.1029/2009GB003637

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2004. Applied linear statistical models. McGraw-Hill, IRWIN, Chicago.

Lee, K., Tong, L.T., Millero, F.J., Sabine, C.L., Dickson, A.G., Goyet, C., Park, G.H., Wanninkhof, R., Feely, R.A., Key, R.M., 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. Geophys. Res. Lett 33, L19605.

Lewis, E., Wallace, D., Allison, L., 1998. Program developed for CO₂ system calculations. ORNL/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

Miller, A.J., 1984. Selection of Subsets of Regression Variables. *Journal of the Royal Statistical Society. Series A (General)* 147, 3, 389–425. doi: 10.2307/2981576

Millero, F., 2007. The marine inorganic carbon cycle. *Chemical Reviews-Columbus* 107, 2, 308–341.

Millero, F.J., 1995. Thermodynamics of the carbon dioxide system in the oceans. *Geochimica et Cosmochimica Acta* 59, 4, 661–677. doi: 10.1016/0016-7037(94)00354-O

Millero, F.J., Lee, K., Roche, M., 1998. Distribution of alkalinity in the surface waters of the major oceans. *Marine Chemistry* 60, 1-2, 111–130. doi: 10.1016/S0304-4203(97)00084-4

Redfield, A.C., 1934. On the proportions of organic derivatives in sea water and their relation to the composition of plankton. *James Johnstone memorial volume* 176, 92.

Riebesell, U., Fabry, V.J., Hansson, L., Gattuso, J.P., 2010. Guide to best practices for ocean acidification research and data reporting. Publications Office of the European Union, Luxembourg.

Rintoul, S.R., 2011. The Southern Ocean in the Earth System. *Science Diplomacy: Antarctica, Science, and the Governance of International Spaces* 175–188.

Sabine, C.L., Key, R.M., Kozyr, A., Feely, R.A., Wanninkhof, R., Millero, F.J., Peng, T.S., Bullister, J.L., Lee, K., 2005. Global Ocean data Analysis Project (GLODAP): Results and data. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee ORNL/CDIAC-145, NDP-083, 110 pp.

Sarle, W.S., 1997. Neural Network FAQ. comp.ai.neural-nets.

Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., Van Heuven, S., Velo, A., Lin, X., Kozyr, A., Alvarez, M., Bakker, D.C.E., Brown, P., Falck, E., Jeansson, E., Lo Monaco, C., Olafsson, J., Perez, F.F., Pierrot, D., Rios, A.F., Sabine, C.L., Schuster, U., Steinfeldt, R., Stendardo, I., Anderson, L.G., Bates, N.R., Bellerby, R.G.J., Blindheim, J., Bullister, J.L., Gruber, N., Ishii, M., Johannessen, T., Jones, E.P., Köhler, J., Körtzinger, A., Metzl, N., Murata, A., Musielewicz, S., Omar, A.M., Olsson, K.A., de la Paz, M., Pfeil, B., Rey, F., Rhein, M., Skjelvan, I., Tilbrook, B., Wanninkhof, R., Mintrop, L., Wallace, D.W.R., Key, R.M., 2009. CARINA Data Synthesis Project. ORNL/CDIAC-157, NDP-091, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 37831-6335.

Tanhua, T., van Heuven, S., Key, R.M., Velo, A., Olsen, A., Schirnick, C., 2010. Quality control procedures and methods of the CARINA database. *Earth System Science Data* 2, 1, 35–49. doi: 10.5194/essd-2-35-2010

van Heuven, S., Pierrot, D., Lewis, E., Wallace, D., 2009. MATLAB Program developed for CO₂ system calculations. ORNL/CDIAC-105b, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

Velo, A., Vázquez-Rodríguez, M., Padín, X.A., Gilcoto, M., Ríos, A.F., Pérez, F.F., 2010. A multiparametric method of interpolation using WOA05 applied to anthropogenic CO₂ in the Atlantic. *Scientia Marina* 74, S1, 21–32. doi: 10.3989/scimar.2010.74s1021

Wallace, D.W.R., 1995. Monitoring global ocean carbon inventories. Ocean Observing System Development Panel, Texas A&M University, College Station, TX, p. 54.

Wolf-Gladrow, D.A., Zeebe, R.E., Klaas, C., Körtzinger, A., Dickson, A.G., 2007. Total alkalinity: The explicit conservative expression and its application to biogeochemical processes. *Marine Chemistry* 106, 1-2, 287–300. doi: 16/j.marchem.2007.01.006

Wootton, J.T., Pfister, C.A., Forester, J.D., 2008. Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset. *Proceedings of the National Academy of Sciences* 105, 48, 18848–18853. doi: 10.1073/pnas.0810079105

Zeebe, R.E., Wolf-Gladrow, D.A., 2001. CO₂ in seawater: Equilibrium, Kinetics, Isotopes. Elsevier Oceanography Series, Amsterdam.