Reconstructing the Carbon Dioxide Absorption Patterns of World Oceans Using a Feed-Forward Neural Network: Software Implementation and Employment Techniques

Jiye Zeng, Hideaki Nakajima, Yukihiro Nojiri, Shin-Ichiro Nakaoka

Reconstructing the Carbon Dioxide Absorption Patterns of World Oceans using a Feed-Forward Neural Network: Software Implementation and Employment Techniques

Jiye Zeng[1], Hideaki Nakajima[1], Yukihiro Nojiri[1], and Shin-ichiro Nakaoka[1]

[1]National Institute for Environmental Studies, Tsukuba, Japan.

zeng@nies.go.jp

Abstract. Oceans play a major role in the global carbon budget, absorbing approximately 27% of anthropogenic carbon dioxide (CO2). As the degree to which an ocean can serve as a carbon sink is determined by the partial pressure of CO2 in the surface water, it is critical to obtain an accurate estimate of the spatial distributions of CO2 and its temporal variation on a global scale. However, this is extremely challenging due to insufficient measurements, large seasonal variability, and short spatial de-correlation scales. This paper presents an open source software package that implements a feed-forward neural network and a back-propagation training algorithm to solve a problem with one output variable and a large number of training patterns. We discuss the employment of the neural network for global ocean CO2 mapping.

Keywords: CO2 · climate · neural network · ocean · software

1 Introduction

Recent changes in the global climate are closely related to increasing atmospheric levels of anthropogenic greenhouse gases [1], particularly carbon dioxide (CO2). According to the global carbon budget 2013 report [2], the total global carbon emissions increased from 3.9 GtC/yr in 1959 to 10.7 GtC/yr in 2012, of which approximately 43% remained in the atmosphere, 30% was taken by terrestrial sinks, and 27% was absorbed by the oceans. The uncertainty of the magnitude of the ocean sink is largely due to the uncertainty in estimating the CO2 at the ocean's surface, given that the flux at the air–water interface is determined by the difference in the partial pressure of CO2 between the two mediums, the transfer coefficient, and the wind at the surface [3]. Although the latter two parameters encompass some uncertainty, the CO2 in the surface ocean determines whether the oceans are a carbon sink or source.

The most challenging obstacle in estimating the capacity of the oceans as a carbon sink is the insufficiency of measurement data. Using the 1×1 gridded dataset from the Surface Ocean CO2 Atlas[1] (SOCAT) version 2 [4], which is the most complete and

---

[1] http://www.socat.info/

quality-controlled dataset, we calculated that no more than 4% of the surface waters of the world's oceans had been sampled in any single month between 1990 and 2011. This motivated us to examine robust and reliable methods for interpolating the available CO2 data, both spatially and temporarily. However, most existing models use a basin scale (e.g., [5,6,7,8,9,10,11,12,13,14]) as the spatial de-correlation length scale of CO2 on the order of 100 km, which is about 10 times smaller than the marine atmosphere CO2 [15], and the seasonal changes in the surface ocean CO2 can be 100 μatm or more [16]. The most used product to estimate the global ocean CO2 levels is the monthly climatology data of [17], which uses a 5×4 degree mesh.

This work presents open-source software that implements a feed-forward neural network (FNN) specifically developed for modeling environmental problems such as reconstructing the global ocean CO2. The software has been used by [18] to produce monthly maps of the climatology of CO2 with a spatial resolution of 1×1 degrees. We discuss possible techniques to extend the work of [18] to reconstruct the time variant distribution of the global ocean CO2 from 1990 to 2011.

2 Method

2.1 Software Package

Our software package[2] implements the FNN shown Figure 1. The FNN comprises three layers: input, hidden, and output. A neuron in the input layer takes the value of an input variable and passes it to all neurons in the hidden layer. A neuron in the hidden or output layer adds a bias to the weighted sum of all of the outputs from its upstream neurons and transforms the sum to yield its output:

$$y = f(x) = \frac{1}{1+e^{-x}}$$

$$x = b + \sum_{i=1}^{M} w_i y_{i,upstream}$$

(1)

The sigmoid transform function was chosen because it increases monotonically from 0 to 1 for x from $-\infty$ to $+\infty$ and the derivative of y with respect to x yields a simple form that affiliates the so-called back-propagation [19] algorithm for training the FNN. In training, the algorithm changes the weight and bias parameters proportionally according to the negative gradient of the error cost function, i.e., the sum of squared error between the network output y and the target d:

$$F(\mathbf{w}) = \frac{1}{2}\sum_{l}^{P}(d^l - y^l)^2 = \frac{1}{2}\mathbf{e}^T\mathbf{e}$$

$$\mathbf{e}^T = [e^1 \quad e^2 \quad ... \quad e^P]$$

$$e^l = d^l - y^l$$

(2)

---

where P is the total number of training patterns (a pattern comprises a set of data of all input variables). In the simplest form, the algorithm uses the first-order derivative of the cost function as the gradient $\mathbf{g}$ to find the minima in the error space:

$$\mathbf{g} = \nabla F(\mathbf{w}) = \mathbf{J}^T \mathbf{e} \tag{3}$$

where the vector $\mathbf{w}$ includes both weight and bias parameters and $\mathbf{J}$ is the Jacobian matrix. Assuming $\mathbf{w}$ has N elements, $\mathbf{J}$ has the form of

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial e^1}{\partial w_1} & \dfrac{\partial e^1}{\partial w_2} & \cdots & \dfrac{\partial e^1}{\partial w_N} \\ \dfrac{\partial e^2}{\partial w_1} & \dfrac{\partial e^2}{\partial w_2} & \cdots & \dfrac{\partial e^2}{\partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial e^P}{\partial w_1} & \dfrac{\partial e^P}{\partial w_2} & \cdots & \dfrac{\partial e^P}{\partial w_N} \end{bmatrix} \tag{4}$$

Regarding the output neuron and a training pattern, the partial differentiation by the weights and bias for the jth hidden neuron can be derived as

$$\frac{\partial e}{\partial w_{o,j}} = -\frac{\partial y_o}{\partial w_{o,j}} = -\frac{\partial y_o}{\partial x_o}\frac{\partial x_o}{\partial w_{o,j}} = -\delta_o y_j$$

$$\frac{\partial e}{\partial b_o} = -\frac{\partial y_o}{\partial b_o} = -\frac{\partial y_o}{\partial x_o}\frac{\partial x_o}{\partial b_o} = -\delta_o \tag{5}$$

$$\delta_o = \frac{\partial y_o}{\partial x_o} = (1 - y_o)y_o$$

and the partial differentiations by the weights and bias of the jth hidden neuron for the ith input neuron can be expressed as

$$\frac{\partial e}{\partial w_{j,i}} = -\frac{\partial y_o}{\partial w_{j,i}} = -\frac{\partial y_o}{\partial x_o}\frac{\partial x_o}{\partial y_j}\frac{\partial y_j}{\partial x_j}\frac{\partial x_j}{\partial w_{j,i}} = -\delta_j y_i$$

$$\frac{\partial e}{\partial b_j} = -\frac{\partial y_o}{\partial b_j} = -\frac{\partial y_o}{\partial x_o}\frac{\partial x_o}{\partial y_j}\frac{\partial y_j}{\partial x_j}\frac{\partial x_j}{\partial b_j} = -\delta_j \tag{6}$$

$$\delta_j = \frac{\partial y_o}{\partial x_o}\frac{\partial x_o}{\partial y_j}\frac{\partial y_j}{\partial x_j} = \delta_o w_j (1 - y_j) y_j$$

Updating neuron parameters for the tth training epoch (a complete loop through all training patterns) can be expressed as

$$\mathbf{w}(t) = \mathbf{w}(t-1) + \Delta\mathbf{w}(t) = \mathbf{w}(t-1) + \alpha\mathbf{g} \tag{7}$$

where α is the learning rate with a value between 0 and 1. The convergence of training the FNN by the gradient expressed in equation (3) is slow. The Levenberg–Marquardt algorithm (see [20,21]) improves the convergence speed by considering the second-order derivative of the cost function, resulting in a new estimation for the gradient as

$$\mathbf{g} = (\mathbf{J}^T \mathbf{J} + \mu\mathbf{I})^{-1} \mathbf{J}^T \mathbf{e} \tag{8}$$

where μ is the regularization constant and **I** is a unit matrix. In our implementation, we set μ proportional to the root mean square error:

$$\mu = \beta \sqrt{\frac{1}{P}\sum_{l}^{p}(d^l - y^l)^2}$$

(9)

We call β the regularization coefficient.

We implemented several methods in the software to achieve efficient computing and simple usage. First, the method of [21] was used to eliminate the necessity of storing a large Jacobian matrix in the memory. Second, the software scales all input variable data to have a zero mean and unit standard deviation and scales the target variable data to vary from 0.1 to 0.9. This method not only improves convergence speed but also makes it possible to use a default α and β for a variety of cases. Third, the software allows a user to use a text file and select any given number of columns in any order as input and output variables.
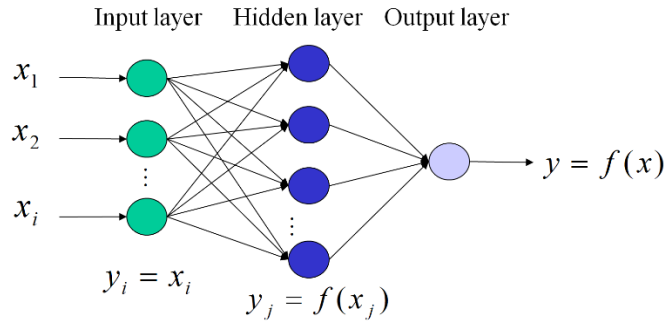

Fig. 1. Illustration of the feed forward neural network. Neurons in different layers are fully connected, but no connection exists among neurons of the same layer.

2.2 CO2 Model

Our basic model equation defines the fugacity of CO2 (fCO2) as a nonlinear function of month (MON), latitude (LAT), and longitude (LON):

$$fCO2 = F(MON, LAT, LON)$$

(10)

Because of insufficient measurements, we introduced sea surface temperature (SST), sea surface salinity (SSS), and chlorophyll-a concentration (CHL) as proxy variables in space and time. Therefore, the model equation becomes

$$fCO2 = F(MON, LAT, LON, SST, SSS, CHL)$$

(11)

In a global setting for long-term modeling, MON and LON must be transformed to account for their circular properties. As a result, the complete model equation becomes

$$fCO2 = F(CMON, SMON, LAT, CLON, SLON, SST, SSS, CHL) \qquad (12)$$

where

$$CMON = \cos(MON)$$
$$SMON = \sin(MON)$$
$$CLON = \cos(LON)$$
$$SLON = \sin(LON)$$

The model equation establishes the relationship between the eight input variables and the one output variable of the neural network.

2.3   Employment Issues

Two issues often cloud the proper use of an FNN: landing on a local minimum in training and choosing the "right" number of hidden neurons. The first issue can be detected using a different random number seeds to initialize the weight and bias parameters, or using different learning rates and regularization parameters. We have experimented with the Levenberg-Marquardt algorithm with these approaches and found that landing on a local minimum rarely occurred in our case.

The second issue is notorious. It is well known that an FNN can approximate any finite function given a sufficiently large number of hidden neurons [22,23]. That means that if the fCO2 values are unique in corresponding to unique values of MON, LAT, LON, SST, SSS, and CHL, equation (12) can be fitted accurately with a larger number of hidden neurons. An FNN configured as such remembers all training patterns and thus loses the capability to filter out erroneous information in measurement data. Unfortunately, there is no theoretically approved rule for choosing the right number of hidden neurons [24]. Our criterion for selecting the number is that the standard deviation of the difference between model outputs and observations (std1) is no larger than the standard deviations of repeated fCO2 measurements averaged over all months and grid boxes (std2) by 20%, i.e., |std1-std2|/std2<0.2.

An FNN configuration needs validation before it can be used to make predictions. Our approach uses 90% of the randomly selected data points for training and the remaining 10% for validation [18]. Once a configuration is validated, we used all data points to train the FNN to obtain a final product for prediction.

3   Results

We succeeded in using the FNN to model the mean global rate of increase in CO2 fugacity at the ocean's surface [18], yielding a value of 1.5 μatm/yr, which is in agreement with the commonly accepted value. Prior to this work, the rate was esti-

mated from a collection of rates [17] that vary greatly. Using our rate, we were able to obtain the monthly climatology maps of CO2 for the reference year of 2000 [18]. The results agree well with both measurements and the values reported by [17], which is the most used resource for calculating ocean sinks.

Extending the work of Zeng et al. [18], we reconstructed monthly CO2 maps for 1990–2011 (see the examples in Fig. 2). Two new steps were taken to improve the mapping. First, we excluded the LAT and LON variables in equation (12) to obtain a first guess of CO2 for areas where measurements were scarce, i.e., the Southern Pacific Ocean. As a large spatial gap may lead to overestimating the nonlinear interpolation of the FNN, we used the first guess for gap-filling every 10 degrees. Second, a version of the FNN without CHL was used to estimate CO2 for areas where CHL data were not available. This step completes the coverage of the global oceans. The yielded product provides an alternative for estimating the time-variant ocean carbon sinks for 1990–2011.
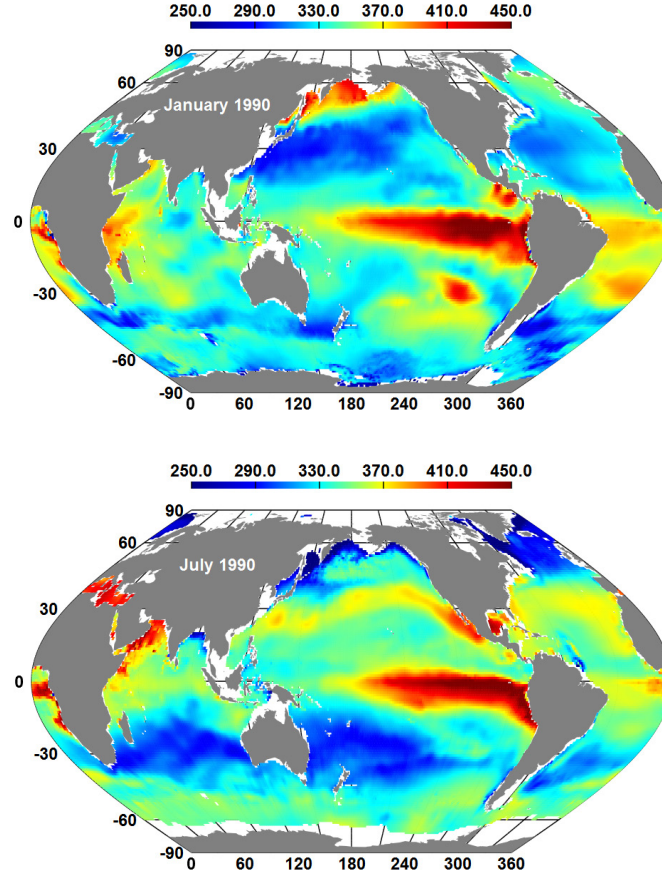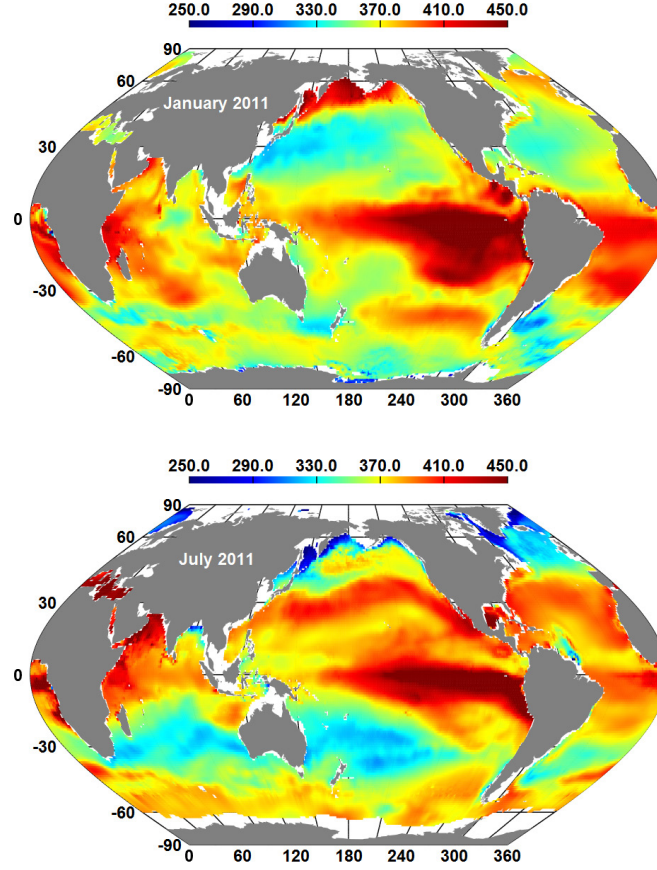
Fig. 2. Distributions of CO2 fugacity (μatm) in 1990 and 2011. The blank areas were either covered by ice or the depth is less than 500 m, which is our criterion for identifying open oceans.

4   Conclusions

Our results show that the FNN is an effective model for reconstructing global ocean CO2 sinks. As the model only establishes the dependence of the output variable on the input variables, and makes no assumption on the explicit expression of the model equation, it's straightforward to employ in practice. However, users must be aware of its nonlinear property and take appropriate steps to prevent over-fitting and over-shooting.

We made efforts to minimize the user interaction required, and the FNN converges quickly and steadily. In our experiments, the error of the cost function decreased steadily and changed very slightly after about 100 epochs for over 100,000 training patterns. The software only requires an ordinary PC to obtain results in 10 to 30 min

for large datasets. We expect the software to be used for solving environmental problems similar to the carbon sink example discussed herein.

References

1. Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann, C. Jones, C. Le Quéré, R.B. Myneni, S. Piao and P. Thornton, 2013: Carbon and Other Biogeochemical Cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
2. Quéré, C. Le, G.P. Peters, R.J. Andres, R. M. Andrew, T.A. Boden, P. Ciais, P. Friedlingstein, R.A. Houghton, G. Marland, R. Moriarty, S. Sitch, P. Tans, A. Arneth, A. Arvanitis, D.C.E. Bakker, L. Bopp, J.G. Canadell, L.P. Chini, S.C. Doney, A. Harper, I. Harris, J.I. House, A.K. Jain, S.D. Jones, E. Kato, R.F. Keeling, K. Klein Goldewijk, A. Körtzinger, C. Koven, N. Lefèvre, F. Maignan, A. Omar, T. Ono, G.-H. Park, B. Pfeil, B. Poulter, M.R. Raupach, P. Regnier, C. Rödenbeck, S. Saito, J. Schwinger, J. Segschneider, B.D. Stocker, T. Takahashi, B. Tilbrook, S. van Heuven, N. Viovy, R. Wanninkhof, A. Wiltshire, and S. Zaehlem, 2014: Global Carbon Budget 2013. Earth System Science Data, 6:235-263. doi:10.5194/essd-6-235-2014.
3. Wanninkhof, R., 1992: Relationship between Wind-Speed and Gas-Exchange over the Ocean. J. Geophys. Res.-Oceans, 97, 7373-7382.
4. Bakker, D. C. E., B. Pfeil, K. Smith, S. Hankin, A. Olsen, S. R. Alin, C. Cosca, S. Harasawa, A. Kozyr, Y. Nojiri, K. M. O'Brien, U. Schuster, M. Telszewski, B. Tilbrook, C. Wada, J. Akl, L. Barbero, N. R. Bates, J. Boutin, Y. Bozec, W.-J. Cai1, R. D. Castle, F. P. Chavez, L. Chen, M. Chierici, K. Currie, H. J. W. de Baar, W. Evans, R. A. Feely, A. Fransson, Z. Gao, B. Hales, N. J. Hardman-Mountford, M. Hoppema, W.-J. Huang, C. W. Hunt, B. Huss, T. Ichikawa, T. Johannessen, E. M. Jones, S. D. Jones, S. Jutterström, V. Kitidis, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. B. Manke, J. T. Mathis, L. Merlivat, N. Metzl, A. Murata, T. Newberger, A. M. Omar, T. Ono, G.-H. Park, K. Paterson, D. Pierrot, A. F. Ríos, C. L. Sabine, S. Saito, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, R. Sieger, I. Skjelvan, T. Steinhoff, K. F. Sullivan, H. Sun, A. J. Sutton, T. Suzuki, C. Sweeney, T. Takahashi, J. Tjiputra, N. Tsurushima, S. M. A. C. van Heuven, D. Vandemark, P. Vlahos, D. W. R. Wallace, R. Wanninkhof, and A. J. Watson, 2013: An update to the Surface Ocean CO2 Atlas (SOCAT version 2). Earth System Science Data Discussions, 6, 465-512, doi:10.5194/essdd-6-465-2013.
5. Zeng, J. Y., Y. Nojiri, P. P. Murphy, C. S. Wong, and Y. Fujinuma, 2002: A comparison of Delta pCO2 distributions in the northern North Pacific using results from a commercial vessel in 1995-1999. Deep-Sea Res. Part II-Top. Stud. Oceanogr, 49, 5303-5315.
6. Lefevre, N., A. J. Watson, and A. R. Watson, 2005: A comparison of multiple regression and neural network techniques for mapping in situ pCO2 data. Tellus Ser. B-Chem. Phys. Meteorol., 57, 375-384.
7. Chierici, M., A. Fransson, and Y. Nojiri, 2006: Biogeochemical processes as drivers of surface fCO(2) in contrasting provinces in the subarctic North Pacific Ocean. Glob. Biogeochem. Cycle, 20.
8. Sarma, V. V. S. S., T. Saino, K. Sasaoka, Y. Nojiri, T. Ono, M. Ishii, H. Y. Inoue, and K. Matsumoto, 2006: Basin-scale pCO2 distribution using satellite sea surface temperature,

Chla, and climatological salinity in the North Pacific in spring and summer. Glob. Biogeochem. Cycle, 20.

9. Jamet, C., C. Moulin, and N. Lefevre, 2007: Estimation of the oceanic pCO2 in the North Atlantic from VOS lines in-situ measurements: parameters needed to generate seasonally mean maps. Ann. Geophys., 25, 2247-2257.

10. Friedrich, T., and A. Oschlies, 2009: Neural network-based estimates of North Atlantic surface pCO(2) from satellite data: A methodological study. J. Geophys. Res.-Oceans, 114.

11. Telszewski, M., and Coauthors, 2009: Estimating the monthly pCO2 distribution in the North Atlantic using a self-organizing neural network. Biogeosciences, 6, 1405-1421.

12. Takamura, T. R., H. Y. Inoue, T. Midorikawa, M. Ishii, and Y. Nojiri, 2010: Seasonal and Inter-Annual Variations in pCO2sea and Air-Sea CO2 Fluxes in Mid-Latitudes of the Western and Eastern North Pacific during 1999-2006: Recent Results Utilizing Voluntary Observation Ships. Journal of the Meteorological Society of Japan, 88, 883-898.

13. Landschützer, P., N. Gruber, D. C. E. Bakker, U. Schuster, S. Nakaoka, M. R. Payne, T. Sasse, and J. Zeng, 2013: A Neural Network-based Estimate of the Seasonal to Inter-annual Variability of the Atlantic Ocean Carbon Sink. Biogeosciences Discuss., 10, 8799-8849.

14. Nakaoka, S., M. Telszewski, Y. Nojiri, S. Yasunaka, C. Miyazaki, H. Mukai, and N. Usui, 2013: Estimating temporal and spatial variation of ocean surface pCO2 in the North Pacific using a Self Organizing Map neural network technique. Biogeosciences, 10, 6093-6106.

15. Li, Z., D. Adamec, T. Takahashi, and S. C. Sutherland, 2005: Global aurocorrelation scales of the partial pressure of oceanic CO2. J Geophys. Res., 110, C08002, doi:10.1029/2004JC002723.

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

17. Takahashi, T., S. C. Sutherland, R. Wanninkhof, C. Sweeney, R. A. Feely, D. W. Chipman, B. Hales, G. Friederich, F. Chavez, C. Sabine, A. Watson, D. C.E. Bakker, U. Schuster, N. Metzl, H. Y. Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, A. Körtzinger, T. Steinhoff, M. Hoppema, J. Olafsson, T. S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. Bellerby, C.S. Wong, B. Delille, N. R. Bates, H. J. W. de Baar, 2009: Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans (vol 56, pg 554, 2009). Deep-Sea Res. Part I-Oceanogr. Res. Pap., 56, 2075-2076.

18. Zeng, J., Y. Nojiri, P. Landschützer, M. Telszewski, and S. Nakaoka. 2014: A Global Surface Ocean fCO2 Climatology Based on a Feed-Forward Neural Network. Journal of Atmospheric and Oceanic Technology, 31:1838-1849

19. Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning Representations by Back-Propagating Errors. Nature, 323, 533-536.

20. LeCun, Y., L. Bottou, G. B. Orr, and K. R. Muller, 1998: Efficient backprop. Neural Networks: Tricks of the Trade, 1524, 9-50.

21. Wilamowski, B. M., and H. Yu, 2010: Improved Computation for Levenberg-Marquardt Training. Ieee Transactions on Neural Networks, 21, 930-937.

22. Blum, E. K., and L. K. Li, 1991: Approximation-Theory and Feedforward Networks. Neural Networks, 4, 511-515.

23. Hornik, K., 1991: Approximation Capabilities of Multilayer Feedforward Networks. Neural Networks, 4, 251-257.

24. Svozil, D., V. Kvasnicka, and J. Pospichal, 1997: Introduction to multi-layer feed-forward neural networks. Chemometrics and Intelligent Laboratory Systems, 39, 43-62.