Technical note: Evaluation of three machine learning models for surface ocean CO2 mapping

Jiye Zeng[1], Tsuneo Matsunaga[1], Nobuko Saigusa[1], Tomoko Shirai[1], Shin-ichiro Nakaoka[1], Zheng-Hong Tan[2]

[1]National Institute for Environmental Studies, Tsukuba, Japan
[2]Department of Environmental Science, Hainan University, China

Correspondence to: Jiye Zeng (zeng@nies.go.jp)

Abstract. Reconstructing surface ocean CO_2 from scarce measurements plays an important role in estimating oceanic CO_2 uptake. There are varying degrees of differences among the 14 models included in the Surface Ocean CO_2 Mapping (SOCOM) inter-comparison initiative, in which five models used neural networks. This investigation evaluates two neural networks used in SOCOM, self-organization map and feedforward neural network, and introduces a machine learning model called support vector machine for ocean CO_2 mapping. The technique note provides a practical guide to selecting the models.

1 Introduction

The global ocean is a major sink for anthropogenic carbon and therefore an important contributor for slowing down the human-induced global warming (Stocker et al., 2013). For calculating the oceanic CO_2 uptake, various models have been used to interpolate scarce CO_2 measurements in the surface ocean spatially and temporally to obtain basin-wide (e.g. Zeng et al., 2002; Lefevre et al., 2005; Chierici et al., 2006; Sarma et al., 2006; Jamet et al., 2007; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Takamura et al., 2010; Landschützer et al., 2013; Nakaoka et al., 2013; Iida et al., 2015) and global ocean CO_2 maps (Takahashi et al., 2002 and 2009; Park et al., 2010. Rödenbeck et al., 2013; Sasse et al., 2013; Jones et al., 2015; Zeng et al., 2015). The Surface Ocean CO_2 Mapping (SOCOM) inter-comparison initiative revealed varying degrees of differences among 14 models (Rödenbeck et al., 2015), of which 5 used neural networks. They include self-organizing maps (SOM) and feedforward neural networks (FNN). The SOM has a long history in CO_2 mapping (Lefevre et al., 2005; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Nakaoka et al., 2013). Recently, the FNN is gaining popularity in this field (Landschützer et al., 2015; Zeng et al., 2014 and 2015). In this investigation we introduce a machine learning model called support vector machine (SVM) for ocean CO_2 mapping and compare the SVM with the SOM and FNN. We intend to provide a practical guide for using these machine learning models.

2 Model Equations

The machine learning models included in this study cannot directly model the long term trend of CO_2. Therefore, we express the dependence of CO_2 fugacity (fCO_2) on year (YR), month (MON), latitude (LAT), and longitude (LON) as the sum of a nonlinear static component and a linear trend component:

$$fCO_2 = F_{static}(LAT, LON, MON) + F_{trend}(YR). \qquad (1)$$

As available observations are scarce with respect to the biogeochemical properties of the surface ocean, we used sea surface temperature (SST), sea surface salinity (SSS), chlorophyll-a concentration (CHL), and mixed layer depth (MLD) as the proxy variables of space and time. These proxy variables were commonly used by models included in the SOCOM. The model

10   equation became

$$fCO_2 = F_{static}(LAT, SST, SSS, CHL, MLD, dSST) + F_{trend}(YR), \qquad (2)$$

where dSST denotes the difference between the monthly and annual means of SST. Here we excluded LON and MON. They have a circular property and therefore cannot be used directly. Zeng et al. (2014 and 2015) circumvented this problem by using their sine and cosine transformed components. The approach could unintentionally enhance the influence of LAT and MON on

15   fCO_2 as one more derived variable from each of them were added to the model. We excluded LON for the belief that the combination of SST, SSS, CHL, and MLD contains sufficient spatial information, but retained LAT for its different seasonal and geophysical meanings in the northern and southern hemispheres. Replacing MON by dSST also improves expressing the seasonal variable continuously, especially for those measurements taken near the start or end of a month.

3 Data

20   We extracted monthly fCO_2 from the track-gridded database of the Surface Ocean CO_2 Atlas (SOCAT) version 3.0[1] (Pfeil et al., 2013; Sabine et al., 2013; Bakker et al. 2013). The database has a 1°×1° spatial resolution and includes global measurements from 1970 to 2014. Similar to Zeng et al. (2014), we excluded some data points by these criteria: (i) fCO_2 values smaller than 250 µatm or larger than 550 µatm, (ii) ocean depth smaller than 500 m, and (iii) salinity smaller than 25.0. A total of 158,052 data points were extracted with these conditions.

25

---

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

The monthly SST data of 1990 to 2015 were extracted from the Optimum Interpolation (OI) V2 product[2] of NOAA (Reynolds et al., 2002). The monthly SSS climatology was extracted from the World Ocean Atlas 2013 (WOA13) product[3] (Boyer et al., 2013), which contains the monthly mean SSS from June 27, 1896 to December 25, 2012. The monthly CHL climatology was calculated using the MODIS Aqua and SeaWiFS climatology[4], which covers the period of 2012 to 2015. The mean of the two

5 CHLs was used as the CHL climatology. The mixed layer data were derived from the Monthly Isopycnal and Mixed-layer Ocean Climatology[5] of NOAA (Schmidtko et al., 2013), which includes the period of 1955 to 2012.

4 Machine Learning Models

The Appendix summarizes the algorithms of the three models. Here we focus on discussing their usage in CO_2 mapping.

10 For a given dataset, the SVM requires a prior step to find the optimal values for the parameter γ in Eq.(A10) and the parameter σ in Eq.(A15). To shorten the training time, we used 10% of the measurement data in this step and obtained 10 for γ and 0.06 for σ. Note that these values are dependent of data scaling, which is necessary in this case to avoid overflow problem in solving Eq.(A12). We scaled all variable values to the range of 0 and 1 for the SVM, i.e.,

$$v = \frac{v - v_{min}}{v_{max} - v_{min}}. \quad (3)$$

Data scaling is not necessary for the FNN, but can improve its performance. Following Zeng et al. (2014), we scaled the input variables by their mean and standard deviation as

$$v = \frac{v - \bar{v}}{s}. \quad (4)$$

The target variable $f\mathrm{CO}_2$ is scaled by

20
$$v = 0.1 + 0.8 \frac{v - v_{min}}{v_{max} - v_{min}}. \quad (5)$$

This confines the scaled $f\mathrm{CO}_2$ between 0.1 and 0.9 for better response to changes of input variables. We also used 64 hidden neurons as Zeng et al. (2014) did.

---

[2] http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html

[3] https://www.nodc.noaa.gov/OC5/woa13/

[4] http://oceancolor.gsfc.nasa.gov/cgi/l3

[5] http://www.pmel.noaa.gov/mimoc/

Data scaling is critical for the SOM, as the Euler distance in Eq.(A1) would be affected by variable units. We used Eq.(4) to scale input variables in training the SOM. Based on preliminary studies, we applied a factor of 2 to enhance the influence of *SST* and *CHL* on the distance.

5   One more step between training and prediction is needed in using SOM for CO_2 mapping: CO_2 labelling. The training usually uses all available input data to obtain a feature map; then CO_2 measurements are assigned to the best matching neuron cells (BMC) according Eq.(A1). Making prediction for an input is realized by finding the labelled BMC and extracting its mean CO_2 value. It is not difficult to see that the feature map size affects the labelling and hence the prediction. Unfortunately, there is no guideline for choosing the size. Based on previous studies (Telszewski et al., 2009 and Nakaoka et al. 2013), we used

10   20,000 neuron cells, roughly one neuron cell for one 1x1 grid cell of sampled areas.

Zeng et. al. (2014) presented a method to use the FNN to model the linear component in Eq. (2). Instead of repeating the process, we used their annual rate of 1.5 μatm to normalize *f*CO_2 to the reference year 2005 to model the nonlinear component. Similarly, we applied their log transform to *CHL* prior to data scaling discussed above, i.e.,

$$CHL = \log_{10}(1.0 + CHL). \qquad (6)$$

5 Model Validation

We examined the goodness of fitting by randomly selecting 10% to 50% of the data points to train the FNN and SVM, and to label the SOM; and then calculated the correlation coefficient between modelled and observed CO_2.

20   The SOM yields the best correlation in the case of 10% of data points and the correlation decreases with the number of data points (Fig.1). The reason is that for a given feature map size, the fewer the data points, the less possible a neuron cell will be labelled by multiple measurements and the more likely that the prediction will find the same CO_2 value used for labelling. Therefore, the goodness of fit does not necessary mean good SOM modelling.

25   The correlations obtained by the SVM and FNN do not vary much with the number of data points. While the SVM's correlation decreases monotonically, even though by only a little, with the number of data points, the FNN's correlation obtained with 75000 data points is larger than that with 60000 data points. The FNN is known for not being able to find the global optimum in training. This case could be an indication of an imperfect training. The FNN appears inferior to SVM in all case. However, imperfect training does not account for all the differences. If we use the number of model parameters to be determined by the

30   training as the indicator of the dimension of the model space, the FNN's dimension is far smaller than that of the SVM. The former is determined by the number of hidden neurons and input variables, whereas the latter is determined by the number of

training data. For 6 input variables, 15000 training, and 64 hidden neurons, the number of parameters is 509 for the FNN and 15001 for the SVM.

A better indicator for the performance of the models would be the goodness of prediction. To emulate the situation that the
5   sampled area was only a small portion of the global ocean, we evaluated the goodness of prediction by training FNN and SVM and labelling SOM with 10% of the data to make prediction for the rest of the data. Fig.2 shows that the SVM yielded the best correlation (R^2=0.72), the FNN fell behind (R^2=0.67), and the SOM performed the worst (R^2=0.54). The differences between predicted and observed CO_2 are 0.1±17.4 μatm for SVM, 0.1±18.9 μatm for FNN, and 0.2±23.3 μatm respectively. These differences are small compared with measurement uncertainties, for example the uncertainty caused by seawater temperature
10 correction in an underway measurement (Murphy et al., 2001). The uncertainty is only a little larger than the average standard deviation of repeated measurements. By sorting all normalized CO_2 measurements in July by latitude and longitude and calculating the standard deviation for those grids that have more than 5 measurements, we obtained an average standard deviation of 13.9 μatm.

6 Differences

15 Figure 3 shows CO_2 maps in February and July, 2005, which is the reference year for normalization. In the mapping, we randomly selected 50% of the data to train the FNN and SVM and to label the SOM. All models captured the major features of spatial distribution of observed CO_2. The SOM exhibits obvious discontinuity because of its discrete characteristics of picking up CO_2 values from a labelled feature map. For year 2015, the mean CO_2 difference is -0.05±12.73 μatm for FNN-SVM and -0.6±18.80 for SOM-SVM. The statistics indicates that FNN agrees better with SVM than SOM does.

20

Although the differences among models might be on the order of 10 to 20 μatm, the effect on the global ocean CO_2 flux estimate is small (Fig.4). The fluxes are calculated using the wind speed from ECMWF's interim product (Deea et al., 2011). Our estimate for the oceanic uptake is on the higher end among those in Wanninkhof et al. (2013) and Le Quéré et al. (2015). For example, Wanninkhof et al. (2013) reported that the median sea–air anthropogenic CO_2 fluxes centered on year 2000
25 ranged from 1.9 to 2.5 PgC yr^{-1} among the seven models. In comparison, our estimates by the three models are about 2.4 PgC yr^{-1}. The mean difference of CO_2 flux is 0.02 PgC yr^{-1} between the FNN and the SVM (FNN-SVM) and 0.06 PgC yr^{-1} between the SOM and the SVM (SOM-SVM). They are small in comparison with those differences among the models in Wanninkhof et al. (2013) and Le Quéré et al. (2015).

30 On the spatial scale of tens of degrees, the three models show good mutual agreement for modelled CO_2 distributions among them. However, each model shows distinguished fine structures, which are determined by the biogeochemical processes in the

ocean, by model parameters obtained from training, and by the characteristics of the models. We believe that the modelled monthly CO_2 distributions are true to the degree given by the model validations.

7 Summary

The main features of the three machine models are listed in Table 1. The SVM is recommended when the computer has enough
5 memory to store the matrix in Eq.(A12), which is proportional to the square of the number of training data. The SVM performs the best, but the training time could become very long when the number of training data is too large to be handled by a computer without using virtual memory. For any given dataset, using the SVM requires a prior steps to find the optimal value for the parameter γ in Eq.(A10) and the parameter σ in Eq.(A15).

10 The FNN model does not perform as well as the SVM, but the number of training data does not affect its training as much as the SVM's, The training time can become long when a large number of hidden neurons are used and many iterations are needed to achieve convergence. It takes longer time to train the FNN than the SVM for a small number of data points. However, the FNN is simpler to use as it requires no prior step.

15 The SOM is recommended only when the other two models have over fitting or over interpolation problems. The SOM performs the worst and is not as straightforward as the others as its results depends too much on data scaling and the feature map size. An advantage of the SOM is that once its feature map is obtained, re-labelling the map with new CO2 measurements and making a new prediction is fast. Although the SOM does not have the over interpolation problem of the other two, it may produce nonsense predictions due to its strong dependence on data scaling.

20 **Appendix**

A.1 Self-Organization Map

A self-organizing map (SOM) is a type of artificial neural network that is trained using unsupervised learning to project the input space of training samples to a feature space (Kohonen, 1984), which is usually represented by grid points in two dimensional space. Each grid point, also called neuron cell, is associated with a weight vector $\mathbf{w}$ having the same number of
25 components as the vector $\mathbf{v}$ of input data. In our study, the components of a data vector comprise a set of values for LAT, SST, SSS, CHL, MLD, and dSST.

We used the batch learning algorithm (Abe et al., 2002) to train SOM as the resulting features in the SOM do not depend on the input order of training samples. The weight components were initialized with random numbers between -1 and 1. In each
30 iterative training loop, a best matching cell (BMC) is searched for each data vector by minimizing the distance

$$d = |\mathbf{f}(\mathbf{w} - \mathbf{v})|, \qquad (A1)$$

where $\mathbf{f}$ is a scale matrix that we introduced to scale the input variables and its components are all zero except for those on the diagonal, which are set to 1 by default. The scale matrix can be used to change the influence of certain variables on the distance. After all data vectors are matched, the weight vector is updated by

$$w_i = \frac{\sum_k h_{ik} v_k}{\sum_k h_{ik}}, \qquad (A2)$$

where i and k denote indexes of neuron cells and data vectors, respectively. The neighbourhood function that determines the weight factor h is defined as

$$h_{ik} = \exp(-\frac{|\mathbf{r}_{ik}|}{q}), \qquad (A3)$$

where $|\mathbf{r}_{ik}|$ denotes the geographic distance on the feature map between the ith neuron cell and the BMC for the kth data vector and q is a factor that decreases linearly with iteration loop. For a given neuron cell, the procedure adjust its weight factors toward those data vectors whose BMC are close to it and the adjustment decreases exponentially with the geographic distance between neuron cells and linearly with the training loop.

A.2 Feedforward Neural Network

A feedforward neural network (FNN) is an artificial neural network that is trained using supervised learning to map the input variable space to a high dimensional space nonlinearly. Our FNN comprises three layers: An input layer, a hidden layer, and an output layer. The number of neurons in the input layers is determined by the number of input variables. The output layer has only one neuron for fCO2. For a given vector $\mathbf{x}$ of input data, neurons in the input layer pass variables' values to all neurons in the hidden layer. A neuron in the hidden layer uses the following kernel function to transform those values to generate a value for feeding to the output neuron:

$$y = \frac{1}{1+\exp(-(b+\mathbf{w}\cdot\mathbf{x}))}, \qquad (A4)$$

where $\mathbf{w}$ is the weight vector and b is the offset parameter. The output neuron does the same transformation for inputs from the hidden layer to produce an output. The training updates the offset and weight parameters by minimizing the cost function

$$f(\mathbf{w}) = \frac{1}{2}\mathbf{e}\cdot\mathbf{e} = \frac{1}{2}|\mathbf{y}_m - \mathbf{y}_o|. \qquad (A5)$$

where $\mathbf{y}_m$ and $\mathbf{y}_o$ stand for modelled and observed y, respectively; and $\mathbf{w}$ includes both weight and offset parameters.

In the gradient descent training algorithm, updating w at the training iteration t can be written as

$$\mathbf{w}(t) = \mathbf{w}(t-1) - \alpha\mathbf{g}, \qquad (A6)$$

where α is the learning rate and $\mathbf{g}$ the first-order derivative of the cost function:

$$\mathbf{g} = \nabla f(\mathbf{w}) = \mathbf{J}^T\mathbf{e}, \qquad (A7)$$

where $\mathbf{J}$ is the Jacobian matrix whose components are derivatives of $\mathbf{e}$ by $\mathbf{w}$ using back propagation method. We used the efficient Levenberg-Marquardt algorithm (Wilamowski et al., 2010), which derives the gradient as

$$\mathbf{g} = (\mathbf{J}^T \cdot \mathbf{J} + \mu \mathbf{I})^{-1} \mathbf{J}^T \mathbf{e}, \qquad (A8)$$

where μ is a constant.

A.3 Support Vector Machine

A support vector machine (SVM) is a supervised learning model that was conceptualized in the in 1960s for classifications

5   problems and later extended to regression analysis (Basak et al., 2005). Similar to FNN, a SVM uses a kernel function to

map the input variable space to a high dimensional space. The SVM for regression that uses the same optimization algorithm

as the SVM for classification is not suitable for our purpose. As the training seeks to maximize the distance between the

hyperplanes that divide different classes and uses those vectors on the hyperplane as support vectors for making prediction,

the resulting CO_2 is biased, i.e., the difference between model and observation is statistically not zero; and the bias could

10   vary significantly with a different selection of SVM parameters.

We used the so called least-square support vector machine (LS-SVM) for regression (Pelckmans et al., 2002) which, similar

to FNN, seeks to minimize the error between model outputs and measurements. Using a kernel function $\varphi(\mathbf{x})$ to map the

space of independent variables to a high dimensional space, the LS-SVM models a target variable as

15    $$\mathbf{y}(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \qquad (A9)$$

and minimizes the cost function

$$F(\mathbf{w}) = \frac{1}{2}(\mathbf{w}^T \mathbf{w} + \gamma |\boldsymbol{e}|) \qquad (A10)$$

subjecting to the constraint

$$\mathbf{y}(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b + \boldsymbol{e}. \qquad (A11)$$

20   By applying the Lagrangian multiplier, the optimization problem eventually becomes solving the linear equation of

$$\begin{bmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \boldsymbol{\Omega} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \qquad (A12)$$

where $\mathbf{u}$ is a vector with all components being 1, and the components of $\boldsymbol{\Omega}$ are

$$\Omega_{ij} = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j). \qquad (A13)$$

Once Eq.(a12) is solved, making a prediction is done by

25    $$y_i = \sum_j^n \alpha_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \qquad (A14)$$

In this investigation, we used the radial basis kernel function, i.e.,

$$\varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \qquad (A15)$$

where σ is a parameter whose optimal value depends on training data.

Acknowledgements

We thank all scientists and institutes who contributed to the SOCAT database.

References

Abe, T., S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, T. Ikemura, 2002. A Novel Bioinformatic Strategy for Unveiling Hidden Genome Signatures of Eukaryotes: Self-Organizing Map of Oligonucleotide Frequency. Genome Informatics 13: 12–20.

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. Cai, R. D. Castle, F. P. Chavez, L. Chen, M. Chierici, K. Currie, H. J. W. de Baar, W. Evans, R. A. Feely, A. Fransson, Z. Gao, B. Hales, N. J. Hardman-Mountford, M. Hoppema, W.-J. Huang, C. W. Hunt, B. Huss, T. Ichikawa, T. Johannessen, E. M. Jones,
10 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, N. Tsurushima, S. M. A. C. van Heuven, D. Vandemark, P. Vlahos, D. W. R. Wallace, R. Wanninkhof, and A. J. Watson, 2014. An update to the Surface Ocean CO2 Atlas (SOCAT version 2).
15 Earth Syst. Sci. Data, 6: 69–90, doi:10.5194/essd-6-69-2014

Boyer, T.P., J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T.D. O'Brien, C.R. Paver, J.R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, 2013. World Ocean Database 2013, NOAA Atlas NESDIS 72, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp.

20

Chierici, M., A. Fransson, and Y. Nojiri, 2006. Biogeochemical processes as drivers of surface fCO2 in contrasting provinces in the subarctic North Pacific Ocean. Glob. Biogeochem. Cycle, 20, GB1009, doi:10.1029/2004GB002356.

Deea, D. P., S. M. Uppalaa, A. J. Simmonsa, P. Berrisforda, P. Polia, S. Kobayashib, U. Andraec, M. A. Balmasedaa, G.
25 Balsamoa, P. Bauera, P. Bechtolda, A. C. M. Beljaarsa, L. van de Bergd, J. Bidlota, N. Bormanna, C. Delsola, R. Dragania, M. Fuentesa, A. J. Geera, L. Haimbergere, S. B. Healya, H. Hersbacha, E. V. H´olma, L. Isaksena, P. K°allbergc, M. K¨ohlera, M. Matricardia, A. P. McNallya, B. M. Monge-Sanzf , J.-J. Morcrettea, B.-K. Parkg, C. Peubeya, P. de Rosnaya, C. Tavolatoe, J.-N. Th´epauta and F. Vitarta. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137: 553–597.
30

Friedrich, T., and A. Oschlies, 2009. Neural network-based estimates of North Atlantic surface pCO2 from satellite data: A methodological study. J. Geophys. Res. 114, C03020, doi:10.1029/2007JC004646.

Iida, Y., A. Kojima, Y. Takatani, T. Nakano, H. Sugimoto, T. Midorikawa and M. Ishii, 2015. Trends in pCO_2 and sea-air CO_2 flux over the global open oceans for the last two decades, J. Oceanogr.1-25.

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

Jones, S. D., C. L. Quéré, C. Rödenbeck, A. C. Manning, and A. Olsen, 2015. A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data, J. Adv. Model. Earth Syst., 7: 1942–2466.

10   Kohonen, T., Self-Organization and Associative Memory, Springer, Berlin, 1984.

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.

Landschützer, P., N. Gruber, F. Haumann, C. Rödenbeck, D. Bakker, S. van Heuven, M. Hoppema, N. Metzl, C. Sweeney, T. Takahashi, B. Tilbrook, and R. Wanninkhof, 2015. The reinvigoration of the Southern Ocean carbon sink, Science, 349: 1221–1224.

20   Le Quéré1, C., R. Moriarty, R. M. Andrew, J. G. Canadell, S. Sitch, J. I. Korsbakken, P. Friedlingstein, G. P. Peters, R. J. Andres, T. A. Boden, R. A. Houghton, J. I. House, R. F. Keeling, P. Tans, A. Arneth, D. C. E. Bakker, L. Barbero, L. Bopp, J. Chang, F. Chevallier, L. P. Chini, P. Ciais, M. Fader, R. A. Feely, T. Gkritzalis, I. Harris, J. Hauck, T. Ilyina, A. K. Jain, E. Kato, V. Kitidis, K. Klein Goldewijk, C. Koven, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, I. D. Lima, N. Metzl, F. Millero, D. R. Munro, A. Murata, J. E. M. S. Nabel, S. Nakaoka, Y. Nojiri, K. O'Brien, A. Olsen, T. Ono, F. F. Pérez, B.
25   Pfeil, D. Pierrot, B. Poulter, G. Rehder, C. Rödenbeck, S. Saito, U. Schuster, J. Schwinger, R. Séférian, T. Steinhoff, B. D. Stocker, A. J. Sutton, T. Takahashi, B. Tilbrook, I. T. van der Laan-Luijkx, G. R. van der Werf, S. van Heuven, D. Vandemark, N. Viovy, A. Wiltshire, S. Zaehle, and N. Zeng, 2015. Global Carbon Budget 2015. Earth Syst. Sci. Data, 7: 349–396, doi:10.5194/essd-7-349-2015.

30   Lefevre, N., A. J. Watson, and A. R. Watson, 2005. A comparison of multiple regression and neural network techniques for mapping in situ pCO(2) data. Tellus Ser. B-Chem. Phys. Meteorol., 57: 375-384.

Murphy P., Y. Nojiry, Y. Fujinuma, C.S. Wong, J. Zeng, T. Kimoto, and H. Kimoto, 2001. Measurements of surface seawater fCO2 from volunteer commercial ships: Techniques and experiences from Skaugran. Journal of Atmospheric and Oceanic Technology, 18: 1719-1734.

5   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.

Park, G.-H., R. Wanninkhof, S.C. Doney, T. Takahashi, K. Lee, R.A. Feely, C.L. Sabine, J. Triñanes, and I.D. Lima, 2010.
10   Variability of global net sea-air CO2 fluxes over the last three decades using empirical relationships, Tellus B, 62: 352-368. 10.1111/j.1600-0889.2010.00498.x.

Pelckmans, K., J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B. Hamers, D. Moor, and J. Vandewalle, 2002. LS-SVMlab: a MATLAB/C toolbox for Least Squares Support Vector Machines, http://www.esat.kuleuven.ac.be/sista/lssvmlab, presented
15   at Neural Information Processing Systems (NIPS 2002).

Pfeil, B., A. Olsen, D. C. E. Bakker, S. Hankin, H. Koyuk, A. Kozyr, J. Malczyk, A. Manke, N. Metzl, C. L. Sabine, J. Akl, S. R. Alin, N. Bates, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, A. J. Fassbender, R. A. Feely, M. González-Dávila, C. Goyet, B. Hales, N. Hardman-Mountford, C. Heinze, M. Hood, M. Hoppema,
20   C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, S. D. Jones, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Ríos, J. M. Santana-Casiano, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, J. Tjiputra, D. Vandemark, T. Veness, R.
25   Wanninkhof, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue, 2013. A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT). Earth Syst. Sci. Data, 5: 125-143, doi:10.5194/essd-5-125-2013

Reynolds, R. W., N.A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002. An Improved In Situ and Satellite SST Analysis for Climate. J. CLIMATE, 15: 1609-1625.
30

Rödenbeck, C., R.F., Keeling, D.C.E. Bakker, N. Metzl, A. Olsen, C. Sabine, and M. Heimann, 2013. Global surface-ocean pCO2 and sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Ocean Sci., 9: 193–216.

Rödenbeck, C., D. C. E. Bakker, N. Gruber, Y. Iida, A. R. Jacobson, S. Jones, P. Landschützer, N. Metzl, S. Nakaoka, A. Olsen, G.-H. Park, P. Peylin, K. B. Rodgers, T. P. Sasse, U. Schuster, J. D. Shutler, V. Valsala, R. Wanninkhof, and J. Zeng, 2015. Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO2 Mapping intercomparison (SOCOM). Biogeosciences, 12: 7251–7278.

Sabine, C. L., S. Hankin, H. Koyuk, D. C. E. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, J. Malczyk, J. Akl, S. R. Alin, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, R. A. Feely, M. Gonz´alez-D´avila, C. Goyet, N. Hardman-Mountford, C. Heinze, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton1, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Ríos, J. Salisbury, J. M. Santana-Casiano, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, D. Vandemark, T. Veness, A. J. Watson, R. Weiss, C. S.Wong, and H. Yoshikawa-Inoue, 2013. Surface Ocean CO2 Atlas (SOCAT) gridded data products. Earth Syst. Sci. Data, 5: 145-153, doi:10.5194/essd-5-145-2013

Sasse, T. P., B.I., McNeil, and G. Abramowitz, 2013. A new constraint on global air-sea CO2 fluxes using bottle carbon data, Geophys. Res. Lett., 40: 1594-1599.

20 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, doi:10.1029/2005GB002594, 2006.

Schmidtko, S., G. C. Johnson, and J. M. Lyman, 2013. MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers. Journal of Geophysical Research, 118: 1658–1672, doi: 10.1002/jgrc.20122.

Stocker, T., Qin, D., and Platner, G.-K., 2013. Climate Change 2013 The Physical Science Basis, Cambridge University Press, Cambridge, United Kingdom.

30 Takahashi, T., S.C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R.A. Feely, C. Sabine, J. Olafsson, Y. Nojiri, 2002. Global sea-air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects. Deep-Sea Research Part II, 49, 1601-1622.

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. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, A. Körtzinger, T. Steinhoff, M. Hoppema, J. Olafsson, T.S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. Bellerby, C.S. Wong, 2009. Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans.
5   Deep-Sea Research Part II, 56: 554-577.

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.

Telszewski, M., A. Chazottes, U. Schuster, A.J. Watson, C. Moulin, D.C.E. Bakker, M. González-Dávila, T. Johannessen, A. Körtzinger, H. Lüger, A. Olsen, A. Omar, X. A. Padin, A. F. Ríos, T. Steinhoff, M. Santana-Casiano, D. W. R. Wallace, and R. Wanninkhof, 2009. Estimating the monthly pCO2 distribution in the North Atlantic using a self-organizing neural network. Biogeosciences, 6: 1405-1421, doi:10.5194/bg-6-1405-2009.

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, doi:10.5194/bg-10-1983-2013.

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

Zeng, J. Y., Y. Nojiri, P. P. Murphy, C. S. Wong, and Y. Fujinuma, 2002. A comparison of Delta pCO(2) distributions in the northern North Pacific using results from a commercial vessel in 1995-1999. Deep-Sea Research Part II, 49: 5303-5315.
25

Zeng, J., Y. Nojiri, P. Landschützer, M. Telszewski, and S. Nakaoka, 2014. A global surface ocean fCO2 climatology based on a feedforward neural network, J. Atmos. Ocean Technol., 31: 1838–1849.

Zeng, J., Y. Nojiri, S. Nakaoka, H. Nakajima, and T. Shirai, 2015. Surface ocean CO2 in 1990-2011 modelled using a feed-
30   forward neural network, Geoscience Data Journal, 2: 47–51

Table 1 Feature comparison of the three machine learning models.

SVM	FNN	SOM
Maps the input space to a high dimensional space that is proportional to the training sample size.	Maps the input space to a high dimensional space that is proportional to the number of number of hidden neurons and input variables.	Maps the input space to a feature map whose size is determined by the user.
Makes prediction by continuous interpolation.	Makes prediction by continuous interpolation.	Makes prediction by picking up labelling samples that have the closet feature to the input.
May have overfitting and over interpolation problem.	May have overfitting and over interpolation problem.	Has discontinuity problem.
Data scaling helps solving the linear equation, but has no effect on the result.	Data scaling helps the convergence of training and its effect on the results is usually insignificant.	Sensitive to the scaling of input data.
The parameter values for regularization and kernel function affects the results.	The number of hidden neurons affects the results.	The size of the feature map affects the result.

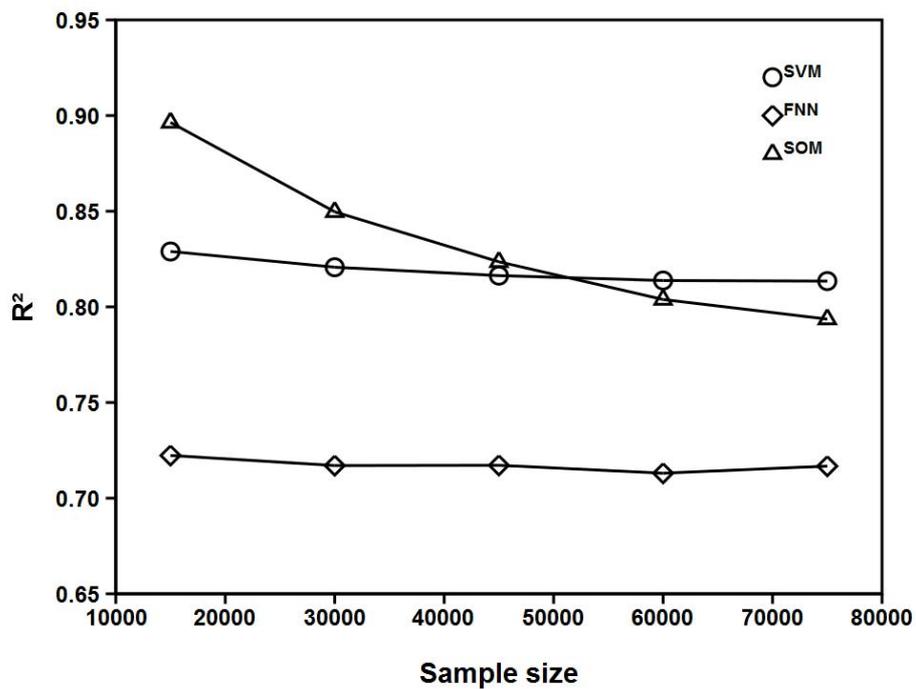
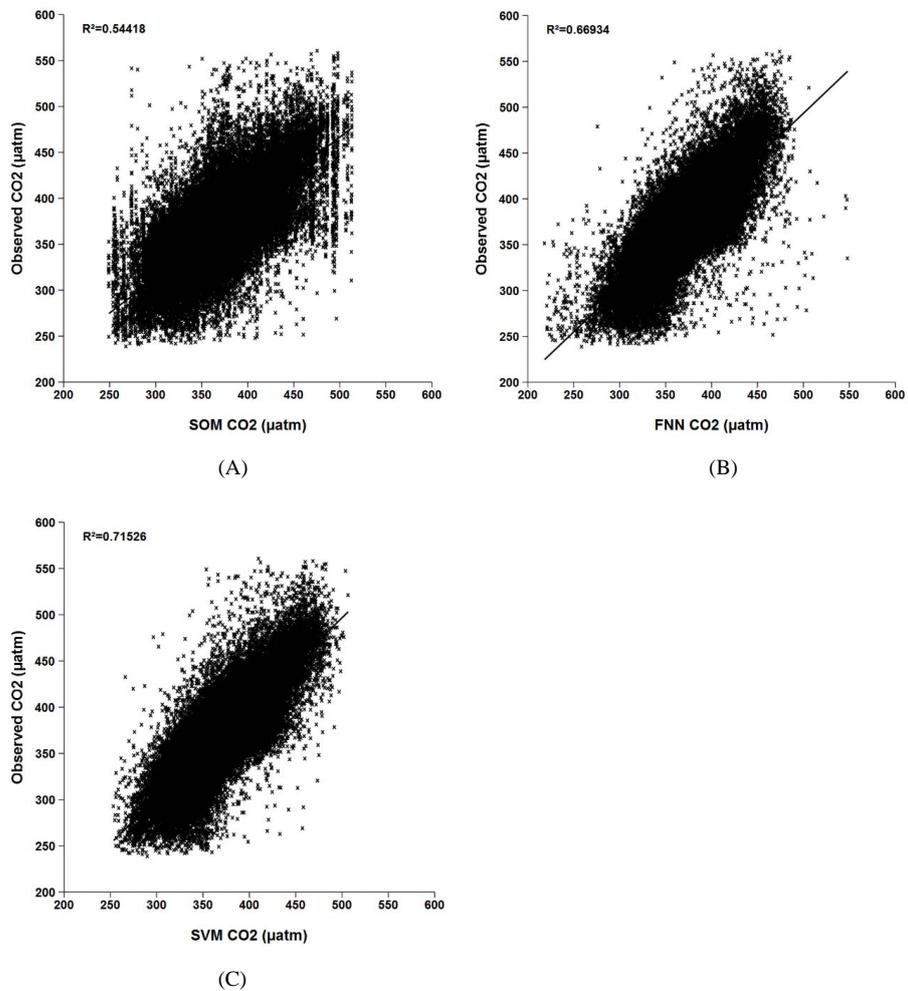



Figure 1: Correlation coefficient between modelled and observed fCO2. The sample size is the number of data points randomly selected to train FFN and SVM and to label SOM.

5   Figure 2: Predicted vs observed fCO2 (µatm). Ten percent of data points was selected randomly to train FNN and SVM and
     to label SOM, and the rest was used for validation.

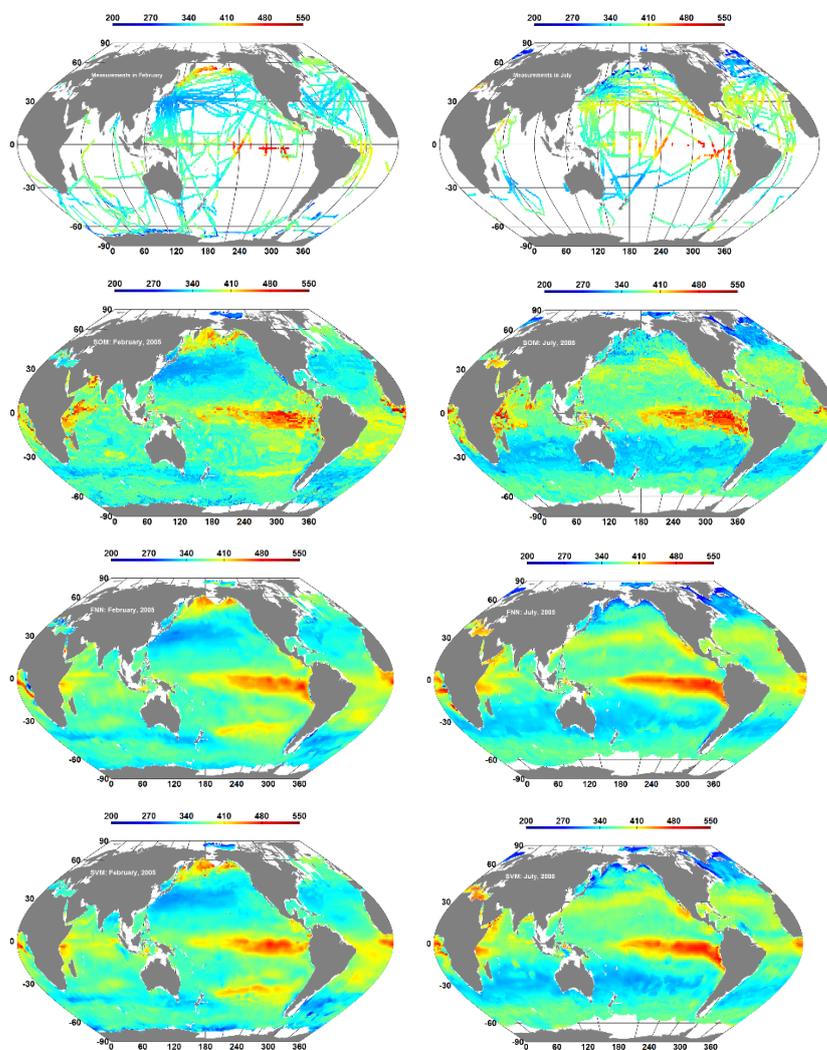



Figure 3:Distributions of modelled and observed CO2. The CO2 map for observations are the composite map with all CO2
normalized to 2005. Half of the data points were used to train FNN and SVM and to label SOM to make prediction.
The left panels show February and the right panels show July.

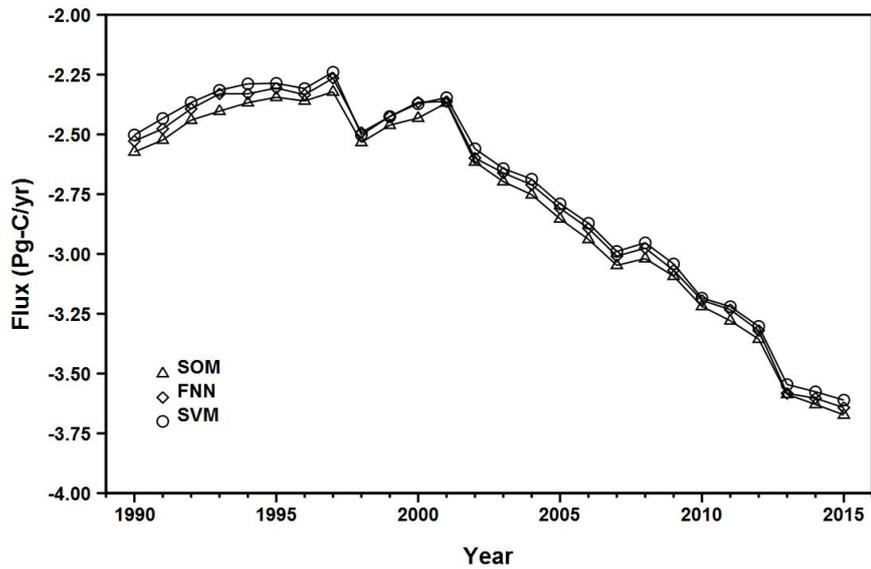



Figure 4: Modelled global CO2 fluxes. A negative value indicates oceanic uptake.