Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem through neural networks (A$_T$pH-NN): The case of Ría de Vigo (NW Spain) between 1992 and 2019

5   Daniel Broullón[1], Fiz F. Pérez[1], and María Dolores Doval[2]

[1]Instituto de Investigaciones Marinas, CSIC, Eduardo Cabello 6, 36208 Vigo, Spain

[2]Instituto Tecnológico para el Control del Medio Marino de Galicia (INTECMAR), Peirao de Vilaxoan, 36611, Vilagarcía de Arousa, Spain

10   **Abstract**

Short and long-term variability of seawater carbon dioxide (CO$_2$) system shows large differences between different ecosystems which are derived from the characteristic processes of each area. The high variability of coastal ecosystems, their ecological and economic significance, the anthropogenic influence on them and their behavior as sources or sinks of atmospheric CO$_2$, highlight the relevance to better understand the

15   processes that underlie the variability and the alterations of the CO$_2$ system at different spatiotemporal scales. To confidently achieve this purpose, it is necessary to have high-frequency data sustained over several years in different regions. In this work, we contribute to this need by configuring and training two neural networks with the capacity to model the weekly variability of pH and total alkalinity (A$_T$) in the upper 50 m of the water column of the Ría de Vigo (NW Spain), with an error of 0.031 pH units and 10.9

20   µmol kg^{-1} respectively. With these networks, we generated weekly time series of pH and A$_T$ in seven locations of the Ría de Vigo in three depth ranges (0-5 m, 5-10 m and 10-15 m), which adequately represent independent discrete measurements. In a first analysis of the time series, a high short-term variability is observed, being larger for the inner stations of the Ría de Vigo. The lowest values of pH and A$_T$ were obtained for the inner zone, showing a progressive increase towards the outer/middle zone of the ría. The

25   mean seasonal cycle also reflects the gradient between both zones, with a larger amplitude and variability for both variables in the inner zone. On the other hand, the long-term trends derived from the time series of pH show a higher acidification than that obtained for the open ocean, with surface trends ranging from -0.020 pH units per year in the outer/middle zone to -0.032 pH units per year in the inner zone. In addition, positive long-term trends of A$_T$ were obtained ranging from 0.39 µmol kg^{-1} per year in the outer/middle

30   zone to 2.86 µmol kg^{-1} per year in the inner zone. The results presented in this study show the changing conditions both in the short and long-term variability as well as the spatial differentiation between the inner and outer/middle zone to which the organisms of the Ría de Vigo are subjected. The neural networks and the database provided in this study offer the opportunity to evaluate the CO$_2$ system in an environment of

high ecological and economic relevance, to validate high-resolution regional biogeochemical models and
35 to evaluate the impacts on organisms of the Ría de Vigo by refining the ranges of the biogeochemical
variables included in experiments.

1 Introduction

Anthropogenic activities as consumption of fossil fuels, cement production and land use change have
increased the atmospheric concentration of carbon dioxide (CO_2) from 277 to 410 parts per million (ppm)
40 since 1750, stablished as the beginning of the Industrial Era (Ciais et al., 2013). This anthropogenic CO_2
circulates through the different carbon reservoirs modifying the natural carbon cycle (Friedlingstein et al.,
2019). The ocean is the largest carbon reservoir and dissolved inorganic carbon (DIC) is its predominant
pool (Ciais et al., 2013). Absorption of anthropogenic CO_2 by seawater increases the DIC pool and modifies
the natural carbon cycle in the ocean, causing chemical changes as decreasing pH and calcium carbonate
45 saturation states (Rhein et al., 2013). These changes define the concept of ocean acidification (Doney et al.,
2009) and are a major concern for marine ecosystems (Orr et al., 2005). The rate of surface ocean
acidification process detected in open ocean for the last decades is between -0.0017 and -0.0026 pH units
per year (Bates et al., 2014). In contrast to acidification rates in open ocean, positive to negative long-term
trends of pH are exhibit in different coastal ecosystems (Carstensen and Duarte, 2019), evidencing the
50 diversity of processes that can drive ocean acidification beyond the absorption of anthropogenic CO_2.

Several studies in coastal environments characterized the processes behind the variability of the CO_2 system
variables and the acidification process (Cai et al., 2021; Waldbusser and Salisbury, 2014 and references
therein). The freshwater balance between seawater and atmosphere modifies total alkalinity (A_T) and DIC
concentrations through the concentration or dilution of the ions that make up these two variables (Turk et
55 al., 2010). Similarly, continental discharge of freshwater also contributes to the variability of the CO_2
system variables (Joesoef et al., 2017; Salisbury et al., 2009). In addition, these fluxes from land to sea have
very different concentrations of A_T, DIC, nutrients and multiple dissolved and particulate organic
substances that play a role in the variability of the CO_2 system (Carstensen et al., 2018; Carstensen and
Duarte, 2019; Joesoef et al., 2017). Biological processes as respiration, primary production or calcium
60 carbonate precipitation are important drivers of the variability of the CO_2 system at different time scales,
mainly because of the consumption and production of CO_2 (Cai et al., 2020; Cyronak et al., 2018;
Waldbusser and Salisbury, 2014). Upwelling regions have an extra source of variability since this process
brings waters to the surface with low pH, high CO_2 and high nutrient concentration (Cai et al., 2020; Feely
et al., 2010). The interaction between all the previous processes, their variability at different time scales
65 and the anthropogenic alterations are reflected in the complex variability of the CO_2 system variables.

Coastal ecosystems have gained attention in the last years in the evaluation of the impacts of ocean
acidification on marine life (e.g., Barton et al., 2012, 2015; Hendriks et al., 2015; Sunday et al., 2017).
Nevertheless, a previous effort to better understand the changing carbon cycle in these environments of
high variability and for this obtain sustained high-frequency measurements of CO_2 chemistry system
70 variables is still needed to elucidate which processes and how they drive both the short-term variability and
long-term trends in coastal CO_2 chemistry and properly address the impacts (Hofmann et al., 2011;
Waldbusser and Salisbury, 2014). The manner at which physical, chemical, biological, geological and

anthropogenic processes interact, generates totally different characteristics among coastal ecosystems that
highlights the need to study the peculiarities of each one to understand the variability of the CO_2 system in

75   detail.

The present study is centered in the Ría de Vigo (NW Spain), a coastal embayment affected by the Canary
Current upwelling system. The upwelling season typically occurs between March and October (Figueiras
et al., 2002), bringing high CO_2, low pH and nutrient-rich waters from the Eastern North Atlantic Central
Water (Fraga, 1981; Ríos et al., 1992) to the Ría de Vigo and promoting a positive estuarine circulation.

80   From November to March, poleward winds change the dynamics of the outer part of the ría to a negative
estuarine circulation and the inner part conserves the positive circulation forced by continental freshwater
inputs (Álvarez-Salgado et al., 2000; Piedracoba et al., 2005). The nutrient enrichment because of
upwelling, but also because of terrestrial run off (Nogueira et al., 1997), is translated into high primary
production rates supporting a variety of fisheries in the area, which are very important for local economy

85   (Fernández et al., 2016).

The main processes affecting the Ría de Vigo and its interaction are drivers of the variability of the CO_2
system, as previously showed by the studies cited for other coastal locations. Very few studies have focused
on analyzing the variability of the CO_2 system in the Ría de Vigo. Essentially, the scarce studies have
assessed the short-term variability of partial pressure and fluxes of CO_2 (Gago et al., 2003; Pérez et al.,

90   1999). The large short-term variability and the long-term trends presented by the drivers of the CO_2 system,
as the weakening of the upwelling intensity (Lemos and Sansó, 2006; Pardo et al., 2011; Pérez et al., 2010),
warming (Lemos and Sansó, 2006; Pérez et al., 2010), increased nutrients (Doval et al., 2016) and
deoxygenation (Padin et al., 2020), show the need to obtain high-frequency data over a long period of time
to robustly evaluate the variability of the system at different time scales.

95   In this study, we reconstruct time series of in situ pH on the total hydrogen ion scale and A_T at seven
locations in the Ría de Vigo in three depth ranges (0-5 m, 5-10 m and 10-15 m) with a weekly frequency
between 1992 and 2019 to provide a database for the detailed analysis of the CO_2 system in this coastal
environment of great ecological and economic significance. The non-linearity of the CO_2 system leads us
to use the ability of neural networks to capture the nonlinear relationships among different variables and

100   thus obtain the time series with a low uncertainty in this highly dynamic system. Based on the good results
of several studies to model variables of the CO_2 system using feed-forward neural networks both in open
ocean (Bittig et al., 2018; Broullón et al., 2019, 2020a; Landschützer et al., 2013; Velo et al., 2013) and in
coastal environments (Li et al., 2020a, 2020b), and the higher performance showed among other methods
(ocean: Velo et al., 2013; coastal: Li et al., 2020a), we selected this type of neural network for the obtention

105   of the time series.

2 Methodology

2.1 Neural Networks

Feed-forward neural networks are powerful tools to model complex nonlinear functions (Hagan et al.,
2014). A neural network is fed with data pairs of inputs and a target in order to obtain a global relation

110   between both type of variables to compute the target variable from independent data with considerably low

error. The network structure is made up of different layers with specific tasks (Fig S1a). The first layer (input layer) stores the input data and sends them to the next layer. In the second layer (hidden layer), a tunable number of neurons receive the data from the previous layer after being weighted by a different coefficient for each input variable and for each neuron in which the information goes in. Inside each neuron (Fig S1b), the weighted data is summed and passed through an activation function, which typically is a sigmoid function, to obtain an output. The network can have several hidden layers depending on the problem to be solved. Finally, if a single target variable is being modeled, the final layer (output layer) contains one neuron with a linear activation function to obtain the desired output, following the process explained for a neuron of the hidden layer.

The goal of a neural network like the previously described is to reach output values closest to the target ones, that is, with a low difference between them. To achieve the goal, this difference is backpropagated through the network to adjust the weights in order to reduce the error through the minimization of a cost function in multiple iterations of the procedure described so far. This process is known as training and its reliability must be tested in independent data to avoid the overfitting of the training data, that is, the output of the network on independent data should have a low error (good generalization). A full description on how the error is reduced through the backpropagation algorithm and the different stopping conditions of the training process can be consulted in Hagan et al. (2014).

In the present study, multiple feed-forward neural networks with one hidden layer were configured to extract the relations between different combinations of input variables: temperature (T), salinity (S), phosphate (P), nitrate (N), silicate (Si), dissolved oxygen (O), position (LL: latitude and longitude; D: depth) and time (Y: year; W: week); and the target ones: pH and A_T. These variables were used as inputs to the networks in the combinations as follows, given their influence in the variability of the target variables, which have been shown in different studies (Sect. 1): DTS, DTSN, DTSO, DTSON, DTSOPN, DTSPN, DTSPNSi, DTSPNSiO, DTSPNSiOW, DTSPNSiOY, DTSPNSiOYW, DTSPNSiW, DTSPNSiY, DTSPNSiYW, LLDTSPNSi, LLDTSPNSiO, LLDTSPNSiOW, LLDTSPNSiOY, LLDTSPNSiOYW, LLDTSPNSiW, LLDTSPNSiY and LLDTSPNSiYW. The position and time were included to disentangle the different processes occurred in different locations of the Ría de Vigo, which may not always be fully captured by biogeochemical variables (Broullón et al., 2020a), such as continental runoff (Gago et al., 2005; Perez et al., 1992). The periodicity of the input week was represented by:

$$cW = \cos\left(\frac{\pi}{53/2} \cdot week\ number\right) \qquad (1)$$

$$sW = \sin\left(\frac{\pi}{53/2} \cdot week\ number\right) \qquad (2)$$

For each combination of the input variables, the tested number of neurons in the hidden layer were: 28, 34, 40, 46, and 52. The number of neurons was selected increasing and decreasing by six the first number tested and testing in which direction the error in the independent dataset decreases to select the following number to test. A sigmoid activation function was selected for the hidden layer, since together with the linear function of the output layer make the network able to fit almost any continuous function (Hagan et al., 2014). The Levenberg-Marquardt algorithm (Hagan and Menhaj, 1994; Levenberg, 1944; Marquardt, 1963)

was used for the training based on several good results obtained in similar studies (Broullón et al., 2019, 2020a; Landschützer et al., 2013; Li et al., 2020a, 2020b; Velo et al., 2013). The training process with this algorithm was carried out in the MATLAB software with the *trainlm* function, detailed in Beale et al. (2018).

The dataset used to train the networks (Sect. 2.2) was divided in three sets: data used to assess the generalization of the network (testing dataset; 10% of the total data); data used to train the network (90% of the total data), which was divided in: training (85%) and validation (15%) datasets. The training dataset is used to extract the relations between the input variables and the target. The validation dataset is used to stop the training process when the performance on this dataset does not improve during six consecutive iterations of the training process.

2.2 Data

The neural networks designed in the present study were trained with data collected by different cruises carried out by the Instituto de Investigaciones Marinas (IIM), dependent of the Consejo Superior de Investigaciones Científicas (CSIC) from 1976 to 2018. Temperature, salinity, phosphate, nitrate, silicate, dissolved oxygen, pH and A_T were measured in discrete samples from multiple profiles around the waters off the Northwestern Iberian Peninsula, mostly collected in the Ría de Vigo (Fig .1). We selected the data inside the Ría de Vigo and in the upper 50m of the water column to specifically capture the processes that control the variability of A_T and pH in the area of study (Fig .1). A full description of this database can be consulted in Padin et al. (2020), where sampling techniques are widely described.

The relations extracted with the neural networks from the previously mentioned database were used to create the time series of pH and A_T at seven locations in the Ría de Vigo (Fig. 1). The data used to feed the trained neural networks and to obtain the pH and A_T values were measured by the Instituto Tecnolóxico para o Control do Medio Mariño (INTECMAR) in a weekly basis from 1992 to 2019 (INTECMAR-database). Temperature, salinity, phosphate, nitrate, silicate and dissolved oxygen were measured in three different depth ranges: 0-5 m, 5-10 m and 10-15 m. Temperature, salinity and oxygen were averaged from the CTD SBE 25 with an SBE 43 oxygen sensor that recorded data in each 0.1 m. Sampling for the analysis of inorganic nutrients (phosphate, nitrate and silicate) was performed using a PVC hose that collected water from three depth intervals (0-5 m, 5-10 m and 10-15 m). Analytical techniques were based on colorimetric methods of continuous segmented flow analysis (CFA). The precisions are 0.1 µM for nitrate, 0.01 µM for phosphate and 0.06 µM for silicate. More details of this database can be consulted in Doval et al., (2016), where the sampling and analytical techniques were extensively described. Time series of these variables were compared with discrete data from the IIM-database to assess their consistency (Appendix A). Temporal gaps in nutrients time series were filled using neural networks (Appendix B). Furthermore, several inconsistencies between the dissolved oxygen in the INTECMAR-database and the one in IIM-database led us to reconstruct the time series of this variable (Appendix C).

2.3 Time series reconstruction

185

Weekly time series of pH and A_T were obtained passing through the neural networks the inputs from the INTECMAR-database. The measured pH and A_T from the IIM-database were used to validate the neural network derived time series. The validation data used for each time series station is depicted in Fig. S2. Distances from time series stations to validation points were selected based on a trade-off between spatial

190 variability and the availability of samples with high temporal resolution around the stations. As an example, station V1 (Fig. S2) has a wider validation area than others to capture discrete samples with both monthly and weekly time resolution.

The seasonal cycle of pH and A_T was obtained averaging the time series to describe the annual amplitude, the variability and the differences among stations. The long-term trends were also obtained for the two

195 variables to show the possible future conditions in the Ría de Vigo. A seasonal detrending to remove the seasonality following the detrending method used by Bates et al. (2014) was applied to obtain the trends. The uncertainty of the trends was obtained based on a Monte Carlo simulation. To properly compute the aforementioned values, the outliers were removed from each time series based on the scaled median absolute deviation (sMAD; Hampel, 1974), which has been proposed as a robust outlier estimator (e.g.,

200 Leys et al., 2013), following Eq. (3):

$$sMAD = \frac{1}{0.6745} * median(|x_i - median(x)|) \qquad (3)$$

where x_i is each value of the time series and x the full time series. Outliers were those values that exceeded more than three sMAD away from the median.

3 Results and discussion

205 **3.1 Neural network mapping**

Figure 2 shows the RMSE and the determination coefficient (r^2) obtained for each combination of inputs for pH and A_T. In general, these statistics improve with increasing the number of inputs to the network, with no significant differences among the tested number of neurons. Considering these averaged statistics of the three networks with the best performance (lowest RMSE in the test dataset) for each number of

210 neurons, the networks with the combinations of inputs LLDTSPNSiOWY for pH and LLDTSPNSiWY for A_T reach the best fits (Figs. 2b and 2c). The inclusion of dissolved oxygen as an input was evaluated (Figs. 2b and 2d) because of the large differences found between the measured dissolved oxygen by IIM and by INTECMAR (Appendix A), and therefore to evaluate the need of reconstruct the time series of dissolved oxygen (Appendix C). For pH, the inclusion of dissolved oxygen considerably improves the mapping for

215 the different combinations (Fig. 2b). This result is probably because of the high correlation between the variability of dissolved oxygen and pH through the organic matter formation and degradation, this process being more remarkable in productive coastal environments (Baumann and Smith, 2018; Cai et al., 2011; O'Boyle et al., 2013; Padin et al., 2020). For A_T, the addition of the input dissolved oxygen does not result in a general improvement (Fig. 2d). A_T variability is mainly controlled by freshwater fluxes and therefore

220 reflected in salinity changes (Beldowski et al., 2010; Millero et al., 1998), with a minor contribution of the

biological activity and dissolution and precipitation of biogenic calcium carbonate (Brewer et al., 1975; Brewer and Goldman, 1976; Cai et al., 2010).

When each network is individually analyzed, other combinations of inputs show better statistics (Table 1) than the ones of the previously mentioned combinations, where averages of the three best networks per each number of neurons were analyzed. Based on these results, a trade-off between the RMSE and r^2 in the test dataset and the most accurate representation of the validation data when constructing the time series of pH and A_T (Sect. 3.2), was followed to select the networks which best represent the weekly variability of pH and A_T from 1992 to 2019. Considering this evaluation, DTSPNSiOYW trained with 28 neurons in the hidden layer for pH (hereinafter referred to as pH_NN) and LLDTSPNSiYW trained with 52 neurons in the hidden layer for A_T (hereinafter referred to as A_T_NN) were selected to reconstruct the time series (Table 1).

The modeled pH and A_T with pH_NN and A_T_NN show the largest differences with the measured values in samples where temperature or dissolved oxygen are higher than 16 ºC or 300 µmol kg^{-1} respectively, and in those where salinity, nitrate or depth are lower than 35, 2.5 µmol kg^{-1} or 15 m respectively (Fig. 3). This result suggests the same biogeochemical conditions are similarly affecting the modeling of the two variables. Nevertheless, most of the differences in these samples are not very large, considering the high variability in the Ría de Vigo regarding both the physical (Nogueira et al., 1997) and the biogeochemical (Doval et al., 2016; Nogueira et al., 1997) context. In samples where the measured pH is lower than 8 and higher than 8.1 and A_T is lower than 2300 µmol kg^{-1}, the computed values are slightly biased (Fig. 3). These biases are mainly determined by samples where absolute differences between the measured and modeled values are higher than 3 times the root mean square error (3RMSE), which represent 1% of the samples used to create each network. For pH, 48% of the samples with differences higher than 3RMSE in the biased ranges are in the 0-15 m layer with nitrate concentrations lower than 2.5 µmol kg^{-1}. For A_T, 86% of the samples with differences higher than 3RMSE in the biased range are in the 0-10 m layer with salinities lower than 35 and most of them with nitrate concentrations lower than 2.5 µmol kg^{-1}. Although it is important to know the conditions where the neural networks present the largest differences with the measured data, these samples only represent 1.7% for pH and 5.7% for A_T of the total samples in the mentioned ranges, suggesting that pH and A_T values with low uncertainties can be obtained for very different conditions.

The importance of each input variable extracted with the connection weight approach (Olden and Jackson, 2002) (Table 2) reveals that pH is mainly modeled through salinity, temperature and year. Depth, sW and dissolved oxygen are the second group of variables with the largest influence on the modeled pH. Nutrients and cW are the input variables with the lowest relative importance. On the other hand, the modeled A_T is highly influenced by silicate, with a secondary importance of phosphate and with temperature and year in the third place of relevance. The other input variables have a low influence. Although surface A_T is highly correlated with salinity worldwide (e.g., Millero et al., 1998), the low influence of this input variable in the modeled A_T does not reduce the high correlation between the two variables (measured data $r = 0.96$;

modeled data $r = 0.97$) nor the linear regression (measured data $A_T = 61.9 \cdot S + 132$; modeled data $A_T = 62 \cdot S + 129$).

260 **3.2 Time series reconstruction**

Time series of pH and A_T obtained with the neural networks over the INTECMAR-database are depicted in Figs. 4 and 5 for the seven stations in the 0-5 m range. The other two depth ranges have a similar pattern but there is a general pH decrease and A_T increase with depth, together with a reduction of the amplitude of the annual cycles (Figs. S3, S4, S5 and S6). In general, the time series agree with the validation data for

265 both variables (Figs. 4, 5, S3, S4, S5 and S6), although there are some extreme values of A_T which are not fully captured and few outliers (not shown in the figures) in samples where one or more input variables have extreme values or fall out of the range of the training data. In addition to the error of the networks, some of the differences between the measured and the computed data come from the high spatiotemporal variability of the area, since validation data are not exactly from the same location and day of the year as

270 data from the INTECMAR-database. The large temporal variability can be seen from the validation data of V1 between 1994 and 1995, where sampling with a 3-4 days frequency shows changes of approximately 0.2 pH units at that frequency (Fig. 4).

pH time series present a clear annual cycle in all the stations, which is similar year by year (Fig. 4), whereas the annual cycle of the A_T time series is quite variable mainly because of the different magnitude of short

275 events of strong decrease in A_T in some years (Fig. 5). In general, strong decreases of A_T are coincident with strong decreases of salinity, remarking the high correlation between both variables. Validation data in years with a weekly or monthly frequency allow to show how both the amplitude and timing of the annual cycle are adequately represented by the computed time series (Figs. 4, 5, S3, S4, S5 and S6). All the stations present a high short-term variability for the two variables (Table 3). This rate of change is larger in the

280 inner stations than in the outer ones in the 0-5 m layer (Table 3), showing the higher variability of the inner zone of the Ría de Vigo. In the 10-15 m layer, the rate of change presents the mentioned difference between zones for A_T but to the contrary for pH (Table 3).

Differences in the magnitude of the values can be appreciated between the outer and the inner zones (Fig. Figs. 4, 5, S3, S4, S5 and S6). The inner zone presents lower values of pH and A_T than the outer one. Lower

285 pH values could be partially explained by the stronger upwelling intensity in the inner zone (Álvarez-Salgado et al., 2008). Upwelled waters present low pH values as well as dissolved oxygen concentrations because of the remineralization of organic matter, as can be noted from the time series of dissolved oxygen (Fig. C2). The lower values of A_T in the inner zone can be essentially explained by the high positive correlation between this variable and salinity, since the presence of a river in the innermost part of the Ría

290 de Vigo generates a salinity gradient with the lowest values in the inner zone.

3.2.1 Seasonal cycle

The surface mean seasonal cycle of pH and A_T depicted in Figs. 6 and 7 allows to distinguish a clear difference in both the amplitude and the variability of the two variables for the different stations, being larger for the inner ones (Table 4). The higher variability of the physical and biogeochemical processes in

295 the inner zone, which is reflected in both the amplitude of the seasonal cycle and the variability for the

input variables (Table S1), is transferred to the computed variables resulting in the aforementioned difference between zones. In general, a reduction of both the amplitude and the variability with depth is obtained for pH and A_T (Table 4). The reduction of the amplitude of the seasonal cycle is also obtained for all the input variables, except for dissolved oxygen in the inner stations (Table S1).

300 Maxima of pH occur in March and April and minima in October and November (Fig. 6). In general, the seasonal cycle is highly correlated with that of dissolved oxygen (Fig. S7), which is mainly driven by the net balance of production and respiration of organic matter. High photosynthetic activity of the phytoplankton community is maintained in the Ría de Vigo from early spring to late summer thanks to influence of the upwelling events (Cermeño et al., 2006; Tilstone et al., 1999), which in average occur

305 between March and October (Figueiras et al., 2002), fostering high values of dissolved oxygen (Fig. S7) and increasing the pH through the removal of dissolved CO_2. The steepest decline of pH in the inner stations is probably due to the positive estuarine circulation, which brings deep waters to the surface with low pH and low dissolved oxygen because of the remineralization of the organic matter through the water column in the outer zone of the Ría de Vigo (Alvarez-Salgado et al., 1993). The slower decline in the outer/middle

310 zone reflects the contribution of the photosynthetic activity in the transit of the surface inner waters with low pH to the outer zone of the Ría de Vigo.

Maxima of A_T seasonal cycle emerge in August and September, but minima are not clearly defined since low concentrations are maintained between November and May (Fig. 7). This feature is determined by the correlation between A_T and salinity, reflecting the net freshwater input through the year, with high salinity

315 values in summer due to the excess of evaporation and low values in winter and spring determined by precipitation and continental runoff (Fig. S8). The higher amplitude and variability in the inner zone (Table 4 and Fig. 7) are mainly determined by the hydrological cycle of the river which flows into the inner zone of the Ría de Vigo.

Spatial differences in the magnitude of the seasonal cycle of pH and A_T follow a longitudinal gradient,

320 showing higher values in the outer zone with a progressively decrease towards the inner zone (Fig. 8). As previously remarked, these differences mainly reflect the dissolved oxygen gradient for pH and the salinity gradient for A_T, since both input variables decrease towards the inner zone (Table S1 and Figs. S7 and S8). These spatial differences in the CO_2 system variables generate a gradient in the buffer capacity between the outer and the inner zones. The ratio DIC/A_T, which can be seen as an indicator of the sensitivity of seawater

325 to changes in CO_2 since the Revelle factor is proportional to it (Egleston et al., 2010), is higher in the inner zone (0.91) than in the outer one (0.89), indicating a lower capacity of the seawater of the inner zone to buffer changes because of the increasing anthropogenic CO_2.

3.2.2 Long-term trends

pH trends reveal a stronger acidification in the Ría de Vigo than the one found in open ocean. The trends

330 range between -0.019 and -0.037 pH units per year, except for station V7 where the unique statistically significant trend is positive (Table 5). Spatial differences are also obtained, with the inner zone presenting the higher trends. The acidification process is increased with depth for all the stations where the pH trends are statistically significant. For comparison, two ocean time series stations located in the North Atlantic

335 Ocean present pH trends of -0.017 and -0.018 pH units per year in BATS and ESTOC sites respectively (Bates et al., 2014). In coastal environments, higher acidification rates can be found in very different ecosystems (e.g., Kapsenberg et al., 2017; Wootton and Pfister, 2012), although there are very different trends, including positive ones, depending on the specific conditions of the area (Carstensen et al., 2018; Carstensen and Duarte, 2019).

340 An analysis of the average pH for each decade reveals that the acidification process is stronger for all the stations between the 90s and the 00s (Fig. 9). Furthermore, the lack of data in 1990 and 1991 would likely increase the pH difference between both decades because of the probably lower pH values for those two years that in the following ones. It can also be appreciated that the fact that pH trends are larger in the inner zone than in the outer/middle one (Table 5) increases the difference in the average pH between both zones from 90s to 10s from 0.082 to 0.110 pH units.

345 Higher long-term trends and lower pH values in coastal environments than open ocean have been attributed to different processes, as upwelling and eutrophication (Cai et al., 2011; Feely et al., 2010; Wallace et al., 2014). The trends obtained here for the Ría de Vigo are probably explained by the two mentioned processes. Upwelling brings waters to the surface with low pH and high nutrients concentration. In addition, nutrients from other sources fertilize the area promoting the increase of primary production and the consequent
350 remineralization, decreasing dissolved oxygen concentrations and pH. Fernández et al. (2016) categorized that allochthonous sources of inorganic nitrogen correspond to coastal upwelling in 76%, wastewater in 22% and continental and atmospheric sources in 2%. Positive and statistically significant long-term trends of nitrate in all the stations (except for V7) ranged between 0.013 and 0.024 μmol kg^{-1} reveal that some or all of these sources have increased their contribution year after year in the analyzed period. In addition,
355 negative and statistically significant long-term trends of dissolved oxygen in all the stations (except for V7, where trends are positive) ranged between -0.13 and -0.47 μmol kg^{-1} show a deoxygenation, probably because of a higher remineralization activity. The increased remineralization could be a consequence of the weakening of the upwelling intensity in the region (Pérez et al., 2010; Pardo et al., 2011), driving a higher renewal time of the waters of the Ría de Vigo (Álvarez-Salgado et al., 2008). This explanation is only a
360 first approximation, since a driver analysis in future studies will elucidate and quantify the importance of each process behind the long-term trends obtained in the present study.

Most of A$_T$ trends indicate an annual increase of this variable, although one statistically significant negative trend is obtained in V7 in the 10-15 m range, showing the decrease of A$_T$ in the period from 2007 to 2019 (Table 5). As for pH, the largest trends are in the surface of the inner zone and decrease outwards. The A$_T$
365 trends cannot be explained by salinity trends (except in V7), contrary to the expected by the correlation between both variables. The higher trends in the inner zone suggest that an increase in the riverine A$_T$ could be contributing to generate them. This process has been associated to positive A$_T$ trends in other locations (Kapsenberg et al., 2017). Nevertheless, it is necessary a detailed analysis to find what processes cause the trends, but this is beyond the scope of the present study.

370

4 Conclusions

A reconstruction of time series of pH and A_T in the Ría de Vigo has been presented. Neural networks trained with discrete samples from the region of study for both variables allowed to add high accurate values of pH

375 and A_T to the INTECMAR-database, which is composed by weekly oceanographic measurements at seven locations in the Ría de Vigo from 1992 to 2019.

A first analysis of the product exposed differences between the inner and the outer/middle zone of the Ría de Vigo, where a larger variability and amplitude of the seasonal cycle and lower values for pH and A_T were obtained for the inner zone. Furthermore, significant negative long-term trends were obtained for pH

380 with a higher magnitude than the ones typically found for open ocean, revealing a higher acidification for this coastal ecosystem. On the other hand, significant positive long-term trends were obtained for A_T. A spatial gradient was also revealed for the magnitude of the trends presented by the two variables, increasing towards the inner zone.

The database provided in the present study offers an opportunity to evaluate the variability of the CO_2

385 system in an ecologically and economically important ecosystem as the Ría de Vigo. It can be very useful in regional and high-resolution modeling, in driver analysis for different variables of the CO_2 system, as some of high relevance for calcifying organisms like calcium carbonate saturation state, and to stablish more accurate conditions in experiments evaluating the impacts of climate change in the organisms living in the Ría de Vigo.

390 **Appendix A: Consistency assessment**

Time series of the input variables from the INTECMAR-database have been compared to discrete measurements from the IIM-database to assess the consistency between both datasets. This analysis is necessary because of the differences in the methods used to measure some of the variables. For example, dissolved oxygen in IIM-database was measured using the Winkler method, as others described in Padin et

395 al. (2020), which generally provide more accurate values than the measured in the INTECMAR-database through a SBE 43 dissolved oxygen sensor integrated in a CTD.

Temperature comparison shows a good agreement between the two datasets (Fig. A1). The typical annual cycle with the highest values in summer and the lowest in winter is depicted in the two datasets. Furthermore, episodes of temperature decrease in the warm season because of upwelling events are well

400 defined in the two datasets.

In general, salinity from the two datasets also shows a good agreement (Fig. A1). Nevertheless, there are some discrepancies in all the stations in the winter of 2018, where salinity in the IIM-database shows a homogeneous increase from winter to spring and in the INTECMAR-database a high short-term variability with an extremely high peak (which is still higher with depth) was obtained. These differences are probably

405 because of some problem (e.g., bad calibration) in the conductivity sensor of the CTD used by INTECMAR and are probably translated to the computed values of pH and A_T.

Nutrients are very consistent between the two databases (Fig. A1), probably because of the high accurate methods used by the two institutions to measure them. Nevertheless, very few extremely high peaks are

present in the INTECMAR-database in both nitrate and silicate time series, which differ considerably from

410   the largest values in the IIM-database. For this reason, pH and A_T derived from these samples were flagged as questionable in the database generated from this study. It should be noted the lack of measurements of phosphate before 1995 and several gaps in both nitrate and silicate time series in the INTECMAR-database, which led us to fill the time series using neural networks (Appendix B).

Dissolved oxygen is the variable that shows the largest discrepancies (Fig. A1). On one hand, data from the

415   INTECMAR-database present considerably higher values for several years before 2004 than the ones measured in the following 16 years of the time series. These values are also higher than the ones available in the IIM-database before 2004. On the other hand, the INTECMAR-database contains lower values of dissolved oxygen than the ones measured in the IIM-database in 2018 and 2019. Considering that dissolved oxygen data in the IIM-database were obtained through accurate methods and the ones in the INTECMAR-

420   database were obtained from an oxygen sensor in a CTD, we decided to discard the dissolved oxygen data from the INTECMAR-database and obtain new values using neural networks (Appendix C).

Appendix B: Gap filling in nutrient time series

Time series of nutrients measured by INTECMAR have some gaps which reduce the time range and the time resolution (Fig. A1). To fill these gaps, different neural networks were configured. As nutrient

425   measurements in the INTECMAR-database are consistent with the ones in the IIM-database (Fig. A1), we decided to train the neural networks with data from the two databases to obtain highly representative relations of the sampling sites between inputs and targets and properly fill the gaps in the time series. The inputs used for the three modeled nutrients were position (latitude, longitude and depth), temperature, salinity, nitrate (to model phosphate and silicate since it was measured where there are gaps in these two

430   variables), silicate (to model nitrate since it was measured where there are gaps in this variable) and week of sampling (as in Eq. (3)). The position and time were included to disentangle the different processes occurred in different locations of the Ría de Vigo, which may not always be fully captured by biogeochemical variables, such as continental runoff (Pérez et al., 1992; Gago et al., 2005). We decided to take as inputs all the possible variables related to nutrients variability based on the results of Sect. 3.1,

435   where the increase in the number of input variables improves the neural network fitting. The number of neurons tested were: 10, 16, 22, 28, 34, 40 and 46; following the same method described in Sect. 2.1. The division of the dataset to train the network, the training algorithm and the stopping rule were the same as mentioned in Sect. 2.1.

The RMSE and r^2 in the test dataset for the number of neurons tested were between: 0.13-0.14 µmol kg^{-1}

440   and 0.66-0.73 for phosphate; 1.5-1.6 µmol kg^{-1} and 0.70-0.84 for nitrate; and 1.5-1.7 µmol kg^{-1} and 0.75-0.8 for silicate respectively. There are not significant differences among the RMSE for each number of neurons and slight differences in the r^2. We decided to select the networks with the best statistics regarding these two parameters. The selected networks show that the inclusion of a nutrient as an input is relevant to obtain accurate values of the modeled nutrients as it is showed by the relative importance of the input

445   variables (Table B1).

In general, all the networks adequately represent the time series from the INTECMAR-database (Fig. B1), except the extremely high peaks described in Appendix A, which are not fully represented. The filling of phosphate time series before 1995 captures the annual cycles and their amplitude as it is showed by the consistency with the discrete samples from the IIM-database (Fig. B1). The gaps filled in both the nitrate

450   and silicate time series also represent the annual cycles and their amplitude with a similar magnitude than in the time series from the INTECMAR-database (Fig. B1). Unfortunately, there are no samples from the IIM-database where there are gaps in these time series, but the filling through the neural networks is clearly consistent with the INTECMAR-database, which have been shown in Appendix A the high quality of the nutrient measurements.

455   **Appendix C: Reconstruction of dissolved oxygen time series**

The inconsistencies described in Appendix A between time series of dissolved oxygen from the INTECMAR-database and discrete data from the IIM-database, together with the correlation between dissolved oxygen and pH (Baumann and Smith, 2018; Cai et al., 2011; O'Boyle et al., 2013; Padin et al., 2020), show the need to obtain higher-quality time series of dissolved oxygen. For this purpose, we trained

460   different neural networks using the IIM-database and the same configuration described in Sect. 2.1 for the division of the dataset, the training algorithm and the stopping rule. The input variables were tested in the following combinations: DTS, DTSN, DTSPN, DTSPNSi, DTSPNSiW, DTSPNSiY, DTSPNSiYW, LLDTSPNSi, LLDTSPNSiW, LLDTSPNSiY and LLDTSPNSiYW. The number of neurons tested were: 10, 16, 22, 28, 34, 40 and 46 following the same method described in Sect. 2.1.

465   As in Sect. 3.1, the RMSE and the r^2 improves with increasing number of inputs, when the statistics of the three networks with the best performance in the test dataset are averaged (Fig. C1). For the best combination of inputs, that is, LLDTSPNSiYW, we selected one of the ten networks trained with 28 neurons to reconstruct the time series of dissolved oxygen, since the computed values with this network showed the highest correlation with the measured data in the test dataset (RMSE: 16.2 µmol kg^{-1}; r^2: 0.85). The relative

470   importance of each input variable (Table C1) shows that salinity and phosphate are the main variables used by the selected network to compute dissolved oxygen.

The reconstructed time series of dissolved oxygen (Fig. C2) were computed used as nutrient inputs the ones of the filled time series in Appendix B so as not to reduce the time range and resolution. In addition, the gaps in dissolved oxygen time series in several months in the years 2000 and 2001 were filled in thanks to

475   the neural network and the availability of measurements of the input variables.

The comparison between the time series of dissolved oxygen from the INTECMAR-database, those reconstructed with the neural network created in this section and the discrete data from the IIM-database shows that the reconstructed time series better represent the discrete measurements for all the stations (Fig. C2). More precisely, the good correlation between the discrete measurements of dissolved oxygen and the

480   computed time series for those dates where data from the INTECMAR-database present anomalies (Appendix A) clearly reflects the bad quality of the dissolved oxygen from the INTECMAR-database on those dates. The unreliability of the positive anomalies in the INTECMAR-database before 2004 can be seen at V1 and V6 around 1995 (Fig. C2) when compared with discrete data from the IIM-database.

Similarly, the unreliability of the negative anomalies can be seen at V5 and V6 around 2018 (Fig. C2). With
485 these results, we demonstrated that the new dissolved oxygen time series may be useful in obtaining the
time series of the target variables of this study: pH and A_T.

Code availability

The MATLAB code used in the present study is available through the data repository of the Spanish
National Research Council (CSIC; http://dx.doi.org/10.20350/digitalCSIC/12642; Broullón et al., 2020b).
490 The two neural network objects used to compute pH and A_T are also included.

Data availability

The database used to train the neural networks and a complete dataset including all the input variables from
the INTECMAR-database and the computed pH and A_T is offered through the data repository of the Spanish
National Research Council (CSIC; http://dx.doi.org/10.20350/digitalCSIC/12642; Broullón et al., 2020b).
495 It should be noted that pH and A_T data are flagged with 0: acceptable and 1: questionable. Questionable
data includes both outliers detected though Eq. (3) and data obtained from inputs where at least one of them
is outside of the range of the training dataset.

Author contributions

DB and FFP designed the study. The manuscript was written by DB and revised and discussed by all the
500 authors. The dataset and the neural networks were created by DB and MD.

Competing interests

The authors declare that they have no conflict of interest.

References

505 Alvarez-Salgado, X. A., Roson, G., Perez, F. F. and Pazos, Y.: Hydrographic variability off the
Rias Baixas (NW Spain) during the upwelling season, J. Geophys. Res., doi:10.1029/93jc00458, 1993.

Álvarez-Salgado, X. A., Gago, J., Míguez, B. M., Gilcoto, M. and Pérez, F. F.: Surface Waters of
the NW Iberian Margin: Upwelling on the Shelf versus Outwelling of Upwelled Waters from the Rías
Baixas, Estuar. Coast. Shelf Sci., 51(6), 821–837, doi:10.1006/ecss.2000.0714, 2000.

510 Álvarez-Salgado, X. A., Labarta, U., Fernández-Reiriz, M. J., Figueiras, F. G., Rosón, G.,
Piedracoba, S., Filgueira, R. and Cabanas, J. M.: Renewal time and the impact of harmful algal blooms on
the extensive mussel raft culture of the Iberian coastal upwelling system (SW Europe), Harmful Algae,
7(6), 849–855, doi:10.1016/j.hal.2008.04.007, 2008.

Barton, A., Hales, B., Waldbusser, G. G., Langdon, C. and Feelyd, R. A.: The Pacific oyster,
515 Crassostrea gigas, shows negative correlation to naturally elevated carbon dioxide levels: Implications for
near-term ocean acidification effects, Limnol. Oceanogr., doi:10.4319/lo.2012.57.3.0698, 2012.

Barton, A., Waldbusser, G., Feely, R., Weisberg, S., Newton, J., Hales, B., Cudd, S., Eudeline, B., Langdon, C., Jefferds, I., King, T., Suhrbier, A. and McLauglin, K.: Impacts of Coastal Acidification on the Pacific Northwest Shellfish Industry and Adaptation Strategies Implemented in Response, Oceanography, 25(2), 146–159, doi:10.5670/oceanog.2015.38, 2015.

Bates, N., Astor, Y., Church, M., Currie, K., Dore, J., Gonaález-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J. and Santa-Casiano, M.: A Time-Series View of Changing Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO2 and Ocean Acidification, Oceanography, 27(1), 126–141, doi:10.5670/oceanog.2014.16, 2014.

Baumann, H. and Smith, E. M.: Quantifying Metabolically Driven pH and Oxygen Fluctuations in US Nearshore Habitats at Diel to Interannual Time Scales, Estuaries and Coasts, 41(4), 1102–1117, doi:10.1007/s12237-017-0321-3, 2018.

Beldowski, J., Löffler, A., Schneider, B. and Joensuu, L.: Distribution and biogeochemical control of total CO2 and total alkalinity in the Baltic Sea, J. Mar. Syst., doi:10.1016/j.jmarsys.2009.12.020, 2010.

Beale, M. H., Hagan, T. M., and Demuth, H. B.: Deep Learning Toolbox™, User's Guide, Release 2018a, The MathWorks, Inc., Natick, Massachusetts, US, available at: https://es.mathworks.com/help/pdf_doc/deeplearning/nnet_ug.pdf, last access: 20 August 2018.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A. and Gattuso, J.-P.: An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks, Front. Mar. Sci., 5(328), doi:10.3389/fmars.2018.00328, 2018.

Brewer, P. G. and Goldman, J. C.: Alkalinity changes generated by phytoplankton, Limnol. Oceanogr., 21(1), 108–117, doi:10.4319/lo.1976.21.1.0108, 1976.

Brewer, P. G., Wong, G. T. F., Bacon, M. P. and Spencer, D. W.: An oceanic calcium problem?, Earth Planet. Sci. Lett., 26(1), 81–87, doi:10.1016/0012-821X(75)90179-X, 1975.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., González-Dávila, M., Jeansson, E., Kozyr, A. and Van Heuven, S. M. A. C.: A global monthly climatology of total alkalinity: A neural network approach, Earth Syst. Sci. Data, doi:10.5194/essd-11-1109-2019, 2019.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., Santana-Casiano, J. M. and Kozyr, A.: A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach, Earth Syst. Sci. Data, 12(3), 1725–1743, doi:10.5194/essd-12-1725-2020, 2020a.

Broullón, D., Pérez, F. F., and Doval, M. D.: Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem: The case of Ría de Vigo (NW Spain) between 1992 and 2019 (Dataset), DIGITAL.CSIC, http://dx.doi.org/10.20350/digitalCSIC/12642, 2020b.

Cai, W.-J., Hu, X., Huang, W.-J., Jiang, L.-Q., Wang, Y., Peng, T.-H. and Zhang, X.: Alkalinity distribution in the western North Atlantic Ocean margins, J. Geophys. Res., 115(C8), C08014, doi:10.1029/2009JC005482, 2010.

555 Cai, W.-J., Hu, X., Huang, W.-J., Murrell, M. C., Lehrter, J. C., Lohrenz, S. E., Chou, W.-C., Zhai, W., Hollibaugh, J. T., Wang, Y., Zhao, P., Guo, X., Gundersen, K., Dai, M. and Gong, G.-C.: Acidification of subsurface coastal waters enhanced by eutrophication, Nat. Geosci., 4(11), 766–770, doi:10.1038/ngeo1297, 2011.

Cai, W.-J., Feely, R. A., Testa, J. M., Li, M., Evans, W., Alin, S. R., Xu, Y.-Y., Pelletier, G., 560 Ahmed, A., Greeley, D. J., Newton, J. A. and Bednaršek, N.: Natural and Anthropogenic Drivers of Acidification in Large Estuaries, Ann. Rev. Mar. Sci., 13(1), doi:10.1146/annurev-marine-010419-011004, 2021.

Cai, W. J., Xu, Y. Y., Feely, R. A., Wanninkhof, R., Jönsson, B., Alin, S. R., Barbero, L., Cross, J. N., Azetsu-Scott, K., Fassbender, A. J., Carter, B. R., Jiang, L. Q., Pepin, P., Chen, B., Hussain, N., 565 Reimer, J. J., Xue, L., Salisbury, J. E., Hernández-Ayón, J. M., Langdon, C., Li, Q., Sutton, A. J., Chen, C. T. A. and Gledhill, D. K.: Controls on surface water carbonate chemistry along North American ocean margins, Nat. Commun., 11(1), 1–13, doi:10.1038/s41467-020-16530-z, 2020.

Carstensen, J. and Duarte, C. M.: Drivers of pH Variability in Coastal Ecosystems, Environ. Sci. Technol., 53(8), 4020–4029, doi:10.1021/acs.est.8b03655, 2019.

570 Carstensen, J., Chierici, M., Gustafsson, B. G. and Gustafsson, E.: Long-Term and Seasonal Trends in Estuarine and Coastal Carbonate Systems, Global Biogeochem. Cycles, 32(3), 497–513, doi:10.1002/2017GB005781, 2018.

Cermeño, P., Marañón, E., Pérez, V., Serret, P., Fernández, E. and Castro, C. G.: Phytoplankton size structure and primary production in a highly dynamic coastal ecosystem (Ría de Vigo, NW-Spain): 575 Seasonal and short-time scale variability, Estuar. Coast. Shelf Sci., 67(1–2), 251–266, doi:10.1016/j.ecss.2005.11.027, 2006.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. Le, Myneni, R., Piao, S. and Thornton, P.: Carbon and Other Biogeochemical Cycles, in Climate Change 2013 - The Physical Science Basis, edited by 580 Intergovernmental Panel on Climate Change, pp. 465–570, Cambridge University Press, Cambridge., 2013.

Cyronak, T., Andersson, A. J., D'Angelo, S., Bresnahan, P., Davidson, C., Griffin, A., Kindeberg, T., Pennise, J., Takeshita, Y. and White, M.: Short-Term Spatial and Temporal Carbonate Chemistry Variability in Two Contrasting Seagrass Meadows: Implications for pH Buffering Capacities, Estuaries and Coasts, 41(5), 1282–1296, doi:10.1007/s12237-017-0356-5, 2018.

585 Doney, S. C., Fabry, V. J., Feely, R. A. and Kleypas, J. A.: Ocean Acidification: The Other CO_2 Problem, Ann. Rev. Mar. Sci., 1(1), 169–192, doi:10.1146/annurev.marine.010908.163834, 2009.

Doval, M. D., López, A. and Madriñán, M.: Temporal variation and trends of inorganic nutrients

in the coastal upwelling of the NW Spain (Atlantic Galician rías), J. Sea Res., 108, 19–29, doi:10.1016/j.seares.2015.12.006, 2016.

590   Egleston, E. S., Sabine, C. L. and Morel, F. M. M.: Revelle revisited: Buffer factors that quantify the response of ocean chemistry to changes in DIC and alkalinity, Global Biogeochem. Cycles, 24(1), GB1002, doi:10.1029/2008GB003407, 2010.

Feely, R. A., Alin, S. R., Newton, J., Sabine, C. L., Warner, M., Devol, A., Krembs, C. and Maloy, C.: The combined effects of ocean acidification, mixing, and respiration on pH and carbonate saturation in
595   an urbanized estuary, Estuar. Coast. Shelf Sci., 88(4), 442–449, doi:10.1016/j.ecss.2010.05.004, 2010.

Fernández, E., Álvarez-Salgado, X. A., Beiras, R., Ovejero, A. and Méndez, G.: Coexistence of urban uses and shellfish production in an upwelling-driven, highly productive marine environment: The case of the Ría de Vigo (Galicia, Spain), Reg. Stud. Mar. Sci., 8, 362–370, doi:10.1016/j.rsma.2016.04.002, 2016.

600   Figueiras, F. G., Labarta, U. and Fernández Reiriz, M. J.: Coastal upwelling, primary production and mussel growth in the Rías Baixas of Galicia, Hydrobiologia, 484, 121–131, doi:10.1023/A:1021309222459, 2002.

Fraga, F.: Upwelling off the Galician coast, northwest Spain. In: Richardson FA (Ed.), Coastal Upwelling. American Geophysical Union, Washington., pp. 176–182, 1981.

605   Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., DBakker, O. C. E., Canadell1, J. G., Ciais1, P., Jackson, R. B., Anthoni1, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer,
610   E., Kaplan, J. O., Kato, E., Goldewijk, K. K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., Van Der Werf, G. R., Wiltshire, A. J. and Zaehle, S.: Global
615   carbon budget 2019, Earth Syst. Sci. Data, 11(4), 1783–1838, doi:10.5194/essd-11-1783-2019, 2019.

Gago, J., Gilcoto, M., Pérez, F. . and Ríos, A. .: Short-term variability of fCO2 in seawater and air–sea CO2 fluxes in a coastal upwelling system (Ría de Vigo, NW Spain), Mar. Chem., 80(4), 247–264, doi:10.1016/S0304-4203(02)00117-2, 2003.

Gago, J., Álvarez-Salgado, X. A., Nieto-Cid, M., Brea, S. and Piedracoba, S.: Continental inputs
620   of C, N, P and Si species to the Ría de Vigo (NW Spain), Estuar. Coast. Shelf Sci., 65(1–2), 74–82, doi:10.1016/j.ecss.2005.05.008, 2005.

Hagan, M. T. and Menhaj, M. B.: Training Feedforward Networks with the Marquardt Algorithm, IEEE Trans. Neural Networks, doi:10.1109/72.329697, 1994.

Hagan, M. T., Demuth, H. B., Beale, M. H., and De Jesús, O.: Neural network design, ISBN 978-0971732117, available at: http://hagan.okstate.edu/nnd.html (last access: 26 July 2018), 2014.

Hampel, F. R.: The Influence Curve and its Role in Robust Estimation, J. Am. Stat. Assoc., 69(346), 383–393, doi:10.1080/01621459.1974.10482962, 1974.

Hendriks, I. E., Duarte, C. M., Olsen, Y. S., Steckbauer, A., Ramajo, L., Moore, T. S., Trotter, J. A. and McCulloch, M.: Biological mechanisms supporting adaptation to ocean acidification in coastal ecosystems, Estuar. Coast. Shelf Sci., 152, A1–A8, doi:10.1016/j.ecss.2014.07.019, 2015.

Hofmann, G. E., Smith, J. E., Johnson, K. S., Send, U., Levin, L. A., Micheli, F., Paytan, A., Price, N. N., Peterson, B., Takeshita, Y., Matson, P. G., Crook, E. D., Kroeker, K. J., Gambi, M. C., Rivest, E. B., Frieder, C. A., Yu, P. C. and Martz, T. R.: High-Frequency Dynamics of Ocean pH: A Multi-Ecosystem Comparison, edited by W.-C. Chin, PLoS One, 6(12), e28983, doi:10.1371/journal.pone.0028983, 2011.

Joesoef, A., Kirchman, D. L., Sommerfield, C. K. and Cai, W.-J.: Seasonal variability of the inorganic carbon system in a large coastal plain estuary, Biogeosciences, 14(21), 4949–4963, doi:10.5194/bg-14-4949-2017, 2017.

Kapsenberg, L., Alliouane, S., Gazeau, F., Mousseau, L. and Gattuso, J.-P.: Coastal ocean acidification and increasing total alkalinity in the northwestern Mediterranean Sea, Ocean Sci., 13(3), 411–426, doi:10.5194/os-13-411-2017, 2017.

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

Lemos, R. T. and Sansó, B.: Spatio-temporal variability of ocean temperature in the Portugal Current System, J. Geophys. Res. Ocean., doi:10.1029/2005JC003051, 2006.

Levenberg, K.: A Method for the solution of certain non-linear probles in least squares., Q. Appl. Math., II(2), 164–168, 1944.

Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol., 49(4), 764–766, doi:10.1016/j.jesp.2013.03.013, 2013.

Li, X., Bellerby, R. G. J., Ge, J., Wallhead, P., Liu, J., and Yang, A.: Retrieving monthly and interannual total-scale pH (pHT) on the East China Sea shelf using an artificial neural network: ANN-pHT-v1, Geosci. Model Dev., 13, 5103–5117, https://doi.org/10.5194/gmd-13-5103-2020, 2020a.

Li, X., Bellerby, R. G. J., Wallhead, P., Ge, J., Liu, J., Liu, J. and Yang, A.: A neural network-based analysis of the seasonal variability of surface total alkalinity on the East China Sea shelf, Front. Mar. Sci., 7, 219, doi:10.3389/fmars.2020.00219, 2020b.

Marquardt, D.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, J. Soc. Ind. Appl. Math., 11(2), 1963.

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

Nogueira, E., Pérez, F. F. and Ríos, A. F.: Seasonal Patterns and Long-term Trends in an Estuarine
Upwelling Ecosystem (Ría de Vigo, NW Spain), Estuar. Coast. Shelf Sci., 44(3), 285–300,
doi:10.1006/ecss.1996.0119, 1997.

O'Boyle, S., McDermott, G., Noklegaard, T. and Wilkes, R.: A Simple Index of Trophic Status in
Estuaries and Coastal Bays Based on Measurements of pH and Dissolved Oxygen, Estuaries and Coasts,
36(1), 158–173, doi:10.1007/s12237-012-9553-4, 2013.

Olden, J. D. and Jackson, D. A.: Illuminating the '" black box "': a randomization approach for
understanding variable contributions in artificial neural networks, Ecol. Modell., 154, 135–150,
doi:10.1016/S0304-3800(02)00064-9, 2002.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A.,
Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P.,
Mouchet, A., Najjar, R. G., Plattner, G. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R.,
Slater, R. D., Totterdell, I. J., Weirig, M. F., Yamanaka, Y. and Yool, A.: Anthropogenic ocean acidification
over the twenty-first century and its impact on calcifying organisms, Nature, 437(7059), 681–686,
doi:10.1038/nature04095, 2005.

Padin, X. A., Velo, A., and Pérez, F. F.: ARIOS: An acidification ocean database for the Iberian
Upwelling Ecosystem (1976–2018), Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-65,
accepted, 2020.

Pardo, P., Padín, X., Gilcoto, M., Farina-Busto, L. and Pérez, F.: Evolution of upwelling systems
coupled to the long-term variability in sea surface temperature and Ekman transport, Clim. Res., 48(2),
231–246, doi:10.3354/cr00989, 2011.

Perez, F. F., Alvarez-Salgado, X. A., Roson, G. and Rios, A. F.: Carbonic-calcium system,
nutrients and total organic nitrogen in continental runoff to the Galician Rias Baixas, NW Spain, Oceanol.
Acta, 15(6), 595–602, 1992.

Pérez, F. F., Ríos, A. F. and Rosón, G.: Sea surface carbon dioxide off the Iberian Peninsula (North
Eastern Atlantic Ocean), J. Mar. Syst., 19(1–3), 27–46, doi:10.1016/S0924-7963(98)00022-0, 1999.

Pérez, F. F., Padín, X. A., Pazos, Y., Gilcoto, M., Cabanas, M., Pardo, P., Doval, M. D. and Farina-
Busto, L.: Plankton response to weakening of the Iberian coastal upwelling, Glob. Chang. Biol., 16(4),
1258–1267, doi:10.1111/j.1365-2486.2009.02125.x, 2010.

Piedracoba, S., Álvarez-Salgado, X. A., Rosón, G. and Herrera, J. L.: Short-timescale
thermohaline variability and residual circulation in the central segment of the coastal upwelling system of
the Ría de Vigo (northwest Spain) during four contrasting periods, J. Geophys. Res. Ocean., 110(C3), 1–
15, doi:10.1029/2004JC002556, 2005.

Rhein, M., S.R. Rintoul, S. Aoki, E. Campos, D. Chambers, R.A. Feely, S. Gulev, G.C. Johnson, S.A. Josey, A. Kostianoy, C. Mauritzen, D. Roemmich, L.D. Talley and F. Wang: Observations: Ocean. 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, 2013.

Ríos, A. F., Pérez, F. F. and Fraga, F.: Water masses in the upper and middle North Atlantic Ocean east of the Azores, Deep Sea Res. Part A. Oceanogr. Res. Pap., 39(3–4), 645–658, doi:10.1016/0198-0149(92)90093-9, 1992.

Salisbury, J., Vandemark, D., Hunt, C., Campbell, J., Jonsson, B., Mahadevan, A., McGillis, W. and Xue, H.: Episodic riverine influence on surface DIC in the coastal Gulf of Maine, Estuar. Coast. Shelf Sci., 82(1), 108–118, doi:10.1016/j.ecss.2008.12.021, 2009.

Sunday, J. M., Fabricius, K. E., Kroeker, K. J., Anderson, K. M., Brown, N. E., Barry, J. P., Connell, S. D., Dupont, S., Gaylord, B., Hall-Spencer, J. M., Klinger, T., Milazzo, M., Munday, P. L., Russell, B. D., Sanford, E., Thiyagarajan, V., Vaughan, M. L. H., Widdicombe, S. and Harley, C. D. G.: Ocean acidification can mediate biodiversity shifts by changing biogenic habitat, Nat. Clim. Chang., 7(1), 81–85, doi:10.1038/nclimate3161, 2017.

Tilstone, G., Figueiras, F., Fermín, E. and Arbones, B.: Significance of nanophytoplankton photosynthesis and primary production in a coastal upwelling system (Ría de Vigo, NW Spain), Mar. Ecol. Prog. Ser., 183(Mann 1993), 13–27, doi:10.3354/meps183013, 1999.

Turk, D., Zappa, C. J., Meinen, C. S., Christian, J. R., Ho, D. T., Dickson, A. G. and McGillis, W. R.: Rain impacts on CO 2 exchange in the western equatorial Pacific Ocean, Geophys. Res. Lett., 37(L23610), doi:10.1029/2010GL045520, 2010.

Velo, A., Pérez, F. F., Tanhua, T., Gilcoto, M., Ríos, A. F. and Key, R. M.: Total alkalinity estimation using MLR and neural network techniques, J. Mar. Syst., 111–112, 11–18, doi:10.1016/j.jmarsys.2012.09.002, 2013.

Waldbusser, G. G. and Salisbury, J. E.: Ocean Acidification in the Coastal Zone from an Organism's Perspective: Multiple System Parameters, Frequency Domains, and Habitats, Ann. Rev. Mar. Sci., 6(1), 221–247, doi:10.1146/annurev-marine-121211-172238, 2014.

Wallace, R. B., Baumann, H., Grear, J. S., Aller, R. C. and Gobler, C. J.: Coastal ocean acidification: The other eutrophication problem, Estuar. Coast. Shelf Sci., doi:10.1016/j.ecss.2014.05.027, 2014.

Wootton, J. T. and Pfister, C. A.: Carbon System Measurements and Potential Climatic Drivers at a Site of Rapidly Declining Ocean pH, edited by W.-C. Chin, PLoS One, 7(12), e53396, doi:10.1371/journal.pone.0053396, 2012.

730 Table 1. Statistics of the ten networks with the best performance in the test dataset for pH and A_T. Units of RMSE are pH units for pH and micromoles per kilogram (μmol kg^{-1}) for A_T.

	Combination	Number of neurons	Network number	Training dataset		Test dataset	
				RMSE	r^2	RMSE	r^2
pH	LLDTSPNSiOYW	34	3	0.027	0.89	0.03	0.86
	LLDTSPNSiOYW	34	6	0.028	0.89	0.031	0.86
	LLDTSPNSiOYW	52	2	0.028	0.89	0.031	0.85
	DTSPNSiOYW	**28**	**3**	**0.029**	**0.88**	**0.031**	**0.85**
	LLDTSPNSiOYW	40	4	0.026	0.9	0.031	0.85
	DTSPNSiOYW	28	4	0.028	0.88	0.031	0.85
	DTSPNSiOYW	52	10	0.027	0.89	0.032	0.85
	DTSPNSiOYW	28	7	0.028	0.89	0.032	0.85
	DTSPNSiOYW	46	9	0.027	0.9	0.032	0.85
	LLDTSPNSiOYW	28	4	0.027	0.89	0.032	0.85
A_T	DTSPNSiY	28	4	14.7	0.96	10.5	0.94
	DTSPNSiY	28	10	14.9	0.96	10.6	0.94
	DTSPNSiOW	46	6	13.9	0.97	10.7	0.97
	DTSPNSiY	28	3	13.8	0.97	10.7	0.94
	LLDTSPNSiYW	**52**	**8**	**10.3**	**0.98**	**10.9**	**0.97**
	LLDTSPNSiYW	52	7	14.7	0.96	10.9	0.97
	DTSPNSiOY	40	2	10.5	0.98	10.9	0.96
	LLDTSPNSiYW	52	4	10.8	0.98	11	0.97
	LLDTSPNSiYW	34	10	9.4	0.99	11.1	0.96
	LLDTSPNSiOYW	52	2	10.3	0.98	11.4	0.97

735

Table 2. Relative importance of each input variable for pH_NN and A_T_NN. Lat: latitude; Lon: longitude; D: depth; T: temperature; S: salinity; P: phosphate; N: nitrate; Si: Silicate; O_2: dissolved oxygen; Y: year; cW: Eq. (1); sW: Eq. (2).

Variable	Inputs relative importance (%)											
	Lat	Lon	D	T	S	P	N	Si	O_2	Y	cW	sW
pH	-	-	11	15.4	22.5	3.2	4.1	6.2	7.3	16	4.4	9.9
A_T	2.4	3.1	1.9	10.1	2.2	23	2.6	42.5	-	8.1	3	1.2

740   Table 3. Average weekly rate of change for pH and A_T.

		Station						
	Depth range	Inner			Outer/Middle			
		V3	V4	V2	V7	V1	V6	V5
pH (week^{-1})	0-5 m	0.056	0.052	0.051	0.050	0.043	0.040	0.041
	5-10 m	0.038	0.042	0.045	0.048	0.042	0.038	0.038
	10-15 m	0.033	0.036	0.041	0.048	0.047	0.045	0.040
A_T (µmol kg^{-1} week^{-1})	0-5 m	127.7	63.8	52.9	34.3	35.1	34.5	30.9
	5-10 m	61.8	36.4	31.4	24.9	25.1	23.5	22.9
	10-15 m	38.2	29.7	27.3	20.3	22.5	20.2	19.3

Table 4. Maxima, minima, amplitude and variability of the seasonal cycle of pH and A_T.

Zone	Station	Depth	pH				A_T (µmol kg^{-1})			
			Max	Min	Amplitude	Variability	Max	Min	Amplitude	Variability
Inner	V3	0-5 m	8.08	7.92	0.16	0.064 ± 0.013	2284.3	2116.6	167.7	118.7 ± 59.6
		5-10 m	8.06	7.92	0.15	0.054 ± 0.011	2302	2239.3	62.7	56.2 ± 21.4
		10-15 m	8.05	7.92	0.13	0.053 ± 0.011	2316.8	2263.8	53	44.7 ± 17.3
	V4	0-5 m	8.10	7.93	0.18	0.061 ± 0.015	2295.7	2207.5	88.2	69.9 ± 30.2
		5-10 m	8.08	7.94	0.15	0.054 ± 0.014	2311.6	2275.5	36.1	42.1 ± 14.3
		10-15 m	8.06	7.94	0.13	0.053 ± 0.012	2326.9	2297.4	29.5	34.5 ± 11.3
	V2	0-5 m	8.12	7.95	0.17	0.058 ± 0.016	2297.6	2210.9	86.7	64.4 ± 31
		5-10 m	8.10	7.96	0.14	0.053 ± 0.016	2313.3	2279.6	33.7	38.3 ± 12.9
		10-15 m	8.08	7.96	0.12	0.052 ± 0.014	2325.7	2299.9	25.7	32.3 ± 10.1
Outer/Middle	V7	0-5 m	8.12	7.98	0.14	0.053 ± 0.019	2327.1	2286.3	40.8	39.2 ± 15.3
		5-10 m	8.09	7.99	0.10	0.053 ± 0.022	2328.8	2307.6	21.2	28.6 ± 10.8
		10-15 m	8.08	7.97	0.11	0.053 ± 0.021	2331.2	2308.9	22.2	22.7 ± 8.2
	V1	0-5 m	8.14	8.02	0.12	0.050 ± 0.014	2308.8	2259.6	49.2	50.1 ± 24.7
		5-10 m	8.13	8.02	0.11	0.050 ± 0.018	2319.9	2279.2	40.8	34.8 ± 14
		10-15 m	8.10	8.02	0.08	0.053 ± 0.019	2327.8	2301.7	26.1	27 ± 8.1
	V6	0-5 m	8.15	8.04	0.11	0.050 ± 0.016	2316.3	2276.9	39.4	43.7 ± 21.1
		5-10 m	8.13	8.03	0.10	0.048 ± 0.016	2324.1	2284.8	39.3	32.9 ± 13.8
		10-15 m	8.11	8.03	0.08	0.052 ± 0.017	2331	2297.2	33.8	26 ± 10
	V5	0-5 m	8.15	8.05	0.11	0.049 ± 0.013	2324.9	2287.9	36.9	39.2 ± 15.3
		5-10 m	8.13	8.04	0.09	0.048 ± 0.014	2331.8	2294.5	37.3	28.6 ± 10.8
		10-15 m	8.11	8.03	0.08	0.049 ± 0.015	2337	2310.8	26.2	22.7 ± 8.2

745

Table 5. Long-term trends of pH and A_T between 1992 and 2019 (2007-2019 for V7). Bold trends are statistically significant (p-value<0.01).

Zone	Station	Depth	pH (yr^{-1})	A_T (µmol kg^{-1} yr^{-1})
Inner	V3	0-5 m	**-0.0032 ± 0.0001**	**2.86 ± 0.04**
		5-10 m	**-0.0036 ± 0.0001**	**1.01 ± 0.04**
		10-15 m	**-0.0037 ± 0.0001**	**0.95 ± 0.04**
	V4	0-5 m	**-0.0030 ± 0.0001**	**1.04 ± 0.04**
		5-10 m	**-0.0032 ± 0.0001**	**0.75 ± 0.04**
		10-15 m	**-0.0035 ± 0.0001**	**0.46 ± 0.04**
	V2	0-5 m	**-0.0026 ± 0.0001**	**1.13 ± 0.04**
		5-10 m	**-0.0027 ± 0.0001**	**0.66 ± 0.04**
		10-15 m	**-0.0028 ± 0.0001**	**0.27 ± 0.03**
Outer/Middle	V7	0-5 m	0.0001 ± 0.0003	-0.98 ± 0.14
		5-10 m	**0.0021 ± 0.0003**	-0.66 ± 0.12
		10-15 m	0.0012 ± 0.0004	**-0.77 ± 0.13**
	V1	0-5 m	**-0.0019 ± 0.0001**	**0.76 ± 0.04**
		5-10 m	**-0.0022 ± 0.0001**	**0.35 ± 0.04**
		10-15 m	**-0.0024 ± 0.0001**	**0.23 ± 0.04**
	V6	0-5 m	**-0.0022 ± 0.0001**	**0.62 ± 0.04**
		5-10 m	**-0.0022 ± 0.0001**	**0.39 ± 0.04**
		10-15 m	**-0.0024 ± 0.0001**	**0.25 ± 0.04**
	V5	0-5 m	**-0.0020 ± 0.0001**	**0.39 ± 0.04**
		5-10 m	**-0.0022 ± 0.0001**	0.24 ± 0.04
		10-15 m	**-0.0024 ± 0.0001**	0.08 ± 0.04

Table B1. RMSE, r^2 and the relative importance of each input for the three selected networks to fill the gaps in the time series of nutrients from the INTECMAR-database.

Variable	Neurons	RMSE (µmol kg^{-1})*	r^2*	Inputs relative importance (%)								
				Lat	Lon	D	T	S	N	Si	cW	sW
Phosphate	46	0.13	0.73	2.8	15.3	0.6	4.8	19.1	47.1	-	3.4	6.9
Nitrate	28	1.4	0.84	6.2	3.4	5.2	11.4	16.4	-	46.6	5.5	5.3
Silicate	28	1.5	0.8	7.8	8	10.3	10.5	23.7	14.9	-	15.2	9.5

*Statistics computed in the test dataset.

755   Table C1. Relative importance of each input variable for the neural network designed to compute dissolved oxygen.

Variable	Neurons	RMSE (µmol kg^{-1})*	r^{2}*	Inputs relative importance (%)										
				Lat	Lon	D	T	S	P	N	Si	Y	cW	sW
Oxygen	28	16.2	0.85	8.1	6.3	11.7	2.4	28.1	19.3	0.8	3.8	1.3	4.9	13.2

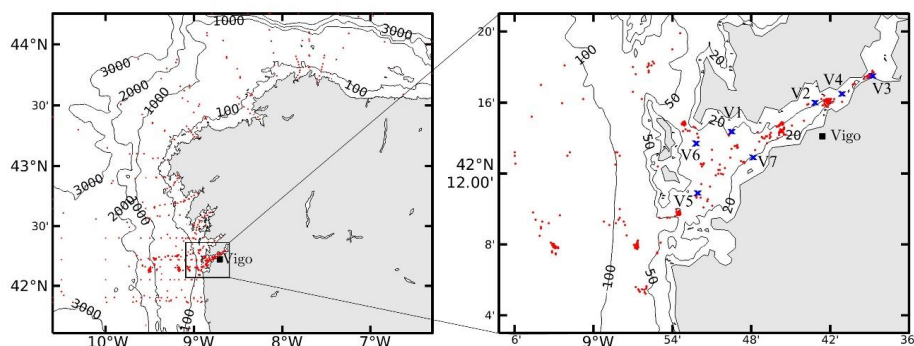



Figure 1. Location of the study area: (a) Galicia (NW Spain) and (b): Ría de Vigo. Red dots: IIM-database profiles.
760   Blue crosses: INTECMAR-database stations. Stations V2, V3 and V4 are considered as inner zone. Stations V1, V5, V6 and V7 are considered as outer/middle zone. The bathymetric data for the Ria de Vigo was provided by the Oceanographic Observatory of the Iberian Margin (RAIA Observatory; www.marnaraia.org).

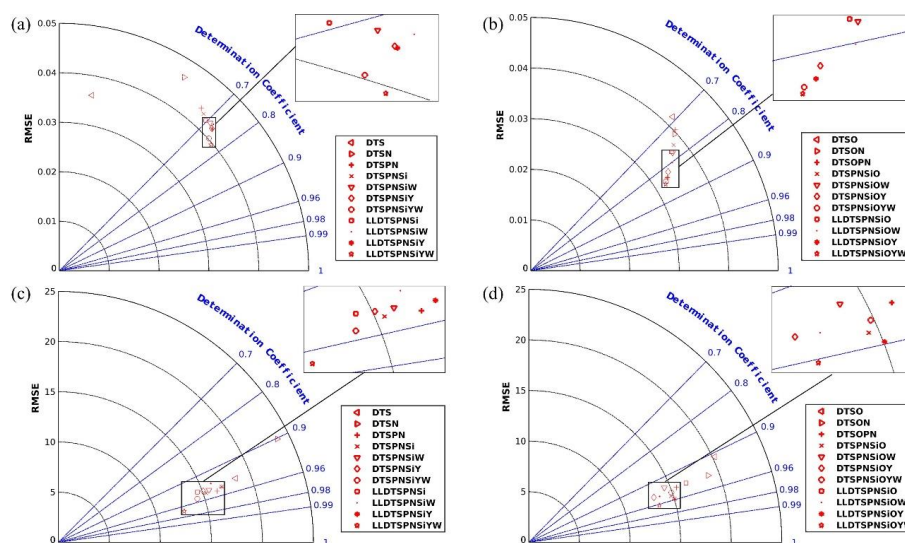



Figure 2. Statistics of the networks for the different combinations of the input variables. These statistics are obtained
from the average of the three networks with the best performance in the test dataset, for each input combination and for
each number of neurons tested. (a) and (b) are for pH with the RMSE in pH units and (c) and (d) for A_T with the RMSE
in micromoles per kilogram (μmol kg^{-1}).

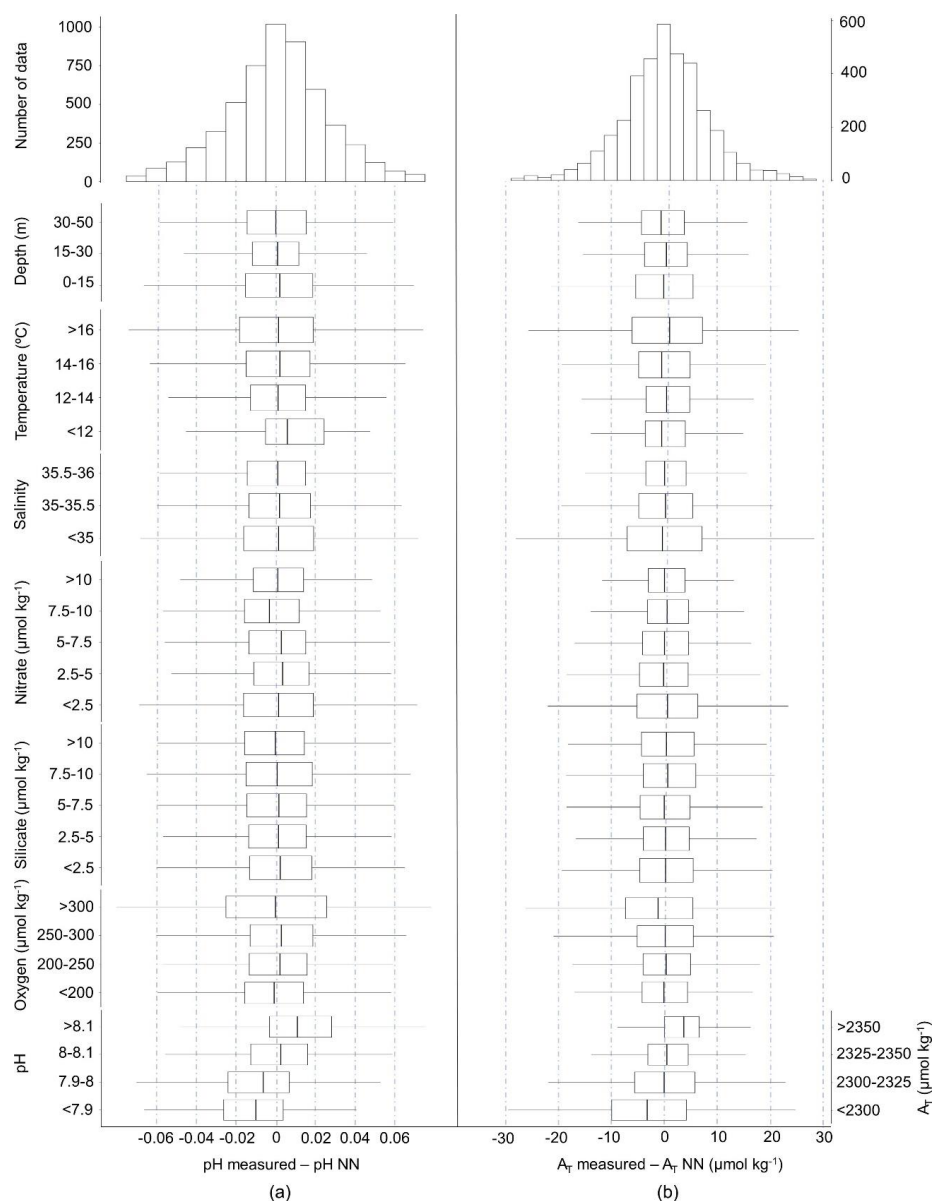



Figure 3. Histograms of data density and box plots of differences between measured and computed pH and A_T against different input variables and the target ones. Units of the differences are pH units for pH and micromoles per kilogram (μmol kg^{-1}) for A_T.

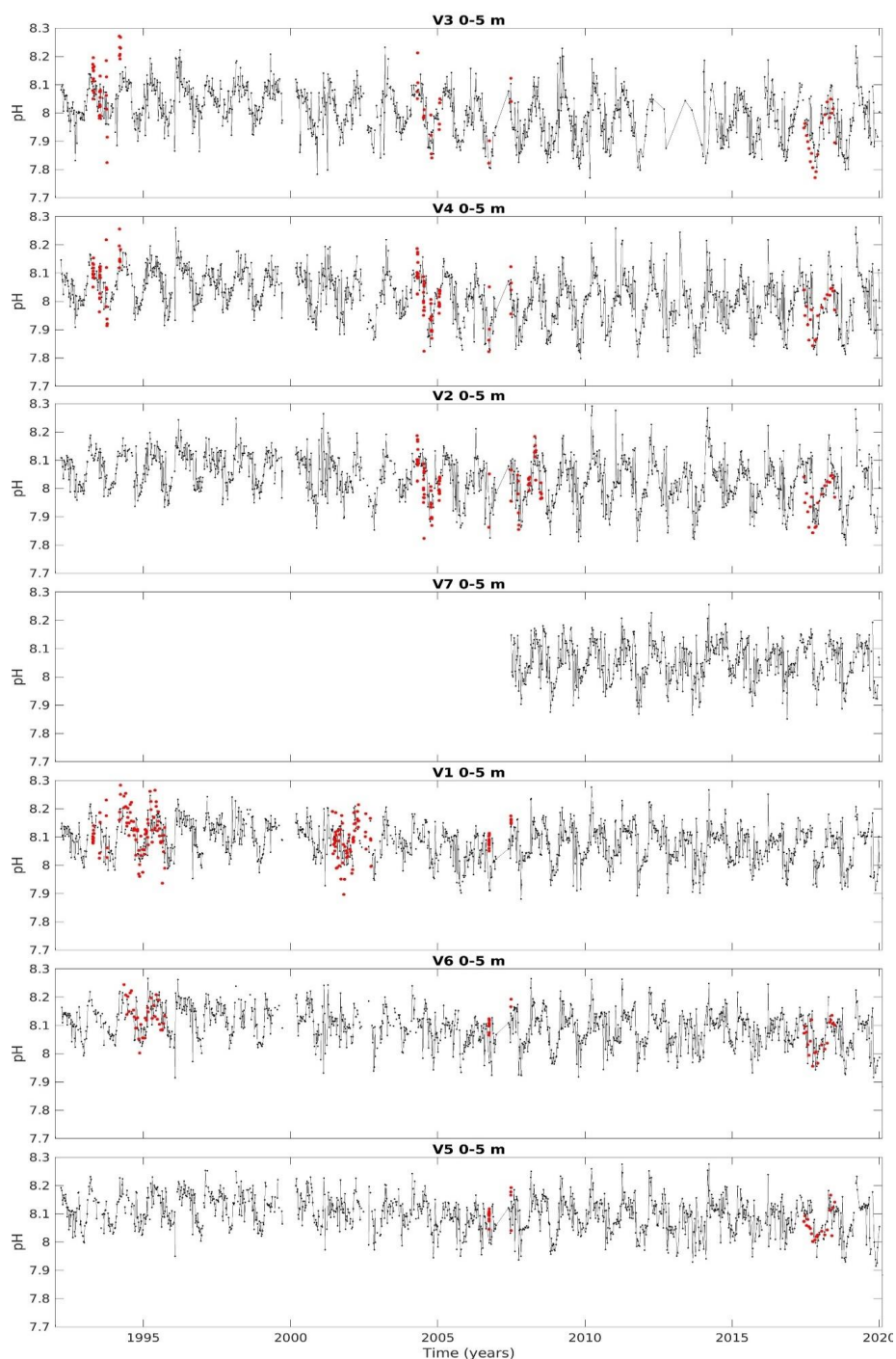



Figure 4. pH time series in the 0-5 m range. Red dots: validation points from the IIM-database. Stations are placed in the figure in such a way to highlight the differences between the inner and the outer/middle zone. The same holds for all figures and tables in the present manuscript when all the stations are showed for the same variable.

775

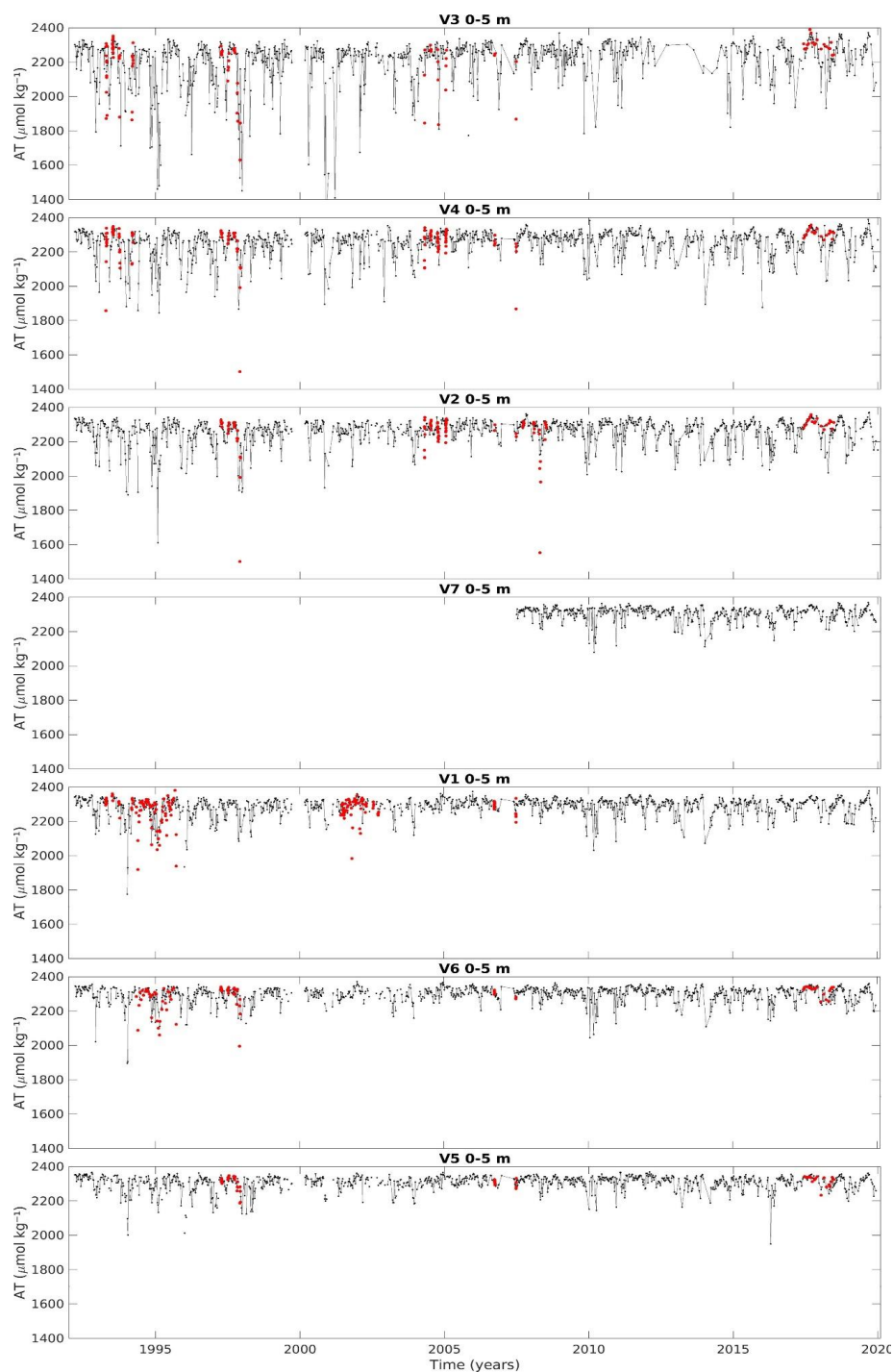



Figure 5. A$_T$ time series in the 0-5 m range. Red dots: validation points from the IIM-database.

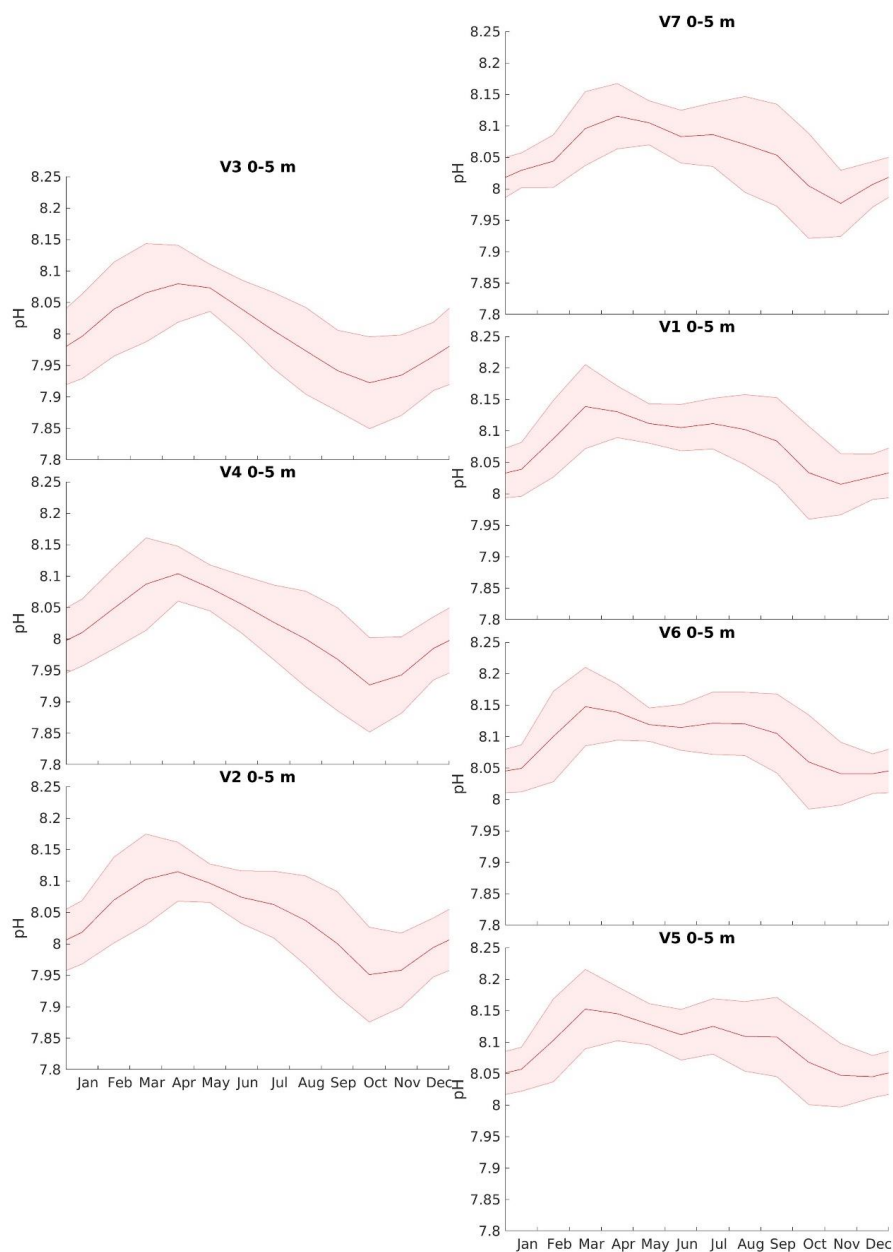



Figure 6. Seasonal cycle of pH in the 0-5 m range. Shadow area represents one standard deviation from the mean of the full time series for each month.

780

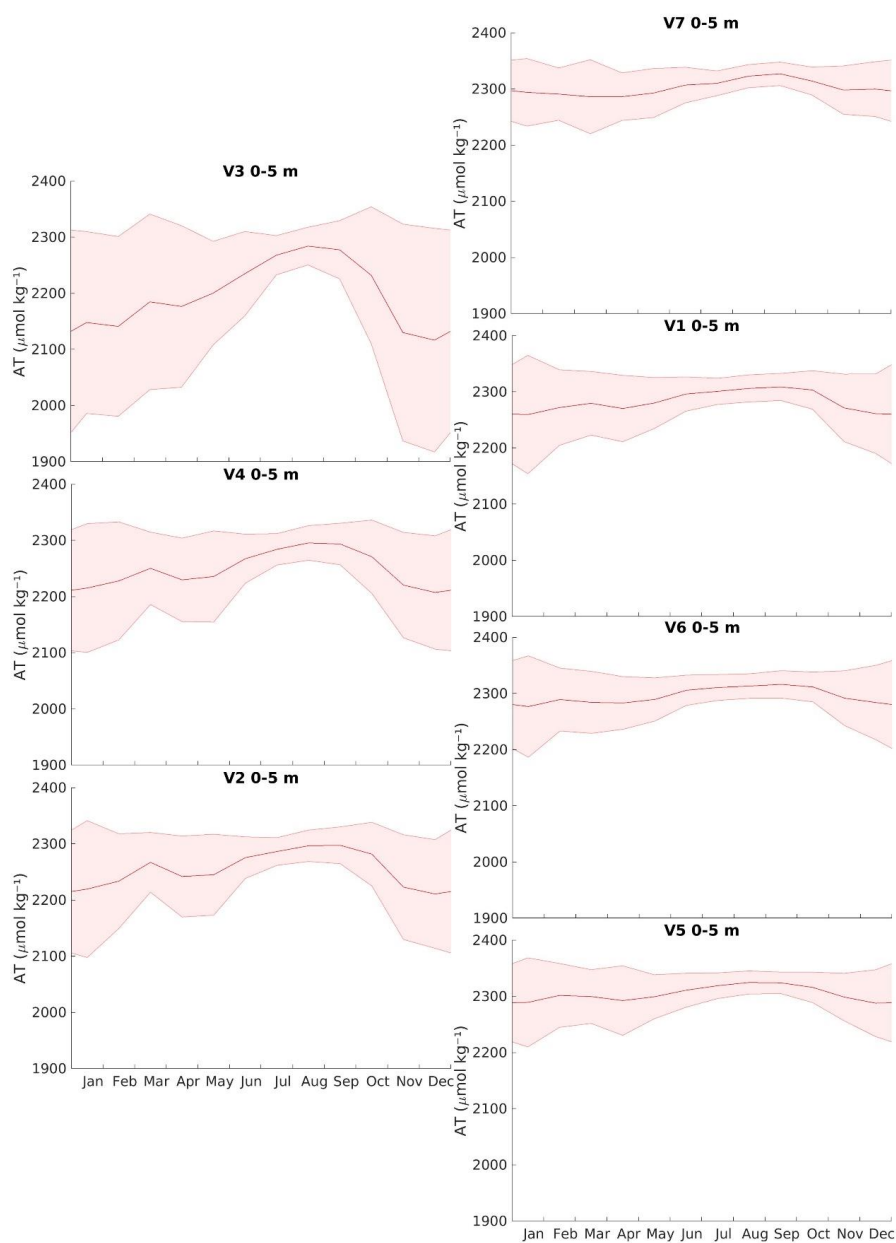
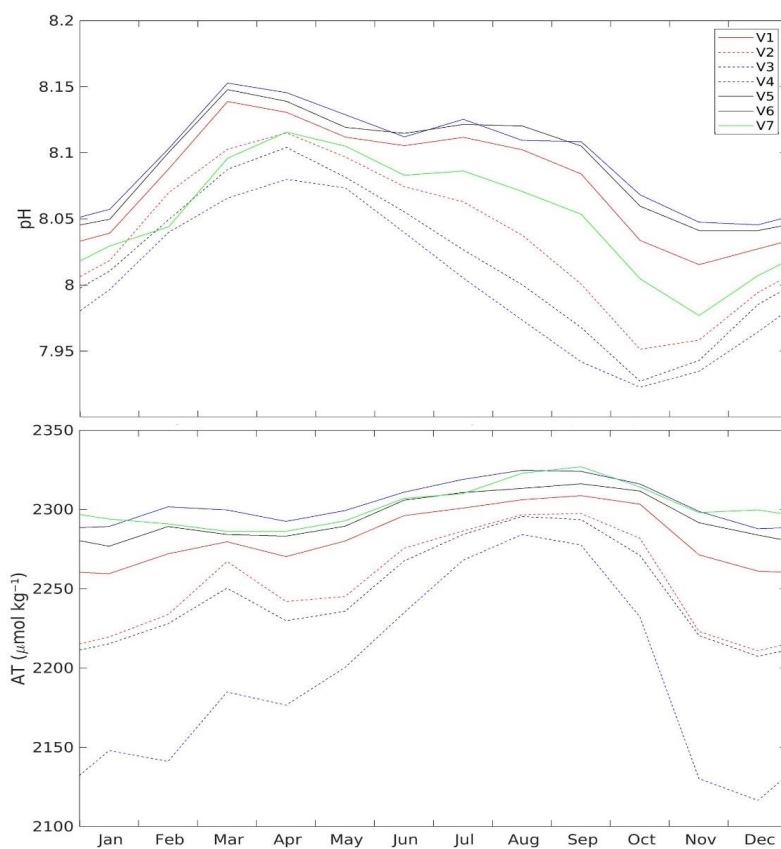



Figure 7. Seasonal cycle of A_T in the 0-5 m range. Shadow area represents one standard deviation from the mean of the full time series for each month.

785   Figure 8. Seasonal cycle of pH and A_T in the 0-5 m range for all the stations. Dashed lines: inner stations. Solid lines: outer/middle stations.

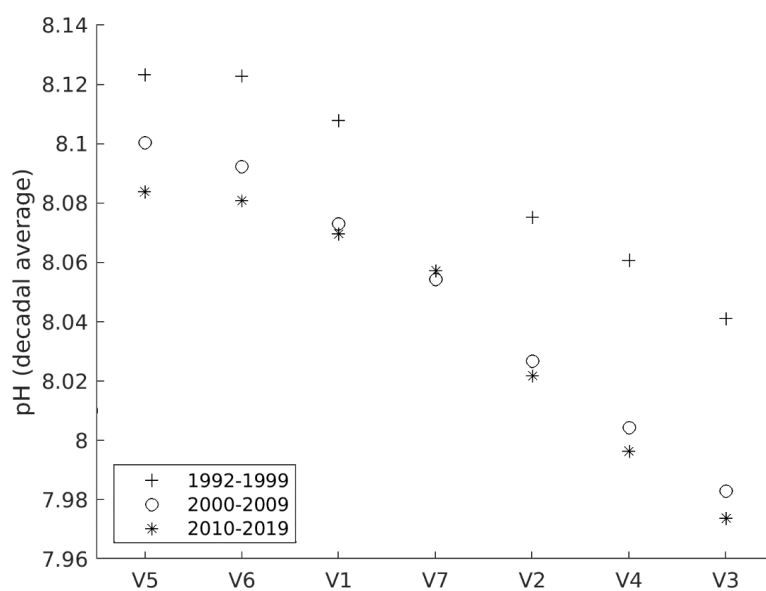



Figure 9. Average pH by decade and station. Note that V7 2000-2009 is the average from 2007 to 2009.

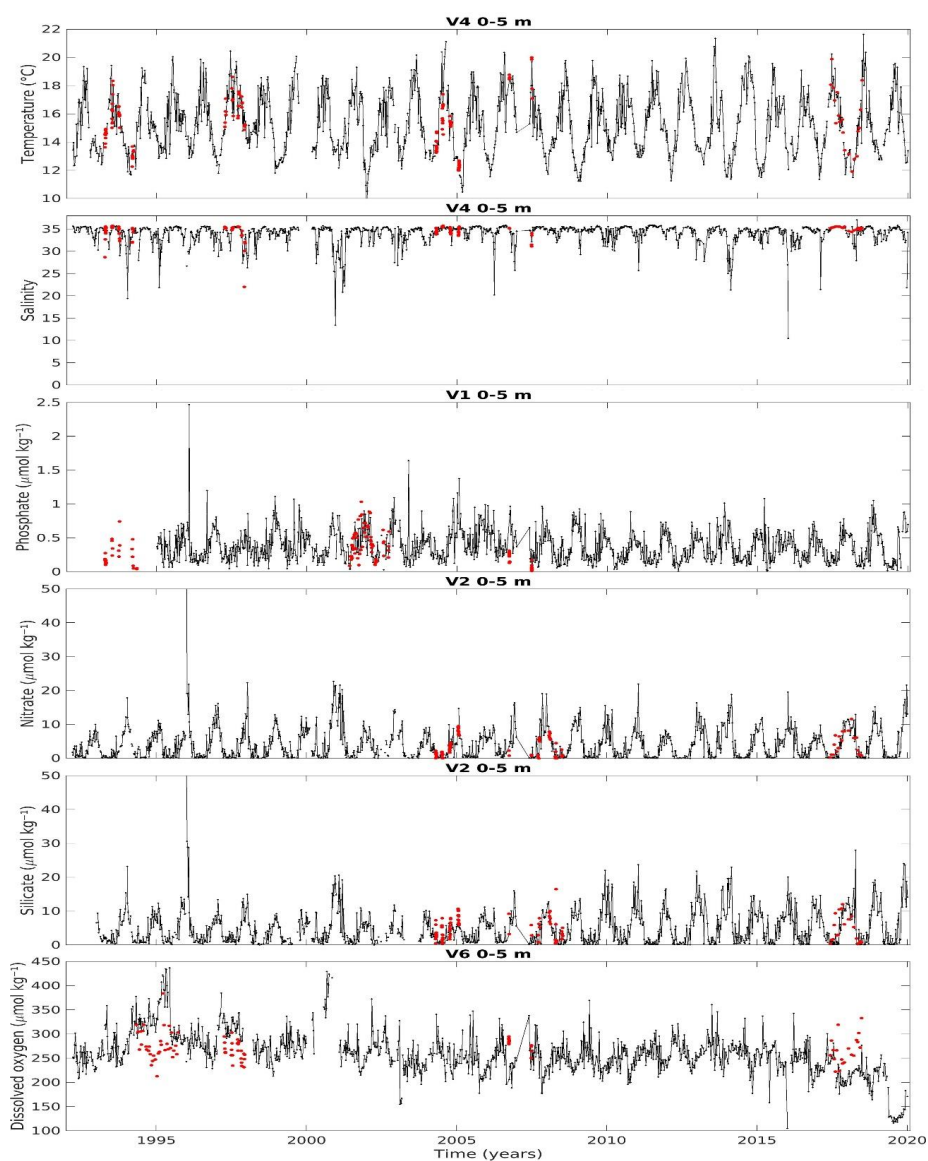
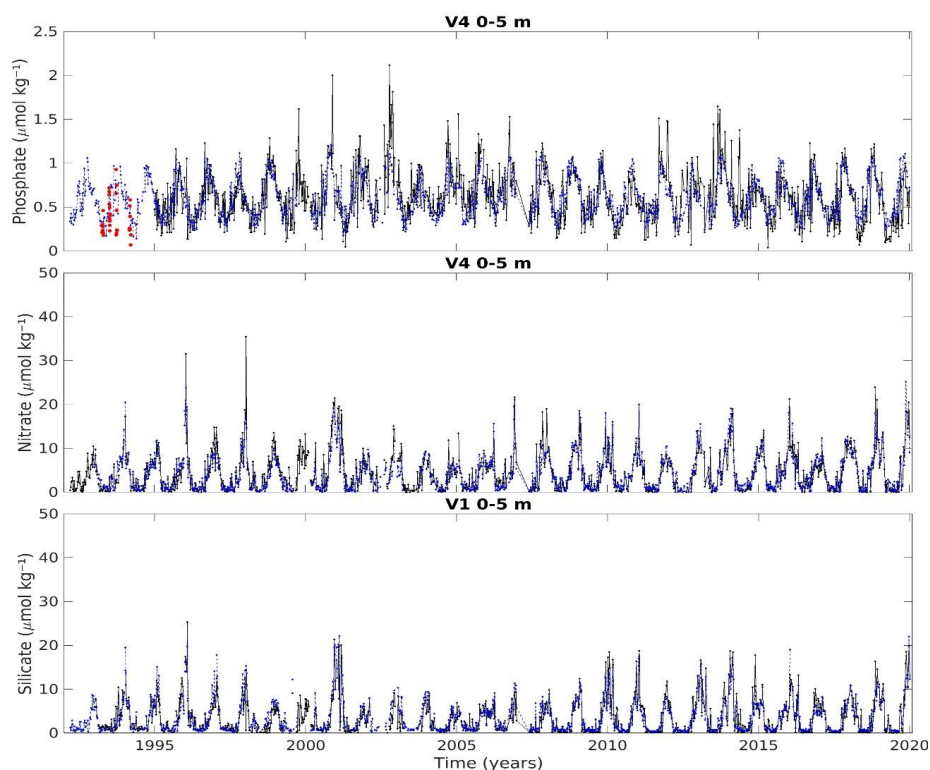



Figure A1. Time series of the input variables from the INTECMAR-database (black line and dots) and discrete measurements from the IIM-database (red dots). Selected stations are shown to represent the features described in the main text, which are common to other stations. Comparisons for the different stations and depths can be easily obtained from the datafiles attached to this study (see Sect. Data availability).

795   Figure B1. Time series of nutrients from the INTECMAR-database (black line and dots) and from neural networks
configured in Appendix B (blue dashed line and dots). Selected stations are shown to illustrate the gap filling. Discrete
samples from the IIM-database are shown in phosphate time series before 1995 to show the good reconstruction of the
time series using the neural network. Comparisons in the different stations and depths can be easily obtained from the
datafiles attached to this study (see Sect. Data availability).

800

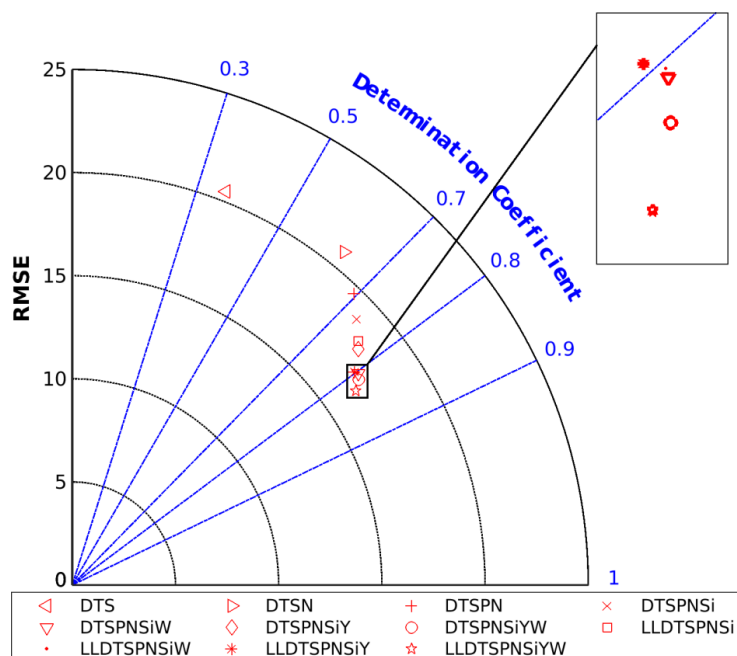
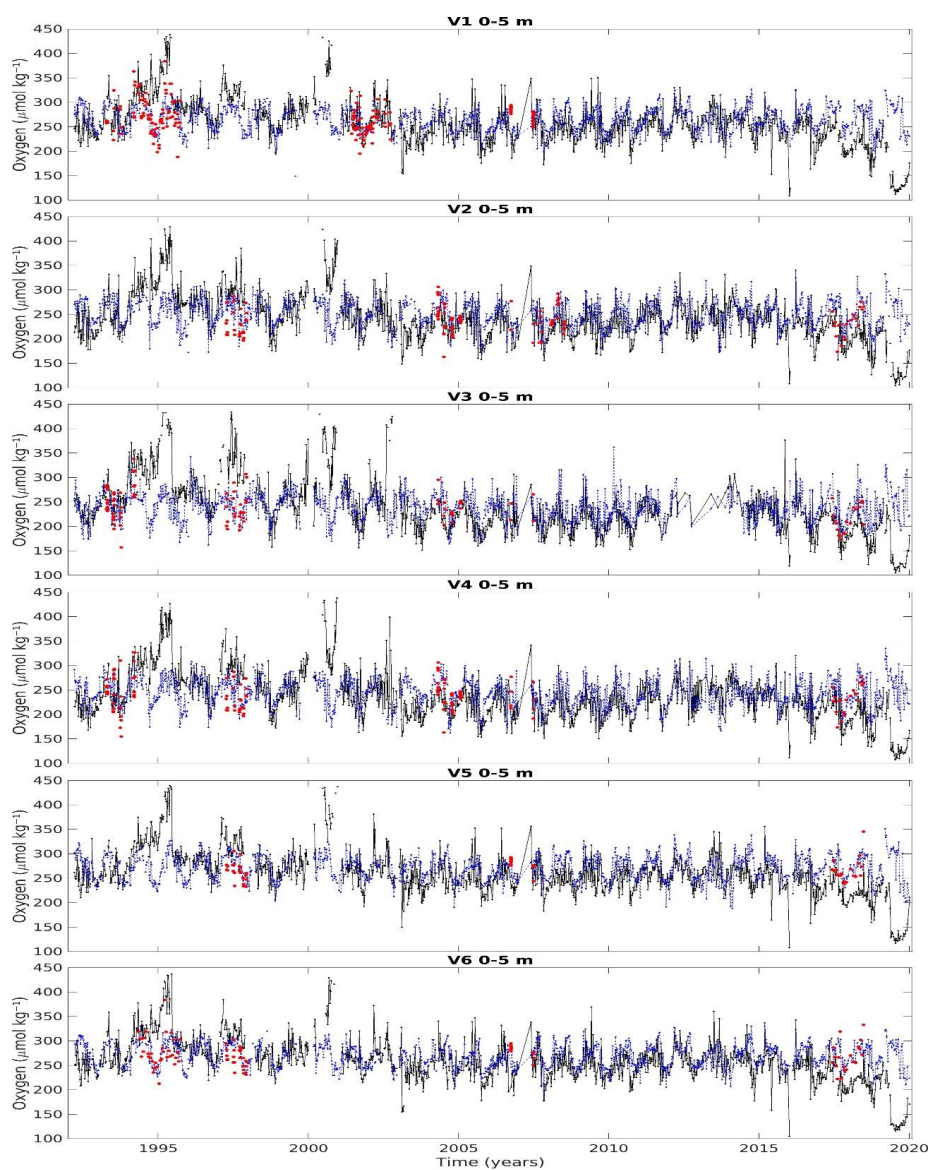



Figure C1. Statistics of the networks for the different combinations of the input variables. These statistics are obtained from the average of the three networks with the best performance in the test dataset, for each input combination and for each number of neurons tested. Units of RMSE are micromoles per kilogram ($\mu mol\,kg^{-1}$).

805 Figure C2. Time series of dissolved oxygen from the INTECMAR-database (black line and dots) and from the neural network configured in Appendix C (blue dashed line and dots). Red dots are validation points from the IIM-database. Station V7 is not showed since it does not have validation points. Comparisons for the other two depth ranges can be easily obtained from the datafiles attached to this study (Sect. Data availability).