# scientific reports

OPEN

# pH trends and seasonal cycle in the coastal Balearic Sea reconstructed through machine learning

Susana Flecha[1,2✉], Àlex Giménez-Romero[3], Joaquín Tintoré[2,4], Fiz F. Pérez[5], Eva Alou-Font[4], Manuel A. Matías[3] & Iris E. Hendriks[2]

The decreasing seawater pH trend associated with increasing atmospheric carbon dioxide levels is an issue of concern due to possible negative consequences for marine organisms, especially calcifiers. Globally, coastal areas represent important transitional land-ocean zones with complex interactions between biological, physical and chemical processes. Here, we evaluated the pH variability at two sites in the coastal area of the Balearic Sea (Western Mediterranean). High resolution pH data along with temperature, salinity, and also dissolved oxygen were obtained with autonomous sensors from 2018 to 2021 in order to determine the temporal pH variability and the principal drivers involved. By using environmental datasets of temperature, salinity and dissolved oxygen, Recurrent Neural Networks were trained to predict pH and fill data gaps. Longer environmental time series (2012–2021) were used to obtain the pH trend using reconstructed data. The best predictions show a rate of $-0.0020 \pm 0.00054$ pH units year$^{-1}$, which is in good agreement with other observations of pH rates in coastal areas. The methodology presented here opens the possibility to obtain pH trends when only limited pH observations are available, if other variables are accessible. Potentially, this could be a way to reliably fill the unavoidable gaps present in time series data provided by sensors.

Atmospheric carbon dioxide ($CO_2$) emissions are exponentially increasing since the industrial revolution, principally due to fossil fuel use, industry and land-use change. Around a 46% of this $CO_2$ remains in the atmosphere while the rest is captured by natural compartments: the terrestrial biosphere and the ocean[1]. At present, the oceans have absorbed around an estimated 26% of the total anthropogenic $CO_2$ released from 2011 to 2020[1]. Once $CO_2$ dissolves in seawater, a sequence of chemical reactions occurs that derives in an increase of [H$^+$] ions, which results in a decrease in seawater pH. This process, a consequence of increasing atmospheric $CO_2$, is termed Ocean Acidification (OA)[2]. In addition to the pH decrease, [H$^+$] ions react with carbonate ions [$CO_3^{2-}$] to form [$HCO_3^-$], leading to a reduction of the [$CO_3^{2-}$] ion levels[3]. Low carbonate levels affect the saturation state of calcium carbonate minerals, increasing difficulties in shell-forming for calcifying marine organisms (e.g., plankton, mollusks, echinoderms and corals). Consequences of OA are an important threat to marine ecosystems visible in higher levels of the trophic chain, with complex and wide-ranging impacts on the physiology of different species and therefore on their metabolism[4,5]. These metabolic effects will have numerous consequences at an organism scale, in particular, they can cause a decrease in growth, locomotion, reproductive capacity and homeostasis if they are not capable to control the conditions for calcification[6]. Negative effects of this magnitude could cause an unexpected cascade effect impacting on the structure and functions of ecosystems and trophic networks[7] and cannot be easily generalized.

Also, ocean $CO_2$ uptake and derived OA are not homogeneous at the global scale, with some areas more affected. For instance, the Mediterranean Basin is an area where effects are stronger compared to the global ocean[8]. The Mediterranean Sea, constituting only a 0.82% of the surface and 0.32% of the volume of the global ocean, is cataloged as one of the most complex marine ecosystems, defined as a "miniature ocean"[9], inhabited

[1]Instituto de Ciencias Marinas de Andalucía (ICMAN-CSIC), Polígono Río San Pedro s/n, 11519 Cádiz, Puerto Real, Spain. [2]Instituto Mediterráneo de Estudios Avanzados (IMEDEA-UIB-CSIC), Miquel Marquès 21, 07190 Esporles, Spain. [3]Instituto de Física Interdisciplinar y Sistemas Complejos, (IFISC-UIB-CSIC), Campus UIB, 07122 Palma, Spain. [4]Balearic Islands Coastal Observing and Forecasting System (SOCIB), Parc Bit, Naorte, Bloc A 2op. pta. 3, 07121 Palma, Spain. [5]Instituto de Investigaciones Marinas (IIM-CSIC), Eduardo Cabello 6, 36208 Vigo, Spain. ✉email: susana.flecha@csic.es

by an extensive and diverse biota that represents between 4 and 18 % of the world's total marine species[10] and serves as a model[9] to anticipate the responses of the global ocean to different types of pressures. It has been also defined as a climate change "hot spot"[8], whit OA and its derived consequences characterized as one of the climatic threats with the greatest potential impact, followed by the temperature and UV radiation increase[11]. The temperature rise in this semi-enclosed sea is expected to be two to four-fold times higher than that in the global ocean[12,13]. In addition, the sixth assessment report (AR6) of the IPPC7, places a high level of confidence on the increase in frequency of heatwaves and ongoing ocean acidification[14]. Recent studies have confirmed that there is a trend of around 0.34 °C warming per decade in the Mediterranean Outflow Water (MOW) through the Strait of Gibraltar towards the Atlantic Ocean[15], associated with decreasing values of pH. Furthermore, in the Mediterranean Sea, due to its biogeochemical and hydrodynamic characteristics, such as the high alkalinity of its waters and the active thermohaline circulation[16], there is a larger absorption of atmospheric $CO_2$ and an intense transport of this $CO_2$ from the oceanic surface to deep areas[17,18], already observed in the MOW[19,20], with estimated OA trends of − 0.0044 pH units per year in the Strait of Gibraltar[19] and ranging from − 0.0017 to − 0.003 in the Mediterranean Basin[21,22].

The Mediterranean Sea has an extensive coastline, which extends for 46,000 km and is shared by 21 countries[23]. Coastal zones, as transitional areas, are inherently complex systems due to the strong biogeochemical-physical coupling, occurring relevant biogeochemical exchanges. Interactions in coastal areas involve terrestrial inputs of nutrients and particulate matter from river runoff and groundwater discharges, oceanic forcing (waves, tides, and currents), and atmospheric exchange of aerosols and trace gases, all of them which are influenced by the intense human activity in the coastline[24]. Hence, processes related to the carbon system in coastal areas are more dynamic and complex than in the open ocean[25], and the range of pH change between − 0.023 and 0.023 pH units per year[26] is therefore ∼ 35 times larger than in the open ocean with − 0.0013 to − 0.0026 pH units year$^{-1}$[27]. In particular, anthropogenic $CO_2$ inputs appear to play a minor role compared to other sources of variability in coastal zones[28]. Therefore, it is difficult to foresee how the pH conditions in the coastal areas in the year 2100 will differ from the present, due to the lack of knowledge on precise current pH values in the different coastal ecosystems and their variability obtained from long time series. Carbonate chemistry and in particular pH fluctuations are characterized by a wide spatial heterogeneity and temporal variability (daily and seasonal oscillations) in coastal ecosystems[28–30]. The variability of pH is determined by a wide range of physical and biogeochemical processes, from mesoscale hydrological processes to small-scale metabolic processes[31].

The primary production in the western Mediterranean Sea is characterized by a seasonal variability induced by the increase of the surface layer nutrients by the winter vertical mixing in the water column[32]. In addition, the presence of macrophytes[33] in the coastal areas of the northern Mediterranean Sea, mainly the endemic *Posidonia oceanica* whose meadows extend from the surface to 30–40 m depth, are defined as highly productive habitats. In these ecosystems, variability tends to follow daily and seasonal cycles, since biological metabolism is responsible for variations in the concentrations of oxygen ($O_2$) and $CO_2$[29,34]. Thus, increasing pH values are expected for autotrophic ecosystems (production > respiration) during daylight hours. Indeed, recent studies indicate that seagrass meadows can locally alleviate low pH conditions for extended periods of time with important implications for the conservation and management of coastal ecosystems[35].

Nevertheless, changes in pH can appear idiosyncratic and display a diversity of patterns depending on the coastal area under consideration, as many drivers of the carbon system can influence these variable ecosystems, including temperature variability, biological activity and terrestrial and open ocean inputs[26]. Therefore, the properties of the carbon system have to be evaluated while taking into account the different interactions in every area. To the present day, there is still a lack of understanding of how coastal areas behave and how they contribute to the global carbon budget, also in part due to the intensive effort necessary to obtain representative time series of the carbon system data according to standard practices. The Global Ocean Acidification Observing Network (GOA-ON) defined that the accuracy included in the "weather goal" should be better than 0.02 and for the "climate goal" < 0.003 pH units[36]. Instrumentation for autonomous pH measurements have improved in recent years and production costs have come down, but they remain complex and relatively expensive. For climate change studies commercial oceanographic instrumentation barely accomplish the GOA-ON "climate goal" accuracy recommendation, with only spectrophotometric devices and Ion Sensitive Field-Effect Transistors (ISFETs) based pH probes reaching the standards. In this sense, the SAMI-pH sensor (Submersible Autonomous Moored Instrument, Sunburst Sensors, LCC), based on spectrophotometric techniques has been denoted as an excellent pH sensor for OA studies[19]. However, the maintenance of oceanographic time series stations entails several operational and non-operational difficulties, involving financial costs, meteorological risks (i.e. bad conditions for navigation, instrumentation loss, etc.), deployment in areas with high transit, issues essentially related to the sensor itself (i.e. instrumental failure) and possible human errors. Therefore, the appearance of data gaps is common, implying the lack of pH data obtained using high-quality instrumentation for global carbon studies.

Currently, novel computational methods based on Machine Learning (ML) are allowing to tackle these data absence difficulties. Machine Learning is a part of Artificial Intelligence that has attained a mature status in the last decade or so, particularly through the so-called Deep Learning (DL)[37], with major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years[38]. In particular, some DL techniques are useful in time series forecasting[39] and also in the reconstruction of coupled time series[40], such as Recurrent Neural Network (RNN) architectures like Long Short-Term Memory (LSTM)[41] or Gated Recurrent Unit (GRU).

Nowadays, there is an increasing number of studies that use DL to understand the processes involved in the carbon system variability, but mainly focused on the open ocean[42–46], while relatively few studies focused on coastal seas[47,48] and none specifically in the Mediterranean coastal Sea, perhaps because of the complexity and heterogeneity of the basin and its continental shelves. Therefore, the main objective of this study is to obtain the trend for pH decrease in the coastal Balearic Sea by applying Machine Learning techniques. In addition, this
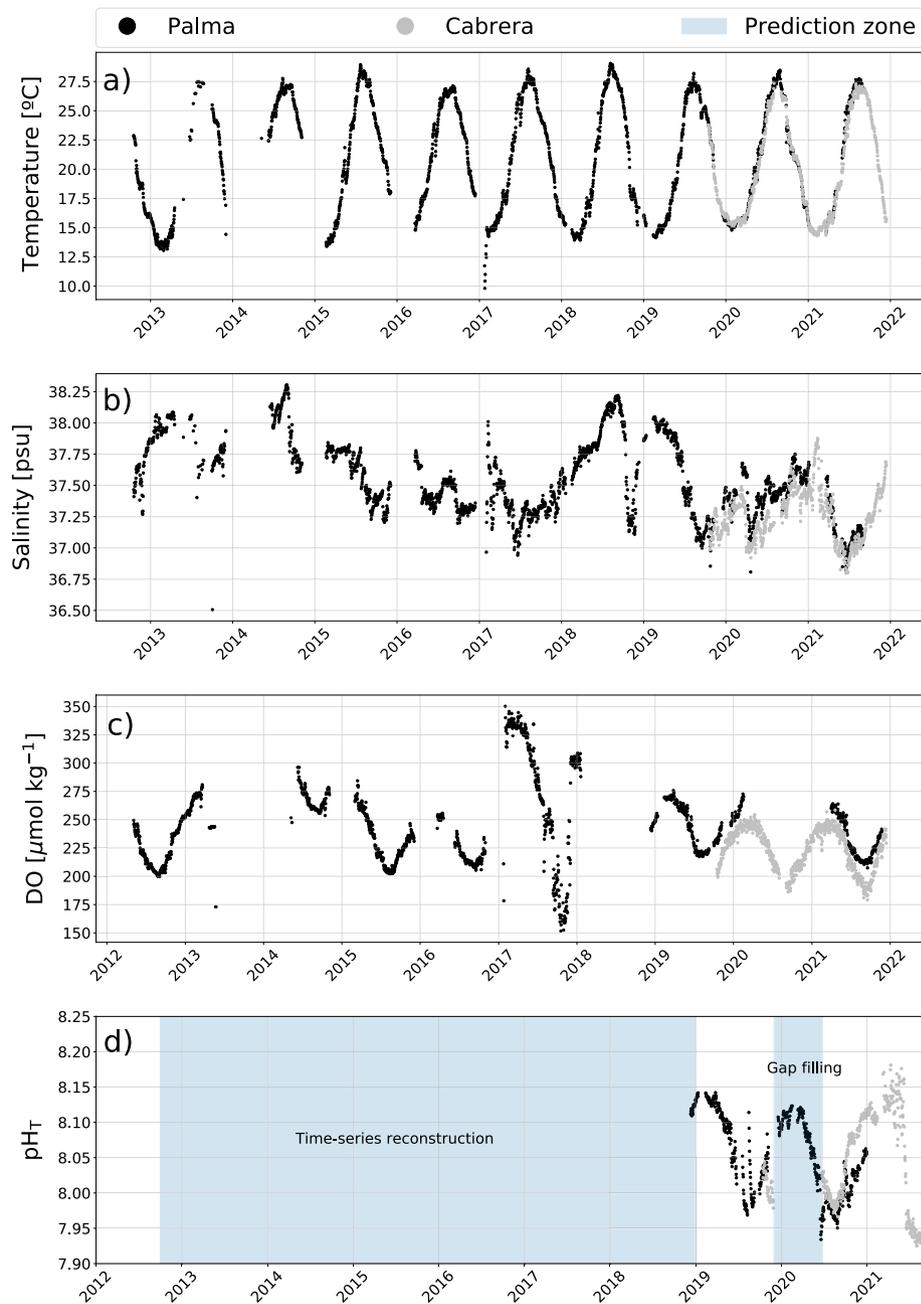
**Figure 1.** Daily averaged time series data from the Bay of Palma (black dots) and Cabrera stations (grey dots): (**a**) temperature (°C), (**b**) salinity (psu), (**c**) dissolved oxygen (DO) ($\mu$mol kg$^{-1}$) and (**d**) pH$_T$ in pH units. The pH time series of the Bay of Palma will be reconstructed in the period 2012–2021 while only gaps will be filled in Cabrera, as marked in blue in the figure.

study aims to provide a useful tool to fill gaps in pH time series and to reconstruct pH data when additional environmental variables are available.

## Results

**Time series data.** The collection of pH values, in total scale (pH$_T$), started in December 2018 in the Bay of Palma, recording data almost continuously until the end of 2021. In the Cabrera station, pH$_T$ was obtained from November 2019 to December 2021, with a relevant data gap from December 2019 to June 2020 (Fig. 1d) due to a sensor malfunction with a reparation prolonged for an extended period of time owing to the Covid-19 lockdown. Additional environmental parameters like temperature, salinity and dissolved oxygen (DO) concentration are available from the Bay of Palma station since 2012, while only a limited time series of these variables (since 2019) is available for Cabrera (Fig. 1a–c).

In both stations, temperature ranged from a minimum of 12.99 °C to maximum values of 29.07 °C from 2012 to 2021, with no observed differences between the stations in Cabrera and the Bay of Palma (Fig. 1a). The surface water temperatures are a clear representation of the typical Mediterranean climate seasonality with mild winters and warm to hot summers. Salinity did not show a repetitive seasonal pattern between years in either stations. However, in Cabrera salinity is slightly lower than in the Bay of Palma. During the data acquisition period, the lowest salinity value of 36.83 was found in Cabrera and highest of 38.30 in the Bay of Palma (Fig. 1b).

The surface water of the coastal sites in the Balearic Sea in the Palma Bay and the Cabrera stations was highly saturated with oxygen during all the seasons, with DO concentrations up to $348.94\,\mu mol\,kg^{-1}$ during winter and of $150.66\,\mu mol\,kg^{-1}$ during the summer and early autumn (Fig. 1c). $pH_T$ values obtained starting in December, 2018 to December, 2021 increased during winter reaching up to 8.18 pH units at in situ temperature and decreasing to 7.91 pH units in summer, with the highest variability and maximum and minimum values measured in Cabrera (Fig. 1d).

Considering sampling period of the additional (temperature and salinity) and calculated parameters (Total Alkalinity; TA) was larger in the Bay of Palma compared to Cabrera, we evaluate the linear tendencies with time for the Bay of Palma variables. The sea surface temperature in the Bay of Palma increased with a rate of $0.035 \pm 0.008$ °C per year ($R^2 = 0.008$, $p$-value $< 0.001$) from 2012 to 2021, whereas the salinity decreased significantly with $-0.059 \pm 0.002$ psu per year ($R^2 = 0.25$, $p$-value $< 0.001$). The annual trend for TA, clearly related to the decrease in surface salinity, showed a relevant decrease of $-4.0 \pm 0.4\,\mu mol\,kg^{-1}$ ($R^2 = 0.0379$, $p$-value $< 0.001$), supported by the discrete water samples for TA obtained during the period from 2019 to 2021 (Fig. S3).

### Reconstruction pH time series with deep learning.

The amount of available $pH_T$ data from both Palma Bay and Cabrera stations is comparable and relatively short (mostly in Cabrera), but the length of the additional ambient data (temperature, salinity and DO) differs enormously among stations. Thus, there is a need to approach the time series prediction problem for both sites with different objectives. Common to both sites, a DL model with a RNN architecture will be developed to predict the $pH_T$ time series from the accompanying ambient data (temperature, salinity and dissolved oxygen), which are expected to be correlated with $pH_T$[42,47]. To avoid the effect of site-specific correlations between ambient and $pH_T$ time series, the model will be trained independently with the dataset of each location. In this way a proper model calibration is ensured and the prediction power of the model is enhanced. In the Bay of Palma, the model will be used to reconstruct the $pH_T$ time series from 2012, exclusively from the points for which the full set of ambient time series data are available (Fig. 1d). This is not possible in Cabrera, due to the fact that no temperature, salinity and DO concentration is available before 2019. Fortunately, these time series do not have the same gaps that the $pH_T$ time series exhibits. Thus, we will use the model to fill the gaps in the $pH_T$ time series from 2019 to present, as shown in (Fig. 1d).

A BiDireccional Long Short-Term Memory (BD-LSTM) neural network (Fig. S4) was selected as the best recurrent neural network architecture to reconstruct the $pH_T$ time series in the Bay of Palma. The training process was successfully completed with no signs of overfitting achieving less than 1% error in both training and validation sets (Fig. 2a). The BD-LSTM neural network was able to fairly predict the majority of the individual pH data points in the time series, although there are some deviations (Fig. 2b). Furthermore, the time series pattern is perfectly captured by the neural network (Fig. 2c). Notice the gaps in the reconstructed pH points (in red) in (Fig. 2d), that are those for which the full ambient time-series is not available. Finally, the reconstructed pH data using the BD-LSTM model was used to assess the decadal trend of acidification in Palma Bay, which yielded $-0.0020$ units of pH per year (black line in Fig. 2d). Indeed, to further characterize this decadal trend, 1000 independent training-prediction processes were carried out using a BD-LSTM neural network. The results showed a mean slope of $-0.0020 \pm 0.00054$ for the decadal acidification trend (see "Methods").

Regarding the Cabrera data set, with the available ambient time series it is only possible to fill the data gaps, task for which a BD-LSTM neural network was also used. As for the Bay of Palma the training process was successfully completed with no signs of overfitting, yielding less than 1% error in both training and validation dataset (Fig. 3a). The model fairly predicts most of the individual $pH_T$ data points in the training dataset, showing some deviations as usual (Fig. 3b). The tendency of the time series is perfectly captured by the model (Fig. 3c) and thus the gap can be filled with reliable data, red points in (Fig. 3d).

## Discussion

The achievement of long term oceanographic data series suitable to evaluate the effects of climate change constitutes a great operational effort which is unequivocally accompanied by partial data loss due to multiple factors (human and instrumental). The advances in the development of pH sensors are enabling the acquisition of precise pH data without identified drift through highly accurate indicator-based spectrophotometric methods[49]. However, in order to determine OA trends, several years of quality seawater pH data are needed, adding more difficulty to the vicissitudes inherent to field work. Recently, the application of computational methods based on Deep Learning (DL) is becoming a useful tool to fill the gaps due to data loss. Several studies have implemented the DL methodology and successfully predicted bio-optical and biogeochemical parameters[44,46–48,50–54].

Here, the application of a BiDireccional Long Short-Term Memory (BD-LSTM) neural network to predict $pH_T$ from physical data, namely temperature, salinity, and dissolved oxygen, the latter as a key indicator of biological activity, permitted the reconstruction of gaps in the time series of $pH_T$ and allowed the reconstruction of nine years of $pH_T$ data. The BD-LSTM architecture has been proved extremely effective in predicting sequence data, such as time series, as they combine the information for both front and back directions of time (Fig. S4)[55], and is more effective (accurate and stable) compared to unidirectional Long Short-Term Memory neural networks. Therefore, in this study the BD-LSTM offered better estimation results over the other neural networks considered to reconstruct time series but also in the completion of missing data.
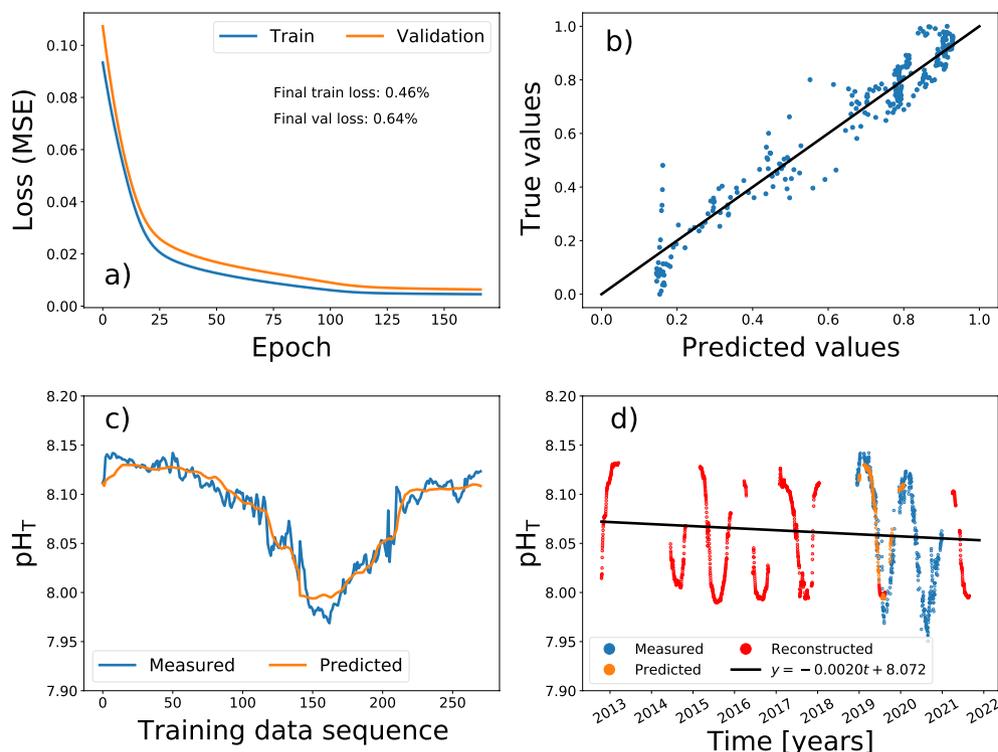
**Figure 2.** Bidirectional LSTM neural network model applied to assess the decadal $pH_T$ trend in the Bay of Palma: (**a**) training process monitoring loss for both training and validation sets, (**b**) predicted $pH_T$ values against their true values where the black line is the reference for a perfect prediction, (**c**) predicted $pH_T$ time series in the training process (orange) and ground truth series (blue) and (**d**) final prediction for the decadal $pH_T$ time series using the output data of the trained model and the measured data. Measured pH data shown in blue, predicted data in the training process is shown in orange and reconstructed data is shown in red. The black line represents the decadal pH trend.

In the Cabrera station, the BD-LSTM permitted a reliable reconstruction of the gaps in $pH_T$ data from December 2019 to June 2020 (Fig. 1d), constituting an advantageous methodology to support the acquisition of long time series data without losing accuracy, as the model can reproduce $pH_T$ data with an error lower than 1% (Fig. 3b), closely following the annual variability of the observations (Fig. 3d).

The ability of the BD-LSTM to reconstruct time series was observed through the reconstruction of nine years of $pH_T$ data in the Bay of Palma station (Fig. 2d). The modeled $pH_T$ data combined with the observations allowed the accomplishment of a long pH time series in order to estimate a pH trend, seasonally adjusted through a sinusoidal fitting, with a rate of decrease of $0.0020 \pm 0.00054$ pH units per year ($R^2 = 0.1$, $p$-value $< 0.001$, Fig. S1), and represents the first estimate of pH trend obtained in the Balearic coastal Sea. Additionally, we applied a linear fit on the reconstructed pH time series obtaining trend of $-0.0025 \pm 0.00053$ $year^{-1}$ ($R^2 = 0.01$, $p$-value $< 0.001$). This fit was discarded, because it was shown to introduce a bias in the pH decrease trend.

The observed pH decrease in the Balearic Sea coastal area is well aligned with OA trends reported for open ocean areas, from $-0.0013$ pH units $year^{-1}$ in the Munida station (New Zealand) to the high trend found in the Cariaco Basin station up to $-0.0026$ pH units per year[27]. The processes associated with the increased pH decline in the Cariaco Basin where related to the upwelling of Subtropical Underwater, rich in dissolved inorganic carbon, thus lowering the pH.

In the Mediterranean Sea, previous annual estimates in open ocean areas ranged from $-0.003$ to $-0.0044$[19,22], reflecting the effect of the hydrodynamical and biogeochemical characteristics of the basin on the seawater pH variability[18,56,57]. However, it can be assumed that differences in physical oceanography and ecological processes between areas may modulate local changes of pH. In a coastal Mediterranean area located in the northwestern basin, close to Villefranche-sur-Mer (France), a rate of pH change of $-0.0028 \pm 0.0003$ pH units $year^{-1}$ was observed[21] and attributed principally to atmospheric forcing and secondly to increased warming. The calculated trend of pH decrease due to the atmospheric $CO_2$ growth during the period of this study, from 2013 to 2021, was of $0.0025 \pm 0.0002$ pH units per year ($R^2 = 0.95$, $p$-value $< 0.001$), consistently related to the seawater pH decline. Therefore, these analyses suggest that the atmospheric forcing is the main driver responsible for the pH decreasing trend found in the surface coastal Balearic Sea. Subsequently, the difference between the seawater pH decreasing trend obtained and the pH trend calculated from the atmospheric levels could be related to natural biogeochemical processes, not distinctly quantifiable with the available length of the Bay of Palma pH time series.

In addition, the effect of temperature on surface ocean pH, occurring directly through the temperature dependence of the seawater $CO_2$ chemistry, as changes in temperature and salinity influence the equilibrium
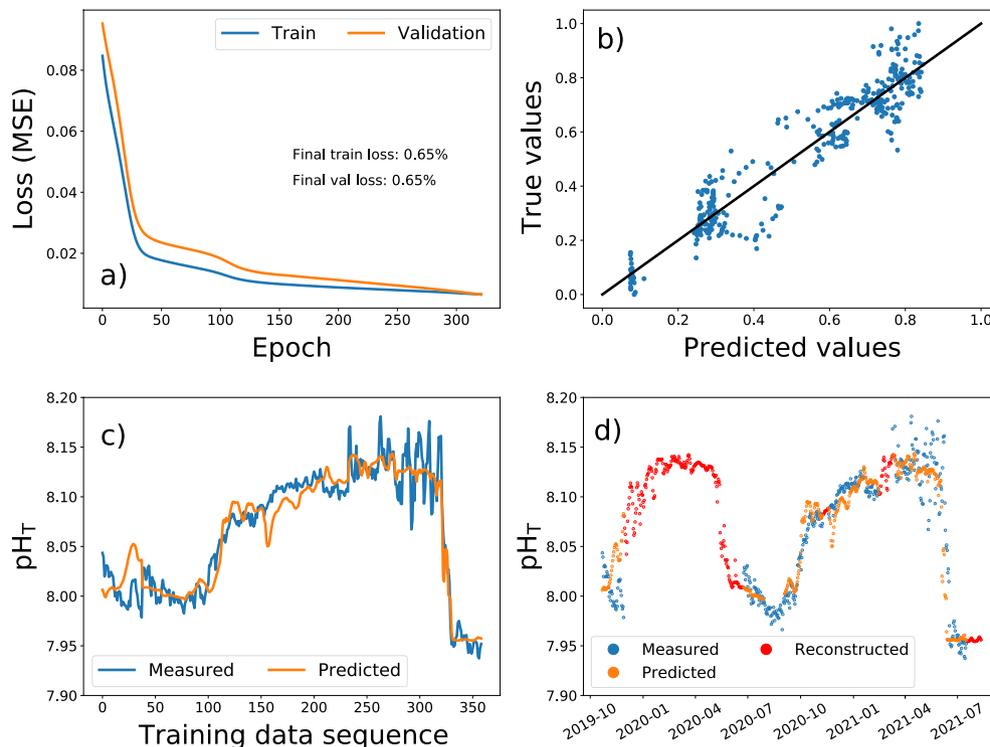
**Figure 3.** Bidirectional LSTM Neural Network model applied to fill the gaps in the $pH_T$ time series in Cabrera: (**a**) training process monitoring loss for both training and validation sets, (**b**) Predicted $pH_T$ values against their true values where the black line is the reference for perfect prediction, (**c**) Predicted $pH_T$ time series in the training process (orange) and ground truth series (blue) and (**d**) Gaps in the $pH_T$ time series filled with the trained model (red), while measured $pH_T$ are shown in blue and predicted data in the training process shown in orange.

constants of the oceanic $CO_2$ system and indirectly through air-sea exchange of $CO_2$, can be considered. The influences of these two temperature processes on surface ocean pH has been found responsible of a 50% of the increase in $[H^+]$ ions, thus a pH decrease, in the surface layers of the Iceland and Irminger Seas[58]. In the Mediterranean Sea northwestern basin, a temperature increase of $0.072 \pm 0.022\,°C\,year^{-1}$ was estimated to be responsible for a 40% of the pH decrease[21]. The obtained temperature variability in the Balearic Sea coastal area during this study was of $0.035 \pm 0.008$ ($R^2 = 0.008$, $p$-value $< 0.001$, Fig. S2), indicating that temperature-driven changes could also be assumed to affect the pH trend.

The observed seasonal variability of the data, presented a $pH_T$ increase from 7.91 during summer up to 8.18 pH units (Fig. 1d) in winter seasons, clearly followed by the TA values (Fig. S3). Seasonal changes in TA levels in the study area are ranging from around 2350 to $2550\,\mu mol\,kg^{-1}$ (Fig. S3), largely overtaking the seasonal differences reported previously in the Balearic Sea of up to $50\,\mu mol\,kg^{-1}$ in total[59]. This discrepancy in variability could be explained by the intense metabolic processes at the coastal location of the Bay of Palma station. This shallow area has a strong coverage of *Posidonia oceanica*, which due to its high ecosystem production[60] could be triggering an increase of pH and TA levels, as seen in salinity normalized TA values (NTA, not shown) during winter-spring, due to the uptake of nitrate and phosphate and the calcium carbonate dissolution[59,61], and during summer, related to the lower community production[62] a NTA-pH decrease[59].

Another result from this study worth to mentioning is the obtained decreasing TA trend in the Bay of Palma of $-4.0 \pm 0.4\,\mu mol\,kg^{-1}$ per year. Although the Western Mediterranean is characterized with lower total alkalinity values in relation to the rest of the basin, resulted from the nearby influence of Atlantic waters, less salty with low-alkalinity water[63,64], was not expected to influence decreasing decadal TA values. In the northwestern basin, TA values increased over time at a rate of $2.08 \pm 0.19\,\mu mol\,kg^{-1}\,year^{-1}$. In the Balearic Sea, the decreasing TA confirm the Atlantic forcing on the alkalinity values and the negligible TA discharges due to rivers in the Balearic Islands. There is a marked south-to-north surface gradient in the western region coupled with the west-to-east gradient of alkalinity in the Mediterranean Sea related to the Atlantic influence[59,65]. Due to a well-established linear relationship of TA and salinity[66] and the calculated origin of our values[65] we cannot neglect the strong TA related to the salinity decrease in the study area of $-0.059 \pm 0.002$ psu per year ($R^2 = 0.25$, $p$-value $< 0.001$). This rate is in agreement with the salinity decrease found at the coastal site at Villefranche-sur-Mer ($-0.0017 \pm 0.0044$ psu $year^{-1}$)[21]. Notwithstanding, the intense salinity decrease observed in the Bay of Palma can be linked to a decrease of the intensity of the southern spreading of the Balearic Current trough the Ibiza channel (located between Ibiza and Mallorca Islands) driven by mesoscale processes, and the prevalence of new Atlantic Water
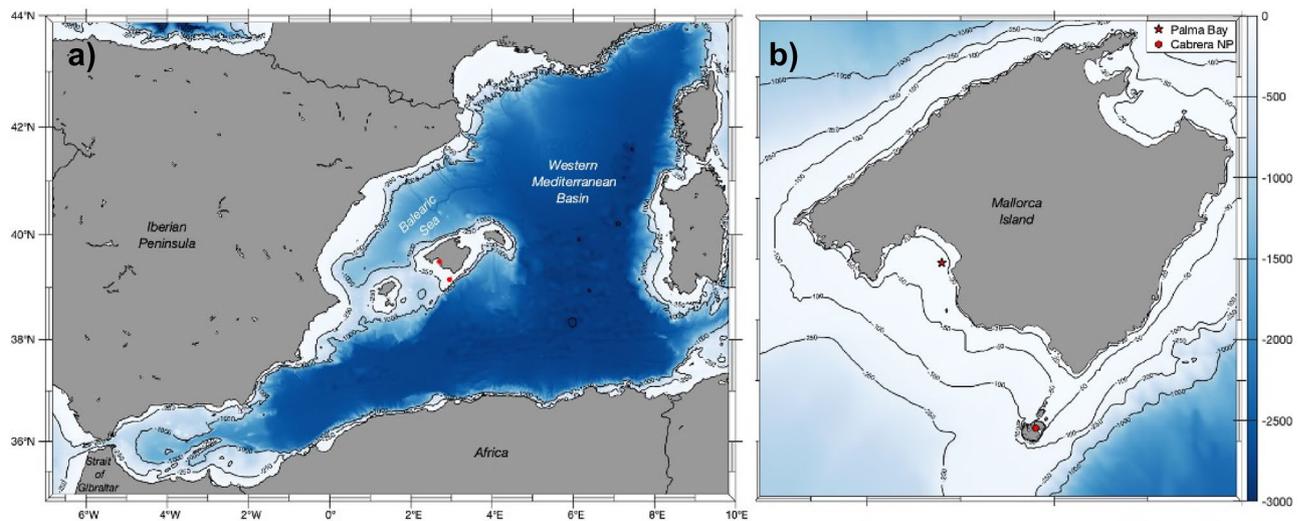
**Figure 4.** (**a**) Map of the stations location in the Western Mediterranean Sea Basin (red dots) and (**b**) detailed location of the Bay of Palma (red star) and the Cabrera National Park (Cabrera NP, red dot) study sites. Maps were developed with the MATLAB® R2010b software (https://mathworks.com) by using the M_Map toolbox[70].

coming from the Strait of Gibraltar[67]. Although, this observation is out of the scope of this study and therefore further investigation is needed.

In summary, this work pointed out the useful use of DL techniques, specifically the BD-LSTM architecture, to reconstruct pH data relevant to evaluate seasonal pH variability and to elucidate the climate change consequences, as the OA effect, in a coastal area of the Balearic Sea, which can be extended to the coastal areas of the Western Mediterranean Sea Basin. Nevertheless, future research is necessary to assess and confirm these regional trends, which highlights the importance of maintaining the time series monitoring networks whose data are the base of this study.

## Methods

### Study area.
We monitored two coastal stations located in the Archipelago of the Balearic Islands in the Western Mediterranean Sea (Fig. 4). One site was positioned within the Bay of Palma (39° 29.57088′ N; 2° 42.02430′ E (Fig. 4b) at a fixed station consisting of an oceanographic buoy managed by the Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB; https://www.socib.eu/). Here meteorological, hydrological and hydrodynamic data are collected with an hourly frequency since October 2012. The buoy is located at the surface over 20m bottom depth. The Bay of Palma is a large bay with a surface area of 217 km$^2$ and approximately 30% seagrass cover[68]. The second station was located at 4 m depth on a mooring line over 8 m bottom depth deployed in the Marine and Terrestrial National Park of the Archipelago of Cabrera (39° 9.08217′ N; 2° 57.04767′ E; Fig. 4b). The mooring line is in a small bay of just under 1 km$^2$ and full protection with the largest meadow of the archipelago, covering 89.1% of the surface area between 0 and 10 m depth[69]. Neither site has important freshwater inputs. Both stations are part of the Balearic Ocean Acidification Time Series (BOATS) network included in the Interdisciplinary Thematic Platform: Water:iOS (https://pti-waterios.csic.es/).

### Data collection.
In both stations a SAMI-pH (Sunburst Sensors LCC) was attached, at 1 m in the Bay of Palma and at 4 m depth in Cabrera. The pH sensors were measuring pH, in the total scale (pH$_T$), hourly since December 2018 in the Bay of Palma and since November 2019 in Cabrera. The sensor precision and accuracy are < 0.001 pH and ± 0.003 pH units, respectively. Monthly maintenance of the sensors was performed including data download and surface cleaning.

Temperature and salinity from the Bay of Palma oceanographic buoy was obtained from October 2012 and for the Cabrera mooring line from November 2019 with a CT SBE37 (Sea-Bird Scientific©) in both stations. Accuracy of the CT is ± 0.002 °C for temperature and ± 0.003 mS cm$^{-1}$ for conductivity. Additionally, oxygen data from a SBE 63 (Sea-Bird Scientific ©) sensor attached to the CT in Cabrera and from a YSI 6600V2-4 Multiparameter Water Quality Sonde with a 6450 ROX DO sensor (Yellow Spring Instruments Inc. ©)[71] and a miniDot (PME, Inc. ©) in the Bay of Palma were used. Accuracy of oxygen sensors is ± 2%, ± 1% and ± 5% for the SBE 63, the YSI and the miniDot, respectively.

Periodically water samplings for dissolved oxygen (DO), pH in total scale at 25 °C (pH$_{T25}$) and total alkalinity (TA) were obtained during the sensor maintenance campaigns. DO and (pH$_{T25}$) samples were collected in order to validate the data obtained by the sensors.

DO concentrations were evaluated with the Winkler method modified by Benson and Krause[72] by potentiometric titration with a Metrohm 808 Titrando with a accuracy of the method of ± 2.9 μmol kg$^{-1}$ and with an obtained standard deviation from the sensors data and the water samples collected of ± 5.9 μmol kg$^{-1}$.

pH$_{T25}$ data was obtained by the spectrophotometric method with a Shimadzu UV-2501 spectrophotometer containing a 25 °C-thermostated cells with unpurified *m*-cresol purple as indicator following the methodology

|  | Hidden layers | Nodes/cells | Window size | Activation function | Output function | Loss | Learning rate | Optimizer |
|---|---|---|---|---|---|---|---|---|
| SRNN | 1 | 3 | 6 | Tanh | Sigmoid | MSE | 0.01 | Adam |
| LSTM | 1 | 3 | 6 | Tanh | Sigmoid | MSE | 0.01 | Adam |
| BD-LSTM | 1 | 3 | 6 | Tanh | Sigmoid | MSE | 0.01 | Adam |
| BD-GRU | 1 | 1 | 6 | Tanh | Sigmoid | MSE | 0.01 | Adam |

**Table 1.** Optimal parameters used for the different RNN architectures.

established by Clayton and Byrne[73] by using Certified Reference Material (CRM Batch #176 supplied by Prof. Andrew Dickson, Scripps Institution of Oceanography, La Jolla, CA, USA). The accuracy obtained from the CRM Batch was of ± 0.0051 pH units and the precision of the method of ± 0.0034 pH units. The mean difference between the SAMI-pH and discrete samples was of 0.0017 pH units.

TA samples were collected in 50 ml Falcom vials and poisoned with 20 µL of $HgCl_2$ and determined by open cell potentiometric titration with a Titrando 808 (Metrohm) following the Standard Operation Procedure (SOP) 3b[74]. TA values were also calculated from the temperature and salinity values obtained in the Bay of Palma from 2012 by using a second-order polynomial model for TA specifically described for the Mediterranean Basin[65].

pH values due to the atmospheric $CO_2$ levels were estimated by using the CO2SYSv3 program[75], with the most internally consistent and preferred carbon[76,77] and sulphate dissociation constants[78] for current surface ocean studies[79], with the Bay of Palma in situ temperature and salinity, the calculated TA values and the atmospheric $CO_2$ levels converted from dry air to wet[80] as inputs. Carbon dioxide ($CO_2$) atmospheric molar fraction used was obtained from the monitoring station of Lampedusa (LMP), Italy of the NOAA (National Oceanic and Atmospheric Administration, USA) monitoring network[81].

**Data processing.** Once data was validated, several processing steps were performed to ensure an optimal training process for the neural network models. First, all the data of the time series were re-sampled by averaging the data points obtaining a daily frequency. Afterwards, a standard feature-scaling procedure (min-max normalization) was applied to every feature (temperature, salinity and oxygen) and to $pH_T$. Finally, we built our training and validations sets as tensors with dimensions ($batch_{size}$, $window_{size}$, $N_{features}$), where $batch_{size}$ is the number of examples to train per iteration, $window_{size}$ is the number of past and future points considered and $N_{features}$ is the number of features used to predict the target series. Temperature values below $T = 12.5$ °C were discarded as they are considered outliers in sensor data outside the normal range in the study area.

**Computing the trend of seasonal data.** The trend of seasonal time-series is often computed by means of statistical methods based on moving averages or more advanced techniques such as the Seasonal Trend Decomposition Loess[82]. Nevertheless, these procedures do not work with gappy time series, so that a different approach is needed. In this work we fitted the following oscillatory function with trend to our data:

$$y(t) = A\sin(\omega t + \phi) + Bt + C \,, \qquad (1)$$

where the parameter $B$ corresponds to the trend of the data.

Moreover, after this fit, the seasonal component ($A\sin(\omega t + \phi)$) can be removed from the original time-series and a standard linear regression can be performed to the transformed data to obtain the trend (which is exactly $B$) with the $R^2$ and $p$-value estimates given by the linear regression (Figs. S1, S2).

**Selecting the best neural network architecture.** Several recurrent neural network (RNN) architectures were considered as candidates to reconstruct the pH time series, including a Simple Recurrent neural network (SRNN), Long-Short Term Memory (LSTM), BiDirectional LSTM (BD-LSTM, (Fig. S4) and BiDirectional Gated Recurrent Unit (BD-GRU).

Initially, manual tests were performed on each architecture to determine the optimal set of parameters that yielded the best possible results. These tests were based on minimizing the errors in both training and validation set while avoiding overfitting. To avoid overfitting, we implemented automated callbacks to stop the training process whenever the validation loss increased or crossed the training loss. During this test we determined the minimum number of nodes, which helps in avoiding overfitting and the minimum window size, which allows to use the most possible number of data points for training and prediction. All the RNNs were trained in batches of size 32. To enhance clarity and accessibility, the optimal values obtained for the more relevant parameters are summarized in Table 1.

In order to identify the best-performing architecture an automated procedure was developed to statistically compare the outputs of each model. Each architecture was trained in 1000 independent processes, ensuring a final training mean-squared error of less than 0.8% while avoiding overfitting implementing the previously mentioned callbacks. The code used for the analysis can be found in[83].

In Table 2, a summary of the statistical results obtained for each architecture is presented. All architectures provide similar training and validation errors and provide similar results for the decadal $pH_T$ trend, predicting a slope of around −0.0020 pH units per year with an intercept of 8.07 pH units. However, the BD-LSTM turns out to be the architecture providing most accurate (smallest training and validation errors) and precise (smallest

| | Slope | Intercept | Training error | Validation error | Training epochs | Training time |
|---|---|---|---|---|---|---|
| RNN | − 0.0021 ± 0.00077 | 8.07 ± 0.006 | 0.54 ± 0.08 | 0.72 ± 0.12 | 293 ± 95 | 15.52 ± 4.75 |
| LSTM | − 0.0018 ± 0.00067 | 8.06 ± 0.005 | 0.49 ± 0.03 | 0.68 ± 0.05 | 245 ± 68 | 17.55 ± 4.21 |
| BD-LSTM | − 0.0020 ± 0.00054 | 8.07 ± 0.004 | 0.46 ± 0.03 | 0.64 ± 0.04 | 167 ± 45 | 15.13 ± 3.00 |
| BD-GRU | − 0.0020 ± 0.00066 | 8.07 ± 0.005 | 0.51 ± 0.07 | 0.74 ± 0.10 | 347 ± 95 | 27.68 ± 6.84 |

**Table 2.** Statistical comparison between different RNN architectures.

statistical error) results (Fig. S4). Thus, we selected the BD-LSTM neural network as the best architecture to reconstruct the $pH_T$ time series. Data corresponding to the Bay of Palma were used in the selection of the best neural network architecture.

The code and data used to determine the best neural network architecture can be found in a GitHub repository[83].

## References

1. Friedlingstein, P. *et al.* Global carbon budget 2021. *Earth Syst. Sci. Data Discuss.* **1–191**, 2021. https://doi.org/10.5194/essd-2021-386 (2021).
2. Caldeira, K. & Wickett, M. E. Anthropogenic carbon and ocean pH. *Nature* **425**, 365. https://doi.org/10.1038/425365a (2003).
3. Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. Ocean acidification: The other $CO_2$ problem. *Ann. Rev. Mar. Sci.* **1**, 169–192. https://doi.org/10.1146/annurev.marine.010908.163834 (2009).
4. Kroeker, K. J. *et al.* Impacts of ocean acidification on marine organisms: Quantifying sensitivities and interaction with warming. *Glob. Change Biol.* **19**, 1884–1896. https://doi.org/10.1111/gcb.12179 (2013).
5. Nilsson, G. E. *et al.* Near-future carbon dioxide levels alter fish behaviour by interfering with neurotransmitter function. *Nat. Clim. Chang.* **2**, 201–204. https://doi.org/10.1038/nclimate1352 (2012).
6. Hendriks, I. E. *et al.* Biological mechanisms supporting adaptation to ocean acidification in coastal ecosystems. *Estuar. Coast. Shelf Sci.* **152**, A1–A8. https://doi.org/10.1016/j.ecss.2014.07.019 (2015).
7. Zunino, S., Libralato, S., MelakuCanu, D., Prato, G. & Solidoro, C. Impact of ocean acidification on ecosystem functioning and services in habitat-forming species and marine ecosystems. *Ecosystems* **24**, 1561–1575. https://doi.org/10.1007/s10021-021-00601-3 (2021).
8. Giorgi, F. Climate change hot-spots. *Geophys. Res. Lett.* **33**, L08707. https://doi.org/10.1029/2006GL025734 (2006).
9. Bethoux, J. P. *et al.* The Mediterranean Sea: A miniature ocean for climatic and environmental studies and a key for the climatic functioning of the North Atlantic. *Prog. Oceanogr.* **44**, 131–146. https://doi.org/10.1016/S0079-6611(99)00023-3 (1999).
10. Bianchi, C. N. & Morri, C. Marine biodiversity of the Mediterranean Sea: Situation, problems and prospects for future research. *Mar. Pollut. Bull.* **40**, 367–376. https://doi.org/10.1016/S0025-326X(00)00027-8 (2000).
11. Micheli, F. *et al.* Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: Assessing current pressures and opportunities. *PLoS One* **8**, e79889. https://doi.org/10.1371/journal.pone.0079889 (2013).
12. Vargas-Yáñez, M. *et al.* Warming trends and decadal variability in the western mediterranean shelf. *Glob. Planet. Change* **63**, 177–184. https://doi.org/10.1016/j.gloplacha.2007.09.001 (2008).
13. Vargas-Yáñez, M. *et al.* Climate change in the Western Mediterranean sea 1900–2008. *J. Mar. Syst.* **82**, 171–176. https://doi.org/10.1016/j.jmarsys.2010.04.013 (2010).
14. Masson-Delmotte, V. *et al.* Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In *Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2021).
15. García-Lafuente, J. *et al.* Hotter and weaker mediterranean outflow as a response to basin-wide alterations. *Frontiers* **8**(613444). https://doi.org/10.3389/fmars.2021.613444 (2021).
16. Álvarez, M. *et al.* The $CO_2$ system in the Mediterranean Sea: A basin wide perspective. *Ocean Sci.* **10**, 69–92. https://doi.org/10.5194/os-10-69-2014 (2014).
17. Hassoun, A. E. R. *et al.* Acidification of the Mediterranean Sea from anthropogenic carbon penetration. *Deep Sea Res. Part 1 Oceanogr. Res. Pap.* **102**, 1–15. https://doi.org/10.1016/j.dsr.2015.04.005 (2015).
18. Palmiéri, J. *et al.* Simulated anthropogenic $CO_2$ storage and acidification of the Mediterranean Sea. *Biogeosciences* **12**, 781–802. https://doi.org/10.5194/bg-12-781-2015 (2015).
19. Flecha, S. *et al.* Trends of pH decrease in the Mediterranean Sea through high frequency observational data: Indication of ocean acidification in the basin. *Sci. Rep.* **5**, 1–8. https://doi.org/10.1038/srep16770 (2015).
20. Flecha, S., Pérez, F. F., Murata, A., Makaoui, A. & Huertas, I. E. Decadal acidification in Atlantic and Mediterranean water masses exchanging at the Strait of Gibraltar. *Sci. Rep.* **9**, 1–11. https://doi.org/10.1038/s41598-019-52084-x (2019).
21. Kapsenberg, L., Alliouane, S., Gazeau, F., Mousseau, L. & Gattuso, J.-P. Coastal ocean acidification and increasing total alkalinity in the northwestern Mediterranean Sea. *Ocean Sci.* **13**, 411–426. https://doi.org/10.5194/os-13-411-2017 (2017).
22. Yao, K. M. *et al.* Time variability of the north-western Mediterranean Sea pH over 1995–2011. *Mar. Environ. Res.* **116**, 51–60. https://doi.org/10.1016/j.marenvres.2016.02.016 (2016).
23. EEA. State and pressures of the marine and coastal Mediterranean environment. *European Environment Agency* 1–44 (1999). https://www.eea.europa.eu/publications/medsea/download.
24. Crossland, C. J. *et al.* The coastal zone-a domain of global interactions. In *Coastal fluxes in the Anthropocene* 1–37 (Springer, 2005).
25. Borges, A. V. & Gypens, N. Carbonate chemistry in the coastal zone responds more strongly to eutrophication than to ocean acidification. *Limnol. Oceanogr.* **55**, 346–353. https://doi.org/10.4319/lo.2010.55.1.0346 (2010).
26. Carstensen, J. & Duarte, C. M. Drivers of pH variability in coastal ecosystems. *Environ. Sci. Technol.* **53**, 4020–4029. https://doi.org/10.1021/acs.est.8b03655 (2019).
27. Bates, N. R. *et al.* A time-series view of changing surface ocean chemistry due to ocean uptake of anthropogenic $CO_2$ and ocean acidification. *Oceanography* **27**, 126–141. https://doi.org/10.5670/oceanog.2014.16 (2014).

28. Hofmann, G. E. *et al.* High-frequency dynamics of ocean pH: A multi-ecosystem comparison. *PLoS ONE* **6**, 25. https://doi.org/10.1371/journal.pone.0028983 (2011).

29. Duarte, C. M. *et al.* Is ocean acidification an open-ocean syndrome? Understanding anthropogenic impacts on seawater pH. *Estuaries Coasts* **36**, 221–236. https://doi.org/10.1007/s12237-013-9594-3 (2013).

30. Mercado, J. M. & Gordillo, F. J. L. Inorganic carbon acquisition in algal communities: Are the laboratory data relevant to the natural ecosystems?. *Photosynth. Res.* **109**, 257. https://doi.org/10.1007/s11120-011-9646-0 (2011).

31. Krause-Jensen, D. *et al.* Macroalgae contribute to nested mosaics of pH variability in a subarctic fjord. *Biogeosciences* **12**, 4895–4911. https://doi.org/10.5194/bg-12-4895-2015 (2015).

32. Goffredo, S. & Dubinsky, Z. *The Mediterranean Sea: Its History and Present Challenges* (Springer, 2013).

33. Murphy, K. *et al.* World distribution, diversity and endemism of aquatic macrophytes. *Aquat. Bot.* **158**, 103127. https://doi.org/10.1016/j.aquabot.2019.06.006 (2019).

34. Hendriks, I. E. *et al.* Photosynthetic activity buffers ocean acidification in seagrass meadows. *Biogeosciences* **11**, 333–346. https://doi.org/10.5194/bg-11-333-2014 (2014).

35. Ricart, A. M. *et al.* Coast-wide evidence of low pH amelioration by seagrass ecosystems. *Glob. Change Biol.* **27**, 2580–2591. https://doi.org/10.1111/gcb.15594 (2021).

36. Newton, J., Feely, R., Jewett, E., Williamson, P. & Mathis, J. Global ocean acidification observing network: requirements and governance plan (2015). https://archimer.ifremer.fr/doc/00651/76343/.

37. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

38. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444. https://doi.org/10.1038/nature14539 (2015).

39. Hewamalage, H., Bergmeir, C. & Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int. J. Forecast.* **37**, 388–427 (2021) (**j.ijforecast.2020.06.008**).

40. Huang, Y., Yang, L. & Fu, Z. Reconstructing coupled time series in climate systems using three kinds of machine-learning methods. *Earth Syst. Dyn.* **11**, 835–853. https://doi.org/10.5194/esd-11-835-2020 (2020).

41. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).

42. Fourrier, M. *et al.* A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean sea CANYON-MED. *Front. Mar. Sci.* **7**, 620. https://doi.org/10.3389/fmars.2020.00620 (2020).

43. Friedrich, T. & Oschlies, A. Basin-scale $pCO_2$ maps estimated from ARGO gfloat data: A model study. *J. Geophys. Res. Oceans* **114**, 1–9. https://doi.org/10.1029/2009JC005322 (2009).

44. Bittig, H. C. *et al.* An alternative to static climatologies: Robust estimation of open ocean $CO_2$ variables and nutrient concentrations from T, S, and $O_2$ data using Bayesian neural networks. *Front. Mar. Sci.* **5**, 328. https://doi.org/10.3389/fmars.2018.00328 (2018).

45. Landschützer, P. *et al.* A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences* **10**, 7793–7815. https://doi.org/10.5194/bg-10-7793-2013 (2013).

46. Broullón, D. *et al.* A global monthly climatology of total alkalinity: A neural network approach. *Earth Syst. Sci. Data* **11**, 1109–1127. https://doi.org/10.5194/essd-11-1109-2019 (2019).

47. Broullón, D., Pérez, F. & Doval, M. D. Weekly reconstruction of pH and total alkalinity in an upwelling-dominated coastal ecosystem through neural networks ($A_{TpH-NN}$): The case of Ría de Vigo (NW Spain) between 1992 and 2019. *Biogeosci. Discuss.* **20**, 1–36. https://doi.org/10.5194/bg-2021-33 (2021).

48. Contractor, S. & Roughan, M. Efficacy of Feedforward and LSTM Neural Networks at predicting and gap filling coastal ocean timeseries: Oxygen, nutrients, and temperature. *Front. Mar. Sci.* **8**, 368. https://doi.org/10.3389/fmars.2021.637759 (2021).

49. Seidel, M. P., DeGrandpre, M. D. & Dickson, A. G. A sensor for in situ indicator-based measurements of seawater pH. *Mar. Chem.* **109**, 18–28 (2008) (**j.marchem.2007.11.013**).

50. Gregor, L., Lebehot, A. D., Kok, S. & Scheel Monteiro, P. M. A comparative assessment of the uncertainties of global surface ocean $CO_2$ estimates using a machine-learning ensemble (csir-ml6 version 2019a)-have we hit the wall?. *Geosci. Model Dev.* **12**, 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019 (2019).

51. Lefévre, N., Watson, A. J. & Watson, A. R. A comparison of multiple regression and neural network techniques for mapping in situ $pCO_2$ data. *Tellus B Chem. Phys. Meteorol.* **57**, 375–384. https://doi.org/10.3402/tellusb.v57i5.16565 (2005).

52. Li, X. *et al.* A neural network-based analysis of the seasonal variability of surface total alkalinity on the East China Sea Shelf. *Front. Mar. Sci.* **7**, 219. https://doi.org/10.3389/fmars.2020.00219 (2020).

53. Sauzède, R. *et al.* Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: A novel approach based on neural networks. *Front. Mar. Sci.* **128**, 20. https://doi.org/10.3389/fmars.2017.00128 (2017).

54. Velo, A. *et al.* Total alkalinity estimation using MLR and neural network techniques. *J. Mar. Syst.* **111**, 11–18. https://doi.org/10.1016/j.jmarsys.2012.09.002 (2013).

55. Graves, A. & Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **18**, 602–610. https://doi.org/10.1016/j.neunet.2005.06.042 (2005).

56. Lee, K. *et al.* Roles of marginal seas in absorbing and storing fossil fuel $CO_2$. *Energy Environ. Sci.* **4**, 1133–1146. https://doi.org/10.1039/C0EE00663G (2011).

57. Schneider, A., Tanhua, T., Körtzinger, A. & Wallace, D. W. High anthropogenic carbon content in the eastern Mediterranean. *J. Geophys. Res. Oceans* **115**, 20. https://doi.org/10.1029/2010JC006171 (2010).

58. Pérez, F. F., Olafsson, J., Ólafsdóttir, S. R., Fontela, M. & Takahashi, T. Contrasting drivers and trends of ocean acidification in the subarctic atlantic. *Sci. Rep.* **11**, 1–16. https://doi.org/10.1038/s41598-021-93324-3 (2021).

59. Cossarini, G., Lazzari, P. & Solidoro, C. Spatiotemporal variability of alkalinity in the Mediterranean Sea. *Biogeosciences* **12**, 1647–1658. https://doi.org/10.5194/bg-12-1647-2015 (2015).

60. Koopmans, D., Holtappels, M., Chennu, A., Weber, M. & De Beer, D. High net primary production of mediterranean seagrass (*Posidonia oceanica*) meadows determined with aquatic eddy covariance. *Front. Mar. Sci.* **7**, 118. https://doi.org/10.3389/fmars.2020.00118 (2020).

61. Barrón, C., Duarte, C. M., Frankignoulle, M. & Borges, A. V. Organic carbon metabolism and carbonate dynamics in a Mediterranean seagrass (*Posidonia oceanica*), meadow. *Estuaries Coasts* **29**, 417–426. https://doi.org/10.1007/BF02784990 (2006).

62. Champenois, W. & Borges, A. V. Seasonal and interannual variations of community metabolism rates of a *Posidonia oceanica* seagrass meadow. *Limnol. Oceanogr.* **57**, 347–361. https://doi.org/10.4319/lo.2012.57.1.0347 (2012).

63. Rivaro, P., Messa, R., Massolo, S. & Frache, R. Distributions of carbonate properties along the water column in the Mediterranean Sea: Spatial and temporal variations. *Mar. Chem.* **121**, 236–245. https://doi.org/10.1016/j.marchem.2010.05.003 (2010).

64. Hassoun, A. E. R. *et al.* Modeling of the total alkalinity and the total inorganic carbon in the Mediterranean Sea. *J. Water Resour. Ocean Sci.* https://doi.org/10.11648/j.wros.20150401.14 *(2015)*.

65. Gemayel, E. *et al.* Climatological variations of total alkalinity and total dissolved inorganic carbon in the Mediterranean Sea surface waters. *Earth Syst. Dyn.* **6**, 789–800. https://doi.org/10.5194/esd-6-789-2015 (2015).

66. Schneider, A., Wallace, D. W. & Körtzinger, A. Alkalinity of the Mediterranean Sea. *Geophys. Res. Lett.* https://doi.org/10.1029/2006GL028842 *(2007)*.

67. Millot, C. Circulation in the Western Mediterranean Sea. *J. Mar. Syst.* **20**, 423–442. https://doi.org/10.1016/S0924-7963(98)00078-5 (1999).

68. Gazeau, F. *et al.* Whole-system metabolism and $co_2$ fluxes in a Mediterranean Bay dominated by seagrass beds (Palma Bay, NW Mediterranean). *Biogeosciences* **2**, 43–60. https://doi.org/10.5194/bg-2-43-2005 (2005).
69. Marbà, N. *et al.* Effectiveness of protection of seagrass (*Posidonia oceanica*) populations in Cabrera National Park (Spain). *Environ. Conserv.* **29**, 509–518 (2002).
70. Pawlowicz, R. M_map: A mapping package for matlab, version 1.4 m. *Computer Software, UBC EOAS.* https://www.eoas.ubc.ca/rich/map.html (2020).
71. Tintoré, J. & Casas Pérez, B. *Buoy Bahía de Palma Physicochemical parameters of sea water data [Data set].* Balearic Islands Coastal Observing and Forecasting System, SOCIB. https://doi.org/10.25704/S6JB-CK61 (2022).
72. Benson, B. B. & Krause, D. The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere. *Limnol. Oceanogr.* **29**, 620–632. https://doi.org/10.4319/lo.1984.29.3.0620 (1984).
73. Clayton, T. D. & Byrne, R. H. Spectrophotometric seawater pH measurements: Total hydrogen ion concentration scale calibration of m-cresol purple and at-sea results. *Deep-Sea Res. Part* **I**(40), 2115–2129. https://doi.org/10.1016/0967-0637(93)90048-8 (1993).
74. Dickson, A. G., Sabine, C. L. & Christian, J. R. *Guide to best practices for ocean $CO_2$ measurements* (North Pacific Marine Science Organization, 2007). https://www.ncei.noaa.gov/access/ocean-carbon-data-system/oceans/Handbook_2007.html.
75. Sharp, J. D. *et al.* CO2SYSv3 for MATLAB. https://doi.org/10.5281/zenodo.3950563 (2020).
76. Mehrbach, C., Culberson, C., Hawley, J. & Pytkowicx, R. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. Oceanogr.* **18**, 897–907. https://doi.org/10.4319/lo.1973.18.6.0897 (1973).
77. Dickson, A. & Millero, F. J. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep Sea Res. Part A Oceanogr. Res. Pap.* **34**, 1733–1743. https://doi.org/10.1016/0198-0149(87)90021-5 (1987).
78. Dickson, A. G. Standard potential of the reaction: AgCl (s)+12H$_2$(g)= Ag (s)+ HCl (aq), and the standard acidity constant of the ion $HSO_4^-$ in synthetic sea water from 273.15 to 318.15 K *J. Chem. Thermodyn.* **22**, 113–127. https://doi.org/10.1016/0021-9614(90)90074-Z (1990).
79. Woosley, R. J. Evaluation of the temperature dependence of dissociation constants for the marine carbon system using ph and certified reference materials. *Mar. Chem.* **229**, 103914. https://doi.org/10.1016/j.marchem.2020.103914 (2021).
80. Weiss, R. & Price, B. Nitrous oxide solubility in water and seawater. *Mar. Chem.* **8**, 347–359 (1980).
81. Dlugokencky, E., *et al.* Atmospheric nitrous oxide dry air mole fractions from the NOAA gml carbon cycle cooperative global air sampling network, 1997–2020. *Tech. Rep., NOAA.* https://doi.org/10.15138/53g1-x417 (2021).
82. Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. Stl: A seasonal-trend decomposition. *J. Off. Stat* **6**, 3–73 (1990).
83. Giménez-Romero, A. Coastal pH variability reconstructed through neural networks the coastal balearic sea case study. *GitHub Repo* (2021). https://github.com/agimenezromero/Coastal-pH-variability-reconstructed-through-neural-networks-the-coastal-Balearic-Sea-case-study.

## Acknowledgements

## Author contributions

Conceptual idea: I.E.H., S.F., and J.T. Data collection: S.F., I.E.H. and E.A.F. Data processing and analyses: A.G.R., S.F., I.E.H., F.F.P. and M.A.M. S.F.: wrote the first version of the manuscript with contributions of I.E.H., A.G.R. and M.A.M. All authors discussed the results and contributed to the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-17253-5.

**Correspondence** and requests for materials should be addressed to S.F.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.