https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Modelling seawater pCO_2 and pH in the Canary Islands region based on satellite measurements and machine learning techniques.

Irene Sánchez-Mendoza, Melchor González-Dávila, David González-Santana, David Curbelo-Hernández, David Estupiñán-Santana, Aridane G. González and J.Magdalena Santana-Casiano.

Instituto de Oceanografía y Cambio Global, QUIMA, Universidad de Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

Correspondence to: Melchor González-Dávila (melchor.gonzalez@ulpgc.es).

Irene Sánchez-Mendoza. Irene.sanchez@ulpgc.es

Melchor González-Dávila*, melchor.gonzalez@ulpgc.es https://orcid.org/0000-0003-3230-8985

David González-Santana, david.gonzalez@fpct.ulpgc.es https://orcid.org/0000-0001-8726-7768

David Curbelo-Hernández, David.curbelo@ulpgc.es https://orcid.org/0000-0002-9826-7437

David Estupiñan, DavidEstupinanSantana@hotmail.com

Aridane G. González Aridane.gonzalez@ulpgc.es https://orcid.org/0000-0002-5637-8841

J. Magdalena Santana-Casiano1, magdalena.santana@ulpgc.es https://orcid.org/0000-0002-7930-7683

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Abstract

The improvement of remote sensing systems, together with the emergence of new model-fitting algorithms based on machine-learning techniques, has allowed the determination of the partial pressure of carbon dioxide (pCO$_{2,sw}$) and pH (pH$_{T,sw}$) in the waters of the Canary Islands. Among all the fitted models, the most powerful one was the bootstrap aggregation (*bagging*), giving a RMSE of 2.0 µatm (R^2 > 0.99) for pCO$_{2,sw}$ and RMSE of 0.002 for pH$_{T,is}$, although the multilinear regression (*MLR*), neural network (*NN*) and categorical boosting (*catBoost*) also have a good predictive performance, with RMSE ranging from 5.4 to 10 µatm for 360 < pCO$_{2,sw}$ < 481 µatm and from 0.004 and 0.008 for 7.97 < pH$_{T,is}$ < 8.07. Using the most reliable model, it was determined that there is an interannual trend of 3.51 ± 0.31 µatm yr^{-1} for pCO$_{2,sw}$ (which surpasses the rate of increase for atmospheric CO_2 of 2.3 µatm yr^{-1}) and an increase in acidity of -0.003 ± 0.001 pH units yr^{-1}. The increase in both, the atmospheric CO_2 and the sea surface temperature of 0.2ºC yr^{-1} observed in the 6-year period, influenced by the unprecedented 2023 marine heat wave, contribute to this important rate. Considering the Canary Islands between 13º-19ºW and 27º-30ºN, the region has moved from a slight CO_2 source of 0.90 Tg CO_2 yr^{-1} in 2019 to 4.5 Tg CO_2 yr^{-1} in 2024. After 2022, eastern locations that acted as an annual sink of CO_2 switched to acting as a source.

Key words: pCO$_2$, modelling, carbon dioxide, seawater, machine-learning, Canary basin, marine heat wave.

1. Introduction

Anthropogenic emissions of carbon dioxide (CO_2) derived from fossil fuels burning, cement production and changes in land use (Siegenthaler and Sarmiento, 1993; Doney et al., 2009; Le Quéré et al., 2009; Zeebe, 2012) since the First Industrial Revolution have led to a sharp increase in this trace gas in the atmosphere. The increase of atmospheric CO_2 is mitigated by terrestrial vegetation and oceanic absorption (Friedlingstein et al., 2025). The North Atlantic Ocean is reported to be one of the major oceanic sinks of the Northern Hemisphere, taking up 2.6 ± 0.4 Pg CO_2 yr^{-1} (i.e., 25% of the total anthropogenic CO_2 absorbed by all oceans) based on the analysis of 18-year dataset (Gruber et al., 2002).

In recent years, considerable effort has focused on quantifying oceanic CO_2 uptake and its implications (e.g., Bange et al., 2024; Gregor et al., 2024). One common approach involves using

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

regression models to estimate surface ocean pCO_2 from environmental variables. However, these models often fall short in capturing the complexity of dynamic regions such as coastal zones and continental shelves (Sun et al., 2021). These areas exhibit intense physical and biogeochemical activity, driven by high rates of primary production, carbon burial, organic matter recycling, and calcium carbonate deposition (Boehme et al., 1998; Gattuso et al., 1998; Borges et al., 2005). Despite their significance, these regions remain poorly represented in global carbon budgets and air-sea CO_2 flux estimates (Takahashi et al., 2002).

Pioneering studies by Borges et al. (2005) and Cai et al. (2006) provided the first global assessments of coastal CO_2 fluxes, emphasizing the spatial heterogeneity and functional diversity of coastal ecosystems in the global carbon cycle. More recent research confirms that these regions act as significant CO_2 sinks, with ingassing estimates ranging from 0.54 to 1.47 Pg CO_2 yr^{-1} (Laruelle et al., 2014; Cao et al., 2020), although newer assessments suggest lower values (Dai et al., 2022; Regnier et al., 2022; Resplandy et al., 2024; Roobaert et al., 2019).

Large-scale latitudinal patterns indicate that sea surface temperature (SST) is a primary driver of surface ocean pCO_2 ($pCO_{2,sw}$), often expressed as CO_2 fugacity ($fCO_{2,sw}$). On smaller spatial scales within latitudinal bands, other factors such as upwelling-driven CO_2 supply and biological uptake of dissolved inorganic carbon (C_T) must also be considered (e.g., Laruelle et al., 2014).

The $pCO_{2,sw}$ is regulated by four interconnected processes: thermodynamic forcing, biological activity, physical mixing, and air-sea CO_2 exchange (Fennel et al., 2008; Ikawa et al., 2013). Typically, one or two of these processes dominate in a given region of the ocean (Bai et al., 2015). The thermodynamic component is primarily influenced by the SST and salinity (SSS), which determine CO_2 solubility in seawater (Weiss, 1970) and affect the dissociation constants of carbonic acid (e.g., Lueker et al., 2000). Biological influences are often represented by surface chlorophyll-a concentrations (Chl *a*) and the diffuse attenuation coefficient of downwelling irradiance at 490 nm ($K_{d,490}$) (Bai et al., 2015; Chen et al., 2019; Lohrenz et al., 2018). Vertical mixing processes, particularly those enriching surface waters with CO_2 from deeper layers, are commonly described using mixed layer depth (MLD) (Chen et al., 2019). Additionally, the continual rise in atmospheric CO_2 ($pCO_{2,atm}$), which drives the air-sea CO_2 gradient, makes it essential to account for $pCO_{2,atm}$ in long-term assessments.

Satellite remote sensing offers valuable spatiotemporal coverage for estimating surface $pCO_{2,sw}$ (Chen et al., 2019). In the open ocean, where variability is low, satellite-based estimates yield RMSEs < 17 µatm. In contrast, coastal regions show much higher errors (> 90 µatm) due to complex physical and biogeochemical processes (Lohrenz et al., 2018; Sun et al., 2021).

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Basic models rely on empirical regressions such as multilinear, MLR, and non linear regression, MNR. Shadwick et al. (2010) applied MLR to the Scotian Shelf ($R^2 = 0.81$; SE = 13 µatm), while Signorini et al. (2013) achieved RMSEs of 22.4–36.9 µatm across the U.S. East Coast. Chen et al. (2016) developed a satellite-based model for the West Florida Shelf with RMSE < 12 µatm.

Machine learning approaches, including neural networks (NN), random forests, and CatBoost, show improved accuracy. Lefèvre and Taylor (2002) reported NN residuals of 3–11 µatm in the subpolar gyre. Telszewski et al. (2009) obtained an RMSE of 11.6 µatm in the North Atlantic. Sun et al. (2021) used CatBoost to achieve an RMSE of 8.25 µatm and $R^2 = 0.946$. Gregor et al. (2024) applied ML with target transformations globally (1982–2022), capturing 15% more CO_2 fluxes, FCO_2, variance than traditional methods.

In coastal areas, Jo et al. (2012) used NN with SST and Chl *a* in the South China Sea (RMSE = 6.9 µatm; r = 0.98). Duke et al. (2024) showed nearshore outgassing reduces net flux in the Northeast Pacific. Roobaert et al. (2024) highlighted seasonal variability driven by open-ocean and intracoastal exchanges. Wu et al. (2024) used ML products in the Gulf of Mexico, estimating a CO_2 uptake of 1.5 TgC yr^{-1}, though long-term trends remain uncertain.

This study focuses on the coastal region of the Canary Islands basin (27.0–30ºN; 13.0–19ºW) (Figure 1), located in the oligotrophic waters of the eastern subtropical North Atlantic gyre (Pelegrí et al., 1996). The area is influenced by the Canary Current (CC) and trade winds, which drive mesoscale features such as cyclonic and anticyclonic eddies. Despite low surface Chl *a* levels, upwelling filaments from the NW African coast, eddies, and dust fertilization can enhance primary production (Davenport et al., 1999). Marine heatwaves (MHWs) (Hobday et al., 2016; Frölicher and Laufkötter, 2018; Holbrook et al., 2019), increasingly linked to climate change, have recently intensified in this region. Varela et al. (2024) reported that 2023 was the warmest year in the Canary Upwelling System (CUS) during the 1982–2023 period, with most months showing record SSTs—likely affecting CO_2 dynamics.

Long-term observations reveal a consistent rise in surface pCO_2 in the region. Takahashi et al. (2009) estimated an increase of 1.8 ± 0.4 µatm yr^{-1} in the North Atlantic (1972–2006). Bates et al. (2014) found a rate of 1.92 ± 0.92 µatm yr^{-1} and a pH decline of –0.0018 ± 0.0002 yr^{-1} at ESTOC (1996–2012). More recently, González-Dávila and Santana-Casiano (2023) reported a $pCO_{2,sw}$ increase of 2.1 ± 0.1 µatm yr^{-1} and a $pH_{T,21}$ decline of –0.002 ± 0.0001 yr^{-1} in the upper 100 m (1995–2023), about 20% higher than rates for 1995–2010.

The aim of this work was to develop and validate an algorithm based on machine learning

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

techniques to compute the $pCO_{2,sw}$, pH_T and FCO_2 on the Canary Basin (NE Atlantic) using satellite-downloaded data of the main variables controlling these parameters and a high-resolution time series of $pCO_{2,sw}$ observations obtained from voluntary observing ships (VOS) and moored oceanographic buoys.

2. Material and methods

2.1. Data

2.1.1. ***In-situ* observations**

The observational dataset was built from data collected by Surface Ocean Observation Platforms (SOOPs) installed in Volunteer Observing Ships (VOS) and moored to oceanographic buoys (Figure 1 and Table S1). Two VOS carry out underway monitoring within their usual routes: (1) the CanOA-VOS-1 on board the *Jona Sophie* (formerly *Renate P.*), a cargo vessel owned by Reederei Stefan Patjens GmbH & Co. KG and operated in Spain by Nisa Marítima, which serves the easternmost part of the Canary Islands archipelago between the ports of Tenerife (S.C. de Tenerife, 28.4867ºN, 16.2284ºW, hereinafter TF), Gran Canaria (Las Palmas de Gran Canaria, 28.1319ºN, 15.4185ºW; GC) and Lanzarote (Arrecife, 28.9682ºN, 13.5294ºW; LZ) and passes northeast of LZ on its way to Barcelona (Spain). (2) The CanOA-VOS-2, using the vessel *Benchijigua Express*, owned by the company *Fred Olsen Express*, which serves the westernmost part of the Canary Archipelago between the ports of Tenerife (Los Cristianos, 28.0486ºN, 16.7163ºW; TF), La Gomera (San Sebastián de La Gomera, 28.0859ºN, 17.1090ºW; GOM) and La Palma (S.C. de La Palma, 28.6751ºN, 17.7666ºW; LP). The VOS line covered by the *Jona Sophie* is part of the Spanish contribution to the Integrated Carbon Observation System (ES-SOOP-CanOA, ICOS-ERIC; https://www.icos-cp.eu/) since 2021 and has been recognized as an ICOS Class 1 Ocean Station. Moreover, two moored oceanographic buoys provide valuable data at strategic coastal locations: (1) MORGAN-1 (Gando, Gran Canaria, 27.9296ºN, 15.3646ºW; González et al., 2024) and (2) ULA-2 (El Hierro, 27.6350ºN, 17.9964ºW).

Autonomous underway monitoring and data acquisition follows the recommendations described by Pierrot et al., (2009) to ensure comparable and high-quality data sets. Detailed description of equipment can be found in Curbelo-Hernández et al. (2021, 2022) and in the Supplementary Material. The number of observations used in this work are shown in Table S1.

Discrete samples for total alkalinity (A_T) and total inorganic carbon (C_T) were collected every three months, covering the different seasons and sites, and analyzed using a VINDTA 3C (Marianda^{TM}) following the procedure detailed by Mintrop et al., (2000). The VINDTA 3C was

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

calibrated by titration of Certified Reference Material (CRMs; provided by A. Dickson at Scripps Institution of Oceanography), with an accuracy of ±1.5 µmol kg^{-1} for A_T and ±1.0 µmol kg^{-1} for C_T. Differences between measured and discrete pCO$_2$(A_T, C_T) data (CO2sys.V2.1.xls, set of carbonic acid constants from Lueker et al., 2000, n=66) were 4 ± 4 µatm and 7 ± 5 µatm for the GO8050 and ProCV systems, respectively. To account for these differences, the observed data were corrected with the appropriate factors.

To compare the data, seven locations across the Canary Archipelago were considered (Figure 1). Site A is located along the LP-LG route at 17.5 ± 0.05ºW. Site B is along the LG-TF route at 16.95 ± 0.05 ºW. Site C lies at the intersection of different routes at 14.65 ± 0.05 ºW. Site D is near the African coast, along the route between LZ and the Iberian Peninsula, at 13.2 ± 0.05 ºW. Site E corresponds to the ULA-2 buoy near El Hierro. Site F corresponds to the MORGAN-1 (Gando) buoy. Site G marks the location of the ESTOC site.

2.1.2. Satellite data

Satellite data for SST, Chl *a*, K_{d490}, MLD were used to develop the $p\mathrm{CO}_{2,sw}$ and pH_T forecast models while wind speed was used for fluxes computation. These data were downloaded from the Copernicus Marine Environmental Monitoring Service (CMEMS) website (https://marine.copernicus.eu/access-data, last accessed 05/27/2025). They were processed to determine each variable at the time and location of the observations to be used in the validation and determination of predictive models (data were averaged daily). The complete daily dataset was used to model and estimate the surface marine carbonate system (MCS) variables in the Canary Islands.

2.2. Variable determination and computational methods

The raw data were processed using MATLAB® (version R2019b) and Python (2023, version 3.13.6). For the VOS data, the $x\mathrm{CO}_{2,sw}$ from GO8050 system was corrected using the four-standard calibration after filtering out all points near the ports that could bias the CO_2 measurements in seawater. To ensure data quality, several filters were applied, using a threshold of 2.5 L min^{-1} water flow and 50 mL min^{-1} for the LICOR© gas flow.

The partial pressure of CO_2 in seawater ($p\mathrm{CO}_{2,eq}$) was calculated (Dickson et al., 2007) from the corrected xCO$_2$ values in dry air. The $p\mathrm{CO}_{2,eq}$ data from both VOS lines were corrected to the intake temperature due to the difference between the termosalinograph/equilibrator temperature and the SST (Takahashi et al., 1993). All $\mathrm{pCO}_{2,sw}$ data for VOS and buoys were processed to determine the real partial pressure $f\mathrm{CO}_{2,sw}$ (Dickson et al., 2007). The discrete seawater samples analysed for A_T with the VINDTA 3C system were used to determine an A_T-SSS relationship for

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

the area (n = 66) that followed the relationship determined in the ESTOC time series (González Dávila et al., 2010). The normalized A_T to a constant salinity $NA_T = A_T/SSS*35$ was 2290 ± 3 µmol kg^{-1}, which is statistically significant at the 99% confidence level (p-value < 0.01; $r^2 = 0.96$). This relationship was then used to compute pH (A_T(SSS),$fCO_{2,sw}$) values in the Canary Region (González Dávila et al., 2010). The data were then averaged daily.

Daily mean atmospheric $xCO_{2,atm}$ were obtained from the atmospheric ship data for the area and compared with those from the Izaña Atmospheric Research Centre (AEMET, 2024), as the data could be overestimated due to ship operations. The $xCO_{2,atm}$ maxima at the end of winter were close in both databases (± 1.5 µatm), while the minima at the end of summer were on average 3 µatm higher in the Izaña atmospheric station than in the 10 m inlet at the ship. To have a longer series of atmospheric data, Izaña data were used in our study. The atmospheric $xCO_{2,atm}$ was then used to compute the corresponding $pCO_{2,atm}$ and $fCO_{2,atm}$ (Dickson et al., 2007).

The flux of CO_2, FCO_2, was determined using the *Eq. 1*:

$$FCO_2 = 0.24\, k\, S\, \Delta pCO_2 \qquad (1)$$

where 0.24 is a conversion factor used to express the flux in mmol m^{-2} d^{-1}, S is the solubility of CO_2 in seawater (Weiss, 1970), ΔpCO_2 is $pCO_{2,sw} - pCO_{2,atm}$ and k is the gas transfer rate determined using the Wanninkhof (2014) parameterisation (*Eq. 2)*

$$k_{Wan} = 0.251\, u^2\, \left({}^{Sc}/_{650}\right)^{-0.5} \qquad (2)$$

where u is the wind speed (m s^{-1}) and Sc is the Schmidt number.

Equation 1 was applied to the daily experimental and modelled data. Daily fluxes were averaged to provide monthly fluxes but expressed as daily average value for the month (in mmol m^{-2} d^{-1}).

Each of the physicochemical variables (y), including $pCO_{2,atm}$ and $fCO_{2,atm}$ were fitted to harmonic functions (Eq. 3, where t is the year fraction for each data). Eq. 4 allows the calculation of the interannual trend for the de-seasonal data, even if the number of years to obtain an accurate trend (5-6 years) is low. The use of seasonal detrended data reduces end-effects in relatively short-term data sets.

$$y = a + c * \sin(2\pi t) + d * \cos(2\pi t) + e * \sin(4\pi t) + \cos(4\pi t) \qquad (3)$$

$$y = a + b * (t - 2019) + c * \sin(2\pi t) + d * \cos(2\pi t) + e * \sin(4\pi t) + \cos(4\pi t)$$

(4)

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

2.3. Models fitting and statistical treatment

The R software was used for the statistical treatment (R Core Team, 2019). Machine learning methods were used to fit the different models. The original datasets were first divided into two subsets with a probability (p) of 0.8 and 0.2, called training and validation datasets, respectively. The first was used to tune the model, while the second was used to validate the results obtained.

The simplest fitted model consisted of a multiple linear regression (*MLR*), following the analytical expression of *Eq. 5.*

$$pCO_{2,sw} = p_0 + \hat{\alpha}\, pCO_{2,atm}(\mu atm) + \hat{\beta}\ \ SST(ºC) + \hat{\gamma}\, Chl(mg\ m^{-3}) + \ \hat{\delta}\ \ K_{d,490}(m^{-1}) + \hat{\varepsilon}\, MLD(m) + \ \vartheta \qquad (5)$$

where $\hat{\alpha},\ \hat{\beta}, \hat{\gamma}, \hat{\delta}\ and\ \hat{\varepsilon}$ are the estimated coefficients for each predictor and ϑ the residuals. A similar equation was considered for the $pH_{T,sw}$ dependence.

Three machine learning techniques were used, a neural network (*NN*, Wang, 2003), categorical boosting (*CatBoost,* Prokhorenkova et al., 2018; Dorogush et al., 2018; Qian et al., 2023) and bootstrap aggregation (*bagging,* Breiman, 1996), which attempt to reduce the variance of predictions.

In the validation of the models, the main statistical parameters were determined, including the coefficient of determination (R^2), the root mean square error (RMSE; *Eq. 6*), the mean absolute error coefficient (MAE; *Eq. 7*), and the square sum of errors expressed on a daily basis (SSE; *Eq. 8*).

$$RMSE\ = \sqrt{\frac{\sum_{i=1}^{N}\left(pCO_{2,i}-\widehat{pCO_{2,\iota}}\right)^2}{N}} \qquad (6)$$

$$MAE\ = \sum_{i=1}^{N}\left|pCO_{2,I} - \widehat{pCO_{2,\iota}}\right|/d \qquad (7)$$

$$SSE\ \ = \ \sum_{i=1}^{N}\left(pCO_{2,I} - \widehat{pCO_{2,\iota}}\right)^2/d \qquad (8)$$

where $pCO_{2,i}$ and $\widehat{pCO_{2,\iota}}$ are the observed and estimated values of the partial pressure of CO_2, N is the number of data and d is the number of days in the database.

The Akaike's information criterion corrected for a finite dataset (AICc) was determined following *Eq. 9*. It allows the evaluation of the trade-off between model goodness-of-fit and complexity (i.e., number of variables involved). A model is considered better if its AICc is the lowest of all the fitted models.

$$AIC_c = 2k - 2\ln(L) + \frac{2k^2+2k}{n-k-1} \quad (9)$$

where k is the number of parameters involved in the model, $ln(L)$ is the log-likelihood for the predicted model and n is the number of data.

To determine the estimated coefficients in each seasonal model and the different confidence intervals, the two assumptions required to achieve predictive ability were tested. The normality of the residuals was determined using the two-Welch Shapiro-Wilk test with a significance level (α) of 0.05 and quantile-quantile plots. The homogeneity of the residual variance (homoscedasticity) was demonstrated using a graphical method. If the assumption of normality of the residual was not met, the bootstrapping method was used to determine the confidence intervals (C.I.) of the linear relationships. To compare the models, the analysis of covariance (ANCOVA) and the analysis of variance (ANOVA) were used to detect significant differences at $\alpha = 0.05$.

3. Results

The observational data made it possible to generate a database from which to model the behaviour of $pCO_{2,sw}$ and pH_T in the Canary Basin. To characterise both the measured and the satellite-derived parameters used in this study, Table 1 summarises the mean values and their associated standard deviations for each season and observation system. The SST (Figure 2) presented its maximum value during the summer (July-September) and its minimum value during the winter (January-March). The maximum SST were found across the westernmost part of the archipelago (between La Palma and Tenerife), which was on average ~1ºC higher than that in the easternmost region (between Gran Canaria and Lanzarote). This seasonal and longitudinal pattern is also observed for $pCO_{2,sw}$ and pH_T (Table 1). The seasonal and annual means of SST calculated with both in situ and satellite data show an average difference of ~0.15ºC.

3.1. Variability of the SST data

Figure 2 shows the monthly means of both the observed and the satellite-derived SST at sites A-F. The SST shows a strong seasonal pattern at these sites, generally reaching its highest value in September (24.20 ± 0.76ºC in the westernmost part at A and B and 23.70 ± 0.68ºC in the easternmost part at C and D) and its lowest value in March (19.47 ± 0.24 and 18.97 ± 0.31ºC for the respective regions). An anomalous maximum of SST was observed in the summer of 2023, reaching more than 25ºC at sites A-C and more than 24ºC at site D. The data give a seasonal amplitude of 4.5 ± 0.5ºC and 4.2 ± 0.4ºC in the route covered by the CanOA-VOS-2 and the CanOA-VOS-1, respectively. Although no significant differences were found between sections belonging to the same region (comparison between A and B and between C and D), the mean

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

SST value at site D (20.59 ± 0.09ºC) is slightly lower than the mean SST at site C (21.00 ± 0.09ºC). The analysis of covariance between observational and satellite-derived SST data shows that there are no significant differences ($p < 0.05$) between the two types of data. The mean of the daily residuals was 0.16ºC with a standard error of 0.12ºC in the western part and 0.12ºC with a standard error of 0.10ºC in the eastern part.

The seasonal cycle was represented at the E site despite the scarcity and temporal gaps of the data collected by the ULA-2 buoy (Figure 2E). The seasonal amplitude (5.10 ± 0.18ºC) was calculated for 2021 (the year with the most data availability) with the highest SST obtained in September (24.70 ± 0.26ºC) and the lowest in March (19.60 ± 0.40ºC). A similar seasonal pattern was observed at site F based on data collected by MORGAN-1 (Figure 2F), with the highest SST obtained in September (23.71 ± 0.47ºC) and the lowest in March (19.46 ± 0.52ºC), giving a seasonal amplitude of 4.22 ± 0.51ºC.

The longitudinal variability of both CanOA-VOS and satellite-derived SST data can be observed in Figure 2 and in Figure S1. In the western region, the observed SST varies from 20.59 ± 0.09ºC in winter to 24.04 ± 0.13ºC in summer and presents an annual average of 22.45 ± 0.11ºC. The seasonal averages agreed (0.1-0.2ºC) with those calculated from the satellite-derived data, with the largest differences between both datasets in summer (0.26ºC). Although SST in the eastern region were lower throughout the year (annual mean 21.02 ± 0.27ºC), related to the influence of the Northwest African upwelling in the region, similar seasonal variations were found (from 19.19 ± 0.24ºC in winter to 22.82 ± 0.25ºC in summer). The differences between observational and satellite data were smaller than in the western region (0.05-0.2ºC). The west-to-east decrease of the SST seemed to remain constant along the longitudinally monitored span in the Canary archipelago, except for the slight increase related to the wake island effect that occurred near the southern coast of Tenerife (monitored by the CanOA-VOS-2 line; Figure S1).

3.2. Predictive models of $p\mathrm{CO}_{2,\mathrm{sw}}$

3.2.1. Multiple linear regression (*MLR*)

The first set of models uses the traditional multiple linear regression statistics and aims to provide a first, simple but useful approximation of the $p\mathrm{CO}_{2,\mathrm{sw}}$ prediction. Five prediction models containing a combination of the five variables considered, including $\mathrm{pCO}_{2,\mathrm{atm}}$, SST, Chl *a*, $\mathrm{K}_{\mathrm{d},490}$ and the MLD, were fitted according to the analytical expression written in Eq. 5. Considering the strong correlation observed between Chl *a* and $\mathrm{K}_{\mathrm{d},490}$ ($R^2 = 0.96$), and, therefore, the non-significance of $\mathrm{K}_{\mathrm{d},490}$, the model considering this variable is not used for the rest of the work, as it does not provide any additional information. The coefficients obtained for each of them are

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

presented in Table 2.

The Akaike's information criterion (AIC_c <2) and the statistical parameters (Table 3) suggest that the prediction model to be chosen is the one that combines the atmospheric CO_2, thermal, physical and biological components ($pCO_{2,atm}$+SST+MLD+Chl-*a*), although the two-variable model (SST and pCO_2) also offered a similar accuracy.

Figure S2 shows the measured vs. predicted variable for training and validation using the $pCO_{2,atm}$ + SST + MLD + Chl *a* model. Although many measured and predicted $pCO_{2,sw}$ showed small differences, there is a large scatter in the predictions, which is reflected in the calculated statistical parameters (Table 3). In the validation data set (Table S2), the statistical parameters obtained were like those of the training set (RMSE = 8.2 µatm, MAE = 7.0 µatm day^{-1}, SSE = 16.7 $µatm^2$ day^{-1}, and R^2 = 0.802).

3.2.2. Machine learning techniques

Table 3 shows a comparison of the different machine learning based approaches obtained using observational $pCO_{2,sw}$ data. All models were developed using the same dataset and input variables.

3.2.2.1 Neuronal network (NN)

The first machine learning method applied to obtain a predictive model of the $pCO_{2,sw}$ consists of a neural network (*NN*). The statistical parameters associated with the different fitted models are given in Table 3. It should be noted that no analytical expression is given, since the knowledge acquired by the training model is hidden in the synoptic weights of its neurons. Considering the computed statistics, there is a closeness between the fits obtained for the three-variable model (SST + MLD + Chl-*a*) and that including the $pCO_{2,atm}$ data while the two-variable models also work closely. The plots of measured vs. predicted variable for both the training and validation datasets, considering the best model, are shown in Fig. S2. Despite the large scatter of the data, the fitness seems to improve at the upper end. The accuracy indicators of the training *vs.* validation datasets were 7.1 vs. 8.4 µatm, 5.0 vs. 5.9 µatm d^{-1}, 16.2 vs. 17.9 $µatm^2$ d^{-1}, and 0.891 vs. 0.862, in terms of RMSE, MAE, SSE, and R^2, respectively.

3.2.2.2 Categorical boosting (*CatBoost*) regression

The second type of machine learning technique used to predict the $pCO_{2,sw}$ in the waters of the Canary archipelago is the *CatBoost*. A total of 500 iterations were used to generate the prediction model. The results associated with the fit of the different models, including the statistics used to determine the accuracy, are presented in Table 3. The $pCO_{2,atm}$ + SST + Chl *a* + MLD model

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

was the most accurate due to the lower RMSE, MAE and SSE and the higher R^2 compared to the other models. The performance of the $pCO_{2,atm}$ + SST + Chl*a* + MLD model (Figure S2), used for the training and validation datasets, showed an R^2 above 0.95 with an RSME of only 3.6 µatm. The training dataset produced the most accurate results, with an MAE of 2.4 µatm day^{-1} and an SSE of 3.0 $µatm^2$ day^{-1}. The results of the validation statistics were consistent with those obtained during the training phase (Table S2).

3.2.2.3 Bootstrap aggregating (*bagging*) regression

A bagging algorithm was applied to predict the $pCO_{2,sw}$ using 200 bootstrap replicates. The computed statistics for the training set, combining the different parameters controlling the $pCO_{2,sw}$ are summarised in Table 3.

From the analysis of the computed statistics, it can be concluded that the model with the best predictive capacity is the one that considers three or four parameters, since it provides lower RMSE, MAE and SSE. As in the previously adjusted models, the models that includes SST + MLD or SST + $pCO_{2,atm}$ also provide a good fit (Table 3). The bagging algorithm appears to be the best predictive treatment with the highest R^2 and the lowest RMSE, MAE and SSE for any combination of variables, even when only SST variable is considered. The plot of measured vs. predicted $pCO_{2,sw}$ obtained for both the training and validation sets using a four-variable model is shown in Figure S2. This model has a low RMSE, MAE, and SSE (2.0 µatm, 1.6 µatm d^{-1}, and 0.8 $µatm^2$ d^{-1}, respectively). In this scenario, the application of the model to the validation set showed greater data dispersion than the training set (Table S2) due to the lower number of data (Figure S2).

3.3. Predictive models for pH_T

pH_T predictions were made based on the computed $pH_T(A_T(SSS), fCO_2)$ using observations and satellite data (interpolated to the time and coordinates of the observations) as input variables. In this case, $pCO_{2,atm}$ was not considered in the predictive model variables to avoid redundancy. Table 4 shows a comparison of the models employed in the machine learning based approaches. It is important to note that all models were developed using the same dataset and input variable.

3.3.1. Multiple linear regression (*MLR*)

The coefficients obtained for each of the four combination models are shown in Table 2, while the statistical performance is shown in Table 4. As was shown for the $pCO_{2,sw}$ fitting, the model including SST + Chl*a* + MLD was the best performing for pH_T, with R^2 of 0.745 and an RMSE of 0.006. The plot of measured vs. predicted pH_T for model training (Figure S3) shows a similar distribution to that for the validation set of data, where the number of data used for the fitting in

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

the validation set was not a limiting factor.

3.3.2. Machine learning techniques

The three techniques provide better correlation coefficients than those using *MLR* (Table 4). The performance of the *NN* was lower than that of *catboost*, while *bagging* showed the best performance for all models. The model considering the three variables (SST + Chl-a + MLD) was the most accurate in predicting pH_T in all cases (Table 4) with an R^2 as high as 0.99 and an RMSE as low as 0.002 for the *bagging* technique. Every combination of satellite data, even when considering only the SST, gave an R^2 greater than 0.95 when Bagging was used. For the *catboost* technique, the three variable model was required to achieve an R^2 greater than 0.93. We compared the accuracy indicators for the training and validation datasets for the three variable models (Table 4 and S3, Figure S3) for the pH_T range of this study, from 7.97 to 8.07. Applying machine learning techniques, and in particular for *bagging*, which always provides the best fit of data, the number of data in the treatment is a factor that improves determination. The RMSE, MAE and SSE indicators for the training and validation data sets are always below 0.01 in pH, with values as low as 0.002 and 0.003, respectively when *bagging* is used.

3.4. Validation of the results

The best prediction models for each class, considering the different statistical parameters calculated, were used to reconstruct the monthly means of $pCO_{2,sw}$ and pH_T at sites A-D and compared. The temporal variation of both observed and predicted values is shown in Figure 3. All models described the seasonal cycle: $pCO_{2,sw}$ reached its maximum and minimum in March and August-September, respectively, while an opposite behavior was observed for pH_T. The predictions showed slight significant differences ($p > 0.05$) when compared with the observations. The linear, neural network and *catBoost* models did not differ significantly from each other ($p <$ 0.05). When comparing the *bagging* model with the observational data, no significant differences were found, confirming that the model that best describes the real data is the *boostrap* aggregation model. The agreement between experimental and predicted concentrations in all sections was better than 1.7 ± 1.8 µatm for $pCO_{2,sw}$ and 0.002 ± 0.001 for pH_T.

Statistical differences ($p > 0.05$) were obtained when the western and eastern sections were compared by ANCOVA. At sites A and B (Figure 3), $pCO_{2,sw}$ (and pH) varied seasonally between 404 ± 18 µatm (8.045 ± 0.012) and 449 ± 19 µatm (8.004 ± 0.010). Seasonal amplitudes were of 47 ± 8 µatm (0.049 ± 0.005). At C and D (Figure 4) the seasonal ranges were between 390 ± 15 µatm (8.069 ± 0.008) and 440 ± 16 µatm (8.028 ± 0.012), with a seasonal amplitude of 52 ± 7 µatm (0.038 ± 0.006).

4. Discussion

Three oceanographic variables, SST, Chl-*a* and MLD, with high satellite resolution for oceanic surface seawater and the atmospheric CO_2 partial pressure were used to model $pCO_{2,sw}$ and pH_T in the Canary archipelago. Salinity was not included in the fitted models because of its minimal role in pCO_2 changes (Sarmiento et al., 2007; Shadwick et al., 2010), and the available satellite data have been found to be very different from the observed data (Yu, 2020). Despite the inclusion of $K_{d,490}$ in the initial models, it is suggested that the lack of statistical significance is because it is correlated with Chl *a* ($R^2 = 0.96$), making it redundant and therefore not significant. The $pCO_{2,atm}$ was not used in the pH_T calculation because it was already considered in the estimation of $pCO_{2,sw}$.

4.1 The Canary region in the period 2019-2024. Observational and modelling data.

In the Canary Islands, the highest temperatures (Figure 2) were recorded in late summer (September), driven by enhanced stratification of the water column and the solar radiation. The lowest temperatures were measured in winter (February-March) due to the convective mixing caused by the surface cooling of the water column. This seasonal behaviour is consistent with the hydrographic conditions described at the ESTOC site, with a seasonal temperature amplitude in the surface waters between 4 and 6ºC, with a maximum and minimum of 18 and 24ºC, respectively, recorded before 2023 (González-Dávila et al., 2010; Santana-Casiano et al., 2007; González-Dávila and Santana-Casiano, 2023). This range is also comparable to the SST observed in the easternmost region covered by the CanOA VOS-1 (Curbelo-Hernández et al., 2021) during 2019-2020.

The statistically significant differences ($p < 0.05$) observed between the different sections distributed to the west and east are related to the distance and proximity to the African continent, with the easternmost part of the archipelago being the most exposed to the upwelling filaments (Davenport, 1999), while the westernmost part is protected by the presence of the islands themselves. This trend is clearly observed in Figure 2 and S1, which show a progressive decrease in SST in the vicinity of the African continent. The validation of the satellite data showed no significant differences ($p < 0.05$), even in areas close to the islands, so that the satellite values could be used for the fitting of the different models and the subsequent determination of the derived parameters.

MORGAN-1 data (site F) shows anomalously high SST in the summer of 2023, consistent with the observation of extreme SST conditions in the CUS in 2023 (Varela et al., 2024). Satellite data in the coastal location of the buoys also showed anomalously high values in summer, but these were on average 0.3ºC lower than those measured by the buoy sensors. In-situ temperatures from June to October 2023 were more than 1ºC higher than those recorded in the previous years. These high temperatures in the summer 2023 were not recorded in 2024, showing that the year 2023 should be considered an anomaly in this area.

It is noteworthy that the SST in February-March 2024 remained high. Winter SST increased in 2024 and was on average 1ºC warmer than in the previous years (average for 2020-2022 was 19.09± 0.20ºC *vs* average for 2023-2024 of 20.01 ± 0.25ºC). These anomalies strongly influence the trends observed in both satellite and observational data.

Harmonic fitting of the temperature (Eq. 4) for the period March 2020 to March 2023, despite the limitation of only three years of data, shows that the warming in seasonal detrended Gando Bay data was 0.03ºC yr^{-1} (González et al., 2024). This is comparable to the warming rates found at the ESTOC site for the October 1995 to March 2023 period (González-Dávila and Santana-Casiano, 2023) and for the 1982–2023 period in the full CUS (Varela et al., 2024).

When considering the full five-year seasonal detrended dataset from Gando Bay (March 2020 to October 2024) the rate of temperature increases shifts to 0.19 ± 0.06ºC yr^{-1} (0.14 ± 0.06ºC yr^{-1} from monthly mean satellite data). This increase in SST was also observed at sites A-D (Fig. 2), where the rate of warming for the six-year period from February 2019 to October 2024 ranges from 0.29 ± 0.03ºC yr^{-1} (sites A-C) to 0.21ºC yr^{-1} (site D). The mean temperature at the western station (ULA-2) was ~1ºC higher (22.12 ± 0.16ºC) than at the eastern station F (MORGAN- 1) in the east of the region (21.13 ± 0.12ºC) related to the influence of the Northwest African upwelling in the area and the coastal upwelling. The ANCOVA performed in both buoys showed no significant differences between the *in-situ* and the satellite-derived SST, with differences of less than 0.19ºC on average.

The $pCO_{2,sw}$ and pH_T were predicted using satellite-derived data. The model with the highest prediction error in this work was the *MLR* (RMSE of 4.9 and R^2 of 0.904 in pCO_2). The neural network model presented similar results (RMSE of 7.1 µatm and R^2 of 0.896). Previous work by Signorini et al., (2013) applied *MLR* to the US coasts and presented RMSE ranging from 22.4 to 36.9 µatm, while other works by Ford et al. (2022) and Friedrich and Oschlies (2009) used *NN* to predict $pCO_{2,sw}$ in the North and South Atlantic Ocean, respectively, and obtained RMSE > 19 µatm and RMSE = 21.68 µatm, respectively. Both models applied in the present study, the simplest *MLR* (with low computational time and an expression that can be used directly in other cases) and the *NN*, behave adequately compared to those used in the published literature. The small area considered in our region and the large amount of observational data contribute strongly to the observed RMSE. In the pH_T estimation, RMSE as low as 0.006 and 0.008 were found for

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

MLR and *NN*, respectively, which are within the experimental error in pH determination.

The *catBoost* empirical algorithm can estimate the $pCO_{2,sw}$ and pH_T in the Canary Islands archipelago with uncertainties of <4 µatm and 0.004 pH units, and $R^2 > 0.93$ for both variables, showing that the model is tolerant to uncertainties in satellite variables dominated by different processes and coastal proximity, proving its applicability in the region. However, the *bagging* approach exhibited exceptional performance for both $pCO_{2,sw}$ and pH_T estimation with uncertainties of 2.0 µatm and 0.002 pH units for the region and the period 2019-2024.

It is suggested that these considerably favourable results and the comparable errors with ocean-scale models are because the variables controlling $pCO_{2,sw}$ and pH_T in the waters of the Canary Islands are well characterised by the thermal component (Takahashi et al., 2002: González-Dávila and Santana-Casiano, 2023). In all cases, the simple model with only SST showed high correlation coefficients ($0.65 < R^2 < 0.94$) and the computed statistics show that, although they are not the best fitted models, there is a good fit when using this single variable. The coefficient estimated in the annual linear regression (10.40 µatm ºC^{-1}, Table 2) showed a certain deviation from the theoretical rate of change for the area in the period 2019-2024 (16 µatm ºC^{-1}), related to the biological and physical effects (i.e., primary production, remineralisation, and water mass mixing) during spring and summer, but in line with values observed in ESTOC (Santana Casiano et al., 2007)

In all four sites, but also in Gando bay, and according to both the observational data and the predicted model treatments, the $pCO_{2,sw}$ increased between 2019 and 2024 at a rate of 3.8 ± 0.6 µatm yr^{-1}, considering only this 6-year period. On the other hand, the pH_T decreases at a rate of -0.004 ± 0.001 for the same period. Previous results at the ESTOC time series from 1995 to 2023 (González-Dávila and Santana-Casiano, 2023) and for the Gando Bay (site F) from 2020 to 2023 (González et al., 2024) showed an increase in $pCO_{2,sw}$ of 2.1 ± 0.1 µatm yr^{-1} and a decrease in pH_T of -0.002 ± 0.001 yr^{-1}. Similar rates for $pCO_{2,sw}$ and pH_T are observed in all selected sites when the period March 2019-March 2023 is selected, without considering the whole year 2023 (the same period considered in González et al., 2024).

4.2 Monthly $pCO_{2,sw}$ and $pH_{T,is}$ gridded maps.

The *Bagging* technique was used to construct gridded monthly maps of $pCO_{2,sw}$ and pH_T (at in situ SST) for the Canary region (13º-19ºW, 27º-30ºN) for the study period and presented in Figure 4 for the year 2023. Monthly experimental averages are shown alongside the predictions to show the accuracy of the estimates. The expected seasonal pattern was observed, with higher $pCO_{2,sw}$ in September and lower in March, with the opposite behaviour for $pH_{T,is}$. A clear

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

longitudinal gradient was observed, with higher $pCO_{2,sw}$ and lower $pH_{T,sw}$ toward the east, mainly due to the thermal effect. The cooler seawater in the east, together with the influence of rich-nutrient, lower pH Northeast African upwelled seawater (Pelegrí et al., 2005), counteract each other, increasing the observed values but decreasing the seasonal amplitude.

Several oceanographic features become apparent. Upwelling filaments, characterised by lower temperature, locally reduce $pCO_{2,sw}$. The leeward island wake zones show warmer water, which increases $pCO_{2,sw}$ and decreases pH_T. The African coastal upwelling signal is especially clear during June and September, with lower $pCO_{2,sw}$ and higher pH_T due to enhanced biological activity that offsets the CO_2-rich upwelled waters.

The monthly mean $pCO_{2,sw}$ and $pH_{T,is}$ in the Canary Basin predicted with *bagging* from 2019 to 2024 is shown in Figure 5 for the whole Canary region. To compute the monthly means, the daily satellite SST, Chl-*a*, and MLD and the values for $pCO_{2,atm}$ were used by the *Bagging* model to calculate $pCO_{2,sw}$ and $pH_{T,is}$ for the region, and then averaged for the area and for each month. During these six years, the mean $pCO_{2,sw}$ was 419.7 ± 16 µatm, with a seasonal amplitude of 55 µatm. The harmonic fit (eq. 4) of the predicted data shows an interannual trend of 3.51 ± 0.31 µatm yr^{-1} for 2019-2024, which is higher than that registered for $pCO_{2,atm}$ (2.3 µatm yr^{-1}).

The pH_T (Figure 5) ranged from 8.015 ± 0.049 in February–March to 7.980 ± 0.058 in September–October, reflecting a 0.04 decrease from winter to summer. High winter values were the result of lower temperatures and convective mixing in the water column. Low summer values were attributed to biological activity and stratification (Santana-Casiano et al., 2001; 2007). The pH decrease was mitigated by the thermal effect, which compensated for 33% of the decrease (the thermal amplitude should be 0.06 units due to the temperature increase of 4.1ºC). This process is evident near the African coast (Figure 8), where the injection of deep, cold, CO_2-rich seawater into the surface waters of the African coastal upwelling decreases the SST and pH, creating a longitudinal gradient in the Canary region.

Figure 5 shows that $pH_{T,is}$ levels in the region declined throughout the study period due to increased ocean acidity, reaching -0.003 ± 0.001 pH units yr^{-1}, determined for the seasonal-detrended data. The strong influence of the MHW effects, described above in summer 2023 and in winter 2023 and 2024 on the observed interannual rate of increase in the two variables is noticeable. The increase in $pCO_{2,atm}$ is also accompanied by an increase in SST of 0.2ºC yr^{-1} over the six-year period. This equates to a cumulative temperature increase from 2019 to 2024 of 1.2ºC. This was a consequence of the anomalous year of 2023, as well as the higher SST in winter 2020 compared to 2019, and the higher SST in winters 2023 and 2024 compared to 2022, when winter temperatures have dropped below 18°C and are now at 19°C. This has resulted in an increase in previous $pCO_{2,sw}$ trends and ocean acidification levels in ESTOC, estimated at 2.1 µatm yr^{-1} and -0.002 pH units yr^{-1} for the period from 1995 to the beginning of 2023 (González-Dávila & Santana-Casiano, 2023). Using only six years of data could also contribute to these rates of change. However, winters with SST over 19ºC and summers with SST over 25ºC had never been recorded at the ESTOC site before 2023.

4.3 Long term model prediction at ESTOC site.

The utility of the bagging predictive model after considering data for the period 2019-2024 was applied to the ESTOC site for the period 2004 to 2024. Previous years were not used as monthly satellite data has a lower resolution. Satellite data for SST, Chl *a*, MLD and atmospheric pCO_2 computed from available xCO_2 data at Izaña (IZO) station were used (https://gml.noaa.gov/aftp/data/trace_gases/co2/flask/surface/txt/co2_izo_surface-flask_1_ccgg_event.txt, last access 26/05/2025). Estimated values at 29º10'N and 15º 30'W were compared with measured data (González-Dávila and Santana Casiano, 2023) updated to 2024 and plotted in Figure 6. The model was able to describe the ESTOC data with average residuals of 1.3 ± 3.1 µatm and with trends for the study period of 1.9 ± 0.1 µatm yr^{-1} determined by both model and seasonal detrended data. It is important to notice that when any of the models without $pCO_{2,atm}$ was considered, the residuals increased to over 2 µatm, and especially during the earliest period, 2004 to 2010, when the residuals were close to 4 µatm. When $pCO_{2,atm}$ is not accounted for in the models for the period 2019-2024, with the presence of the MHW in the year 2023 in the CUS area, the models give the highest weight of the observed trends to SST changes alone. Indeed, analysis of satellite-derived sea surface temperature data at the ESTOC station from 2004 to 2024 reveals that SST exhibited minimal variation during 2004–2019 (0.0012 ± 0.002°C yr^{-1}). In contrast, a significant warming trend was evident over the 2019–2024 period, with SST increasing at a rate of 0.21 ± 0.01°C yr^{-1}, as it was observed in the other selected sites A-F in Figure 1. Consequently, when the model with SST + Chl *a* + MLD was applied to earlier periods, it predicted lower trends. When $pCO_{2,atm}$ was considered in the model fitting, the role played by SST and by $pCO_{2,atm}$ are included in the $pCO_{2,sw}$. Therefore, if the SST rate is low, the model considers the concurrent rise in atmospheric pCO_2, which also influences the surface seawater pCO_2.

4.4 Air-sea CO_2 exchange in the Canary archipelago

The predicted $pCO_{2,sw}$ is highly useful for determining FCO_2 with improved spatial and temporal resolution. Figure 7 shows the FCO_2 calculated using the parametrization given by Wanninkhof (2014) for monthly mean conditions during 2019-2024. The seasonal cycle of FCO_2

is controlled by large seasonal variations in $pCO_{2,sw}$, which modify ΔpCO_2 (since $pCO_{2,atm}$ exhibits shorter seasonal amplitudes), while the effect of wind speed and solubility is lower on a seasonal basis (Landschützer et al., 2014). The region acts as a strong CO_2 sink during winter and spring, whereas during the warm season, the system acts as a source. During the warm period from late May to early September (González-Dávila et al., 2003), when the dominant trade winds impact the Canary Islands, $pCO_{2,sw}$ surpasses $pCO_{2,atm}$. This results in increased wind speeds and reinforces the role of CO_2 supersaturation in global flux estimation, favouring the region's role of as a CO_2 source.

Sites closer to the African continent (C and D) and the coastal waters (F, in the Gando Bay, also in the eastern part of the Canary Islands) are more likely to act as a CO_2 sink (Curbelo et al., 2021) than the westernmost region. This is mainly due to the thermal gradient, with temperatures over one degree lower than to the west, and higher productivity waters. However, Figure 7B shows that, due to the increase in SST across the Canary Islands during the study period, all locations that acted as an annual sink of CO_2 switched to acting as a source after 2022. For the period 2019 to 2024, the Canary region (CR) acted as a slight source of CO_2, at a rate of 0.39 ± 0.17 mol m^{-2} yr^{-1}. Increasing trends in fluxes were observed across all regions, ranging from 0.18 to 0.37 mmol m^{-2} d^{-1} yr^{-1} with an average rate of 0.25 ± 0.02 mmol m^{-2} d^{-1} yr^{-1}. When the Canary region (13-19ºW, 27-30ºN) is considered, with an area of 185,000 km^2 (after removing the island territories), it moves from being a slight source of 0.9 Tg CO_2 in 2019 to being a source of 4.5 Tg CO_2 in 2024, with a maximum of 4.8 Tg CO_2 in 2023. This was also the year of maximum temperature in the area (Figure 2), favouring the highest increase in $pCO_{2,sw}$.

5. Conclusion

This study presents the first predictive model of $pCO_{2,sw}$ and $pH_{T,is}$ for the Canary Islands basin. It demonstrates the usefulness of satellite data in complementing observation platforms such as voluntary observing ships and moored buoys. By combining satellite data from the Copernicus Marine Environmental Monitoring Service with in-situ observations, it was possible to model the behaviour of $pCO_{2,sw}$ and $pH_{T,is}$ in the waters surrounding the Canary Islands archipelago and therefore quantify the air-sea CO_2 flux.

Four types of models, ranging from classical multivariate statistics to more sophisticated machine-learning models were fitted to consider the atmospheric pCO_2, SST, Chl *a*, and MLD variables that control the $pCO_{2,sw}$ and $pH_{T,is}$ in seawater. The multiple linear regression, neural network, and categorical boosting models produced the highest RMSE values. The estimates obtained by these three models did not differ significantly, and the computed statistics (e.g.,

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

RMSE, MAE, and R^2) were comparable to those of models adjusted for oceanic waters. The *bagging* model demonstrated the best fit, with an RMSE < 2.5 µatm (< 0.7%), 0.002 in pH, an R^2 > 0.99, and no significant differences compared to the monthly mean observations.

The application of the *bagging* technique enabled the characterization of the seasonal and longitudinal variability of surface seawater $pCO_{2,sw}$ and total scale pH across the entire marine region of the Canary Islands. The ~1°C longitudinal gradient in SST, driven by the influence of African coastal upwelling and the offshore transport of upwelling filaments, resulted in persistently higher $pCO_{2,sw}$ and lower $pH_{T,is}$ values in the westernmost region (between El Hierro and Tenerife) compared to the easternmost region (between Tenerife and Lanzarote) throughout the year. In terms of air-sea CO_2 exchange, the westernmost area acted as a source throughout the study period, while the easternmost area acted as a weak sink, changing to source behaviour after 2022. The interannual increase in SST in the Canary region during the study period, including the anomalous year of 2023 and the warmer winter waters in 2020, 2023, and 2024, is considered the main factor responsible for the increase in outgassing. The Canary region acted as a source of 0.39 ± 0.17 mol m^{-2} yr^{-1} for the period from 2019 to 2024, with an increasing source trend that emitted 0.9 Tg CO_2 in 2019 to 4.5 Tg CO_2 in 2024, peaking at 4.8 Tg CO_2 in 2023.

The results presented in this study highlight the complexity of modelling the $pCO_{2,sw}$ and pH_T in coastal environments, where physical and biological conditions are more variable than in open ocean waters. The anomalous behaviour of 2023 was confirmed together with the important influence of a prolonged MHW event lasting more than a year within a relatively short trend study. The importance of long-term data series for predicting interannual changes was also highlighted. Despite the satisfactory model results, much longer work is required to constrain $pCO_{2,sw}$ and $pH_{T,is}$ in the Canary Islands waters, especially with regards to their interannual trends, but the combination of long term data set, satellite imagery and machine learning techniques is shown to provide and excellent description for the ocean-atmosphere CO_2 exchange. This requires longer time series to reduce the effects of MHW events and the warmers summers and winters observed in recent years.

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Code and Data Availability Statement

The underway observations provided by the SOOP CanOA-VOS in the Canary Region including the buoys data (February 2019 – December 2024) used in this investigation, are published in open-access at Zenodo (https://doi.org/10.5281/zenodo.16780085) and available since September 2023 at the ICOS Data Portal (https://www.icos-cp.eu/data-products/ocean-release) for the CanOA-VOA-1. The model codes used in this work to apply the different machine learning techniques are published in open-access at Zenodo (https://doi.org/10.5281/zenodo.16780313). All satellite data is available at Copernicus Climate Data Store (https://cds.climate.copernicus.eu/). The IZAÑA $pCO_{2,atm}$ data is available at https://gml.noaa.gov/aftp/data/trace_gases/co2/flask/surface/txt/co2_izo_surface-flask_1_ccgg_event.txt

Supplement link:

Author contribution

All the authors made significant contributions to this research. M. G.-D., J. M. S.-C. and A.G.G. installed and maintained the equipment in the VOS and led the work. Together with D. C-H and D. G.-S. proceeded with data acquisition and processing. I.S.-M. and D.E. developed routines and applied machine learning techniques to data processing. All authors contributed to the writing of the manuscript and supported its submission.

Declaration Competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgement

We thank the JONA SOPHIE ship owner, Reederei Stefan Patjens GmbH & Co. KG, the NISA-Marítima company and the captains and crew members for the support during this collaboration. We also thank the FRED OLSEN EXPRESS shipping company, the captains and crew members for helping in all the operations. Special thanks to the technician Adrian Castro-Álamo, for biweekly equipment maintenance and discrete sampling of total alkalinity aboard the ship. The SOOP CanOA-VOS line is part of the Spanish contribution to the Integrated Carbon Observation System (ICOS-ERIC; https://www.icos-cp.eu/) since 2021 and has been recognised as an ICOS Class 1 Ocean Station. The participation of D. C-H was funded by the PhD grant

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

PIFULPGC-2020-2 ARTHUM-2.

Financial support

This research was supported by the Canary Islands Government and the Loro Parque Foundation through the CanBIO project, CanOA subproject (2019–2024), and the CARBOCAN agreement (Consejería de Transición Ecológica y Energía, Gobierno de Canarias).

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

REFERENCES

AEMET. Centro de Investigación Atmosférica de Izaña. Medidas de CO2 https://gml.noaa.gov/aftp/data/trace_gases/co2/flask/surface/txt/co2_izo_surface-flask_1_ccgg_event.txt (accessed 5.26.2025), 2024

Bai, Y., Cai, W.-J., He, X., Zhai, W., Pan, D., Dai, M., and Yu, P.: A mechanistic semi-analytical method for remotely sensing sea surface pCO2 in river-dominated coastal oceans: A case study from the East China Sea, J. Geophys. Res. Ocean., 120, 2331–2349, https://doi.org/https://doi.org/10.1002/2014JC010632, 2015.

Bange, H. W., Mongwe, P., Shutler, J. D., Arévalo-Martínez, D. L., Bianchi, D., Lauvset, S. K., Liu, C., Löscher, C. R., Martins, H., Rosentreter, J. A., Schmale, O., Steinhoff, T., Upstill-Goddard, R. C., Wanninkhof, R., Wilson, S. T., and Xie, H.: Advances in understanding of air–sea exchange and cycling of greenhouse gases in the upper ocean, Elem. Sci. Anth., 12(1), doi:10.1525/elementa.2023.00044, 2024.

Bates, N. R., Astor, Y. M., Church, M. J., Currie, K., Dore, J. E., González-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santana-Casiano, J. M.: A Time-Series View of Changing Surface Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO_2 and Ocean Acidification, 27(1), 126–141, https://doi.org/10.2307/24862128, 2014.

Boehme, S.E., Sabine, C.L., and Reimers, C.E.: CO2 fluxes from a coastal transect: A time-series approach, Mar. Chem. 63, 49–67, https://doi.org/10.1016/S0304-4203(98)00050-4, 1998.

Borges, A. V., Delille, B., and Frankignoulle, M.: Budgeting sinks and sources of CO2 in the coastal ocean: Diversity of ecosystem counts, Geophys. Res. Lett., 32, 1–4. https://doi.org/10.1029/2005GL023053, 2005.

Breiman, L.: Bagging predictors, Mach. Learn., 24, 123–140, https://doi.org/10.1007/BF00058655, 1996.

Cai, W.J., Dai, M., and Wang, Y.: Air-sea exchange of carbon dioxide in ocean margins: A province-based synthesis, Geophys. Res. Lett., 33, 2–5, https://doi.org/10.1029/2006GL026219, 2006

Cao, Z., Yang, W., Zhao, Y., Guo, X., Yin, Z., Du, C., Zhao, H., and Dai, M.: Diagnosis of CO2 dynamics and fluxes in global coastal oceans, Natl. Sci. Rev., 7, 786–797, https://doi.org/10.1093/nsr/nwz105, 2020.

Chen, S., Hu, C., Barnes, B.B., Wanninkhof, R., Cai, W.J., Barbero, L., and Pierrot, D.: A machine learning approach to estimate surface ocean pCO2 from satellite measurements, Remote Sens. Environ., 228, 203–226, https://doi.org/10.1016/j.rse.2019.04.019, 2019.

Chen, S., Hu, C., Byrne, R.H., Robbins, L.L., and Yang, B. : Remote estimation of surface pCO2 on the West Florida Shelf, Cont. Shelf Res., 128, 10–25. https://doi.org/https://doi.org/10.1016/j.csr.2016.09.004, 2016.

Curbelo-Hernández, D., González-Dávila, M., González, A.G., González-Santana, D., and Santana- Casiano, J.M.: CO2 fluxes in the Northeast Atlantic Ocean based on measurements from a surface ocean observation platform, Sci. Total Environ., 775, 145804. https://doi.org/10.1016/j.scitotenv.2021.145804, 2021.

Curbelo Hernández, D., Santana Casiano, J. M., González González, A., González Santana,

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

D., and González Dávila, M.: Air-Sea CO_2 Exchange in the Strait of Gibraltar, Front. Mar. Sc., 8,745304, https://doi.org/ 10.3389/fmars.2021.745304, 2022.

Dai, M., Su, J., Zhao, Y., Hofmann, E. E., Cao, Z., Cai, W.-J., Gan, J., Lacroix, F., Laurelle, G. G., Meng, F., Müller, J. D., Regnier, P.A. G., Wang, G., and Wang, Z.: Carbon fluxes in the coastal ocean: Synthesis, boundary processes and future trends. Annu. Rev. Earth Planet. Sci., 50(1), 593–626. https://doi.org/10.1146/annurev-earth-032320-090746, 2022.

Davenport, R., Neuer, S., Hernandez-Guerra, A., Rueda, M.J., Llinas, O., Fischer, G., and Wefer, G.: Seasonal and interannual pigment concentration in the Canary Islands region from CZCS data and comparison with observations from the ESTOC, Int. J. Remote Sens. 20, 1419–1433, https://doi.org/10.1080/014311699212803, 1999.

Dickson, A.G., Sabine, C.L. and Christian, J.R. (Edss): Guide to best practices for ocean CO2 measurement. Sidney, British Columbia, North Pacific Marine Science Organization, 191pp, PICES Special Publication 3; IOCCP Report 8, https://doi.org/10.25607/OBP-1342, 2007.

Doney, S.C., Fabry, V.J., Feely, R.A., and Kleypas, J.A.: Ocean acidification: The other CO2 problem, Ann. Rev. Mar. Sci.,1,169–192, https://doi.org/10.1146/annurev.marine.010908.163834, 2009.

Dorogush, A.V., Ershov, V., and Gulin, A.: CatBoost: gradient boosting with categorical features support, ArXiv, 1–7, http://arxiv.org/abs/1810.11363, 2018.

Duke, P. J., Hamme, R. C., Ianson, D., Landschützer, P., Swart, N. C., and Covert, P. A.: High-resolution neural network demonstrates strong CO2 source-sink juxtaposition in the coastal zone, J. Geophys. Res.: Oceans, 129, e2024JC021134, https://doi.org/10.1029/2024JC021134, 2024.

Fennel, K., Wilkin, J., Previdi, M., and Najjar, R.: Denitrification effects on air-sea CO2 flux in the coastal ocean: Simulations for the northwest North Atlantic, Geophys. Res. Lett., 35, https://doi.org/https://doi.org/10.1029/2008GL036147, 2008.

Ford, D. J., Tilstone, G. H., Shutler, J. D., and Kitidis, V.: Derivation of seawater pCO2 from net community production identifies the South Atlantic Ocean as a CO2 source, Biogeosciences, 19(1), 93–115, https://doi.org/10.5194/bg-19-93-2022, 2022.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Arneth, A., Arora, V., Bates, N. R., Becker, M., Bellouin, N., Berghoff, C. F., Bittig, H. C., Bopp, L., Cadule, P., Campbell, K., Chamberlain, M. A., Chandra, N., Chevallier, F., Chini, L. P., Colligan, T., Decayeux, J., Djeutchouang, L., Dou, X., Duran Rojas, C., Enyo, K., Evans, W., Fay, A., Feely, R. A., Ford, D. J., Foster, A., Gasser, T., Gehlen, M., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Kato, E., Keeling, R. F., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Lauvset, S. K., Lefèvre, N., Liu, Z., Liu, J., Ma, L., Maksyutov, S., Marland, G., Mayot, N., McGuire, P., Metzl, N., Monacci, N. M., Morgan, E. J., Nakaoka, S.-I., Neill, C., Niwa, Y., Nützel, T., Olivier, L., Ono, T., Palmer, P. I., Pierrot, D., Qin, Z., Resplandy, L., Roobaert, A., Rosan, T. M., Rödenbeck, C., Schwinger, J., Smallman, T. L., Smith, S., Sospedra-Alfonso, R., Steinhoff, T., Sun, Q., Sutton, A. J., Séférian, R., Takao, S., Tatebe, H., Tian, H., Tilbrook, B., Torres, O., Tourigny, E., Tsujino, H., Tubiello, F., van der Werf, G., Wanninkhof, R., Wang, X., Yang, D., Yang, X., Yu, Z., Yuan, W., Yue, X., Zaehle, S., Zeng, N., and Zeng, J.: Global Carbon Budget 2024, Earth Syst. Sci. Data, 17, 965-

1039, https://doi.org/10.5194/essd-17-965-2025, 2025.

Friedrich, T., and Oschlies, A.: Neural network-based estimates of North Atlantic surface pCO2 from satellite data: A methodological study, J. Geophys. Res.: Oceans, 114(3), https://doi.org/10.1029/2007JC004646, 2009.

Frölicher, T.L., and Laufkötter, C.: Emerging risks from marine heat waves. Nat. Commun., 9 (1), 650, https://doi.org/10.1038/s41467-018-03163-6, 2018.

Gattuso, J.P., Frankignoulle, M., and Wollast, R.: Carbon and carbonate metabolism in coastal aquatic ecosystems, Annu. Rev. Ecol. Syst., 29, 405–434, https://doi.org/10.1146/annurev.ecolsys.29.1.405, 1998.

González, A. G., Aldrich-Rodríguez, A., González-Santana, D., González-Dávila, M., and Santana-Casiano, J. M.: Seasonal variability of coastal pH and CO2 using an oceanographic buoy in the Canary Islands. Frontiers in Marine Science, 11, https://doi.org/10.3389/fmars.2024.1337929, 2024.

González-Dávila, M., Santana-Casiano, J. M.: Long-term trends of pH and inorganic carbon in the Eastern North Atlantic: the ESTOC site, Front. Mar. Sci., 10, https://doi.org/10.3389/fmars.2023.1236214, 2023.

González-Dávila, M., Santana-Casiano, J.M., Rueda, M.J., and Llinás, O.: The water column distribution of carbonate system variables at the ESTOC site from 1995 to 2004, Biogeosciences, 7, 3067–3081, https://doi.org/10.5194/bg-7-3067-2010, 2010.

González-Dávila, M., Santana-Casiano, J.M., Rueda, M.J., Llinás, O., and González-Dávila, E.F.: Seasonal and interannual variability of sea-surface carbon dioxide species at the European Station for time series in the Ocean at the Canary Islands (ESTOC) between 1996 and 2000, Glob. Biogeochem. Cycles, 17, https://doi.org/10.1029/2002gb001993, 2003.

Gregor, L., Shutler, J., and Gruber, N.: High-Resolution Variability of the Ocean Carbon Sink, Glob. Biogeochem. Cycles, 38(8), doi:10.1029/2024GB008127, 2024.

Gruber, N., Keeling, C.D., and Bates, N.R.: Interannual variability in the North Atlantic ocean carbon sink, Science, 298, 2374–2378, https://doi.org/10.1126/science.1077077, 2002.

Hobday, A.J., Alexander, L.V., Perkins, S.E., Smale, D.A., Straub, S.C., Oliver, E.C., and Wernberg, T.: A hierarchical approach to defining marine heatwaves. Prog. Oceanogr., 141, 227-238. https://doi.org/10.1016/j.pocean.2015.12.014, 2016.

Holbrook, N.J., Scannell, H.A., Sen Gupta, A., Benthuysen, J.A., Feng, M., Oliver, E.C., and Wernberg, T.: A global assessment of marine heatwaves and their drivers, Nat. Commun., 10 (1), 2624, https://doi.org/10.1038/s41467-019-10206-z, 2019.

Ikawa, H., Faloona, I., Kochendorfer, J., Paw U, K.T., and Oechel, W.C.: Air-sea exchange of CO2 at a Northern California coastal site along the California Current upwelling system, Biogeosciences 10, 4419–4432, https://doi.org/10.5194/bg-10-4419-2013, 2013.

Jo, Y.-H., Dai, M., Zhai, W., Yan, X.-H., and Shang, S.: On the variations of sea surface pCO2 in the northern South China Sea: A remote sensing based neural network approach, J. Geophys. Res., 117, C08022, https://doi.org/10.1029/2011JC007745, 2012-

Laruelle, G.G., Lauerwald, R., Pfeil, B., and Regnier, P. : Regionalized global budget of the CO2 exchange at the air-water interface in continental shelf seas, Glob. Biogeochem. Cycles, 1199–1214, https://doi.org/10.1111/1462-2920.13280, 2014.

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, Glob. Biogeochemical Cycles, 28(9), 927–949. https://doi.org/10.1002/2014GB004853, 2014.

Lefevre, N., and Taylor, A.: Estimating pCO2 from sea surface temperatures in the Atlantic gyres, Deep–Sea Res. Pt. I, 49(3), 539–554, https://doi.org/10.1016/s0967-0637(01)00064-4, 2002.

Le Quéré, C., Raupach, M.R., Canadell, J.G., Marland, G., Bopp, L., Ciais, P., Conway, T.J., Doney, S.C., Feely, R.A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R.A., House, J.I., Huntingford, C., Levy, P.E., Lomas, M.R., Majkut, J., Metzl, N., Ometto, J.P., Peters, G.P., Prentice, I.C., Randerson, J.T., Running, S.W., Sarmiento, J.L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., Van Der Werf, G.R., and Woodward, F.I.: Trends in the sources and sinks of carbon dioxide, Nat. Geosci., 2, 831–836. https://doi.org/10.1038/ngeo689, 2009.

Lohrenz, S.E., Cai, W.-J., Chakraborty, S., Huang, W.-J., Guo, X., He, R., Xue, Z., Fennel, K., Howden, S., and Tian, H.: Satellite estimation of coastal pCO2 and air-sea flux of carbon dioxide in the northern Gulf of Mexico, Remote Sens. Environ., 207, 71–83. https://doi.org/https://doi.org/10.1016/j.rse.2017.12.039, 2018.

Lueker, T.J., Dickson, A.G., and Keeling, C.D.: Ocean pCO2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: Validation based on laboratory measurements of CO2 in gas and seawater at equilibrium, Mar. Chem., 70, 105–119, https://doi.org/10.1016/S0304-4203(00)00022-0, 2000.

Mintrop, L., Pérez, F.F., González-Dávila, M., Santana-Casiano, J.M., and Körtzinger, A.: Alkalinity determination by potentiometry: intercalibration using three different methods, Ciencias Mar., 26(1), 23–27. https://doi.org/10.7773/cm.v26i1.573, 2000.

Pelegrí, J.L., Arístegui, J., Cana, L., González-Dávila, M., Hernández-Guerra, A., Hernández-León, S., Marrero-Díaz, A., Montero, M.F., Sangrá, P., and Santana-Casiano, M.: Coupling between the open ocean and the coastal upwelling region off northwest Africa: water recirculation and offshore pumping of organic matter, J. Mar. Sys., 54 (1–4) 3-37, https://doi.org/10.1016/j.jmarsys.2004.07.003, 2005.

Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Lüger, H., Johannessen, T., Olsen, A., Feely, R.A., and Cosca, C.E.: Recommendations for autonomous underway pCO2 measuring systems and data-reduction routines, Deep Sea Res., II, 56 (8–10), 512-522, https://doi.org/10.1016/j.dsr2.2008.12.005, 2009.

Qian, L., Chen, Z., Huang, Y., and Stanford, R. J.: Employing categorical boosting (CatBoost) and meta-heuristic algorithms for predicting the urban gas consumption, Urban Climate, 51, https://doi.org/10.1016/j.uclim.2023.101647, 2023.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: CatBoost: unbiased boosting with categorical features. 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada, https://github.com/catboost/catboost, 2018.

Regnier P., Resplandy L., Najjar R. G., and Ciais P.: The land-to-ocean loops of the global carbon cycle, Nature, 603, 401–410, https://doi.org/10.1038/s41586-021-04339-9, 2022.

R Core Team: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/, 2019.

Resplandy, L., Hogikyan, A., Müller, J. D., Najjar, R. G., Bange, H. W., Bianchi, D., Weber, T., Cia. W.-J., Doney, S.C., Fennel, K., Gehlen, M., Hauck, J., Lacroix, F., Landschützer, P., Quéré, C. Le, Roobaert, A., Schwinger, J., Berthet, S., Bopp, L., Chau, T. T. T., Dai, M., Gruber, N., Ilyina, T., Kock, A., Manizza, M., Lachkar, Z., Laurelle, G. G., Liao, E., Lima, E. D., Nissen, C., Rödenbeck, C., Séférian, R., Toyama, K., Tsujino, H., and Regnier, P.: A synthesis of global coastal ocean greenhouse gas fluxes. Glob. Biogeochem. Cycles, 38(1), e2023GB007803, https://doi.org/10.1029/2023GB007803, 2024.

Roobaert, A., Laruelle, G. G., Landschützer, P., Gruber, N., Chou, L., and Regnier, P.: The spatiotemporal dynamics of the sources and sinks of CO_2 in the global coastal ocean, Glob. Biogeochem. Cycles, 33(12), 1693–1714, https://doi.org/10.1029/2019GB006239, 2019.

Roobaert, A., Resplandy, L., Laruelle, G.G., Liao, E., and Regnier, P.: Unraveling the physical and biological controls of the global coastal CO2 sink., Global Biogeochem. Cycles, 38 (3), e2023GB007799, https://doi.org/10.1029/2023GB007799, 2024.

Santana-Casiano, J.M., González.Dávila, M., Laglera-Baquer, L.M., and Rodríguez-Somoza, M.J.: Carbon dioxide system in the Canary region during October 1995, Sci. Mar., 65, 41 – 49. https://doi.org/10.3989/scimar.2001.65s141, 2021.

Santana-Casiano, J.M., González-Dávila, M., Rueda, M.J., Llinás, O., and González-Dávila, E.F.: The interannual variability of oceanic CO2 parameters in the northeast Atlantic subtropical gyre at the ESTOC site, Glob. Biogeochem. Cycles, 21, https://doi.org/10.1029/2006GB002788, 2007.

Sarmiento, J., Gruber, N., and McElroy, M.: Ocean Biogeochemical Dynamics, Phys. Today, 60, 65, https://doi.org/10.1063/1.2754608, 2007.

Shadwick, E.H., Thomas, H., Comeau, A., Craig, S.E., Hunt, C.W., and Salisbury, J.E.: Air-sea CO2 fluxes on the Scotian Shelf: Seasonal to multi-annual variability, Biogeosciences 7, 3851–3867, https://doi.org/10.5194/bg-7-3851-2010, 2010.

Siegenthaler, U., and Sarmiento, J.L.: Atmospheric carbon dioxide and the ocean, Nature 399, 119–125, https://doi.org/10.1038/340301a0, 1993.

Signorini, S.R., Mannino, A., Najjar Jr., R.G., Friedrichs, M.A.M., Cai, W.-J., Salisbury, J., Wang, Z.A., Thomas, H., and Shadwick, E.: Surface ocean pCO2 seasonality and sea-air CO2 flux estimates for the North American east coast, J. Geophys. Res. Ocean, 118, 5439–5460. https://doi.org/https://doi.org/10.1002/jgrc.20369, 2013.

Sun, H., He, Y., Chen., Y., and Zhao, B.: Space-Time Sea Surface pCO2 estimation in the North Atlantic based on CatBoost, Remote Sens., 13(14), 2805, https://doi.org/10.3390/rs13142805, 2021.

Takahashi, T., Olafsson, J., Goddard, J.G., Chipman, D.W., and Sutherland, S.C.: Seasonal variation of CO2 and nutrients in the high-latitude surface oceans: A comparative study, Glob. Biogeochem. Cycles, 7, 843–878, https://doi.org/https://doi.org/10.1029/93GB02263, 1993.

Takahashi, T., Sutherland, S.C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wannikhof, R., Feely, R.A., Sabine, C., Olafsson, J., and Nojiri, Y.: Global air-sea flux of CO2 based on surface ocean pCO2, and seasonal biological and temperature effects, Deep. Res. II, 49, 1601–1622., https://doi.org/10.1016/S0967-0645(02)00003-6, 2002.

Takahashi, T., Sutherland, S.C., Wanninkhof, R., Sweeney, C., Feely, R.A., Chipman, D.W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D.C.E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T.S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C.S., Delille, B., Bates, N.R., and de Baar, H.J.W.: Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans, Deep. Res. II, 56, 554–577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

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

Varela, R., de Castro, M., Costoya, X., Días, J.M., and Gómez-Gesteira, M.: Influence of the cnary upwelling system on SST during the unprecedented 2023 North Atlantic marine heatwave, Sci. Total Environ., 949, 175043, https://doi.org/10.1016/j.scitotenv.2024.175043, 2024.

Wang, S.-C.: Artificial Neural Network BT - Interdisciplinary Computing in Java Programming, edited by: Wang, S.-C., Springer US, Boston, MA, 81–100, https://doi.org/10.1007/978-1-4615-0377-4_5, 2003.

Wanninkhof, R.: Relationship between wind speed and gas exchange over the ocean revisited revisited, Limnol. Oceanogr. Methods, 12, 351–362, https://doi.org/10.4319/lom.2014.12.351, 2014.

Weiss, R.F.: The solubility of nitrogen, oxygen and argon in water and seawater, Deep Sea Res. Oceanogr. Abstr., 17, 721–735, https://doi.org/10.1016/0011-7471(70)90037-9, 1970.

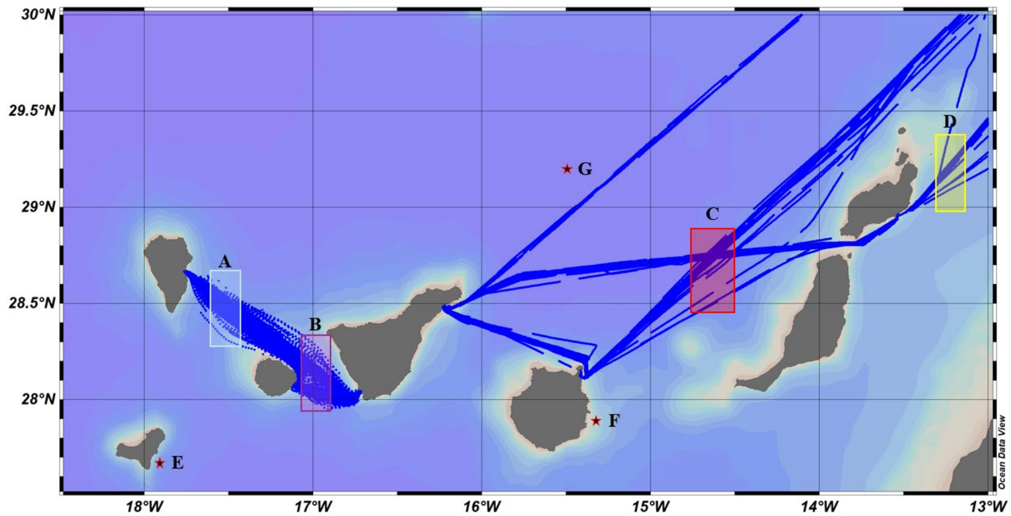
Wu, Z., Vermeulen, A., Sawa, Y., Karstens, U., Peters, W., de Kok, R., Lan, X., Nagai, Y., Ogi, A., and Tarasova, O.: Investigating the differences in calculating global mean surface CO2 abundance: the impact of analysis methodologies and site selection, Atmos. Chem. Phys., 24, 1249–1264, https://doi.org/10.5194/acp-24-1249-2024, 2004.

Yu, L.: Variability and Uncertainty of Satellite Sea Surface Salinity in the Subpolar North Atlantic (2010–2019), Remote Sens., 12, 2092, https://doi.org/10.3390/rs12132092, 2020.

Zeebe, R.E.: History of seawater carbonate chemistry, atmospheric CO2, and ocean acidification, Annu. Rev. Earth Planet. Sci., 40, 141–165, https://doi.org/10.1146/annurev- earth-042711-105521, 2012.

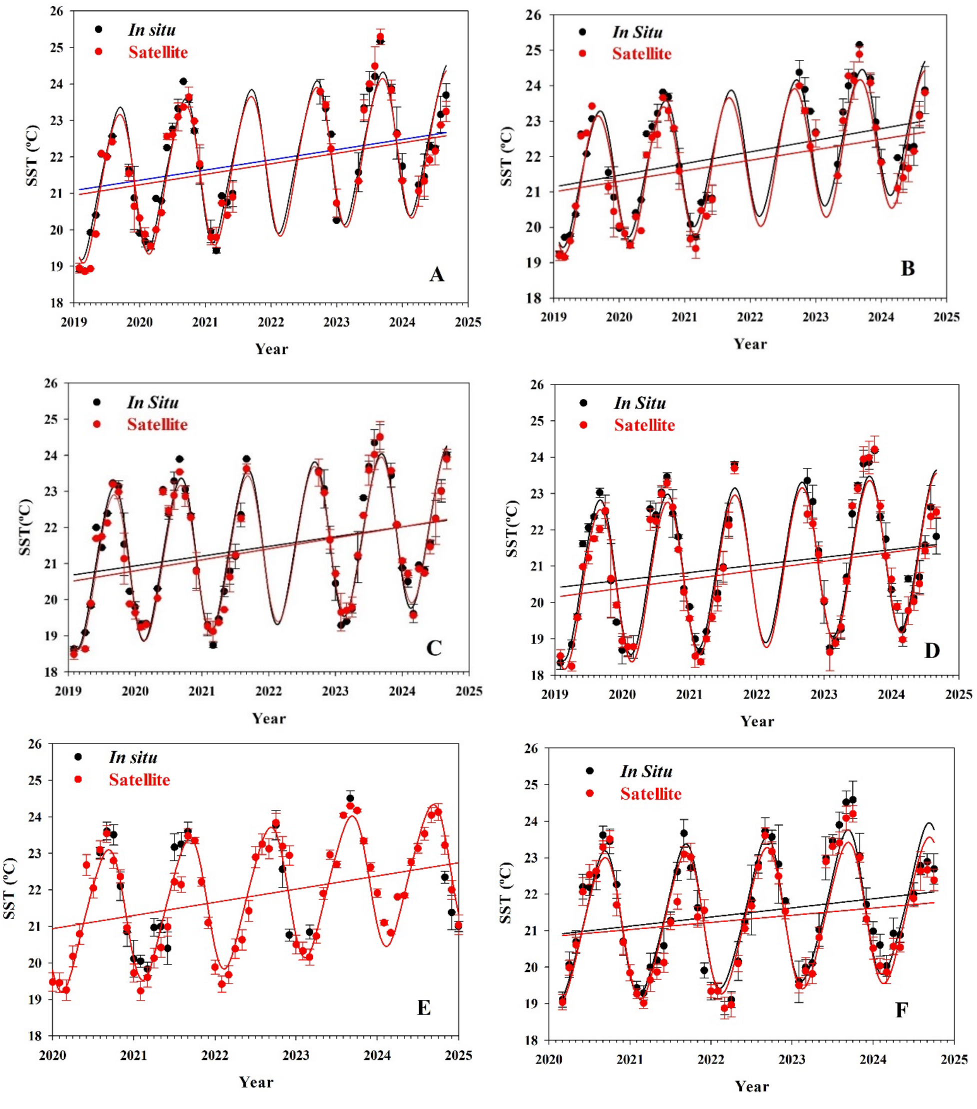
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 1. Map of the region of study in the Canary archipelago with the CanOA-VOS's tracks and the location of A-F sites. The location of the G site (ESTOC site) is also shown.



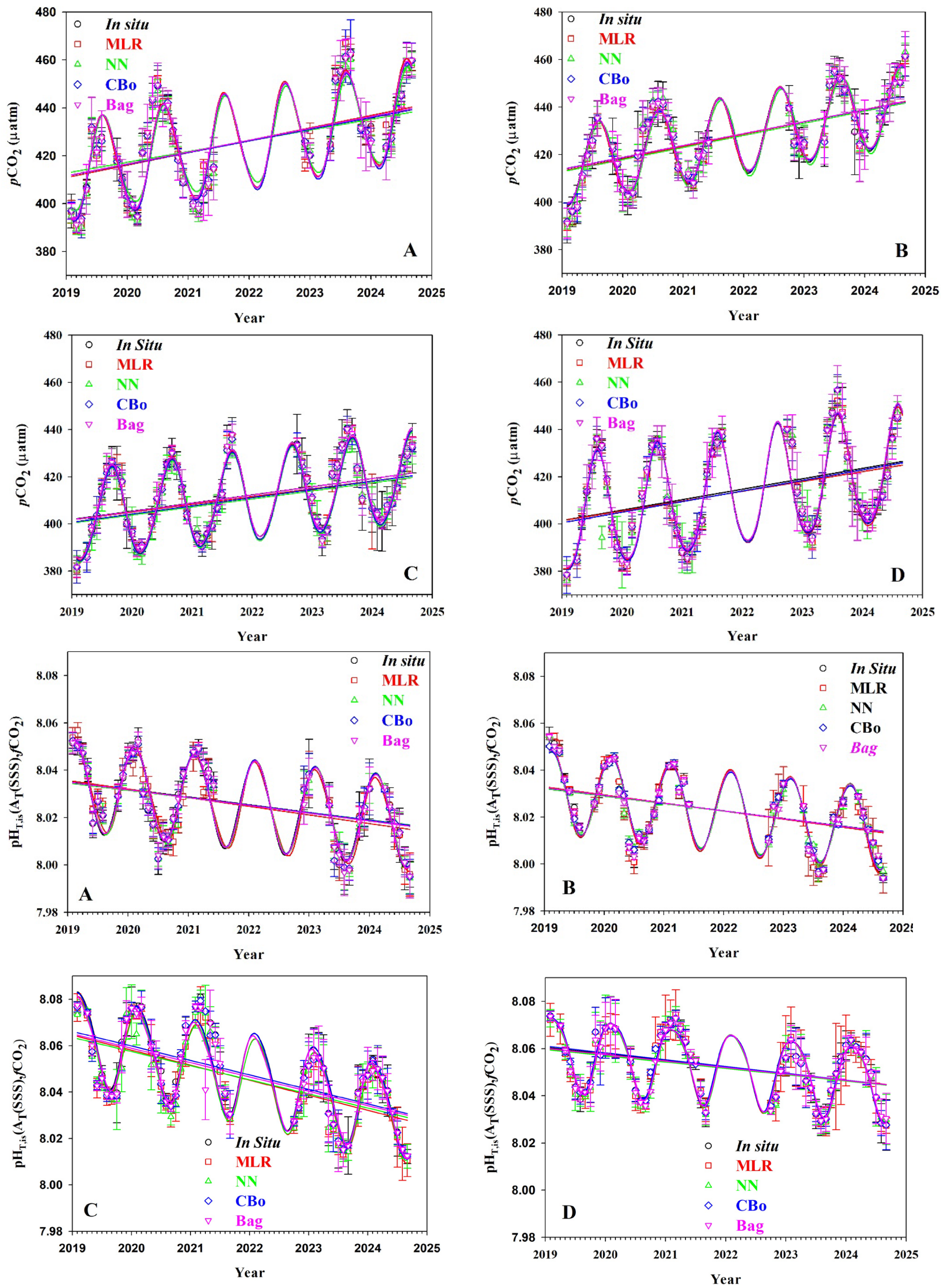
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 2. Monthly mean of observational SST (black) and satellite-based SST (red) at locations A-F. Harmonic fittings (Eq 4) of the data are shown together with the linear fitting for the seasonal detrended data.



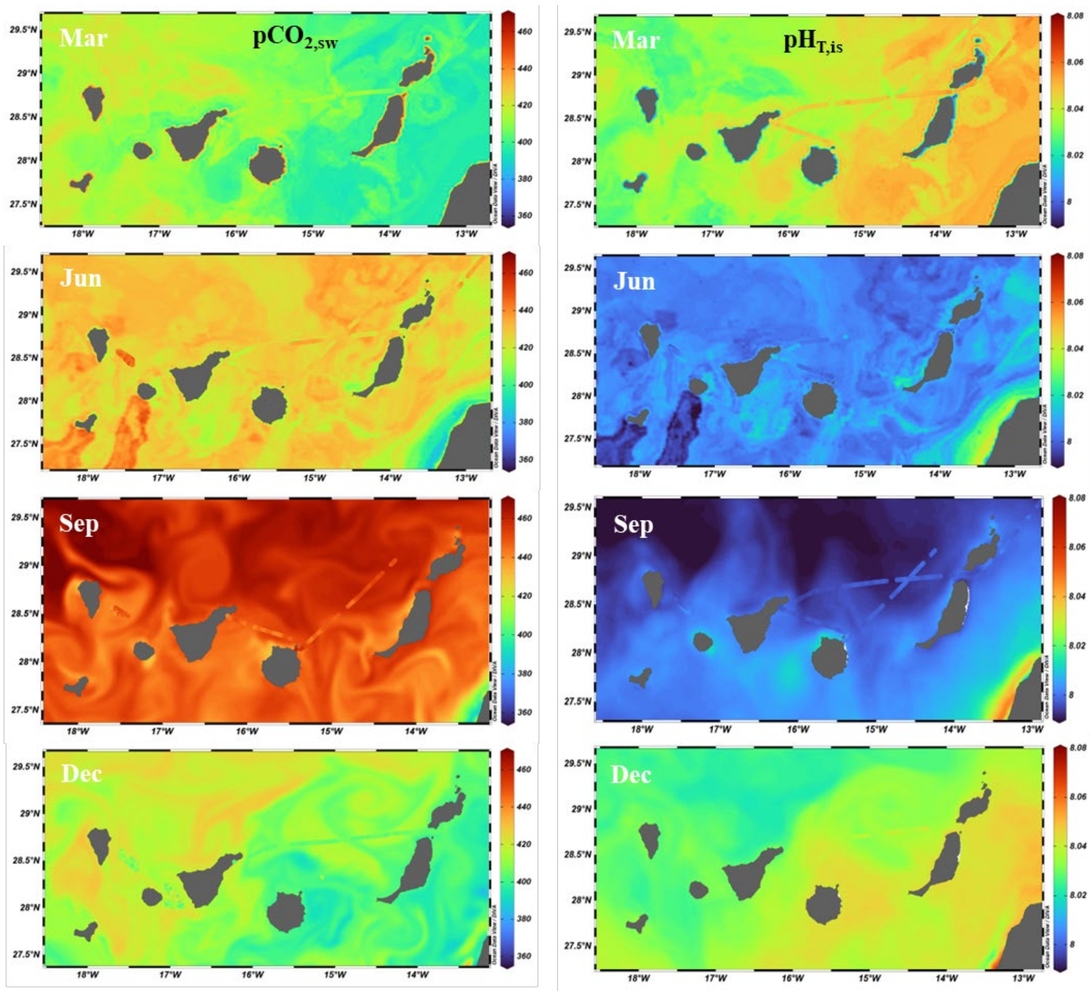
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 3. Monthly means of observational-based and model-predicted $p\mathrm{CO}_{2,\mathrm{sw}}$($\mathrm{pCO}_{2,\mathrm{atm}}$, SST, Chl *a*, MLD) and $\mathrm{pH_T}$(SST,Chl *a*,MLD) at the locations A-D (Figure 1). *MLR* (red) means *multilinear regression*, *NN* (green) means *neural network*, *CBo* (blue) means *CatBoost* and *Bag* (purple) means *Bagging*. Linear fittings for the seasonal detrended data are plotted.



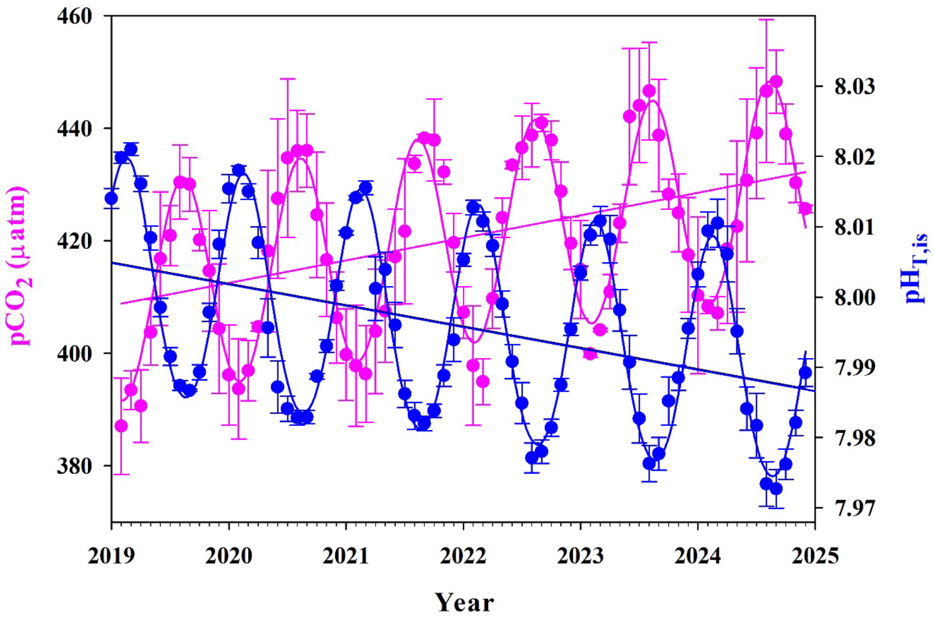
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 4. Gridded maps for $pCO_{2,sw}$ (left) and pH_T (right) predicted with *Bagging* for March (Mar), June (Jun), September (Sep) and December (Dec) 2023 using $pCO_{2,atm}$ and satellite conditions of SST, Chl-*a*, and MLD together with observational data (the same colour code was used). Figure produced with Ocean Data View (Schlitzer et al., 2021; https://odv.awi.de).



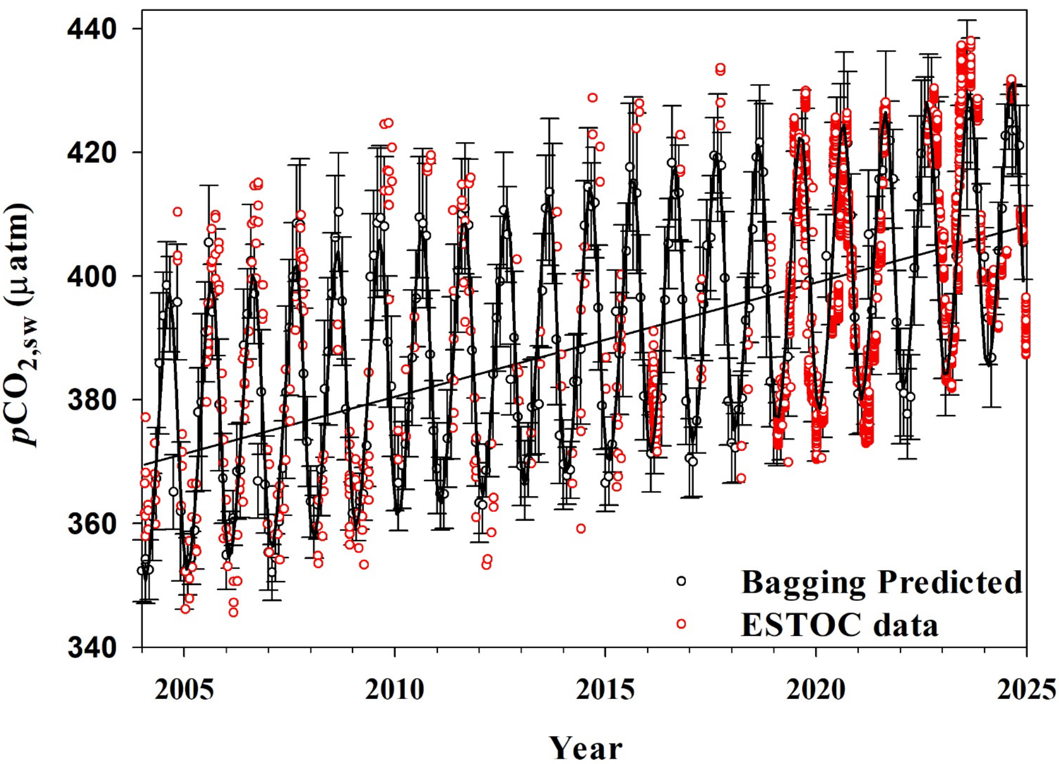
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 5. Monthly means of $pCO_{2,sw}$ (µatm) and $pH_{T,is}$ predicted with *Bagging* for 2019-2024 for the entire Canary region (27º-30ºN, 13º-19ºW). Linear fittings for the seasonal detrended data are also plotted.



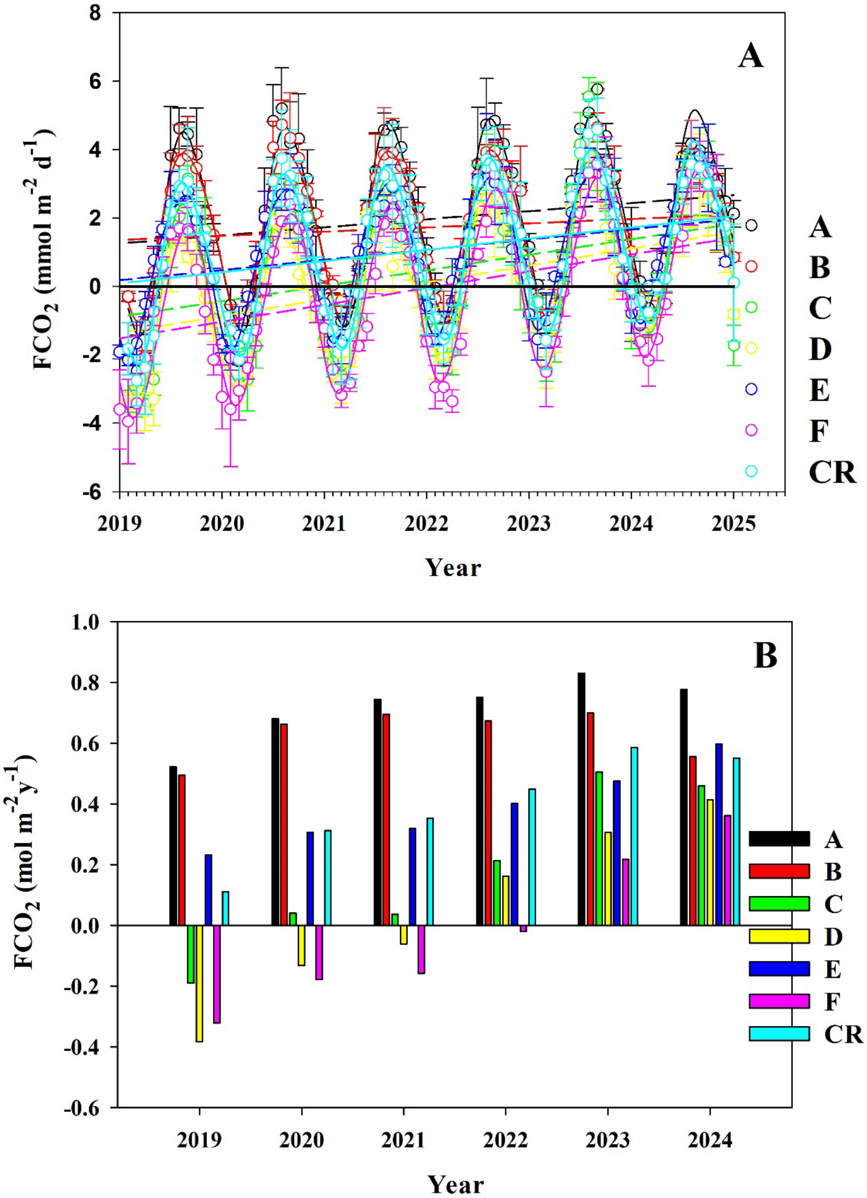
https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 6. Monthly means of $pCO_{2,sw}$ (µatm) predicted with *Bagging* considering $pCO_{2,atm}$, SST, Chl *a*, MLD for the period 2004-2024 at the location of the ESTOC site (G in Figure 1) and measured ESTOC $pCO_{2,sw}$.



https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Figure 7. (a) Monthly means of FCO_2 (mmol $m^{-2}d^{-1}$) in the Canary archipelagic waters predicted with *Bagging* from 2019 to 2024 and (b) net annual FCO_2 (mol $m^{-2}yr^{-1}$). In both plots, FCO_2 was represented at locations A-F and for the entire Canary Region (CR). Linear fittings for the seasonal detrended data are also plotted.



https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Table 1. Summary of the data used in this study by seasons and observing system.

		SST (ºC)	SST Satellite (ºC)	Chl-a Satellite (mg m^{-3})	Kd-490 Satellite (m^{-1})	MLD Satellite (m)	$pCO_{2,sw}$ (µatm)
Bechinjigua Express (LP-TNF)	Winter	**20.05 ± 0.34**	**20.03 ± 0.25**	**0.172 ± 0.041**	**0.041 ± 0.003**	**43.6 ± 17.6**	**402.0 ± 6.6**
	Spring	**21.39 ± 0.47**	**21.08 ± 0.37**	**0.115 ± 0.0217**	**0.035 ± 0.002**	**18.4 ± 6.5**	**419.8 ± 8.3**
	Summer	**23.40 ± 0.51**	**23.31 ± 0.56**	**0.12 ± 0.0214**	**0.036 ± 0.003**	**18.5 ± 6.3**	**440.3 ± 8.1**
	Autumn	**22.80 ± 0.38**	**22.61 ± 0.33**	**0.115 ± 0.0124**	**0.037 ± 0.002**	**39.4 ± 11.4**	**428.8 ± 7.3**
	Annual	**21.91 ± 0.43**	**21.76 ± 0.38**	**0.131 ± 0.024**	**0.037 ± 0.003**	**29.9 ± 10.4**	**422.7 ± 7.6**
Jona Sophie (GC-LNZ)	Winter	**19.39 ± 0.53**	**19.41 ± 0.36**	**0.172 ± 0.029**	**0.034 ± 0.002**	**52.4 ± 13.7**	**395.1 ± 5.9**
	Spring	**20.64 ± 0.46**	**20.44 ± 0.35**	**0.146 ± 0.024**	**0.034 ± 0.002**	**40.8 ± 12.0**	**408.2 ± 8.6**
	Summer	**22.87 ± 0.43**	**22.73 ± 0.39**	**0.122 ± 0.018**	**0.036 ± 0.002**	**41.3 ± 8.9**	**432.8 ± 6.8**
	Autumn	**22.09 ± 0.45**	**21.98 ± 0.37**	**0.106 ± 0.022**	**0.034 ± 0.002**	**32.2 ± 5.6**	**415.3 ± 5.8**
	Annual	**21.25 ± 0.47**	**21.32 ± 0.37**	**0.136 ± 0.023**	**0.034 ± 0.002**	**41.7 ± 10.0**	**412.8 ± 4.3**
MORGAN-1 (GC)	Winter	**21.07 ± 0.30**	**20.99 ± 0.23**	**0.193 ± 0.045**	**0.043 ± 0.004**	**57.0 ± 11.3**	**393.4 ± 1.9**
	Spring	**21.49 ± 0.31**	**20.66 ± 0.25**	**0.129 ± 0.021**	**0.039 ± 0.004**	**25.3 ± 9.4**	**405.1 ± 2.0**
	Summer	**21.50 ± 0.34**	**22.97 ± 0.24**	**0.11 ± 0.016**	**0.04 ± 0.004**	**23.1 ± 5.9**	**431.7 ± 2.8**
	Autumn	**21.53 ± 0.66**	**22.48 ± 0.25**	**0.126 ± 0.019**	**0.042 ± 0.004**	**41.2 ± 10.3**	**423.4 ± 5.9**
	Annual	**21.39 ± 0.40**	**21.78 ± 0.24**	**0.139 ± 0.025**	**0.041 ± 0.004**	**36.7 ± 9.2**	**413.9 ± 3.2**
ULA-2 (HI)	Winter	**19.76 ± 0.38**	**19.73 ± 0.39**	**0.193 ± 0.033**	**0.042 ± 0.003**	**47.7 ± 19.0**	**385.6 ± 3.3**
	Spring	**20.52 ± 0.56**	**20.48 ± 0.52**	**0.155 ± 0.037**	**0.037 ± 0.003**	**24.5 ± 7.9**	**397.9 ± 5.0**
	Summer	**21.92 ± 0.38**	**21.83 ± 0.33**	**0.159 ± 0.039**	**0.041 ± 0.006**	**25.2 ± 8.0**	**429.3 ± 4.6**
	Autumn	**23.29 ± 0.33**	**23.20 ± 0.30**	**0.171 ± 0.035**	**0.042 ± 0.004**	**24.0 ± 6.9**	**409.4 ± 5.9**
	Annual	**21.65± 0.36**	**21.59 ± 0.34**	**0.174 ± 0.035**	**0.041 ± 0.004**	**32.3 ± 11.3**	**405.6 ± 4.7**

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Table 2. Coefficients parameters computed for the multiple linear regression for $pCO_{2,sw}$ (top) and pH_T (bottom) applied to the different predictive models according to Eq. (5).

Variables	p_o **(µatm)**	$\hat{\alpha}$	$\hat{\beta}$ **(µatm °C^{-1})**	$\hat{\delta}$ **(µatm mg^{-1} m^3)**	$\hat{\varepsilon}$ **(µatm m^{-1})**
SST	198.5	-	10.40	-	-
SST + Chl-*a*	257.0	-	9.54	-10.89	-
SST + MLD	262.3	-	7.72	-	-0.17
SST + Chl-*a* + MLD	313.3	-	7.99	-0.31	-0.15
pCO_2 + SST + Chl-*a* + MLD	141.3	0.19	9.08	-1.79	-0.003

Variables	pH_o	$\hat{\alpha}$	$\hat{\beta}$ **(°C^{-1})**	$\hat{\delta}$ **(mg^{-1} m^3)**	$\hat{\varepsilon}$ **(m^{-1})**
SST	8.225	-	-0.009	-	-
SST + Chl-*a*	8.201	-	-0.008	0.069	-
SST + MLD	8.193	-	-0.008	-	0.0002
SST + Chl-*a* + MLD	8.185	-	-0.007	0.001	0.008

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Table 3. Algorithm performance between predicted $pCO_{2,sw}$ (µatm) and measured $pCO_{2,sw}$ (µatm) for each model using the training dataset.

Algorithm	Variables	R^2	RMSE (µatm)	MAE (µatm/day)	SSE (µatm/day)
MLR	SST	0.651	11.6	9.1	23.5
	SST + Chl-a	0.689	11.1	8.5	21.6
	SST + MLD	0.710	10.6	8.2	19.9
	SST + Chl-a + MLD	0.738	10.6	8.0	18.5
	SST + $pCO_{2,atm}$	0.865	6.7	5.0	15.3
	$pCO_{2,atm}$+SST + Chl-a + MLD	**0.904**	**4.9**	**3.5**	**10.3**
Neural Network (NN)	SST	0.740	10.4	7.7	25.6
	SST + Chl-a	0.778	9.4	6.7	19.5
	SST + MLD	0.842	8.1	5.7	18.2
	SST + Chl-a + MLD	0.881	7.2	5.0	17.2
	SST + $pCO_{2,atm}$	0.877	7.8	5.1	17.8
	$pCO_{2,atm}$+SST + Chl-a + MLD	**0.896**	**7.1**	**5.0**	**16.2**
CatBoost	SST	0.737	10.1	7.4	16.2
	SST + Chl-a	0.848	7.7	5.5	9.3
	SST + MLD	0.877	6.9	5.0	7.5
	SST + Chl-a + MLD	0.935	5.4	3.9	4.7
	SST + $pCO_{2,atm}$	0.933	4.2	4.0	5.4
	$pCO_{2,atm}$+SST + Chl-a + MLD	**0.956**	**3.6**	**2.4**	**3.0**
Bagging	SST	0.946	4.7	3.4	3.5
	SST + Chl-a	0.972	3.4	2.3	1.9
	SST + MLD	0.975	3.0	2.1	1.5
	SST + Chl-a + MLD	0.991	2.5	1.6	0.9
	SST + $pCO_{2,atm}$	0.982	2.6	2.085	1.1
	$pCO_{2,atm}$+SST + Chl-a + MLD	**0.991**	**2.0**	**1.6**	**0.8**

https://doi.org/10.5194/egusphere-2025-3699
Preprint. Discussion started: 28 August 2025

Table 4. Algorithm performance between predicted and measured pH_T for each model using the training dataset.

Algorithm	Variables	R^2	RMSE	MAE	SSE
MLR	SST	0.678	0.009	0.008	0.056
	SST + Chl-a	0.713	0.009	0.007	0.040
	SST + MLD	0.733	0.009	0.007	0.028
	SST + Chl-a + MLD	**0.745**	**0.006**	**0.005**	**0.013**
Neural Network (NN)	SST	0.751	0.009	0.007	0.050
	SST + Chl-a	0.805	0.009	0.006	0.027
	SST + MLD	0.819	0.008	0.005	0.013
	SST + Chl-a + MLD	**0.853**	**0.008**	**0.009**	**0.009**
CatBoost	SST	0.756	0.008	0.008	0.041
	SST + Chl-a	0.866	0.006	0.004	0.006
	SST + MLD	0.898	0.005	0.004	0.009
	SST + Chl-a + MLD	**0.934**	**0.004**	**0.003**	**0.002**
Bagging	SST	0.954	0.004	0.002	0.015
	SST + Chl-a	0.982	0.003	0.002	0.002
	SST + MLD	0.983	0.002	0.002	0.005
	SST + Chl-a + MLD	**0.991**	**0.002**	**0.001**	**0.001**