Review for "Modelling seawater pCO2 and pH in the Canary Islands region based on satellite measurements and machine learning techniques"

Overview:

The authors used satellite data in combination with in situ and moored data to train machine learning algorithms to estimate $pCO2$ and pH in seawater surrounding the Canary Islands region. They compared different machine learning methods and used the results to estimate an annual trend for the two variables of interest. They also discussed the implications of a marine heat wave and shifts in air-sea $CO2$ flux for the region, finding increasing $pCO2$ and decreasing pH of seawater, accompanied by increasing source behavior of seawater.

Strengths:

The authors conducted a strong intercomparison of different machine learning methods and used a more-than-sufficient set of criteria to evaluate each method. They also did a good job emphasizing the overall importance of this work in the introduction and describing the factors influencing $pCO2$. The aim of the study was also clearly stated in the introduction. Methods for model selection and application were robust overall. Although I have provided many minor suggestions (see the line commentary), overall this manuscript will add significantly to the knowledge gaps for coastal marine carbonate chemistry. Please read all comments with the knowledge that, overall, I think this is a unique and important study which contributes well to the field of research.

Weaknesses:

The article could benefit from language editing with a focus on improved clarity. There is also lack of clarity in which stations and cruise tracks align with which locations on the map (Figure 1), and some essential data about these stations is missing. Figure 1 and Table S1 need a lot more information to be useful. Some methodological details and rationale are also omitted.

We sincerely thank you for your thoughtful and constructive feedback on our manuscript. We greatly appreciate your recognition of the relevance of our study. Your comments have been carefully considered and have played a key role in improving the overall quality, clarity, and robustness of the revised manuscript. Below, we provide a point-by-point response to each of your comments.

Line by Line Commentary:

Abstract

L22-37: Were the final algorithms based solely upon satellite data? It was interesting that the buoy and VOS-based time series weren't mentioned here, might be worth including.

Thank you for your comment. The final algorithms were based solely on satellite-derived data,

which provide high-resolution spatial coverage across the Canary Islands. We agree that mentioning the buoy and VOS-based time series adds valuable context, and this information has now been included in the revised manuscript accordingly.

The correction is: "The improvement of remote sensing systems, together with the emergence of new model-fitting algorithms based on machine-learning techniques, has allowed the estimation of the partial pressure of carbon dioxide ($pCO_{2,sw}$) and pH ($pH_{T,sw}$) in the waters of the Canary Islands (13-19ºW; 27-30ºN). Continuous time series data from moored buoys and Voluntary Observing Ships (VOS) between 2019 and 2024 were used to train and validate the models, providing an observational foundation for the satellite-based estimations. 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 $pH_{T,sw}$, 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 $7.97 < pH_{T,sw} < 8.07$. 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}. Over the 6-year period (2019-2024), the rise in the atmospheric CO_2 and the increase in sea surface temperature, which reached 0.2 ºC per year under the influence of the unprecedented 2023 marine heat wave, contribute to this important rate. Considering the Canary Islands, 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"

L23: Is this determination, or more estimation of pCO_2 and pH?
We thank the reviewer for the comment. The reported values represent estimations of pCO_2 and pH. This clarification has been added to the manuscript accordingly.

L32: The comma after "both" is unnecessary and should be removed.

Thank you for noticing this. The comma after "both" has been removed in the revised manuscript.

L32-32: This sentence is awkwardly worded and a bit confusing.

We apologize for the confusion caused by the original phrasing. The manuscript has been revised, and the sentence has been rewritten for clarity. The correction is: "Over the 6-period (2019-2024), the rise in atmospheric CO_2 and the increase in sea surface temperature, which reached 0.2 ºC per year under the influence of the unprecedented 2023 marine heat wave, contribute to this important rate"

L33: It may be useful to list the 6-year period timeframe.

The 6-year period timeframe has been added to the manuscript as recommended.

L34: The geographic location of the Canary Islands can be mentioned when "waters of the Canary Islands" is first mentioned in L24.

Thank you for the comment. We agree, and the geographic location of Canary Islands has now been included when the "waters of the Canary Islands" are first mentioned in the manuscript.

Introduction

L43: “fossil fuels burning” can be singular (“fossil fuel burning”) when used in this context.

The phrase has been corrected to “fossil fuel burning” in the revised manuscript.

L49-50: This should read: “based on the analysis of an 18-year dataset” (“an” is missing).

The sentence has been corrected to “based on the analysis of an 18-years dataset” in the revised manuscript.

L67: Although fCO_2 is basically equivalent to pCO_2, in an article like this it may be best to quickly mention that fugacity considers non-ideal gas behaviors upon first mention, since simply stating that pCO_2 is “expressed as CO_2 fugacity” omits this important distinction which some marine chemists may notice.

We have revised the manuscript to clarify the distinction between pCO_2 and fCO_2. The updated text now reads:

“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 ($fCO_{2,sw}$). While fCO_2 is numerically very similar to pCO_2, it accounts for non-ideal gas behavior due to molecular interactions, typically being slightly lower than pCO_2 (Wanninkhof et al., 2022).”

Wanninkhof, R., Pierrot, D., Sullivan, K., Mears, P., & Barbero, L. (2022). Comparison of discrete and underway CO2 measurements: Inferences on the temperature dependence of the fugacity of CO2 in seawater. *Marine Chemistry*, *247*, 104178.

L82-85: How was this inherent error of satellite remote sensing considered in the analyses? L86-96: Were these approaches based on satellite data?

We thank the reviewer for the comment. The inherent uncertainties associated with satellite-derived variables (SST, Chl-a, and MLD) are implicitly incorporated into the model training process. Therefore, their errors are propagated through the regression process and are reflected in the model’s resulting metrics. The model accounts for the variability and measurement uncertainty inherent to the satellite data. We have clarified this in the revised manuscript accordingly.

L115: Define ESTOC upon first use.

ESTOC has been defined upon first use in the revised manuscript.

L199: This should read “of the Canary Basin”, not “on the Canary Basin”.

The text has been corrected to read “of the Canary Basin” in the revised manuscript.

Methods

L126-143: It is unclear which stations correspond to which moored buoys and which ships correspond to which tracks on Figure 1. This should be clearly phrased in the Fig. 1 caption. It is mentioned in Lines 157-162, but it is still somewhat unclear.

We agree that the original figure was unclear. Figure 1 has been revised, and the caption has been updated to clearly indicate which stations correspond to which moored buoys and which ships correspond to which tracks.

L148-156: Similarly, at what locations were these discrete samples collected, and by whom (e.g., authors, VOS personnel, or others) and at what depth? It would be useful to show these on the map in Fig. 1.

Discrete samples were taken by the scientific group at the same seawater line feeding the pCO_2 system to assure similar conditions. The locations of the samples are indicated in Fig. 1

L153: Does "measured" pCO_2 data refer to that from moored and VOS systems?

This has been clarified.

L55: Similarly, was "observed" (corrected) data the one measured from moored and VOS systems?

This has been clarified

L158: Is the LG site the same as GOM (La Gomera)? I don't see where LG is defined.

GOM has been corrected in the revised manuscript.

L165: What method was used to obtain wind speed, and at what height?

We thank the reviewer for the comment. Wind speed data were provided by the Spanish State Meteorological Agency (AEMET). Measurements were taken at the following stations and heights: La Palma Airport (33 m), La Gomera Airport (15 m), Fuerteventura Airport (25 m), Lanzarote Airport (14 m), Gran Canaria Airport (24 m), and El Hierro Airport (32 m), corresponding respectively to sections A–F of the study area. Winds data were referenced to a 10 m height following $u_2 = u_z \left(\frac{4.87}{\ln(67.8z-5.42)}\right)$ (Allen et al., 1998). These aspects have been now added.

L178: Is $pCO_{2,sw}$ or $pCO_{2,eq}$ the partial pressure of CO_2 in seawater, as defined here?

Along the text $pCO_{2,sw}$ is the only one used. The measured $xCO_{2,eq}$ was corrected by gas calibration and by seawater gradients between intake and equilibrator as indicated in the experimental section. We compute first pCO_2 at the temperature in the equilibrator and then the pCO_2 at in situ temperature is computed.

L182: Instead of "real partial pressure $fCO_{2,sw}$", should this read "fugacity of seawater $fCO_{2,sw}$"?.

Done.

L182-188: I worry that the observations and averaging used for these calculations were enough to accurately capture AT, and hence pH, dynamics, especially over a span of six years with seasonal variability too. A paragraph during the discussion on uncertainties in calculations and assumptions would benefit the authors' methods here. Also, the depth and analysis of discrete samples should be described.

Discrete samples are taken in the seawater line used for feeding the pCO_2 system. At the selected locations, the 500 ml bottles were washed three times, overfilled, poisoned with 100 ul of concentrated $HgCl_2$, sealed and kept in darkness until their analysis on land, just after arriving. Salinity, temperature and location was taken from the equipment data record. The normalized to a constant salinity of 35 was inside ± 3 µmol kg^{-1}, without any significant seasonal variability after normalization, as it has been observed in the region at the ESTOC site for decades. This has been indicated in the new version.

L187: Is this pH on the total scale?

Yes, pH is expressed on the total scale, as clarified in the revised manuscript: "This relationship was then used to compute pH (AT(SSS), $fCO_{2,sw}$) values in the Canary Region (González Dávila et al., 2010), with pH expressed on the total scale. The data were then averaged daily."

L190-191: I don't understand the phrase, "as the data could be overestimated due to ship operations." Please explain. Also, where, geographically, does the Izaña Atmospheric Research Centre measure $xCO_{2,atm}$?

Shipboard measurements can occasionally be overestimated due to contamination from ship exhaust and deck operations. The Izaña Atmospheric Research Centre (AEMET, 2024), which is located on Tenerife Island (28°18′N, 16°29′W, 2373 m a.s.l.), provides measurements of $xCO_{2,atm}$. These clarifications have been added to the revised manuscript.

L203-204: Does this mean that fluxes were calculated separately for experimental (mean of ship and buoys?) and modelled data? Similarly, were average monthly fluxes averages of both experimental and modelled data?

We appreciate the reviewer's comment. Fluxes (FCO_2) were computed for the modelled datasets using Equation 1. After checking our model was able to describe $pCO_{2,sw}$ with high accuracy compared to measured pCO_2 from buoy and VOS data, FCO_2 was computed for the resulting model output. Daily fluxes were averaged to obtain monthly means, which are expressed as daily averaged values (mmol m^{-2} d^{-1}). This clarification has been added to the revised manuscript.

L215: What is meant by "statistical treatment"? I have never heard this phrase, do the authors mean "statistical analyses"?

We agree with the reviewer. The phrase "statistical treatment" has been replaced by "statistical analyses" to improve clarity and align with standard scientific terminology.

L216-229: The MLR and machine learning models were devised based solely upon satellite data? This is not as clear as it could be. Also, was the final model that was applied trained on both training and validation data, or only training data? L217-218: I suggest adding information on how data were divided to avoid autocorrelation to the supplement. This is not necessary information but may help future researchers using similar methods

We thank the reviewer for the comment. To clarify, the predictor variables for both the MLR and machine learning models were derived solely from satellite observations (SST, Chl-a, Kd_{490}, MLD), while the response variables ($pCO_{2,sw}$ and $pH_{T,sw}$) came from in situ measurements collected by buoys and VOS.

The original datasets were divided into training (80%) and validation (20%) subsets to minimize temporal and spatial autocorrelation. The training dataset was used to tune the models, and the validation dataset was used to assess predictive performance. After determining the optimal model parameters, the final models were retrained on the entire dataset (training + validation) to provide the estimates presented in this study.

These clarifications have been added to the revised manuscript to make the methodology more explicit.

L229: What was the rationale for choosing these three methods? A bit of background information on each method would be useful.

We express our gratitude to the reviewer for this insightful comment. Three machine learning methodologies were selected: neural networks (NN), categorical boosting (CatBoost), and bootstrap aggregation (*bagging*). This were selected due to their complementary nature, which effectively balances flexibility, interpretability, and model robustness. Neural networks are particularly adept at capturing complex nonlinear relationships among oceanographic variables. CatBoost, a gradient-boosting algorithm, is proficient in handling categorical features and reducing prediction bias, thereby ensuring high accuracy and stability. In contrast, *bagging*, an ensemble technique, enhances model reliability by reducing variance and mitigating overfitting. In accordance with the suggestion, we have incorporated this justification and a more comprehensive description of the background of each method into the revised manuscript.

L230: "the main statistical parameters" is vague and may vary depending on the study. Perhaps omit?

We agree that the phrase "the main statistical parameters" was vague. The text has been revised to specify the exact evaluation metrics used, which now reads:

"Model performance was evaluated using the validation dataset by calculating the coefficient of determination (R^2), root mean square error (RMSE; Eq. 6), mean absolute error (MAE; Eq. 7), and the daily sum of squared errors (SSE; Eq. 8)."

L248: What were the "two assumptions required to achieve predictive ability"?

The "two assumptions required to achieve predictive ability" referred to the normality and homoscedasticity of residuals. This section has been clarified in the revised manuscript, which now states: To estimate the coefficients in each seasonal model and determine confidence intervals, two assumptions were tested: (1) the residuals are normally distributed, assessed using the Shapiro-Wilk test ($\alpha = 0.05$) and quantile-quantile plots, and (2) the residuals have homogeneous variance (homoscedasticity), assessed graphically. If the normality assumption was not met, bootstrapping was used to determine confidence intervals. Model comparisons were performed using analysis of covariance (ANCOVA) and analysis of variance (ANOVA) to detect significant differences at $\alpha = 0.05$.

Results

L264-265: Which SST was used (satellite or in situ)? Also was in situ from moored, discrete, and vessel observations combined? If so, how was data density considered in data averaging?

Both satellite-derived and in situ SST datasets were used, but they were not combined. Instead, they were compared to assess their consistency and agreement before applying the satellite-derived SST to train the predictive models for $pCO_{2,sw}$ and $pH_{T,sw}$. The comparison, summarized in Table 1 and shown in Figure 2, indicated an average difference of only ~0.15°C between satellite and in situ SST, confirming good agreement between both datasets. This clarification has been added to the revised manuscript.

L310: What ere the AICc for each prediction model?

Table added as Table S4 in supplementary material

Variables	**AICc**	**ΔAICc**
SST	127.885	52.654
SST + Chl-a	123.994	48.763
SST + MLD	113.823	38.592
SST + Chl-a + MLD	109.954	34.723
$pCO_{2,atm}$+ SST + Chl-a + MLD	**75.231**	**0.00**

L321: Why is “Table 3” italicized here but not elsewhere?

Changed

L324: Section 3.2.2.1, Neuronal network (NN): I believe the correct term is “Neural network”?

We thank the reviewer for the observation. The term has been corrected to “Neural network (NN)” in the revised manuscript.

L332-333: It looks almost the opposite to me, that the fitness of the model (spread of the data) in figure S2 is worse at higher pCO_2 value.

We agree with the reviewer that the model performance decreases slightly at higher pCO_2 values, as reflected by the increased scatter in Figure S2. The text has been revised accordingly to clarify

this point and now reads: “Despite the overall good agreement, the scatter of the data tends to increase at higher pCO_2 values, indicating a slightly poorer fitness in that range.”

L33-335 The accuracy indicators of training versus validation datasets can most likely be listed within the supplement, there is no need to list them here.

We agree. As data is indicated in the Tables, the indications were removed.

L363-364: I am also wondering again how data density was considered when incorporating the various forms of data (ship track, moored, discrete) into the models. Also, what type of interpolation and from satellite data within what vicinity/window of the in situ data should be included in supplement.

Daily averaged measured values of $pCO_{2,sw}$ and SST for the different platforms and the corresponding chlorophyl and MLD satellite data at the same location as the measured values were used to train the models. The density was, therefore, one data for each location per day where data were available. When it was confirmed the measured SST and satellite SST was in agreement, the model was run for the estimation of $pCO_{2,sw}$ with satellite SST, chlorophyll and MLD for the full canary region.

L369-374: What was the AIC for these models? Did it support the other stats?

Table added as Table S5 in supplementary material

Variables	**AICc**	**ΔAICc**
SST	193.129	0.000
SST + Chl-a	196.109	172.629
SST + MLD	197.860	298.023
SST + Chl-a + MLD	201.856	473.035

L380-381: Why is bagging sometimes italicized and sometimes capitalized?

We thank the reviewer for noticing this inconsistency. The formatting of bagging has been corrected throughout the manuscript to ensure consistency.

L394-395: Were the experimental or predicted values higher?

The experimental pCO_2,sw values were slightly higher than the predicted ones. We have clarified this point in the revised manuscript which now reads: “In general, the experimental values of pCO_2,sw were slightly higher than the predicted ones, with an overall agreement better than 1.7 ± 1.8 µatm for pCO_2,sw and 0.002 ± 0.001 for pH_t.”

L406: The results should mention the trends shown in Figure 4, and whether these trends were significant. Also, how were seasons detrended?

The reference to Figure 4 was a mistake. It is Figure 3.

Discussion

L411-412: For what variable were the satellite data different from observed data, and what is meant by "very different"?

We thank the reviewer for the comment. The text has been clarified to specify that salinity is the variable for which satellite data differ from observations. Satellite-derived salinity exhibits significant discrepancies and high variability compared to in situ measurements, making it unreliable for modelling in this region. This clarification has been added to the revised manuscript:

Salinity was excluded from the fitted models due to its negligible impact on $pCO_{2,sw}$ variability (Sarmiento et al., 2007; Shadwick et al., 2010). Furthermore, satellite-derived salinity data have been reported to exhibit significant discrepancies from in situ measurements, characterized by high variability and uncertainty (Yu, 2020), rendering them unreliable for modelling in this region.

L453-462: Was the full five-year dataset in situ or satellite-based?

The in situ data was indicated.

L460-462: Earlier it was stated that the in situ and satellite data were different (presumed for SST). Please explain.

This has been clarified. 0.19ºC was the average difference between measured and satellite at the two buoys while 0.15ºC was determined for the full data set for the Canary region.

L501 (4.2): Once again, pHT,is isn't defined.

We apologize for the confusion. In the original manuscript, both $pH_{T,sw}$ and $pH_{T,is}$ were used interchangeably, although they refer to the same variable. This has now been corrected, and the notation $pH_{T,sw}$ is used consistently throughout the manuscript.

L504-505: I don't see the monthly experimental averages in Figure 4.

Thank you for your comment. The monthly experimental averages are indeed included in Figure 4; however, they overlap with the predicted values, indicating a very close agreement between observations and model predictions. This explains why the experimental averages are in most of the areas distinguishable in the Figure. The values can be seen in September in both variables, but with values inside the estimated error for the model.

L516: Sometimes the term bagging is capitalized and sometimes it is not. Please be consistent.

Done

L529: There is no figure 8, although it is cited here.

We apologize for the confusion. The reference to Figure 8 was a mistake; it should correctly refer to Figure 5. This has been corrected in the manuscript.

L544-545: It would be interesting to see the ESTOC long-term data, or a summary of it, here.

We agree with reviewer about this indication. We have included reference to the last ESTOC study where all data set are considered and also in section 4.3 this has been included and detailed. We considered the relationship of our findings and those at ESTOC are well described.

L594: It would be nice to see a brief caveats section, especially discussing how only surface water quantities are accurately described by satellite data. –

We agree with reviewer and we have included that in the paper. It should be noted that the estimates presented here primarily rely on surface water measurements, particularly those obtained from satellite data. While these datasets provide high-resolution spatial coverage and accurately capture surface trends, they do not resolve subsurface processes or vertical gradients in CO_2 and temperature.

Conclusions

L610-612: I am still left unsure what final data was included in the final model for predictions.

Thank you for the comment. This aspect has been clarified along the text. We started with measured values for $xCO_{2,sw}$ converted in $pCO_{2,sw}$ and SST together with satellite chlorophyll and MLD values to run the models. Then, after checking that measured SST and satellite SST agreed better than ±0.15ºC for the region, the final model was tested with satellite SST, Chl-a and MLD to predict the measured $pCO_{2,sw}$. Therefore, the final model predictions of $pCO_{2,sw}$ and $pH_{T,sw}$ across the Canary Islands were based exclusively on satellite-derived datasets, data that can be obtained on a daily resolution for the full region. These high-resolution satellite data allowed us to capture the seasonal and spatial variability of surface seawater properties across the entire region. We have clarified this point in the revised manuscript which now reads: 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 ($pH_{T,sw}$) across the entire marine region of the Canary Islands. After checking the measured SST and satellite SST agreed better than ±0.15ºC for the region, the model was trained with measured values for $xCO_{2,sw}$ converted in pCO_{2sw} and satellite SST, chlorophyll and MLD values, which provided high-resolution coverage and allowed the capture of spatial and seasonal variability throughout the study area.

Figures

Figure 1: The caption should describe what sorts of data were collected at what sites and tracks. It is confusing to read independently now. E.g., where were moored T, S data collected? At what locations were discrete samples collected? A subplot indicating the location of the study area

relative to a larger-scale map may also be beneficial. Why are some stations represented by stars and others represented by rectangles? Also, this map has no scale bar.

Thank you for your suggestions regarding Figure 1. We have updated the figure and its caption accordingly. The symbols (stars versus rectangles) have been clarified, the routes have been specified, a scale bar has been added, and a subgraph has been included to show the location of the study area in relation to a larger-scale map. The location of discrete sampling has been included.

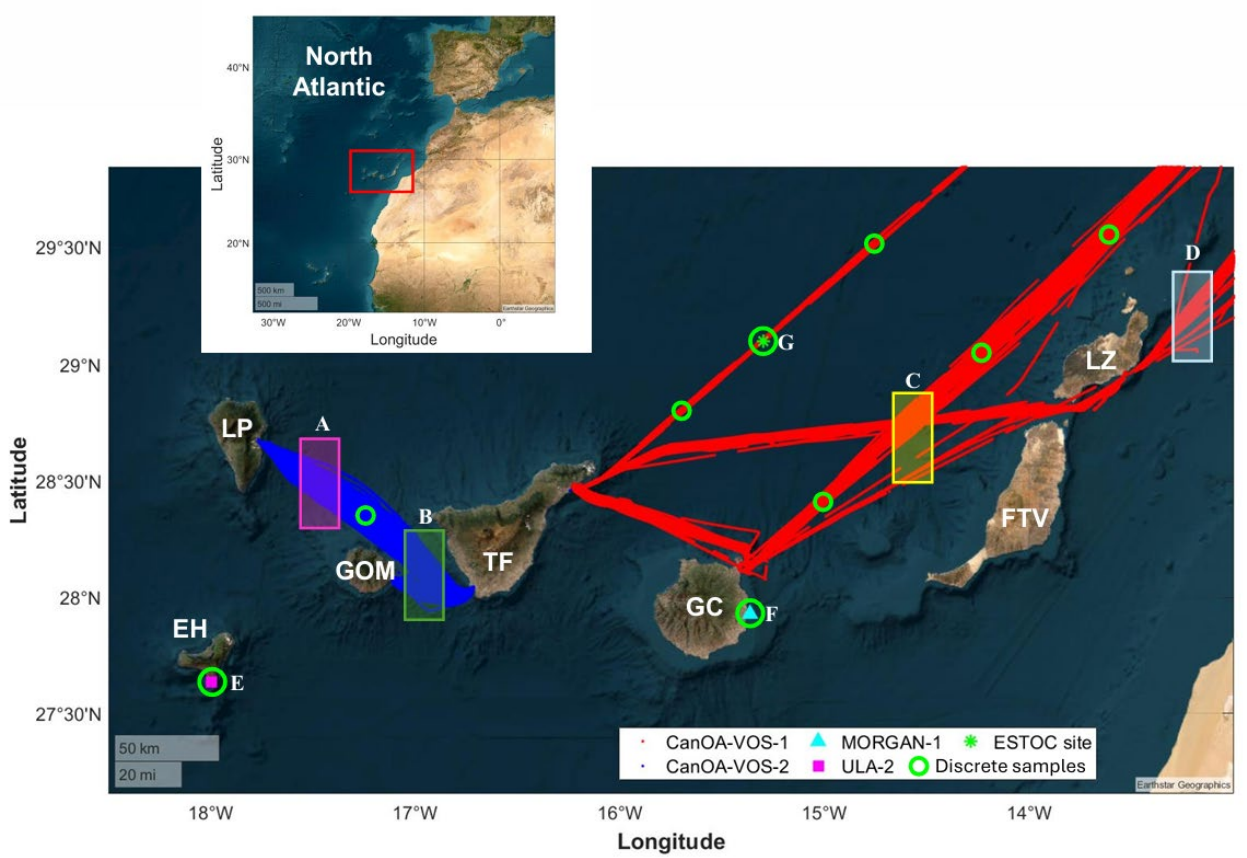

Figure 1. Map of the Canary Islands showing the CanOA-VOS tracks (CanOA-VOS-1 Jona Sophie, in red; and CanOA-VOS-2 Benchijigua Express, in blue), the locations of the moored oceanographic buoys (MORGAN-1, cyan triangle; ULA-2, purple square), the ESTOC site (green star), and green circles indicate discrete sampling locations. The positions of sites A–F are also indicated, with site E corresponding to the ULA-2 buoy, site E to the MORGAN-1 buoy, and site G to the ESTOC site. The island acronyms are included (EH: El Hierro, LP: La Palma, GOM: La Gomera, TF: Tenerife, GC: Gran Canaria, FTV: Fuerteventura, LZ: Lanzarote).

Figure 2: Is in situ SST from ship vessel track observations, or discrete samples, or moored buoys, or a combination? It is unclear. Also, A appears to have a blue trend line for in situ data, while all others have a black trend line. How seasonal detrending was done should be described in the methods. It should also mention that standard deviations are shown by the bars.

Thank you for your comment. The in situ SST data presented in Figure 2 come from the ship-based observations (A-D) and moored buoys (E and F), as now clarified in the figure caption. Seasonal detrending procedure has now been indicated in the Methods section, after Eq. 3 and 4, and standard deviations are indicated by the error bars.

Tables

Table 1: Are the statistics reported in Table 1 mean +/- standard deviation? If so, this should be noted in the title of this table. Also, the geographic location of the titles in the first column are unclear. Do they correspond to any locations in Figure 1? This needs to be stated clearly. Was the $pCO_{2,sw}$ shown here calculated from satellite data only, or from both observed and satellite data? How was the satellite data that corresponds to each ship or moored platform chosen within the window of proximity to that ship/platform?

Thank you for your comment. The statistics in Table 1 are indeed reported as mean ± standard deviation, which has now been clarified in the table title. The locations listed in the first column correspond to the two ship routes (CanOA-VOS-1 Jona Sophie and CanOA-VOS-2 Benchijigua Express) and the two moored buoys (MORGAN-1 and ULA-2). The pCO_2,sw values shown in the table are based on in situ observations from these ships and buoys. For satellite data used in the modelling, values were selected based on the latitude and longitude of the corresponding in situ measurements to match the proximity of each ship or moored platform. Seasonal mean and standard deviation for the observational data (SST and $pCO_{2,sw}$) and satellite data (SST, Chl-a, $Kd_{,490}$ and MLD) used in this study.

Table 2: Is the final row of the first table's pCO_2 coefficient atmospheric pCO_2? This should be clarified within-table. Also, was this the final result for the model which includes all data (training + validation), or only the training data?

Thank you for your comment. We confirm that the final row of the pCO_2 table corresponds to atmospheric pCO_2, and this has now been clarified in the table caption.

Table 4: The gray shading of this table seems random. Is there supposed to be a pattern that is being pointed out?

Thank you for pointing this out. The gray shading in Table 4 was an error during table preparation and has now been removed in the revised manuscript.

Supplement

Which stations align with buoys should also be described in the supplement text and in Table S1, relative to those in Fig. 1. It would also be nice to know the frequency and timespan of observations in Table S1.

Done

Figure S2: The x-axis is lacking labels. Also, the size of each dataset could be listed in the caption for enhanced comprehensibility.

Done

Figure S3: I would also like to know the size of each dataset here, and x-axis labels are needed.

Done