The authors have produced a short and focused study relating measured pCO2,sw in the vicinity of the Canary Islands to predictors that can be obtained from remote sensing. The core aim of the paper is modest, but worthwhile. The execution seems to have some mistakes. The paper should be returned to the authors for revisions. In addition to revisions associated with the primary recommendation below, I would urge the authors to reduce or summarize the statements comparing temperatures in various locations. These are often presented without context such that I found myself wondering why so much text was devoted to discussion of how temperature varies spatially and temporally. I feel that limiting this text could help shorten and strengthen the paper.

My primary recommendation for this paper is, perhaps ironically, the same as in the review that I erroneously submitted earlier (and apologies again for my mistake). When fitting machine learning or regression models, it is insufficient to divide the training data randomly by measurement, as appears to have been done here. This is because the measurements that are collected by seagoing work are usually nearly synoptic and are highly correlated both in space and time. Therefore, the relationships that reconstruct the training measurements along a cruise or transect almost invariably do a fantastic job of reconstructing other 'validation' measurements made along the same cruise or transect… even while failing to reconstruct the patterns of variability found at other times and locations. This tendency can be even more pronounced when using ML models with many degrees of freedom. I'm not positive, but I believe an example of this can be clearly seen in figure 4 where the ML model seems to have optimized a specific relationship for the transect with data that does not at all extend spatially into the rest of the ocean. The fix for this is pretty simple: divide up all of your measurements randomly by "cruise" or whatever identifier is appropriate for a given boat making measurements with a given instrument in a given year. Then partition the data between training and validation using random selections of these collections of data. Ideally, use k-fold validation to ensure that all data are included in both the training and validation data at various times. I would expect that the bagging routine's performance will be much more in line with that of the other approaches after this is done.

Stylistically, I'll note that the writing struggles at times (see non exhaustive comments below), and the notation is sufficiently inconsistent that it appears to have been written piecemeal by multiple people. Please homogenize the notation.

We express our gratitude to the reviewer for their insightful comment and concur that, when utilizing in situ data collected along cruises or transects, it is imperative to exercise caution to prevent spatial and temporal autocorrelation between the training and validation datasets. This concern has been explicitly addressed in the new version. The dataset comprised two ship-based transects (cruises) and two moored buoys, with each cruise dataset containing multiple measurements along a specified transect. Prior to model fitting, we eliminated missing values and outliers and subsequently selected satellite-based predictors. For the purposes of model training and validation, the data were randomly divided at the cruise level; specifically, entire cruise datasets were allocated to either the training or validation subsets. An 80% of the data (randomly selected by cruise) were utilized for training, while 20% were reserved for validation. This random division was executed at each model run, ensuring that the partitioning was both unbiased and representative of the available data. The Bagging model introduces an

additional layer of randomness: each tree within the ensemble is trained on a distinct random subset (bootstrap sample) of the training data, and their outputs are aggregated to yield a stable and robust final prediction. Despite the stochastic nature of the training process, the model consistently provides reliable predictions for the same input data, thereby confirming the efficacy of the randomization procedure. Consequently, we affirm that the data division was conducted not by individual measurements along a transect, but rather randomly by cruise, in alignment with the reviewer's recommendation. We have elucidated this procedure in the revised manuscript to eliminate any potential ambiguity.

Line by line comments:

30: line height formatting error

This is an error the word program does when subindices are included.

32: this sentence has incorrect grammar.

Thank you for noticing this. The sentence has been corrected.

33: what 6 year period?

We thank the reviewer for the suggestion. The 6-year period timeframe has been added to the manuscript as recommended.

58: this reference is almost a quarter of a decade old, so it's not ideal for making the point that this is still a problem, especially as there have been several recent studies aimed at improving coastal pCO2 products.

New references have been added.

67: I suggest italicizing the p in pCO2, especially if you italicize the f in fCO2. This will distinguish it from the "-log10" meaning for p in pH and pe.

Done.

69: IUPAC conformant CT has the C italicized (even though it represents the element carbon)… same for AT later

Thank you for noticing this. Corrected.

86: use parentheses here, otherwise it appears as though MLR is another element in a list with multilinear regression. Also, MLR is already defined in the abstract.

Corrected.

153:here you have italicized the p, definitely be consistent

Corrected.

179: line height error

This is an error the word program does when subindices are included.

189 and 174: inconsistent italicization of x

Thank you for your comment. Finally, they are all in italics.

231: earlier r was not capitalized

Corrected.

245: no need to indent since it's not the start of a paragraph

Corrected.

273: check spacing

Corrected.

403: line height

This is an error the word program does when subindices are included.

413: This could be confirmed by excluding Chl from the fit and confirming that Kd,490 is then selected as a predictor variable

We acknowledge the reviewer's recommendation. The model adjustment was conducted again, this time excluding chlorophyll a as a predictor. As anticipated, Kd490 provides a similar statistically significance that using chlorophyll a. However, error of estimation of Kd490 is higher than Chla in the satellite data base which also makes the use of Chla a primary variable. Both variables were strongly correlated (R^2 = 0.96). This additional analysis reinforces our interpretation that these two variables are redundant, with the model favoring Chl a when both variables were present.

415: I'm confused by this claim. Why does it matter which variables were used to predict pCO2,sw for an algorithm focused on pH? Or are you talking about a calculation for pHT from TA (f(S)) and pCO2sw, in which case why does it matter what the atmospheric value was at all?

We agree with and this sentence is redundant and it has been removed.

432: These sentences are not logically linked. It current reads as though the authors are implying that there is a temporal trend in the distance from the African continent.

This part reads now as "The statistically significant differences ($p < 0.05$) observed between the western and eastern sections are related to the distance from the African continent, with the easternmost part of the archipelago being more exposed to upwelling filaments (Davenport, 1999) and the westernmost part being sheltered by the islands themselves. This spatial pattern, clearly visible in Figures 2 and S1 through the progressive decrease in SST toward the African coast, is well captured by the satellite observations, whose validation showed no significant differences ($p < 0.05$), even near the islands. Therefore, satellite data were deemed suitable for model fitting and subsequent parameter estimation"

445: winter of 2023-2024… or JFM?

We agree with reviewer. We have indicated that winter of 2023-2024 corresponds to the months JFM.

464: it seems odd that the model with the highest prediction error has better validation statistics than an alternative presented immediately afterwards

We thank the reviewer for catching this mistake. This was indeed a typographical error in the text. The model with the highest prediction error is the neural network (RMSE = 7.1 µatm, R^2 = 0.896), whereas the MLR model performed slightly better (RMSE = 4.9 µatm, R^2 = 0.904). The sentence has been corrected in the revised manuscript

483: it is unclear what is meant if a variable controlling something is characterized by a component. Consider "The strong predictive power of this relationship is likely because pCO2sw variability is dominated by thermal changes in this region, and these changes are directly captured by satellite SST records."

We thank the reviewer for this suggestion. We agree that the original phrasing was unclear. The sentence has been revised to clarify that the variability of $pCO_{2,sw}$ and $pH_{T,sw}$ in the Canary Islands region is primarily driven by thermal changes, which are well represented by satellite SST data. The revised text now reads as follows: It is suggested that these considerably favourable results, and the comparable errors with ocean-scale models, arise because the variability of $pCO_{2,sw}$ and $pH_{T,sw}$ in the waters around the Canary Islands is largely dominated by thermal effects (Takahashi et al., 2002; González-Dávila and Santana-Casiano, 2023). In this region, the thermal control on surface carbonate chemistry is directly captured by satellite-derived SST. In all cases, the simple model using only SST showed high correlation coefficients ($0.65 < R^2 < 0.94$), and the computed statistics indicate that, although these are not the best-fitted models, they provide a good representation of the observed variability using a single variable. The coefficient estimated from the annual linear regression (10.40 µatm $°C^{-1}$; Table 2) showed a certain deviation from the theoretical rate of change for the area during 2019–2024 (16 µatm $°C^{-1}$), likely reflecting biological and physical effects (i.e., primary production, remineralization, and water mass mixing) during spring and summer, but remains consistent with values observed at ESTOC (Santana-Casiano et al., 2007).

488: where does this theoretical relationship come from? Also, this relationship is referred to as a rate of change, but there is no temporal component.

We have added the reference for the calculation of the effect of the thermal factor per degree of change.

Figure 4: I might be misunderstanding what I'm seeing, but it appears as though the ML method has found a way to cheat. The sharp discontinuities at the locations where data are available implies that the ML method has created local relationships specific to the times and locations of measurements intended to exactly reproduce the training/validation cruises without allowing those training data to overly affect the overall relationships. This, if I'm understanding correctly, is a strong demonstration of the hazards of not separating your training data from your validation data by transect/occupation. I reiterate that I might be misinterpreting what I'm seeing somehow.

The figure presents the monthly averaged $pCO_{2,sw}$ and pH predicted by the model while the experimental values included in each plot correspond with the observed value for the day/days of that month the ship visited the area that could change along the month. This has been now indicated in the legend for the figure.

527: it is odd to suggest that the thermal effect mitigates the expected effect from the temperature increase. I know what you mean, but many readers won't.

We thank the reviewer for this helpful clarification. We agree that the original phrasing could be misleading. The text has been revised to clarify that the thermal effect *partially offsets* the pH decrease driven by increasing CO_2, since temperature and CO_2 have opposite influence on pH. The revised text now reads: The pH decrease was partially offset by the thermal effect, which compensated for approximately 33% of the total decrease (the thermal contribution corresponds to about -0.06 pH units, associated with a temperature increase of 4.1°C). This compensating effect is evident near the African coast (Figure 5), where the upwelling of deep, cold, CO_2-rich seawater lowers both SST and pH, creating a marked longitudinal gradient across the Canary region.