# Review for "Modelling seawater $p$CO$_2$ 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 $p$CO$_2$ 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 CO$_2$ flux for the region, finding increasing $p$CO$_2$ 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 $p$CO$_2$. 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.

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.

L23: Is this determination, or more estimation of $p$CO$_2$ and pH?

L26: What is the R$^2$ for the pH$_{T,is}$, and what does this represent (pH$_{T,sw}$ is defined in L24, but not pH$_{T,is}$)?

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

L32-34: This sentence is awkwardly worded and a bit confusing.

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

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

### *Introduction*

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

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

L67: Although $f$CO$_2$ is basically equivalent to $p$CO$_2$, 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 $p$CO$_2$ is "expressed as CO$_2$ fugacity" omits this important distinction which some marine chemists may notice.

L82-85: How was this inherent error of satellite remote sensing considered in the analyses?

L86-96: Were these approaches based on satellite data?

L115: Define ESTOC upon first use.

L119: This should read "of the Canary Basin," not "on the Canary Basin".

### *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.

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.

L153: Does "measured" $p$CO$_2$ data refer to that from moored and VOS systems?

L155: Similarly, was "observed" (corrected) data the one measured from moored and VOS systems?

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

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

L178: Is $p$CO$_{2,sw}$ or $p$CO$_{2,eq}$ the partial pressure of CO$_2$ in seawater, as defined here?

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

L182-188: I worry that the observations and averaging used for these calculations were enough to accurately capture A$_T$, 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 analyses of discrete samples should be described.

L187: Is this pH on the total scale?

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 $xCO_{2,atm}$?

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?

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

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.

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

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

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

***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?

L310: What were the $AIC_c$ for each prediction model?

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

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

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 $pCO_2$ values.

L333-335: The accuracy indicators of training versus validation datasets can most likely be listed within the supplement; there is no need to list them here.

33L359-361: Once again, I am left wondering how training and validation datasets were divided.

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.

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

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

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

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

## *Discussion*

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

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

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

L501 (4.2): Once again, $pH_{T,is}$ isn't defined.

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

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

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

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

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

Conclusions

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

## *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.

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.

## *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?

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?

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

## *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.

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

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