

Review for the manuscript: "Exploring the CO₂ Fugacity along the East Coast of South America aboard the Schooner Tara" by Olivier et al.

Time devoted to this review: 12 hours.

Reviewer:

Marcos Fontela (IIM-CSIC)

This manuscript presents an innovative methodology and a timely study, leveraging the capabilities of the *Tara* schooner to monitor CO₂ fugacity (fCO₂) in a key region for the global carbon cycle. The work demonstrates the potential of such expeditions for regional carbon dynamics studies and contributes valuable observations to the limited dataset from this area. Moreover, this represents an excellent initiative to expand the scope of the *Tara* schooner's activities—traditionally centered on marine biology and ecology—toward geoscience observations, as exemplified by the dataset presented in this manuscript. The manuscript is well-written, well-structured, and includes high-quality figures. The article is worthy of publication and appropriate to support the publication of a data set, but the dataset itself exhibits significant shortcomings that must be first addressed. Important modifications are required to ensure the dataset meets the standards of openness, metadata completeness, and long-term usability. I recommend a major revision, primarily to address issues with the data product, as outlined below. Additionally, there are inconsistencies in the framing and presentation that detract from the overall clarity and impact of the work, but they should be easily ameliorated.

The current form of the data product significantly weakens the manuscript's suitability for publication in ESSD, a journal known for its rigorous standards in data quality and accessibility. The dataset lacks metadata, the data file itself it has an untraceable naming ("*CO2Tara.xlsx*" -sic-), fails to adhere to standardized variable names, and does not appropriately specify units. The dataset header lacks key information, such as details about the analytical methods used and the complete list of contributors. It is true that they are included in the Zenodo version, but the file should stand by itself. Additionally, quality control flags for variables should be assigned to enhance the dataset's reliability and usability. Metadata completeness is crucial for long-term usability and re-usability. The file format should be open source (rather than an Excel file) and it would be much better distributed as a NetCDF file as it is interoperable with other platforms. The use of proprietary or non-open-source formats is a critical limitation, so a conversion to a format compatible with community standards is necessary. Since the authors intend to submit this dataset to SOCAT, which I consider an excellent decision as it is the current reference database for fCO₂ measurements, it would be beneficial if the data product presented in ESSD could offer some added value or differentiation.

Specific comments (section and/or lines):

Reframe the title to reflect the true geographic scope, as the analysis is primarily focused on the Amazon River area rather than the entire east coast of South America.

Abstract 26-29: the order of the description should be homogeneous. For example, from river to ocean.

Abstract: North Brazil Current and Brazil Current are concepts not explained in the abstract. Not easy to follow.

Intro. 40. agree with *These regions present much higher temporal and spatial variability*. Therefore, this highly valuable snapshot only informs about spatial variability of a single season.

Intro 45. true sentence. The low number of observations in coastal waters is somewhat unexpected. Considering that coastal zones are, by definition, more accessible than the open ocean and offshore areas, it raises the question: why are these regions underrepresented in observational datasets? While addressing this issue is beyond the scope of this study, including a brief discussion or hypothesis to guide the reader would add valuable context.

Figure 1 is highly effective and well-designed, offering a clear overview of the study area. However, the inclusion of Pacific data skews the color bar and detracts from the manuscript's focus on equatorial South America in the Atlantic Ocean. It would be more appropriate to exclude the Pacific data and revise the figure to better align with the study's regional scope.

58. It represents one of the greatest environmental gradients on “the interface between” land and ocean in the world.

64 as the “Amazon” rainforest sequesters...

69. The references to the ANACONDAS (Mu et al., 2021) and Camadas Finas III (Araujo et al., 2017) campaigns lack both date signatures and spatial context. As a result, these terms may be largely unfamiliar to readers who are not specialized in regional studies, potentially hindering the manuscript's accessibility. Providing additional information about the campaigns' timing and geographical scope would help contextualize these references.

70. The manuscript mentions the link between these two systems, but it is unclear how this connection is made, as the schooner does not sail as far as Óbidos. If the link is indeed established, further clarification is needed, as it is not apparent from the current description. Could you please clarify whether the link is made, or revise the statement to reflect the actual coverage of the study?

71 I would delete “extensively”

75 in advance: The statement in paragraph 75 claiming that the Argo program could address the scarcity of fCO₂ measurements in the medium term is not accurate. Currently, this is technically unfeasible, as the Argo program is designed

primarily for interior ocean monitoring. To measure fCO₂, a surface-intensified approach would be required. The entire paragraph should be reassessed and revised to reflect the current limitations and the specific needs for fCO₂ measurements.

90. As this dataset is based on underway measurements, you use kilometers to express the magnitude of the data, which is a valid option. However, it would be helpful to include additional information regarding the timing of the observations. For example, how many days of data were collected? Furthermore, how many different biomes were crossed during the survey? Providing this contextual information would enhance the understanding of the dataset's temporal and spatial coverage.

108-114. The port-to-port description could be better represented in a table format. This information, while useful, does not add significant value to the narrative and would be more concise and accessible in a tabular form.

In line 103, you mention 14,000 km, which is a considerable distance. However, the schooner sailed a total of 70,000 km. This discrepancy raises the question: why are the remaining 56,000 km not included in the analysis? While there may be valid reasons for excluding these data, offering an explanation would strength the manuscript's transparency.

Figure 2 does not accurately represent the full circuit, as it omits two branches. Please revise the figure to include these missing branches.

However, the dataset lacks integration with existing efforts (e.g., SOCAT) and does not demonstrate sufficient added value over what is already available. It would strengthen the manuscript to articulate why this dataset is unique and necessary in the context of global fCO₂ monitoring.

Section 2.2: You mention an important flow rate, which is often a bottleneck in underway systems on unconventional vessels. Could you please provide the model and specifications of the pump used? This information would be useful for understanding the system's limitations. Does the pump include a filter? Additionally, what is the maximum speed of the schooner at which the pump remains functional? Is the schooner's speed included in the dataset? If not, it should be, as this could be an important variable to consider.

Line 132: The term "accurately" could be removed.

Line 138: It is unclear to me why you do not have this data. In line 140, you mention a temperature difference, are you not talking about that in line 138?

Line 150: The sentences need to be reordered for clarity. It would be more effective to first describe the atmospheric air, then the reference gases, and finally the seawater. This would improve the logical flow of the section.

Line 160: When stating that the system is "cleaned regularly," it would be helpful to include the periodicity of the cleaning process. Additionally, does the intake circuit feature any physical barriers or filters to prevent the introduction of large particles? This should be clarified for completeness.

Line 185: You mention "yellow" in Figure 3, but no yellow is visible (at least to me) in the figure. Please revise the description to match the actual content of the figure.

Figure 3: The straight line in the "Raw" data does not convey meaningful information and appears to be an artifact caused by the connection of data points in the time series. This should be corrected to ensure the figure accurately represents the data.

Figure 4: Please include a scatter subplot that illustrates the 1:1 relationship between the discrete samples and the underway measurements. Use the timing of the bottle closure from the CTD surface samples, applying a time window of 2-3 minutes for the underway. Also, include the uncertainty associated with the discrete samples (5.7 μatm) and the range of underway measurements. The current format of this Figure 4, which combines all values into one plot, makes it difficult to discern patterns due to the large range of values. It would be clearer if the figure were split into three subplots: one for high values, one for the central values around 400, and another for low values.

Figure 5: While I appreciated the broad context provided by Figure 1, I would recommend a closer zoom for Figure 5 to focus specifically on the underway track. The current version leaves a significant amount of blank space that could be better utilized. The figure should be revised to eliminate this excess space and highlight the relevant data more clearly. Additionally, the arrangement of the subplots is unclear, and the "pyramid" layout may not be the most effective. I suggest reconsidering the layout for better clarity and visual coherence before publication.

3. Overview section: Ensure that the variables are always presented in the same order: temperature, salinity, and then fCO_2 .

Figure 7: Please add the locations of Obidos, Belem, and Macapa to the figure. Additionally, label some isobaths of bathymetry for reference.

Figures 8 & 9: It would be more efficient to include the area shown in Table 1 directly in the figures, thereby eliminating the need for the Table 1 itself. In my opinion, this will enhance the visual presentation and reduce redundancy.

Bathymetry: The bathymetry data is sourced from ETOPO2v2 and is colocalized along the ship track. However, including bathymetry values at 0 m or depths of 1-2 m raises some questions, as this would imply the ship was at or near the seabed (which, of course, we hope did not occur!). It may be more appropriate to group these shallow measurements or explore alternative methods for representing this data in a way that better reflects the actual conditions. Another comment about

bathymetry: I would strongly recommend adding the bathymetry data to the dataset. This would enable others to fully reproduce the analyses presented in the manuscript.