

## ***Interactive comment on “A multi-decade record of high-quality fCO<sub>2</sub> data in version 3 of the Surface Ocean CO<sub>2</sub> Atlas (SOCAT)” by Dorothee C. E. Bakker et al.***

**Anonymous Referee #1**

Received and published: 21 June 2016

General comments:

The manuscript by Bakker et al. describes SOCATv3 and its improvements relative to version 2. Inclusion of new flags in this version is discussed in details, in addition to plans for future SOCAT automated system. Impact of SOCAT database for the wider scientific community highlights the critical role of this database.

The manuscript is useful for SOCAT users, providing a clear technical documentations as well as guidelines on how to acknowledge the data generators. As the authors state, the SOCAT data is an important contribution to the carbon cycle community from large number of scientists and technicians, whose continuing fundings depend on proper acknowledgements of their works.

C1

As an end-user of SOCAT who is not all familiar with the data collection and quality control processes involved, there are sections of the manuscripts that I found very informative. Nevertheless, note that I may not be qualified to provide extensive reviews on some sections such as evaluating the correctness or the technicality/criteria described in the quality control or flagging, etc. The paper is well structured and written and I fully support its publication in ESSD. Below, I have several (mostly minor) comments that, if addressed, could further improve the manuscript.

Specific comments:

On Page 5, the paper mentions other pCO<sub>2</sub> data set, namely the LDEO surface pCO<sub>2</sub>. I think it could be useful to elaborate what are the main differences between this and SOCAT database. Is there any plan to merge them together? If not, how one dataset complements the other?

It is clear from the manuscript that in the case of regional studies which significantly use SOCAT data in their analysis, the authors should invite data providers, especially those collecting the data into potential publications. It is less clear however in the case of global studies (e.g., large-ensemble model inter comparison studies involving model-data evaluation, where SOCAT and other physical and biogeochemical data sets and synthesis are used. Please clarify.

The development of SOCAT automation system is certainly appealing and a progressive feature of SOCAT quality control scheme, allowing for annual SOCAT releases, as the authors state. My question is whether it is justified to aim for ‘annual’ SOCAT releases. There are a lot of works going into documenting a data release, e.g., publication, public announcement, etc. But is it really necessary for this? A public release every 2-3 years seems to be sufficient and would allow more time for the data providers to get a first hand to study and analyze them. In addition, I am concern about having too many fragmentation (versions) of the SOCAT database without any clear improvements or distinguishable features (other than simply one more year of extra data) from

C2

one version to the others.

Section 6.3. Does it make sense or feasible to also include measurements of surface wind speed, which would allow for real-time estimate of CO<sub>2</sub> fluxes?

Section 8: Poor coverage in the Indian Ocean and Southern Ocean: Is there any plans or possibilities to address this in the future (e.g., integration with other pCO<sub>2</sub> measurements from other observing platforms not necessarily underway)?

P4, L23-29: For unfamiliar reader, I suggest adding a sentence or two on the mechanisms contributing to these observed large spatial and temporal variabilities. This comment also applies to the ensuing paragraph on pH variability.

P5, L21-23: I think this is better clarified towards the end of the manuscript, but I suggest adding few sentences specifically describing SOCAT's role for the Global Carbon Project.

P7, L16-20: please clarify if the increase in flagA/B data (41% to 49%) is simply due to the higher quality of the newly added data or as a result of improved quality control method applied to the earlier SOCAT version dataset.

P10, L14: It is not clear to me why datasets with less than 50 duplicate get a flag 4 (bad)?

Minor remarks:

Page 4, Line 21: 'surface' pH in the subtropical . . .

P7, L1: WOCE 'flag' of 3 . . .

P8, L26: 'a' good scientific practice.

P12, L28: 0.3C and 5 micro-atm: What are these criteria based on?

P15, L14-15: I am not familiar with the term 'property-property' plots. Can the authors give examples?

C3

Throughout the manuscript, some of the text sections are italicized. I assume this is just a technical error and needs to be fixed.

---

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-15, 2016.

C4