Earth System
Science Open Access
Data Discussions

Interactive comment on "An update to the Surface Ocean CO$_2$ Atlas (SOCAT version 2)" by D. C. E. Bakker et al.

Anonymous Referee #1

Received and published: 13 November 2013

The MS by Bakker et al. presents the updated version of the Surface Ocean CO2 Atlas (SOCAT version 2). SOCAT represents a very important effort to provide a comprehensive and consistent data set for the use in Earth System Sciences, particularly for research on the carbon cycle and ocean acidification. The SOCAT team made considerable efforts with regard to quality checks and the data quality is well documented in the downloadable data sets. I am intrigued by the high degree of organization that must be necessary to accomplish such a big project. The SOCAT data can be viewed or downloaded using different portals to which the user is redirected from the main webpage www.socat.info. This includes: 1) The cruise data viewer (http://ferret.pmel.noaa.gov/SOCAT2_Cruise_Viewer/), which offers a complex mask to query data, but is however still intuitive. 2) The gridded data viewer

(http://ferret.pmel.noaa.gov/SOCAT_gridded_viewer/), which allows the user to get a quick overview over the data by the help of an interactive web map system, time series plots, and plots of correlations between different variables. The user has the possibility to download gridded data for use in desktop applications like Matlab 3) Download complete data set from Pangaea (http://doi.pangaea.de/10.1594/PANGAEA.811776), consisting in tab delimited tables, one for each cruise. 4) Download of complete data set, or large parts of the data set by region, (http://cdiac.ornl.gov/ftp/oceans/SOCATv2/) in one big tab delimited table I am very satisfied with the data download and data quality, and have only some critical remarks to download option 1). The authors did also a great job in presenting SOCAT in their MS. The MS reads very well, is informative and comprehensible. The authors clearly described the evolution of the data set from the older version, the data sources, data coverage and quality, the structure of the data set, how to extract and use the data, and future plans. I have only minor comments on the presented MS and the downloadable data sets, and I recommend publication after minor revisions. In the following, I present first my comments on the MS and then on the data download and downloadable data sets.

Comments on the MS: Abstract: Page 469 Line 8: Replace "have been" by "are".

Introduction Page 470 Line 6: Replace "is" by "was". Line 27: Replace "by" by "for".

Page 473: Maybe you should mention already here in the introduction that SOCAT v2 contains 10.1 million surface water fCO2 values, which represents a substantial increase compared to the 6.3 million values in the older version.

Section 2 Page 475 Line 17: Insert a "for" after "available" Line 21: The historical cruises which are listed as taken at midnight, but maybe were not, are they flagged in a way the user can easily identify the concerned values?

Page 477, Line 11; Page 478, Line 19: Here, I wondered a bit what the data set quality control flags A-D might stand for. I can imagine that A is very good, and D is not so good, but still sufficient, while S and E finally mean that the values are of insufficient

quality. The definition of these quality control flags are not given until page 479. Maybe you should shortly explain the meaning of these flags (like A to D represents sufficient quality, with quality decreasing from A to D) and then state at the first mention of these flags that a definition of how these flags are attributed is given later. Maybe you could give the definitions of these flags as a small table, to provide the reader/user with a quick overview.

Page 481, Line 4: Replace "interrogation" by "querying".

Section 3 Page 483, Line 8: Replace "repeat" by "repeated".

Comments on data download and downloadable data sets: 1) Cruise data viewer (http://ferret.pmel.noaa.gov/SOCAT2_Cruise_Viewer/) If the user selects one variable of interest (like pCO2), she/he can download a nice ready to use table with latitude, longitude, date and time, and the values for the chosen parameter. However, when I tried this several times, the downloaded table listed all samples including those for which the chosen parameter was not assessed ('NAN'), although I did not check the respective option. That might be a bug. If you download 'all variables', you obtain a complex table with a lot of data columns. The variable column headings in the downloaded tables, but also the parameters you can chose for the map viewer (which correspond to the variable column headings), are not well in accordance with the descriptions in the MS (Table 6). With the help of the MS, I could figure out the meaning of most parameter names (column headings). However, a lot of variable names are not consistent between table 6 and the website or tables downloaded from the website. For instance 'fCO2water_equ_wet' in table 6 of the MS might be identical to 'fCO2_water_equi_uatm' in the downloaded table, but I'm not so sure. Similarly, 'pCO2water_SST_wet' in table 6, does this refer to 'pCO2_water_sst_100humidity_uatm' in the downloaded data table. The atmospheric pressure is called 'Pressure_atm' in the data set, but 'PPPP' in table 6. In the data set, I did not find the column 'Algorithm', which could have been interesting. I find some data columns in the data table like 'bottomD' and 'depth', for which I do not know what

they represent; and table 6 does not give any explanations on those. These are only a few examples. As the table 6 in the MS is however consistent with variable names from download options 3) and 4) (see below), maybe the variable names on this website should be renamed accordingly. Alternatively, you could explain each variable column name in the headers of the downloaded tables, like this is done for 3) and 4), or offer the definitions on your website.

2) The gridded data viewer (http://ferret.pmel.noaa.gov/SOCAT_gridded_viewer/) The viewer is well functioning and well documented with several tutorial videos.

3) Download complete data set as zip from Pangaea (http://doi.pangaea.de/10.1594/PANGAEA.811776) The download of the whole data set as zip (295 MB) is fast. Once unzipped, the user finds the data set organized in 2,699 tab-delimited text files, one for each cruise. Here, the column heading are consistent with the descriptions in table 6 in the MS. Further, each variable is shortly explained in the header of each downloadable table. Each table contains only variable columns for which there are variables, which makes them more clearly arranged. I am very satisfied with this download option.

4) Download of complete data set or by region (http://cdiac.ornl.gov/ftp/oceans/SOCATv2/) From here, more compact formats of the whole data set can be downloaded, either the whole package as one table or one table for each region. With regard to variable names (column headings in the table), the same is true as for 3): the names are consistent with those used in the MS (table 6) and the header contains also short descriptions of each column heading. Also the structure of the downloaded table (order of columns) is in harmony with table 6 in the MS. The tables are much more clearly arranged than those downloaded from 1). As the tab delimited files are very large, one should directly import them with a decent data base management system, which works very well. I am very satisfied with this download option.

C185