A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT)


¹Geophysical Institute, University of Bergen, Bergen, Norway
ESSDD
5, 735–780, 2012
Surface Ocean CO₂ Atlas (SOCAT)
B. Pfeil et al.

2 Bjerknes Centre for Climate Research, Bergen, Norway
3 PANGAEA Data Publisher for Earth & Environmental Science, University of Bremen, Bremen, Germany
4 Institute of Marine Research, Bergen, Norway
5 Uni Bjerknes Centre, Bergen, Norway
6 School of Environmental Sciences, University of East Anglia, Norwich, UK
7 Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington, USA
8 Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington, USA
9 Carbon Dioxide Information Analysis Center, Oak Ridge, Tennessee, USA
10 Jetz Laboratory, Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA
11 Université Pierre et Marie Curie, LOCEAN/IPSL, Paris, France
12 CSIRO Wealth from Oceans Flagship, Hobart, Tasmania, Australia
13 Centre for Australian Weather and Climate Research, Hobart, Tasmania, Australia
14 Uni Research AS, Bergen, Norway
15 University of Liège, Chemical Oceanography Unit, Institut de Physique, Liège, Belgium
16 British Antarctic Survey, Cambridge, UK
17 Department of Marine Sciences, University of Georgia, Athens, Georgia, USA
18 Monterey Bay Aquarium Research Institute, Moss Landing, California, USA
19 Institute of Marine Geology and Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan
20 School of Oceanography, University of Washington, Seattle, Washington, USA
21 Universidad de Las Palmas de Gran Canaria, Facultad de Ciencias del Mar, Las Palmas de Gran Canaria, Spain
22 Institut de Modélisation et d’Analyse en Géo-Environnement et Santé, Université de Perpignan, Perpignan, France
23 CSIRO, Marine and Atmospheric Research, Wembley, Western Australia, Australia
24 Intergovernmental Oceanographic Commission, UNESCO, Paris, France
25 Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany
now at: Norwegian Institute for Water Research, Bergen, Norway
** formerly at: Plymouth Marine Laboratory, Plymouth, UK
*** formerly at: International Ocean Carbon Coordination Project (IOCCP), Intergovernmental Oceanographic Commission of UNESCO, Paris, France

Received: 24 June 2012 – Accepted: 18 July 2012 – Published: 14 August 2012

Correspondence to: B. Pfeil (benjamin.pfeil@gfi.uib.no)

Published by Copernicus Publications.
Abstract

A well documented, publicly available, global data set of surface ocean carbon dioxide (CO$_2$) parameters has been called for by international groups for nearly two decades. The Surface Ocean CO$_2$ Atlas (SOCAT) project was initiated by the international marine carbon science community in 2007 with the aim of providing a comprehensive, publicly available, regularly updated, global data set of marine surface CO$_2$, which had been subject to quality control (QC). Many additional CO$_2$ data, not yet made public via the Carbon Dioxide Information Analysis Center (CDIAC), were retrieved from data originators, public websites and other data centres. All data were put in a uniform format following a strict protocol. Quality control was carried out according to clearly defined criteria. Regional specialists performed the quality control, using state-of-the-art web-based tools, specially developed for accomplishing this global team effort. SOCAT version 1.5 was made public in September 2011 and holds 6.3 million quality controlled surface CO$_2$ data points from the global oceans and coastal seas, spanning four decades (1968–2007). Three types of data products are available: individual cruise files, a merged complete data set and gridded products. With the rapid expansion of marine CO$_2$ data collection and the importance of quantifying net global oceanic CO$_2$ uptake and its changes, sustained data synthesis and data access are priorities.

Data coverage and parameter measured

Repository-Reference: doi:10.1594/PANGAEA.767698
Available at: www.socat.info/access.html
Coverage: 90° N to 90° S and 0–360°
Location Name: Global Ocean
Date/Time Start: 16 November 1968
Date/Time End: 31 December 2007
1 Motivation

The net absorption of CO$_2$ by the oceans, caused by rising atmospheric CO$_2$ concentrations since the industrial revolution, has been responsible for removing CO$_2$ equivalent to approximately 50% of the fossil fuel and cement manufacturing emissions or about 30% of the total anthropogenic emissions, including land use change (Sabine et al., 2004). Because of the availability of the carbonate ion, an important species of the dissolved inorganic carbon pool, and carbonate sediments, the oceans have a tremendous CO$_2$ uptake capacity and will, on timescales of ten to hundred thousand years, absorb all but a small fraction of the fossil CO$_2$ that has been and will be emitted (Archer et al., 1997). Meanwhile the changes in ocean CO$_2$ uptake, relying on factors such as ocean circulation and biology, will be among the decisive factors for the evolution of future atmospheric CO$_2$ concentrations and climate development (e.g., Friedlingstein et al., 2006; Riebesell et al., 2009).

Presently there are two types of globally coordinated efforts that seek to resolve the dynamics of ocean CO$_2$ uptake through observations: repeat hydrography and surface ocean CO$_2$ observations (Gruber et al., 2010; Sabine et al., 2010). While repeat hydrography aims to assess variations in the ocean inventory of CO$_2$ on decadal timescales, surface ocean observations may resolve variations on seasonal to interannual timescales due to the higher sampling frequency. This high sampling frequency has been made possible by the advent of autonomous instruments and sensors for the near-continuous determination of surface water CO$_2$, which may be installed on commercial sea going vessels giving an observational repeat rate of a few weeks, depending on ship schedule (Cooper et al., 1998; Pierrot et al., 2009), or on moorings (Merlivat and Brault, 1995; DeGrandpré et al., 2000; Friederich et al., 2008; Wada et al., 2011). Moorings or drifting platforms provide observations on sub-diurnal time scales (e.g., Körtzinger et al., 2008; Leinweber et al., 2009; Merlivat et al., 2009; Parard et al., 2010), while underway observations increase spatial coverage.
These technological developments have led to a rapid increase in new surface ocean CO$_2$ data being collected each year. This is reflected in the number of data underlying the successive surface ocean $p$CO$_2$ (partial pressure of CO$_2$) climatologies of Takahashi et al. (1997, 2002, 2009a, b, 2011), increasing from 0.25 million for the 1997 edition to 5.2 million in 2011. Presently over a million observations are being made each year (Sabine et al., 2010). In order to deal with these data effectively and to maximise their scientific use, the international ocean carbon research community initiated the Surface Ocean CO$_2$ Atlas (SOCAT) project in 2007 (IOCCP, 2007). The aims of SOCAT were threefold. Firstly, SOCAT aimed to merge all available surface ocean CO$_2$ data into one uniformly formatted, quality controlled, publicly available database with regular updates. The second aim of SOCAT was to secure the long-term storage of each data set together with its required documentation (metadata). Finally, the community sought to realise a transparent and traceable approach for the handling, quality control and integration of surface ocean CO$_2$ data, which may be managed by the community on a routine basis in the future.

The first version of SOCAT (version 1.5) was made public on 14 September 2011 during “The ocean carbon cycle at a time of change: Synthesis and Vulnerabilities” meeting at the UNESCO (United Nations Educational, Scientific and Cultural Organization), Paris (Bakker et al., 2012). This SOCAT version compromises 6.3 million surface water CO$_2$ data from 1851 voyages from 1968 to 2007 covering the global oceans and coastal seas (Figs. 1 and 2). Three data products are available: (1) cruise data files of quality controlled surface water $f$CO$_2$ (fugacity of CO$_2$, similar to partial pressure) data and including the reported CO$_2$ values as reported by the investigator, (2) globally and regionally aggregated files of these $f$CO$_2$ data, and (3) a collection of gridded products providing averaged $f$CO$_2$ with minimal interpolation (Sabine et al., 2012). This article describes the history of SOCAT (Sect. 2), the procedures adopted in SOCAT for retrieving data (Sect. 3), for formatting (Sect. 3) and quality controlling these data (Sect. 4). The article introduces SOCAT data products and where they can be accessed (Sect. 5). An accompanying article (Sabine et al., 2012) describes the gridding
procedures. The SOCAT website (www.socat.info) provides documentation on SOCAT, as well as links to sites with SOCAT data products. This article concludes with lessons learned from this first SOCAT version and recommendations for future SOCAT releases (Sect. 6).

2 History and organisation of SOCAT

2.1 History of SOCAT

In the late 1990s attempts were made by the SCOR-IOC (Scientific Committee on Oceanic Research – Intergovernmental Oceanographic Commission) committee on ocean CO₂, the fore-runner of the IOCCP (International Ocean Carbon Coordination Project), to assemble a comprehensive, well documented, publicly available data set of surface ocean $f$CO₂ for the global oceans and coastal seas. Efforts for encouraging data submission to a central location, the Carbon Dioxide Information Analysis Center, were partly successful. In 2004 the marine carbon community agreed on recommendations for the reporting of surface water CO₂ data and metadata (IOCCP, 2004). However, most data gatherers did not strictly follow these. Only a subset of all global surface water CO₂ data were made publicly available via CDIAC, with many data only available via the investigators, institute websites and national or world data centres.

Over the past decades several attempts have been made to establish a global surface ocean CO₂ database. In the late 1990s, Taro Takahashi from Lamont-Doherty Earth Observatory (LDEO) compiled an initial data set and updated this collection in 2002 and every year from 2007 onwards (Takahashi et al., 1997, 2002, 2009a, 2011). The primary reason for this effort was the creation of global climatologies of air-sea CO₂ fluxes (Takahashi et al., 1997, 2002, 2009b). This LDEO database was made public in 2007 and is currently being updated on an annual basis. The data treatment is based upon Takahashi’s long experience. The LDEO database includes $p$CO₂ from discrete and continuous measurements. The most recent version of the LDEO data set
has 5.2 million $p$CO$_2$ data from the global oceans and coastal seas from 1957 to 2010 (Takahashi et al., 2011).

In 2001, Bakker began to assemble a surface ocean CO$_2$ data set by putting public data from CDIAC into a uniform format, as part of the European Union (EU) project OR-FOIS (Origin and fate of biogenic particle Fluxes in the Ocean and their Interaction with the atmospheric CO$_2$ concentration as well as the marine Sediment). Pfeil and Olsen streamlined and expanded this effort within the EU project CarboOcean from 2005 onwards. They compiled public surface ocean CO$_2$ data held at CDIAC, PANGAEA – Data Publisher for Earth & Environmental Science (an International Council for Science (ICSU) World Data Center, formerly the World Data Center for Marine Environmental Sciences, WDC-MARE) and elsewhere into a common format $f$CO$_2$ database based on the recommended formats for data and metadata reporting (IOCCP, 2004).

The Surface Ocean CO$_2$ Atlas was initiated at the Surface Ocean CO$_2$ Variability and Vulnerability (SOCOVV) meeting by the international ocean carbon research community (Table 1) (IOCCP, 2007). The SOCAT project agrees well with the objectives of the joint Carbon Implementation Plan of the Surface Ocean Lower Atmosphere Study (SOLAS) and Integrated Marine Biogeochemistry and Ecosystem Research (IMBER) (IMBER, 2005). SOCAT was given the specific objectives of developing two data products (IOCCP, 2007):

- A quality controlled $f$CO$_2$ data set made publicly available on a regular basis following agreed procedures and regional review;

- A gridded product consisting of monthly surface $f$CO$_2$ means (including number of data points and standard deviation) on a 1° latitude by 1° longitude grid with no interpolation.

A gridded surface ocean $f$CO$_2$ product was deemed to be more useful than air-sea CO$_2$ flux estimates for modelling and other purposes (IOCCP, 2007). Regional groups and a global group for coordination were formed (Table 2). A series of meetings was held in
which SOCAT gradually took shape and in which the regional groups coordinated their work (Table 1) (IOCCP, 2007, 2008, 2009a, b, 2010a, b).

The SOCAT community evaluated existing data compilations and selected the data collection by Pfeil and Olsen as the basis for SOCAT (IOCCP, 2008). The focus for SOCAT has been the assembly of publicly available data (including metadata), standardisation of the file formats, recalculation of consistent and uniform surface water \( f\text{CO}_2 \) data, and basic and secondary level quality control (Sects. 3, 4 and 5).

SOCAT is independent from the LDEO database (Takahashi et al., 2011), but has a large overlap in its original data. SOCAT only includes surface water \( \text{CO}_2 \) values, measured in near-continuous operation or in discrete samples with an equilibrator system or a spectrophotometer and reported as \( x\text{CO}_2 \) (\( \text{CO}_2 \) mixing ratio), \( p\text{CO}_2 \) or \( f\text{CO}_2 \) (Sect. 3). SOCAT does not include \( f\text{CO}_2 \) recalculated from dissolved inorganic carbon, alkalinity or pH.

### 2.2 SOCAT groups

Roughly 45 international, sea-going marine carbon scientists and data managers, from 12 countries actively participated in the assembly and quality control of SOCAT version 1.5. These participants were organised in regional groups and a global group (Table 2). The regional groups were responsible for quality control of the data in their region. Regional groups were formed for the coastal seas (north of 30\(^\circ\) S), the North Atlantic Ocean (north of 30\(^\circ\) N, including the Atlantic Arctic Ocean), the Tropical Atlantic Ocean (30\(^\circ\) N to 30\(^\circ\) S), the North Pacific Ocean (north of 30\(^\circ\) N, including the Pacific Arctic Ocean), the Tropical Pacific Ocean (30\(^\circ\) N to 30\(^\circ\) S), the Indian Ocean (north of 30\(^\circ\) S) and the Southern Ocean (south of 30\(^\circ\) S, including coastal waters). Coastal regions were initially defined by bathymetry (shallower than 200 m) for regions north of 30\(^\circ\) S (IOCCP, 2008). This definition was later replaced by a criterion of distance from a major land mass (less than 400 km) in order to better reflect the environmental significance of these regions as continental margins. Figure 3 shows these oceanic and coastal regions in SOCAT.
SOCAT has been a large, complex undertaking and has involved activities focused on: data retrieval, assembling data in a uniform format, recalculating surface water \( f\text{CO}_2 \) using the same agreed-upon protocol, defining SOCAT QC criteria, developing the QC cookbook and Matlab QC code, making SOCAT available via the Live Access Server (LAS) for QC and public release, data QC, gridding SOCAT, making SOCAT documentation and products available via the web, designing the SOCAT logo, internal communication, organisation of SOCAT meetings, and liaising with the international marine carbon community. Numerous colleagues have played a role in these activities (Table 3). The SOCAT global group initially had five members and has gradually been expanded to reflect the increasing complexity of the tools and products in SOCAT (Table 2).

3  SOCAT data assembly

3.1  Data sources and instrumentation

SOCAT includes 6.3 million \( f\text{CO}_2 \) data points measured in all ocean areas from 1968 to 2007. Most of these data were gathered from the online sources at CDIAC (30 % of the cruises) and PANGAEA® (10 %), as well as from institute and project websites (37 %). The remaining cruises (23 %) were obtained directly from the data originators. Almost half of the cruises (45.7 %) originated in the USA. Other significant contributors are based in Japan (20.1 %), Norway (9.6 %), the United Kingdom (7.4 %), Germany (5.8 %), France (4.5 %), Belgium (2.4 %), Canada (1.6 %), Spain (1.5 %), Australia (1.2 %) and the Netherlands (0.3 %).

The data in SOCAT are a synthesis of 4 decades of seagoing fieldwork by a numerous of scientists from 12 countries. Various instruments have been used to obtain these data and only the basic principles will be summarised here. Further information is available in the metadata, which accompany individual cruise files at PANGAEA® (doi:10.1594/PANGAEA.769638) (Sect. 5.2).
The seawater $f$CO$_2$ values included in SOCAT have been measured according to one of the following two principles: (1) analysis of the CO$_2$ content in an air sample in equilibrium with a large volume of seawater or (2) calculation of the seawater $f$CO$_2$ from the colour response of an acid-base indicator dye (sulfonephtalein) in contact with seawater across a CO$_2$ permeable membrane. The analysis of the CO$_2$ content in an air sample in equilibrium with a large volume of seawater is recommended in the standard work by Dickson et al. (2007). The CO$_2$ concentration in the air sample is determined through either gas chromatography (Weiss et al., 1981) or infrared analysis (Takahashi, 1961). The equilibration of air and water can be carried out in sample bottles (Chipman et al., 1993; Wanninkhof and Thoning, 1993) or in an equilibrator in a flow-through system (Takahashi, 1961; Wanninkhof and Thoning, 1993; Cooper et al., 1998; Pierrot et al., 2009). Flow-through systems combined with a non-dispersive infrared (IR) detector are by far the most common type in operation. Flow-through systems are routinely deployed on commercial vessels (e.g., Cooper et al., 1998; Olsen et al., 2008; Watson et al., 2009), research vessels (e.g., Lefèvre et al., 1994; Skjelvan et al., 1999; Bakker et al., 2008), and on moored platforms (e.g., Friederich et al., 2008; Wada et al., 2011). Intercomparison experiments have taken place on a number of occasions (e.g., Körtzinger et al., 1996, 2000).

The indicator-based, spectrophotometric determination, of $f$CO$_2$, has been developed for moored and drifting platforms (Lefèvre et al., 1993). Prominent examples of these are the CARIOMA (Carbon Interface Ocean Atmosphere) buoy (Merlivat and Brault, 1995) and the SAMI (Submersible Autonomous Moored Instrument) $p$CO$_2$ instrument (DeGrandpré et al., 2000). These instruments have been deployed in many ocean regions (e.g., Hood et al., 1999; Bakker et al. 2001; Körtzinger et al., 2008; Lefèvre et al., 2008).

### 3.2 Data harmonisation and basic quality control

All data files available for SOCAT were first converted to a common file structure. This also included discarding data not directly relevant for surface ocean CO$_2$, e.g.,
meteorological parameters like wind speed and direction, whenever these were supplied in the file. Next, the unit of each parameter was checked and converted into the agreed standard unit, if required (e.g., conversion of atmospheric pressure from atmospheres to hPa, and of latitude and longitude to decimal degrees). For around 10% of the cruises, different versions of the data had been obtained from various sources. In these cases only the most recent version was included in SOCAT in consultation with the data originator.

Basic, primary quality control was carried out at this stage. Outliers and unrealistic values in date, time, position, intake temperature, salinity, atmospheric pressure and surface water CO$_2$ were identified. The criteria were that ship speeds calculated from position should be realistic, that atmospheric pressures should be within 800 hPa and 1100 hPa and that the dates should exist. Rapid changes in intake or equilibrator temperature of several degrees, in salinity of several units or in surface fCO$_2$ of several hundreds of micro-atmospheres were also questioned, except for data in coastal or ice-covered regions. Whenever several such data points were encountered, the data originator was contacted and this often resulted in resubmission of an updated (corrected) version. In some cases several iterations were required, making this a time consuming task. In a few cases interaction with the data originator was not possible, and obviously bad data were removed from the data file.

Further basic QC errors were detected by the regional SOCAT groups (Sect. 4). Data files with numerous (50 or more) outliers or unrealistic data were suspended from SOCAT until a revised file had been submitted. In cases where relatively few bad values (less than 50) were identified, the corresponding recommended fCO$_2$ was assigned the World Ocean Circulation Experiment (WOCE) flag 3 (questionable) or 4 (bad) (see Sect. 4.2.3 for definitions).

### 3.3 fCO$_2$ (re-)calculations

The final stage of the SOCAT data assembly was the (re-)calculation of fCO$_2$ values at sea surface (or intake) temperature in order to ensure a uniform representation of
CO₂ concentration. The conversions from xCO₂ and pCO₂ were carried out using a single set of equations with a clear hierarchy for the preferred CO₂ input parameter (Table 4) (Pfeil and Olsen, 2009). We used the equations recommended by Dickson et al. (2007):

\[ p\text{CO}_2 = x\text{CO}_2^{\text{dry}} (P_{\text{equ}} - \text{pH}_2\text{O}) \] (1)

for the conversion of dry CO₂ mole fraction to partial pressure at 100 % humidity, where \( P_{\text{equ}} \) is the pressure in the equilibrator. The water vapour pressure \( \text{pH}_2\text{O} \) is calculated as:

\[ \text{pH}_2\text{O} = \exp \left( 24.4543 - 67.4509 \left( \frac{100}{T_{\text{equ}}^{\text{K}}} \right) - 4.8489 \ln \left( \frac{T_{\text{equ}}^{\text{K}}}{100} \right) - 0.000544S \right) \] (2)

where \( T_{\text{equ}}^{\text{K}} \) is the measurement (or equilibrator) temperature in Kelvin and \( S \) is sample salinity. For the conversion of \( p\text{CO}_2 \) values into \( f\text{CO}_2 \) the equation is:

\[ f\text{CO}_2 = p\text{CO}_2 \times \exp \left\{ \frac{B(CO_2, T_{\text{equ}}^{\text{K}}) + 2 \left( 1 - x_{\text{wet}}^{\text{equ}} \right)^2 \delta(CO_2, T_{\text{equ}}^{\text{K}})}{R \times T_{\text{equ}}^{\text{K}}} P_{\text{equ}} \right\} \] (3)

The virial coefficients for CO₂, \( B(CO_2, T_{\text{equ}}^{\text{K}}) \) and \( \delta(CO_2, T_{\text{equ}}^{\text{K}}) \) (cm³ mol⁻¹), are given by:

\[ B(CO_2, T_{\text{equ}}^{\text{K}}) = -1636.75 + 12.04087 T_{\text{equ}}^{\text{K}} - 3.2795710^{-2} T_{\text{equ}}^{2\text{K}} + 3.16528 \times 10^{-5} T_{\text{equ}}^{3\text{K}} \] (4)

\[ \delta(CO_2, T_{\text{equ}}^{\text{K}}) = 57.7 - 0.18 T_{\text{equ}}^{\text{K}} \] (5)

Whenever conversion of the measurement (equilibrator) temperature (\( T_{\text{equ}} \)) to the sea surface temperature (SST) was required, we used the equation of Takahashi et
al. (1993) with both temperatures in the same unit:

$$f_{\text{CO}_2}^{\text{SST}} = f_{\text{CO}_2}^{\text{equT}} \exp(0.0423(SST - T_{\text{equT}}))$$  \hspace{1cm} (6)

The Takahashi et al. (1993) temperature correction is preferred, as it does not require knowledge of the alkalinity and dissolved inorganic carbon content of the water and was determined for isochemical conditions, while other temperature corrections (Gordon and Jones, 1973; Weiss et al., 1982; Copin-Montégut, 1988, 1989; Goyet et al., 1993) were not.

Altogether 6 different surface ocean CO$_2$ parameters were reported by the data originators, notably $x_{\text{CO}_2}$, $p_{\text{CO}_2}$ and $f_{\text{CO}_2}$ either at sea surface (or intake) temperature or at equilibrator (or measurement) temperature (Table 4). The (re-)calculations of $f_{\text{CO}_2}$ at sea surface temperature were implemented following these strict guidelines:

1. Whenever possible, (re-)calculate $f_{\text{CO}_2}$,

2. The preferred starting point for the calculations is $x_{\text{CO}_2}$, next $p_{\text{CO}_2}$, and finally $f_{\text{CO}_2}$,

3. Minimize the use of external data required to complete the calculations.

Thus, $f_{\text{CO}_2}$ was recalculated, if $x_{\text{CO}_2}$, $p_{\text{CO}_2}$, and $f_{\text{CO}_2}$, as well as all parameters required to calculate $f_{\text{CO}_2}$ were available in the file. However, $f_{\text{CO}_2}$ was not recalculated, if $f_{\text{CO}_2}$ was reported, but pressure or salinity were not, as Eqs. (1), (2) and (3) could not be applied without resorting to external data. If only surface water $f_{\text{CO}_2}$ at sea surface temperature was provided (as is the case for CARIOCA data and other spectrophotometric measurements), no recalculation was carried out. If $f_{\text{CO}_2}$ was not provided, $f_{\text{CO}_2}$ was always calculated, even if use of external data was necessary. Table 4 lists the parameters that went into the $f_{\text{CO}_2}$ calculations and the preference (or hierarchy) of the different calculation methods. The $f_{\text{CO}_2}$ values, which have been (re-)calculated following the preferred method (lowest index number in Table 4), are
reported as the recommended \( f\text{CO}_2 \ (f\text{CO}_2\text{-rec}) \) values in each SOCAT output file (Table 5). The calculation method is indicated (as \( f\text{CO}_2\text{-source} \)) in the regional and global synthesis files of SOCAT version 1.5 (Table 5).

Two external parameters were used for the recalculations of \( f\text{CO}_2 \), when necessary: climatological monthly mean salinity was obtained from the World Ocean Atlas (WOA) 2005 (Antonov et al., 2006). Sea level pressure (SLP) was acquired from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) project (Kalnay et al., 1996), provided on a 6 hourly, global, 2.5° latitude by 2.5° longitude grid. Whenever NCEP/NCAR SLP or reported atmospheric pressure was used in the calculations (as opposed to equilibrator pressure), 3 hPa were added to account for the slight overpressure normally maintained in ships (Takahashi et al., 2009b). Surface water CO\(_2\) data without accompanying SST were suspended from SOCAT, as \( f\text{CO}_2 \) is highly sensitive to temperature fluctuations.

### 3.4 Naming convention

Each cruise was assigned a unique cruise identifier, an Expocode (Swift, 2008), to remove the ambiguities of the commonly used informal cruise names and to identify duplicate versions of data. The first two characters of a twelve-character Expocode identify the country code of the vessel and are followed by the two-character National Oceanographic Data Center (NODC) vessel code. The final eight characters denote the starting date of the measurements of the cruise (as YYYYMMDD). For instance, 06MT19920510 means that this cruise was conducted on the German (06) research vessel Meteor (MT) and that the first measurement was reported for 10 May 1992. Both the Expocode and the original cruise name are provided in all SOCAT output files (Sect. 5), such that cruises can be retrieved using the Expocode as well as the vessel specific or investigator specific naming convention (M21/3 for the above example). The Expocode has not been used for buoys, since no NODC vessel code is available for these.
4 SOCAT secondary quality control

4.1 SOCAT secondary quality control procedures

An important aim of the SOCAT team has been to establish community agreed secondary quality control procedures. Criteria for secondary QC have been discussed at SOCAT workshops (IOCCP, 2008, 2009b) and have been formalized in the SOCAT cookbook (Table 6) (Olsen and Metzl, 2009). Two types of flags are defined to describe the quality of the $fCO_2$ data: cruise flags (Sects. 4.2.1 and 4.2.2) and WOCE flags for individual $fCO_2$ values (Sect. 4.2.3). The cruise flags provide information on the expected quality of the $fCO_2$ data in a cruise and are based on the methods followed, as well as on the reporting of metadata (Table 6). Only cruises with flags A, B, C or D are included in the SOCAT products (Sect. 5).

A bias assessment and correction, as done for GLODAP (Global Data Analysis Project), CARINA (CARbon dioxide IN the Atlantic Ocean) and PACIFICA (PACific Interior CArbon) (Key et al., 2004; Tanhua et al., 2010; Ishii et al., 2012), is impossible for surface ocean CO$_2$ measurements due the complex processes affecting surface ocean $fCO_2$. For example, even repeat visits to a site (a “cross-over”) within a few days or in a distance of only a few kilometres can have very different $fCO_2$ due to biological activity.

4.2 Cruise flags and WOCE flags

4.2.1 Analytical methods in the cruise flag

Approved methods or standard operating procedures (SOP) are required for cruises to be given a flag A or B, in which recommended $fCO_2$ has an accuracy of 2 µatm or better (Olsen and Metzl, 2009). These criteria follow the recommendations of a 2002 workshop on underway $pCO_2$ systems (Atlantic Oceanographic and Meteorological
Laboratory, 2002). Seven SOP criteria need to be fulfilled for a cruise flag A or B in SOCAT:

1. The data are based on \( x\text{CO}_2 \) analysis, not \( f\text{CO}_2 \) calculated from other carbon parameters, such as pH, alkalinity or dissolved inorganic carbon;

2. Continuous \( \text{CO}_2 \) measurements have been made, not discrete \( \text{CO}_2 \) measurements;

3. The detection is based on an equilibrator system and is measured by infrared analysis or gas chromatography;

4. The calibration has included at least 2 non-zero gas standards, traceable to World Meteorological Organisation (WMO) standards;

5. The equilibrator temperature has been measured to within 0.05 °C;

6. The intake seawater temperature has been measured to within 0.05 °C;

7. The equilibrator pressure has been measured to within 0.5 hPa.

An accuracy of 0.5 hPa for the equilibrator pressure is sufficient for achieving an accuracy of 2 \( \mu \text{atm} \) for seawater \( f\text{CO}_2 \). On installations where only the outside atmospheric pressure is recorded and used for the seawater \( f\text{CO}_2 \) calculation, the required accuracy of 0.5 hPa is never met, because of the overpressure that is normally maintained within ships. In the SOCAT recalculation of \( f\text{CO}_2 \) this has been taken into account by adding 3 hPa (a correction proposed by Takahashi et al., 2009b) (Sect. 3.3), whenever atmospheric pressure, but not equilibrator pressure, was reported in the data files.

A cruise flag A was given to high quality data for which a comparison with other data was available and deemed acceptable (Table 6). Surface water \( \text{CO}_2 \) properties may vary rapidly over time. A measure of cross-over proximity was defined that combined separation in space and separation in time into a single value. The algorithm that was
used treated 1 day of separation in time as equivalent (heuristically) to 30 km of separation in space — i.e., if \( dx \) is the distance between points from two cruises in km, and \( dt \) is the separation between the same two points in days, then the separation between these two points would be given as \( \sqrt{dx^2 + (dt/30)^2} \). The cross-over distance separating two cruises, \( dc \), is the smallest value found comparing all pairs of points between two cruises. If a cross-over distance between two cruises was zero, a cruise had been erroneously duplicated, and the oldest version of the cruise data was suspended in consultation with the data originator. Where the cross-over distance was relatively small, meaningful QC insights were likely to be found by comparing observations from the two cruises. The LAS offered QC operators a quick means to compare cruise pairs with a small cross-over distance between them. No strict criteria were defined for judging the quality and significance of cross-overs.

### 4.2.2 Metadata in the cruise flag

Complete metadata are required for cruise flags A, B and C (Table 6). This information must appear either in the metadata themselves or in a publication cited in the metadata. Complete metadata should provide the following information (Olsen and Metzl, 2009):

- The investigator;
- The vessel;
- The temporal coverage;
- The analytical method;
- The type of reported \( \text{CO}_2 \) data (\( x\text{CO}_2 \), \( p\text{CO}_2 \), \( f\text{CO}_2 \));
- The number of \( \text{CO}_2 \) standards used with their approximate \( \text{CO}_2 \) mixing ratio and traceability;
- A list of sensors and their accuracy, notably for:
– The equilibrator and seawater intake temperature,
– The equilibrator pressure;

Salinity does not need to be highly accurate for meeting the 2 µatm criterion, as the sensitivity of the $x_{CO_2}$ to $f_{CO_2}$ calculation to salinity is small (for example an $x_{CO_2}$ of 360 µmol mol$^{-1}$ at 20°C and 1 atm yields $f_{CO_2}$ of 347.22 µatm and 347.24 µatm at salinity 30 and 35, respectively).

### 4.2.3 WOCE flags for individual $f_{CO_2}$ values

WOCE flags 2 (good), 3 (questionable) or 4 (bad) were assigned to individual recommended $f_{CO_2}$ values (Table 5) (Olsen and Metzl, 2009). While WOCE flags refer only to the values of recommended $f_{CO_2}$, the underlying cause for a flag of 3 or 4 was often in the data required to recalculate $f_{CO_2}$. Examples include outliers in the intake or equilibrator temperatures, in the warming or in salinity and unrealistic positions or time stamps. As a default, all $f_{CO_2}$ values were given a flag 2. The WOCE flags allow the inclusion of cruises with a small number (maximum of 50) of questionable or bad $f_{CO_2}$ data in SOCAT. Revision of cruise data and communication with investigators are time consuming and not always possible for historic cruises. The complete set of $f_{CO_2}$ data with WOCE flags 2, 3 and 4 can be accessed via PANGAEA (Sect. 5.2) and the interactive Cruise Data Viewer (Sect. 5.4), while the global and regional files (Sect. 5.3) only have $f_{CO_2}$ data with flag 2.

### 4.3 Secondary quality control in practice

#### 4.3.1 Secondary quality control by the regional groups

The regional groups had the responsibility for secondary QC of all cruises crossing their region. Regional SOCAT QC operators carried out secondary quality control and assigned flags to each cruise during the QC process upon evaluation of the data and metadata. The recommended $f_{CO_2}$ and supporting data were made available via the
Live Access Server during quality control. Data were evaluated according to the procedures in the SOCAT cookbook (Olsen and Metzl, 2009). The QC was carried out in a variety of ways, either online via the LAS (Sect. 4.3.2) or offline (Sect. 4.3.3).

### 4.3.2 Live Access Server for quality control

The Live Access Server is a web server designed by NOAA PMEL (National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory) to provide access to geo-referenced scientific data (http://ferret.pmel.noaa.gov/LAS). Cruise data and metadata were ingested into a relational database and made available to the regional teams for evaluation through a version of the LAS, which had been enhanced with SOCAT quality control tools. Contents of the database included recommended $fCO_2$ values, ancillary parameters, cruise metadata, and reference variables drawn from other sources (Sect. 5.4). The LAS enabled QC operators to query the data collection using criteria of region, time period, seasonality, cruise and ship identifiers, and ranges of data values. The scientists could select data from one or more cruises, evaluate the data within the LAS and/or download subsets as compressed files for offline QC. The LAS offered QC evaluation tools, such as interactive property-property plots and co-inspection of cruises identified by the cross-over analysis (Sect. 4.2.1). The LAS provided access to the cruise metadata, which had to be evaluated as part of the QC. It also allowed uploading of ancillary documentation about the cruises and QC findings. The QC operators entered cruise flags and WOCE flags with comments explaining the rationale for their evaluations on the LAS during quality control. The flags and comments are available via the Cruise Data Viewer (Sect. 5.5). The system alerted QC operators, when conflicting QC evaluations had been entered, allowing SOCAT scientists to evaluate and resolve these conflicts.
4.3.3 Offline quality control

A set of Matlab® routines for data evaluation was available for offline QC (Olsen and Pierrot, 2010). These routines create a series of property-property plots, enabling QC operators to compare data from cruises in the same region. The $f_{CO_2}$ is plotted and colour coded according to the input parameter used ($x_{CO_2}$, $p_{CO_2}$, $f_{CO_2}$) in the (re-)calculation of recommended $f_{CO_2}$ (Sect. 3.3). Examples include a figure comparing the $f_{CO_2}$ versus sea surface temperature of a particular cruise to that for other cruises in the region. A second plot compares the monthly average and spread of the data in a box plot.

4.3.4 Suspended cruises and conflicting cruise flags

During the primary and secondary quality control, many cruises were suspended from SOCAT (cruise flag “S” in Table 6), as minor and major flaws in the CO$_2$ data or in the data necessary for the (re-)calculation of $f_{CO_2}$ became apparent. Data contributors were informed of these suspensions and were invited to resubmit their data upon making relevant corrections to the original data. In many cases data were resubmitted to SOCAT. If revised data were made available before the SOCAT quality control had been completed and were deemed of good quality, the data were included in version 1.5. Other resubmitted data will be included in the quality control for future SOCAT versions.

Most cruises cross multiple regions, e.g., the coastal region and the North Atlantic Ocean. In SOCAT QC, a cruise needed to receive a cruise flag for each region that it crosses. A final check in the quality control consisted of checking conflicting cruise flags (Bakker). Most “conflicting” cruise flags reflected the absence of quality control in one region. These conflicts were resolved by carrying out appropriate QC and entering the missing cruise flags. Few truly conflicting cruise flags were encountered and in all cases a satisfactory solution was found.
5 SOCAT products and tools

5.1 SOCAT cruises, versions, and time stamps

SOCAT data are publicly available via the SOCAT website (www.socat.info) as individual cruise data files (SOCAT version 1.4) (Sect. 5.2) and as regional and global, concatenated files (SOCAT version 1.5) (Sect. 5.3). SOCAT versions 1.4 and 1.5 include all cruises with a cruise flag A, B, C or D. A table of these cruises is available at doi:10.1594/PANGAEA.769638 and provides information about the investigator, research vessel, Expocode, original cruise naming, metadata (as reported by the investigator), and temporal and geographical coverage. Through PANGAEA® SOCAT is fed into the ICSU World Data System (WDS). The Global Earth Observation System of Systems (GEOSS), which is being built by the Group on Earth Observations (GEO), makes SOCAT available to other research communities.

The individual cruise data files (version 1.4) record observation time stamps at a resolution of integer minutes, rounding off the seconds, when they were available. Some cruises have multiple recommended $f$CO$_2$ values for a given time stamp (around 5% of the observations). Individual cruise data files (Sect. 5.2) contain all recommended $f$CO$_2$ data, including multiple values per minute. However, handling multiple entries for the same time stamp can be problematic for some software programs. The SOCAT global group decided to average multiple entries within a given minute for the regional and global synthesis files (Sect. 5.3) as a pragmatic solution to this issue.

Table 5 lists the contents of the SOCAT files in versions 1.4 and 1.5. Matlab code by Pierrot and Landschützer for reading these files is available via the SOCAT website or directly at CDIAC (http://cdiac.ornl.gov/ftp/oceans/SOCATv1.5/).

5.2 Individual cruise data files (version 1.4)

Individual cruise data files (version 1.4) with cruise flags A, B, C and D are available via PANGAEA® (doi:10.1594/PANGAEA.767698). These cruise data files include all
recommended $f_{\text{CO}_2}$ data with WOCE flags 2 (good), 3 (questionable) and 4 (bad), without listing these WOCE flags. Cruise data files archived at PANGAEA® have not been averaged to remove multiple entries per minute (Sect. 5.1).

The individual cruise data files provide access to the metadata, the original CO$_2$ parameter(s) (as reported by the investigator), which were used to (re-)calculate $f_{\text{CO}_2}$ (Sect. 3.3), and the (re-)calculated and quality controlled $f_{\text{CO}_2}$ data. The files contain these additional parameters: WOA salinity (Antonov et al., 2006), NCEP/NCAR sea level pressure (Kalnay et al., 1996) and ETOPO2 (2006) bathymetry. Each individual cruise data file has been assigned a digital object identifier (doi) for citation and transparency. Table 5 lists the parameters in the cruise data files.

### 5.3 SOCAT global and regional files (version 1.5)

Regional and global, concatenated files (version 1.5) have been merged from the individual cruise data files for a subset of SOCAT parameters (Table 5). These concatenated files only contain recommended $f_{\text{CO}_2}$ data with a WOCE flag 2 from cruises with a flag A, B, C or D. Table 5 lists the parameters in these regional and global synthesis files. Some changes have been applied relative to SOCAT version 1.4 (Sects. 5.1 and 5.2). Notably, multiple entries with the same time stamp were averaged for the global and regional synthesis files (Sect. 5.1).

Additional parameters have been added to the regional and global, concatenated files. These include Julian day (day of year), interpolated atmospheric $x_{\text{CO}_2}$ extracted from GLOBALVIEW-CO2 (2008), WOA salinity, NCEP/NCAR sea level pressure and ETOPO2 bathymetry. The global and regional files specify which reported CO$_2$ variable was used for (re-)calculation of recommended $f_{\text{CO}_2}$ (Sect. 3.3; Table 5). Every line of the concatenated files contains a doi-string, which provides a link to the individual cruise data file with the original CO$_2$ parameter(s) and metadata at PANGAEA® (Sect. 5.2).
The regional and global, concatenated files (version 1.5) are publicly available as “compressed zip” text files via CDIAC (http://cdiac.ornl.gov/ftp/oceans/SOCATv1.5/), in Ocean Data View (ODV) format (http://odv.awi.de/en/data/ocean/socat_v15_fco2_data/) and via the interactive Cruise Data Viewer (Sect. 5.4). NetCDF files (Eaton et al., 2011) will be made available in the future. The text files exist as one very large global file, and as subset files per region, with no overlap between the regions. The latter means that data of a given cruise may have been divided into several regional files (for example North Atlantic, Tropical Atlantic and Coastal region).

5.4 Cruise Data Viewer (version 1.5)

The LAS Cruise Data Viewer provides interactive access to SOCAT version 1.5 on a Live Access Server. It provides all of the output capabilities described in Sect. 4 as tools for the SOCAT QC-ers, except for the ability to enter QC flags and comments. The Cruise Data Viewer also supplies variables from other sources that provide scientific context useful to users of the fCO₂ data: atmospheric xCO₂ values interpolated from GLOBALVIEW-CO₂ (2008), WOA 2005 salinity, NCEP/NCAR sea level pressure values (Kalnay et al., 1996), and bathymetry from ETOPO2 (2006).

The Cruise Data Viewer allows the inclusion of WOCE flag 3 (questionable) or 4 (bad) data when viewing or downloading data. When subsets are downloaded from the Cruise Data Viewer, each data line contains a doi-string that links directly to the relevant cruise data file with its original reported CO₂ parameters at PANGAEA®. A “Table of Cruises” is available from the Cruise Data Viewer and lists the cruise flags, QC comments and SOCAT QC-er for each cruise. The Cruise Data Viewer can be accessed via the SOCAT website or directly at http://ferret.pmel.noaa.gov/SOCAT_cruise_viewer/.

5.5 Gridded products (version 1.5)

The gridded products provide values at a 1° latitude by 1° longitude resolution using monthly, annual, decadal and monthly climatological time scales, and at a 0.25°
latitude by 0.25° longitude with monthly time resolution for coastal analysis (Sabine et al., 2012). The recommended $f$CO$_2$ with a WOCE flag 2 were gridded by two algorithms: (1) averages giving equal weight to each observation in a cell, and (2) averages giving equal weight to each cruise that passed through a cell. Mean, extremes and standard deviations of $f$CO$_2$ are provided. Other statistical measures include the number of cruises per cell, the number of observations per cell and measures of the degree to which the $f$CO$_2$ averaged values may be biased from the cell centre. The SOCAT version 1.5 gridded products have not been corrected for any temporal increase in surface water $f$CO$_2$. Gridded fields are available as NetCDF files from CDIAC (http://cdiac.ornl.gov/ftp/oceans/SOCATv1.5/SOCATv1.5_Gridded_Dat/) and via the interactive Gridded Data Viewer. For more details, refer to the accompanying paper by Sabine et al. (2012).

5.6 Gridded Data Viewer (version 1.5)

The interactive LAS Gridded Data Viewer enables users to explore the gridded SOCAT fields. The viewer displays maps and time series for the specific region or period selected. Sequences of fields can be viewed as animations. Simple statistics such as means, extremes, variance and counts, may be requested of the data. By requesting counts of the number of observations and cruises, a user is able to explore the global coverage of the SOCAT collection. Figure 4 obtained by this means, illustrates the north-south distribution of cruises in the years 2000 through 2007. The gridded viewer also supplies 1° latitude by 1° longitude marine surface variables from ICOADS (2008) that provide useful scientific context when exploring $f$CO$_2$: surface air temperature, sea level pressure, sea surface temperature, and surface wind speed. The Gridded Data Viewer can be accessed at (http://ferret.pmel.noaa.gov/SOCAT_gridded_viewer/) or via the link on the SOCAT website.
6 Lessons learned and outlook

6.1 Lessons learned

SOCAT has taken four years to be put together and has been a large, international, collaborative effort of the marine carbon research community. SOCAT version 1.5 is the culmination of much hard work in data collection, data assembly and quality control by many seagoing marine carbon scientists around the world.

Lessons learned and improvements for future SOCAT releases have been discussed at the Surface Ocean CO$_2$ Data-to-Flux workshop (IOCCP, 2012). The lessons include a strong need for automating SOCAT with respect to: data submission, metadata submission and quality control. The automation and other improvements will reduce the amount of work required for creating SOCAT data products and SOCAT quality control, while at the same time speeding up the whole process with the aim to provide regular updates.

The SOCAT global group, upon consultation with regional group leaders, has decided to start work on SOCAT version 2, while in parallel automating SOCAT for version 3. Data submission to SOCAT version 2 was closed on 31 December 2011. SOCAT version 2 products will report time in seconds as reported in the original data files to remove the need to calculate averaged data. Regular SOCAT releases are envisaged, e.g., every two years from SOCAT version 3 onwards. Such regular future SOCAT releases will require sustained funding for key players.

Colleagues are strongly encouraged to make public their surface water $f$CO$_2$ data and accompanying documentation from the global oceans and coastal seas, preferably via CDIAC (http://cdiac.ornl.gov/oceans/submit.html) for inclusion in future SOCAT releases. Data and metadata should be reported in the IOCCP (2004) recommended formats, which are also listed on the CDIAC website.
6.2 Automation of SOCAT procedures

Automation of the submission of data and metadata will include prompt feedback to the data originator on unrealistic data and property-property plots of the data, such that the data originator can carry out primary and initial secondary quality control. Such automation will facilitate harmonisation of the data for SOCAT and will strongly reduce the number of cruises suspended from SOCAT during secondary quality control.

In the future, new cruises will be added to the LAS at regular (e.g., two monthly) intervals, enabling QC operators to carry out regular SOCAT QC. The Live Access Server will be modified to automatically generate typical property-property plots for secondary QC. The LAS will be enhanced with features to enter cruise flags and QC comments for multiple cruises (e.g., on the same vessel).

6.3 SOCAT products for assessing the ocean carbon sink

The release of SOCAT version 1.5 represents a milestone in ocean carbon research. Research using SOCAT will highlight the response of surface water $f_{CO_2}$ and the oceanic CO$_2$ sink to increasing levels of atmospheric CO$_2$ in a changing climate. The SOCAT products can be used in studies of spatial and temporal (seasonal, interannual and decadal) variability and trends in surface water $f_{CO_2}$. The SOCAT products will enable validation of model distributions of surface water $f_{CO_2}$ and air-sea CO$_2$ fluxes. SOCAT will aid process studies of oceanic $f_{CO_2}$ variability, e.g., in the North Atlantic, in the Pacific Ocean, in coastal seas, in the Arctic Ocean, in seasonally ice-covered Southern Ocean regions, near remote islands and oceanographic fronts. The SOCAT products may be used to create monthly basin-wide $f_{CO_2}$ maps for the most data-rich basins by a range of techniques such as neural networks, statistical techniques and algorithms (e.g., Lefèvre et al., 2005; Telszewski et al., 2009). These $f_{CO_2}$ maps can be used for calculating basin-wide monthly CO$_2$ air-sea fluxes, which may constrain atmospheric inversions for global atmospheric carbon budgets. Study of length scales of $f_{CO_2}$ variability will provide information on the minimum sampling coverage required.
for quantifying the oceanic CO$_2$ sink with sufficient accuracy (e.g., Lenton et al., 2009). It is expected that the regular SOCAT releases will become a crucial tool in quantification of changes in oceanic CO$_2$ uptake and in global climate research. Increasing the number of surface ocean CO$_2$ data has in the past significantly modified the estimate of the oceanic CO$_2$ sink (e.g., Takahashi et al., 2009b). SOCAT and its future development will contribute to further enhance the reliability of such assessments.

**Acknowledgements.** SOCAT is promoted by IOCCP, the Surface Ocean Lower Atmosphere Study, and the Integrated Marine Biogeochemistry and Ecosystem Research program. Douglas Wallace (Dalhousie University, Canada, and former SOLAS chair), Emilie Brévière (SOLAS executive officer) and Lisa Maddison (IMBER deputy executive officer) have strongly encouraged SOCAT. Support for SOCAT has been received from the Bjerknes Centre for Climate Research, the University of Bergen, Uni Research (Norway), the US National Oceanic and Atmospheric Administration, the University of Washington, Oak Ridge National Laboratory (US), the University of East Anglia (UEA, UK), PANGAEA® – Data Publisher for Earth & Environmental data (Germany), the Alfred Wegener Institute for Polar and Marine Research (Germany), the Centre National de la Recherche Scientifique (France), the Research Council of Norway (CARBON-HEAT), the US National Science Foundation (OCE-1068958), the international Scientific Committee on Oceanic Research (SCOR, US, OCE-0938349), the European Union integrated projects CarboOcean (GOCE 511176-1) and CarboChange (FP7 264879), the UK Ocean Acidification Research Programme (NE/H017046/1; funded by the Natural Environment Research Council, the Department for Energy and Climate Change and the Department for Environment, Food and Rural Affairs) and the UK National Centre for Earth Observation. Support for SOCAT meetings has been received from IOCCP, IMBER, the National Institute for Environmental Studies (NIES, Japan), the Commonwealth Scientific and Industrial Research Organisation (CSIRO, Australia), GEOMAR (Germany) and the European Cooperation in Science and Technology (COST) Action 735 (UK). This is a contribution to the Center for Climate Dynamics (SKD) within the Bjerknes Centre for Climate Research.
References


DeGrandpré, M. D., Baehr, M. W., and Hammar, T. R.: Development of an optical chemical sensor for oceanographic applications: the submersible autonomous moored instrument for


Table 1. Meetings for SOCAT version 1.5. Abbreviations are explained in the text and acknowledgements.

<table>
<thead>
<tr>
<th>Timing</th>
<th>Meeting description</th>
<th>Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/2007</td>
<td>First technical meeting. Discussion of procedures</td>
<td>Delmenhorst and Bremen, Germany</td>
<td>NA</td>
</tr>
<tr>
<td>01/2009</td>
<td>SOCAT Coastal regional workshop</td>
<td>GEOMAR, Kiel, Germany</td>
<td>IOCCP (2009a)</td>
</tr>
<tr>
<td>03/2009</td>
<td>SOCAT Pacific regional workshop</td>
<td>NIES, Tsukuba, Japan</td>
<td>IOCCP (2010a)</td>
</tr>
<tr>
<td>06/2009</td>
<td>SOCAT Atlantic and Southern Ocean regional workshop</td>
<td>UEA, Norwich, UK</td>
<td>IOCCP (2009b)</td>
</tr>
<tr>
<td>02/2010</td>
<td>SOCAT Equatorial Pacific, North Pacific and Indian Ocean regional workshop</td>
<td>Tokyo, Japan</td>
<td>NA</td>
</tr>
<tr>
<td>06/2010</td>
<td>SOCAT Southern Ocean and Indian Ocean regional workshop</td>
<td>CSIRO, Hobart, Australia</td>
<td>IOCCP (2010b)</td>
</tr>
</tbody>
</table>
Table 2. Groups and key participants in SOCAT. Figure 3 shows the SOCAT regions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Area</th>
<th>Lead(s) and Key Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td></td>
<td>Bakker (chair), Olsen, Pfeil, Hankin, Koyuk, Kozyr, Malzcyk, Metzl, Pierrot, Sabine, Teliszewski</td>
</tr>
<tr>
<td>Coastal regions</td>
<td>&lt; 400 km from land; north of 30° S</td>
<td>Borges, Chen</td>
</tr>
<tr>
<td>North Atlantic</td>
<td>north of 30° N, incl. Atlantic Arctic</td>
<td>Schuster</td>
</tr>
<tr>
<td>Tropical Atlantic</td>
<td>30° N to 30° S</td>
<td>Lefèvre</td>
</tr>
<tr>
<td>North Pacific</td>
<td>north of 30° N, incl. Pacific Arctic</td>
<td>Nojiri</td>
</tr>
<tr>
<td>Tropical Pacific</td>
<td>30° N to 30° S</td>
<td>Feely</td>
</tr>
<tr>
<td>Indian Ocean</td>
<td>north of 30° S</td>
<td>Sarma</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>south of 30° S, incl. coastal waters</td>
<td>Tilbrook, Metzl</td>
</tr>
</tbody>
</table>
Table 3. Activities and key participants in SOCAT.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Key Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data retrieval</td>
<td>Pfeil, Hood, Bakker</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>Pfeil</td>
</tr>
<tr>
<td>Uniform format</td>
<td>Olsen, Pfeil</td>
</tr>
<tr>
<td>Defining QC (quality control) criteria</td>
<td>Wanninkhof, Olsen, Schuster, meeting participants (IOCCP, 2008, 2009b)</td>
</tr>
<tr>
<td>QC cookbook</td>
<td>Olsen, Metzl</td>
</tr>
<tr>
<td>Coastal mask</td>
<td>Hales, Olsen, Hankin</td>
</tr>
<tr>
<td>Matlab QC code</td>
<td>Pierrot, Olsen</td>
</tr>
<tr>
<td>Live Access Server (LAS)</td>
<td>Hankin, Malczyk, Koyuk</td>
</tr>
<tr>
<td>Data QC</td>
<td>SOCAT regional groups</td>
</tr>
<tr>
<td>Conflicting cruise flags</td>
<td>Bakker</td>
</tr>
<tr>
<td>Gridding</td>
<td>Sabine, Fassbender, Manke</td>
</tr>
<tr>
<td>Logo</td>
<td>Brown</td>
</tr>
<tr>
<td>SOCAT webpage</td>
<td>Pfeil, Koyuk, Bakker</td>
</tr>
<tr>
<td>Online SOCAT products</td>
<td>Kozyr, Koyuk, Hankin, Pfeil</td>
</tr>
<tr>
<td>ODV (Ocean Data View) for SOCAT</td>
<td>Schlitzer</td>
</tr>
<tr>
<td>SOCAT meetings</td>
<td>Nojiri, Borges, Wallace, Schuster, Bakker, Tilbrook, Hood, Tedesco, Telszewski, Brévière, Maddison</td>
</tr>
<tr>
<td>Internal coordination</td>
<td>Hood, Bakker, Koyuk</td>
</tr>
<tr>
<td>Coordination with marine community</td>
<td>Hood, Tedesco, Telszewski</td>
</tr>
</tbody>
</table>
### Table 4. Surface water CO₂ parameters reported in the original data files, which have been used for the calculation of recommended \( f\text{CO}_2 \) \( (f\text{CO}_2\text{.rec}) \) at sea surface (or intake) temperature (Pfeil and Olsen, 2009). The parameters are listed in order of preference (with index 1 as the favourite). The index has been reported in the SOCAT global and regional output files as “\( f\text{CO}_2\text{-source} \)” (Table 5). Ancillary parameters have been used for NCEP (National Centers for Environmental Prediction) atmospheric pressure (Kalnay et al., 1996) and WOA (World Ocean Atlas) salinity (Antonov et al., 2006) (Sect. 3.3) in cases of incomplete data reporting.

<table>
<thead>
<tr>
<th>Index</th>
<th>Reported CO₂ parameter</th>
<th>Unit</th>
<th>Data percentage (%)</th>
<th>Extra variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\textit{xCO₂water_equi_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>57.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>\textit{xCO₂water_SST_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>\textit{pCO₂water_equi_wet}</td>
<td>( \mu \text{atm} )</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>\textit{pCO₂water_SST_wet}</td>
<td>( \mu \text{atm} )</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>\textit{fCO₂water_equi}</td>
<td>( \mu \text{atm} )</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>\textit{fCO₂water_SST_wet}</td>
<td>( \mu \text{atm} )</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>\textit{pCO₂water_equi_wet}</td>
<td>( \mu \text{atm} )</td>
<td>0.4</td>
<td>NCEP Pressure</td>
</tr>
<tr>
<td>8</td>
<td>\textit{pCO₂water_SST_wet}</td>
<td>( \mu \text{atm} )</td>
<td>13.8</td>
<td>NCEP Pressure</td>
</tr>
<tr>
<td>9</td>
<td>\textit{xCO₂water_equi_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>0.2</td>
<td>WOA Salinity</td>
</tr>
<tr>
<td>10</td>
<td>\textit{xCO₂water_SST_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>1.2</td>
<td>WOA Salinity</td>
</tr>
<tr>
<td>11</td>
<td>\textit{xCO₂water_equi_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>0.0*</td>
<td>NCEP Pressure</td>
</tr>
<tr>
<td>12</td>
<td>\textit{xCO₂water_SST_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>2.2</td>
<td>NCEP Pressure</td>
</tr>
<tr>
<td>13</td>
<td>\textit{xCO₂water_equi_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>0.0*</td>
<td>NCEP Pressure, WOA Salinity</td>
</tr>
<tr>
<td>14</td>
<td>\textit{xCO₂water_SST_dry}</td>
<td>( \mu \text{mol mol}^{-1} )</td>
<td>0.2</td>
<td>NCEP Pressure, WOA Salinity</td>
</tr>
</tbody>
</table>

1 Atmospheric pressure was not reported in the original data file.
2 Salinity was not reported in the original data file.
* Not used for data reporting as an approach with a lower index was available.
Table 5. Content of the individual data cruise files (version 1.4) and the global and regional concatenated files (version 1.5) in SOCAT\(^1\) (Sects. 5.1, 5.2 and 5.3).

<table>
<thead>
<tr>
<th>Event cruise ID</th>
<th>Cruise name</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>yyyy-mm-ddhh:mm (ISO 8601)(^1)</td>
<td>Year</td>
<td>Date/Time</td>
</tr>
<tr>
<td>NA</td>
<td>Mon Month</td>
<td>Month</td>
<td>Date/Time</td>
</tr>
<tr>
<td>NA</td>
<td>Day Day</td>
<td>Day</td>
<td>Date/Time</td>
</tr>
<tr>
<td>NA</td>
<td>Hour Hour</td>
<td>Hour</td>
<td>Date/Time</td>
</tr>
<tr>
<td>NA</td>
<td>Min Minute</td>
<td>Minute(^1)</td>
<td>Date/Time</td>
</tr>
<tr>
<td>Longitude</td>
<td>°E Longitude (0 to 360)</td>
<td>°E</td>
<td>Longitude</td>
</tr>
<tr>
<td>Latitude</td>
<td>°N,°S Latitude (−90 to 90)</td>
<td>°N,°S</td>
<td>Latitude</td>
</tr>
<tr>
<td>Depth Water</td>
<td>M Intake depth(^2)</td>
<td>M</td>
<td>Depth Water</td>
</tr>
<tr>
<td>Temp</td>
<td>°C Sea surface temperature</td>
<td>°C</td>
<td>Temp</td>
</tr>
<tr>
<td>Sal</td>
<td>− Sea surface salinity</td>
<td>−</td>
<td>Sal</td>
</tr>
<tr>
<td>T(_{equ})</td>
<td>°C Temperature at equilibration</td>
<td>°C</td>
<td>T(_{equ})</td>
</tr>
<tr>
<td>PPPP</td>
<td>hPa Atmospheric pressure(^1)</td>
<td>hPa</td>
<td>PPPP</td>
</tr>
<tr>
<td>P(_{equ})</td>
<td>hPa Pressure in the equilibrator(^1)</td>
<td>hPa</td>
<td>P(_{equ})</td>
</tr>
<tr>
<td>Sal interp</td>
<td>hPa Salinity from WOA 2005</td>
<td>hPa</td>
<td>Sal interp</td>
</tr>
<tr>
<td>Press atmos interp</td>
<td>hPa Atmospheric pressure from 6-hourly NCEP/NCAR data</td>
<td>hPa</td>
<td>Press atmos interp</td>
</tr>
<tr>
<td>Bathy depth interp</td>
<td>M Bottom depth from ETOPO2</td>
<td>M</td>
<td>Bathy depth interp</td>
</tr>
<tr>
<td>xCO2water(_{equ,dry})</td>
<td>µmol mol(^{-1}) xCO2 (water) at equilibrator temperature (dry air)(^1)</td>
<td>µmol mol(^{-1})</td>
<td>xCO2water(_{equ,dry})</td>
</tr>
<tr>
<td>fCO2water(_{SST, wet})</td>
<td>µatm /CO2 (water) at sea surface temperature (wet air)(^1)</td>
<td>µatm</td>
<td>fCO2water(_{SST, wet})</td>
</tr>
<tr>
<td>pCO2water(_{SST, wet})</td>
<td>µatm /CO2 (water) at sea surface temperature (wet air)(^1)</td>
<td>µatm</td>
<td>pCO2water(_{SST, wet})</td>
</tr>
<tr>
<td>xCO2water(_{SST, dry})</td>
<td>µmol mol(^{-1}) xCO2 (water) at sea surface temperature (dry air)(^1)</td>
<td>µmol mol(^{-1})</td>
<td>xCO2water(_{SST, dry})</td>
</tr>
<tr>
<td>fCO2water(_{SST, wet})</td>
<td>µatm /CO2 (water) at equilibrator temperature (wet air)(^1)</td>
<td>µatm</td>
<td>fCO2water(_{SST, wet})</td>
</tr>
<tr>
<td>pCO2water(_{SST, wet})</td>
<td>µatm /CO2 (water) at equilibrator temperature (wet air)(^1)</td>
<td>µatm</td>
<td>pCO2water(_{SST, wet})</td>
</tr>
<tr>
<td>fCO2water(_{SST, dry})</td>
<td>µatm Recommended /CO2, calculated for the SOCAT protocol (Table 3)</td>
<td>µatm</td>
<td>fCO2water(_{SST, dry})</td>
</tr>
<tr>
<td>ICO2source</td>
<td>− The algorithm for calculating /CO2 (Index in Table 3)</td>
<td>−</td>
<td>ICO2source</td>
</tr>
<tr>
<td>gCO2source</td>
<td>µmol mol(^{-1}) Atmospheric /CO2 from GLOBALVIEW-CO2 (2008)</td>
<td>µmol mol(^{-1})</td>
<td>gCO2source</td>
</tr>
<tr>
<td>Julian_day GMT</td>
<td>Day Day of year (1 for 1 January)</td>
<td>Day</td>
<td>Julian_day GMT</td>
</tr>
<tr>
<td>WOCE_flag</td>
<td>− WOCE flag for /O2 rec</td>
<td>−</td>
<td>WOCE_flag</td>
</tr>
<tr>
<td>doi</td>
<td>Digital object identifier to the individual cruise file and metadata</td>
<td>doi</td>
<td>doi</td>
</tr>
<tr>
<td>Averaged</td>
<td>− Indicator that data was averaged for version 1.5(^1)</td>
<td>−</td>
<td>Averaged</td>
</tr>
</tbody>
</table>

\(^1\)(refers to data reported by the data originator.

---

\(^2\)Individual cruise data files in version 1.4 may contain multiple entries for a given time stamp. Multiple entries for a given time stamp have been averaged in the global and regional concatenated files in version 1.5 (Sects. 5.1, 5.2 and 5.3).

---

\(^1\)If the intake depth has not been reported by the data originator, we assume an intake depth of 5 m.
Table 6. Criteria for assigning cruise flags, based on the expected quality of the recommended $f$CO$_2$ data (after Olsen and Metzl, 2009). SOP is Standard Operating Procedures (Dickson et al., 2007). QC is quality control.

<table>
<thead>
<tr>
<th>Cruise flag (ID)</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (11)</td>
<td>(1) Followed approved methods/SOP criteria and (2) Metadata documentation complete and (3) 2nd level QC was deemed acceptable and (4) A comparison with other data was deemed acceptable.</td>
</tr>
<tr>
<td>B (12)</td>
<td>(1) Followed approved methods/SOP criteria and (2) Metadata documentation complete and (3) 2nd level QC was deemed acceptable.</td>
</tr>
<tr>
<td>C (13)</td>
<td>(1) Did not follow approved methods/SOP criteria and (2) Metadata documentation complete and (3) 2nd level QC was deemed acceptable (including if possible comparison with other data).</td>
</tr>
<tr>
<td>D (14)</td>
<td>(1) Did or did not follow approved methods/SOP criteria and (2) Metadata documentation incomplete and (3) 2nd level QC was deemed acceptable (including if possible comparison with other data).</td>
</tr>
<tr>
<td>F (15) (Failure)</td>
<td>(1) Did or did not follow approved methods/SOP criteria and (2) Metadata documentation complete or incomplete and (3) 2nd level QC revealed non-acceptable data.</td>
</tr>
<tr>
<td>S (Suspend)</td>
<td>(1) Did or did not follow methods/SOP criteria and (2) Metadata documentation complete or incomplete and (3) 2nd level QC revealed non-acceptable data and (4) data are being updated (part or all of the cruise).</td>
</tr>
<tr>
<td>X (15) (Exclude)</td>
<td>The cruise (data set) duplicates another cruise (data set) in SOCAT.</td>
</tr>
<tr>
<td>N (No flag)</td>
<td>No cruise flag has yet been given to this cruise.</td>
</tr>
<tr>
<td>U (Update)</td>
<td>The cruise data have been updated. No cruise flag has yet been given to the revised data.</td>
</tr>
</tbody>
</table>
Fig. 1. Number of data points per year during the period 1968 to 2007 included in SOCAT version 1.5.
Fig. 2. Distribution of data in SOCAT version 1.5 per decade: (a) 1970s, (b) 1980s, (c) 1990s and (d) 2000s.
Fig. 3. Oceanic and coastal regions in SOCAT (Table 1). The white collar surrounding land masses indicates the coastal regions. Coastal regions north of 30° S were quality controlled by the coastal group, while coastal regions south of 30° S were quality controlled by the Southern Ocean group.
Fig. 4. The number of SOCAT data points per 1° latitude by 1° longitude grid box for the years 2000 to 2007 shown as a function of latitude. This figure highlights the paucity of data in the Southern Hemisphere.