

Surface Ocean CO<sub>2</sub> Atlas (SOCAT) project - SOCAT-2 Meeting  
UNESCO, Paris  
June 16 -17 2008

Co-sponsored by the IOC-SCOR International Ocean Carbon Coordination Project (IOCCP) and the SOLAS-IMBER Carbon Group.

**IOCCP Report No. 9**

TABLE OF CONTENTS

1. Introduction and Status Report .....	2
1.1 Introduction to the SOCAT Project.....	2
1.2 Overview of SOCAT-1; Goals for SOCAT-2.....	3
1.3 Status of the SOCAT Data Set .....	4
2. 2 <sup>nd</sup> level Quality Control Issues.....	6
2.1 Assessment of Uncertainty.....	6
2.2 Data Flags.....	8
2.3 Crossover and Other Checks for Underway Data.....	9
2.4 Quality Controlling Non-Carbon Data.....	11
2.5 Live-Access Server and Shared Data Management Strategies .....	12
2.6 Record-keeping for 2 <sup>nd</sup> level QC Procedures.....	13
2.7 Gridding and Interpolation Techniques.....	13
2.8 Data Sharing Issues.....	14
3. Regional Group Reports.....	15
3.1 Indian Ocean Group.....	15
3.2 Pacific Ocean Group.....	16
3.3 Atlantic Ocean Group.....	17
3.4 Southern Ocean Group.....	17
3.5 Coastal Ocean Group.....	21
4. Summary and Action Item List.....	22
5. Annexes:	
1. Participants List.....	24
2. Agenda .....	25
3. Summary of SOCAT-1 .....	26
4. SOCAT version 1.1 Data Distribution Graphics.....	28
5. Calculations Used for the 1 <sup>st</sup> Level QC Data Set.....	31

---

Support for this workshop was provided by the US National Science Foundation Award No. OCE-0608600 to the Scientific Committee on Oceanic Research (SCOR) for the IOCCP, the EU Integrated Project CarboOcean, the Intergovernmental Oceanographic Commission of UNESCO, the US National Oceanic and Atmospheric Administration, the Surface Ocean-Lower Atmosphere (SOLAS) Project, and the Integrated Marine Biogeochemistry and Ecosystem Research (IMBER) Project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the workshop sponsors or of the U.S. National Science Foundation (NSF).

## **I. INTRODUCTION AND STATUS REPORT**

### **1.1 Background on the SOCAT Project (M. Hood, B. Pfeil, A. Olsen, D.C.E. Bakker)**

At the “Surface Ocean CO<sub>2</sub> Variability and Vulnerability” (SOCOVV) workshop at UNESCO, Paris in April 2007, co-sponsored by IOCCP, SOLAS, IMBER, and the Global Carbon Project, participants agreed to assemble a global surface CO<sub>2</sub> data set of all publicly available ocean surface fCO<sub>2</sub> data in a common format. (The fugacity of carbon dioxide, or fCO<sub>2</sub>, is the partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>) corrected for non-ideal behaviour of the gas.) This is an activity that has been called for by several international groups for many years, and has now become a priority activity for the marine carbon community. This data set will serve as a foundation upon which the community will continue to build in the future, based on agreed data and metadata formats and standard 1<sup>st</sup> level quality-control procedures, building on earlier agreements established at the 2004 Tsukuba workshop on “Ocean Surface pCO<sub>2</sub> Data Integration and Database Development”. This activity also supports the SOLAS and IMBER science plans and their joint carbon implementation plan.

This data set is meant to serve a wide range of user communities and it is envisaged that, in the future, 2 distinct data products will be made available in this Surface Ocean CO<sub>2</sub> Atlas (SOCAT):

- a 2<sup>nd</sup> level quality controlled global surface ocean fCO<sub>2</sub> data set following agreed procedures and regional review, and
- a gridded SOCAT product of monthly surface water fCO<sub>2</sub> means on a 1° x 1° grid with no temporal or spatial interpolation using the 2<sup>nd</sup> level quality-controlled global surface ocean fCO<sub>2</sub> data set.

The extended 1<sup>st</sup> level quality-controlled data set builds on the work started in 2001 as part of the EU ORFOIS project by Dorothee Bakker, which now continues as part of the EU CarboOcean project, where Benjamin Pfeil and Are Olsen have compiled the publicly-available surface CO<sub>2</sub> data held at CDIAC (Carbon Dioxide Information Analysis Center) and elsewhere into a common format database based on the IOCCP recommended formats for metadata and data reporting. This compilation currently includes data from more than 10 countries, producing an initial database composed of more than 1250 cruises from 1968 to 2007 with approximately 4.5 million measurements of various carbon parameters, available in a common format, 1<sup>st</sup> level quality-controlled data set.

A small technical meeting, henceforth called the SOCAT-1 meeting, was held in Bremen, Germany, on 5 December 2007 (associated with the 3<sup>rd</sup> CarboOcean Annual Meeting) to agree on 1<sup>st</sup> level QC for the data set and to decide on a way forward for the 2<sup>nd</sup> level QC issues.

The IOCCP, along with CarboOcean and the SOLAS-IMBER Joint Carbon Group, agreed to hold a 2<sup>nd</sup> technical workshop (SOCAT-2, this meeting) to develop internationally agreed 2<sup>nd</sup> level quality-control procedures and to discuss the coordination

of regional scientific groups to conduct the 2<sup>nd</sup> level quality control (QC) analyses. The goals for this meeting were to:

- Reach international agreement on 2<sup>nd</sup> level QC procedures
- Identify approaches for gridding and interpolation
- Identify major science issues for each basin and globally
- Develop a short report for distribution to all relevant networks.

A list of participants is provided in *Annex 1* and the agenda is in *Annex 2* of this report, and the report of the SOCAT-1 meeting is given in *Annex 3*.

## 1.2 Overview of SOCAT-1; Goals for SOCAT-2 (D.C.E. Bakker)

Dorothee Bakker, chair of the SOCAT-2 meeting, provided an overview of the decisions made in the 1<sup>st</sup> technical workshop (December 2007) and outlined the major issues that need to be discussed for the 2<sup>nd</sup> level quality control.

Establishing surface ocean and atmosphere carbon observing systems for constraining the net annual air-sea CO<sub>2</sub> flux per ocean basin to < 0.2 Pg C yr<sup>-1</sup> (SOLAS/IMBER joint carbon implementation plan) was adopted as an overall goal during the SOCOVV workshop. The idea for a Surface Ocean CO<sub>2</sub> Atlas (SOCAT) emerged during discussions on Observation Strategies (Working Group 2) and Data and Scientific Synthesis (Working Group 3) during SOCOVV. A number of action items were identified to advance this goal, and Bakker provided an update on these actions:

1. Determine which global fCO<sub>2</sub> data sets being compiled by different groups should be considered the global standard data set (Sabine, Pfeil). The analysis provided this summary for the following compilations:

- The Takahashi et al. (2008) data set – quality controls and adjustments are not fully documented (see report: [Takahashi, T., S.C. Sutherland, and A. Kozyr. 2007. Global Ocean Surface Water Partial Pressure of CO<sub>2</sub> Database: Measurements Performed During 1968 - 2006 \(Version 1.0\). ORNL/CDIAC-152, NDP-088, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 20 pp.](#))
- Data set of John Callahan (NOAA/PMEL) – not in a common format and not publicly available.
- Data set of Joellen Russell (Uni. Arizona, USA) – more than 3 million data points; adjustments not well documented and compilation not publicly available.
- Data set of Pfeil and Olsen (Bjerknes Centre, Norway) – more than 4.5 million data points in common format that are documented and traceable, with 1<sup>st</sup> level QC carried out (missing values, outliers) and with calculation of fCO<sub>2</sub>

2. Establishment of a standard, global surface ocean fCO<sub>2</sub> data set, based on agreed data and metadata formats and standard 1<sup>st</sup> and 2<sup>nd</sup> level QC. The decision was made to use the common format, well-documented data set of Pfeil and Olsen with 1<sup>st</sup> level QC based

on agreements from the ad-hoc SOCAT-1 meeting. The 2<sup>nd</sup> level QC issues are being discussed at the SOCAT-2 meeting.

3. Initiation of regular data product development. To be discussed at this meeting.
4. Identify approaches for gridding and interpolation in order to estimate flux and surface CO<sub>2</sub> using satellite and proxy techniques (Sabine). To be discussed at this meeting.
5. Establishment of regional groups to address regional QC and synthesis issues. Specifically, the groups will look at the available data and QC issues, identify missing data, compare methods for creating fCO<sub>2</sub> maps, exchange findings on a global level, coordinate regional QC into a global synthesis, and identify key process-related scientific questions that require large-scale joint synthesis efforts. The following groups were established:
  - Atlantic + Arctic (Schuster; Lefèvre)
  - Pacific (Feely for 30N-30S; others under development)
  - Indian (Sarma)
  - Southern Ocean (Tilbrook)
  - Coastal seas (Chen, Borges)
  - Global focal points (Bakker, Pfeil, Sabine, Metzl, Olsen)

Bakker reminded the group of the goals of this 2<sup>nd</sup> technical workshop (SOCAT-2) and the issues that must be covered in the two days (see Agenda in Annex 2). She also added that the group needs to set goals for a realistic date for the release of the 2<sup>nd</sup> level QC data set, to begin discussions on how the gridded data product will be developed, to decide how the 2<sup>nd</sup> level QC data set would be regularly updated, and to look ahead to the scientific analyses and outcomes that might come from the SOCAT data set and gridded products, with a possible goal of having new results to present at the International Carbon Dioxide Conference – 8, scheduled for September 2009.

### **1.3 Status of the Data Set (B. Pfeil and A. Olsen)**

Benjamin Pfeil and Are Olsen provided a status report on the SOCAT data set. The last updated version (version 1.1, June 12 2008) is now available to data contributors. All the data in the data set are public, and the data set contains 1370 cruises from 1968 to 2007, with approximately 4.5 million CO<sub>2</sub> measurements. Included in the data set are all publicly available underway data from CDIAC, NOAA/AOML, NOAA/PMEL, LDEO, CARINA, WDC-MARE and data delivered directly by PIs. Some data have been excluded from the data set: data from NOAA/AOML from 2006 and 2007 that has not yet been submitted to CDIAC; data with known issues (e.g. problems with the temperature sensor); data with intellectual property rights (most CARBOOCEAN data); and data with only atmospheric CO<sub>2</sub> measurements.

Figures in *Annex 4* show the spatial overview (cruise tracks), spatial distribution (number of points as a function of latitude / longitude), and the temporal distribution (number of data points over time) of the SOCAT data (version 1.1).

Pfeil provided an overview of the 1<sup>st</sup> level QC procedures carried out on the data set. All input data were put in a common file format (tab delimited text file), the time and latitude / longitude position of data points were made uniform, the naming of parameters was standardized, and NaN was used for all missing values. Essential missing parameters like salinity and atmospheric pressure were added from standard data collections since they are needed for re-computations of fCO<sub>2</sub>. Obvious outliers for all non-carbon parameters were flagged or sometimes corrected by examining the neighboring stations. For example, treatment of unrealistic values included correction of negative values (e.g., a salinity value of -35.23) based on neighboring data. Unrealistic values in SST, salinity and atmospheric pressure were either deleted or interpolated from the neighboring data.

Where possible, fCO<sub>2</sub> was recomputed (see *Annex 5* for details of the fCO<sub>2</sub> re-calculations applied for this data set). Cruise labels were standardized using Expocodes.

Pfeil pointed out several known issues with this 1<sup>st</sup> level QC data set that must still be addressed:

- Detailed metadata are missing for some cruises.
- Some cruises may have been reported more than once.
- Expocodes still to be assigned for one vessel.
- Problem with Expocodes for commercial vessels (when vessels are sold, change names, change country flag, etc.).
- Expocodes for moorings and time series need to be decided upon.
- Identification of cruises from iron fertilization experiments needs to be carried out in order to avoid inclusion of purposefully perturbed CO<sub>2</sub> conditions in the data set.
- Incomplete checking of non-carbon parameters for unrealistic values (also see page 11).

Metadata are differentiated between basic metadata for each cruise (cruise inventory table) and the detailed metadata as provided by the PI. The cruise inventory table includes

- Cruise name as reported by the data holder
- Expocode
- Research vessel
- PI
- Data source (e.g. CDIAC, NOAA, WDC-MARE etc)
- Date added to the collection
- Bounding box (spatial distribution of the data)
- Start and end date of the cruise
- Number of observations
- Metadata (links to detailed metadata as reported by the PI)

This is linked to the detailed CDIAC Mercury Metadata report provided by the PI.

An international cruise labeling convention was assigned for all cruises, which consists of

- Country code
- Vessel code
- Start date of the cruise: yyyyymmdd

The advantage of this is easy identification of a cruise. Basic information on the cruise can be obtained by referring to the label.

On the Bjerknes website([http://www.carbon-synthesis.org/front\\_content.php?idcat=371](http://www.carbon-synthesis.org/front_content.php?idcat=371)) are available:

- The SOCAT data set version 1.1 in Ocean Data View
- The SOCAT data set version 1.1 in a tab delimited text file
- all MATLAB scripts
- documentation
- the cruise inventory table and detailed metadata per cruise

#### Discussions:

The group agreed that the Expocodes for commercial vessels should be updated in the data set every time the ship's IMO identification changes as a result of change of flag, name and/or ownership. However, the EXPOCODE should not be changed for data already collected. For moorings and time series stations, it was agreed that the OceanSITES formats would be adopted.

The group also brought up the issues of how to distinguish different data sets from the same cruise, and how to deal with future data variables (e.g., Total CO<sub>2</sub>). Pfeil noted that it was straightforward to re-write the scripts to add new variables as they become available.

ACTION ITEM 1: Coordinate with OceanSITES on the Platform Names / Codes for time series stations (Pfeil, Olsen, Kozyr, Hood).

## **II. 2<sup>nd</sup> LEVEL QUALITY CONTROL ISSUES**

### **2.1 Assessments of Uncertainty (A. Körtzinger and C. Sabine)**

Arne Körtzinger provided an overview of previous intercomparison experiments for surface fCO<sub>2</sub> systems to provide insight into the sources of uncertainty in surface CO<sub>2</sub> data from underway ships. He pointed out that when systems are run side-by-side in identical, carefully controlled conditions, it is possible for different systems to agree to within 2 µatm. This has been demonstrated in three international intercomparison experiments. Often, the major source of uncertainty is bad temperature data, resulting from sensors in the seawater inlet or the equilibrator that were not calibrated accurately. However, the reality is that these systems are not run in identical or controlled conditions, and it is likely that the difference between systems is around 4-5 µatm.

From these intercomparison experiments, we can now identify potential sources of error, some of which are minor, others of which are very difficult to address. Known sources of error include

- Calibration error of (secondary) gas standards;
- Suboptimal  $x\text{CO}_2$  (mixing ratio) calibration routine (calibration function, number of standards, cell pressure measurement in the non-dispersive infrared analyser);
- Differences in data reduction algorithms;
- Inconsistent calibration of in-situ and equilibrator temperature (T) sensors;
- Temporal mismatch of in-situ and equilibrator  $T$  measurements (adds spurious variability);
- Improper calibration of inlet temperature sensor, often a thermosalinograph (comparison with CTD casts);
- Improper location of intake temperature sensor (e.g. downstream of pump);
- Loss in accuracy with moist gas  $x\text{CO}_2$  measurement (e.g. with LI-COR 6262 or 7000);
- Contamination problem with dry air  $x\text{CO}_2$  measurements through equilibrator venting (slight continuous under-pressure);
- Existence of undetected gas leaks (e.g., with Nafion dryers);
- Incomplete drying of sample air (exhaustion of desiccants, malfunction of dryers, etc.);
- Uncertainty in equilibrator pressure ;
- Deterioration of signal with drop in seawater and/or gas flow rates;
- Respiration of organic matter accumulating in particle filters; and
- Respiration of organic matter in the seawater line (has been demonstrated to be significant for  $\text{O}_2$  (oxygen) on research vessels and VOS (voluntary observing ships);

Körtzinger highlighted some things we can do to address these errors:

- Make sure complete metadata information is available;
- Perform 1<sup>st</sup> level quality control (time, location, parameter range, missing parameters, ‘flyers’ etc.);
- Exclude data sets on the basis of justifiable minimal criteria (as in Taro Takahashi’s approach);
- Calculate  $p\text{CO}_2$  or  $f\text{CO}_2$  for all data sets from  $x\text{CO}_2$  using a uniform procedure;
- Compare *in-situ* temperature measurements (thermosalinograph) with calibrated temperature measurements from co-located CTD casts;
- Compare atmospheric  $x\text{CO}_2$  with Globalview- $\text{CO}_2$  data set for plausibility; try to correlate steep changes in atmospheric  $x\text{CO}_2$  with changes in meteorological parameters (e.g., barometric pressure, wind direction);
- Inspect key variables vs. time or latitude/longitude for each data set (e.g.,  $\Delta T$  in situ – equilibrator,  $\Delta P$  equilibrator – barometric) and correlations between them (very labor intensive);
- Inspect data visually data set by data set (very labor intensive).

Körtzinger also pointed out that it would be difficult to use crossovers between cruises for offset identification, because the spatiotemporal variability scales of  $p\text{CO}_2$  are such that the match in space and time would have to be extremely close.

#### Discussions:

The group agreed that the uncertainty for most surface  $f\text{CO}_2$  data from VOS ships is probably at least 4-5  $\mu\text{atm}$ .

The group discussed possible ways of identifying offsets and potentially bad data. The main points of this discussion have been merged with those from the discussion on page 10 in the section on Crossover and Other Checks.

## **2.2 Data Flags (R. Wanninkhof)**

Rik Wanninkhof provided an overview of flag issues, which build on experience from GLODAP and CARINA as well as on agreements about data and metadata reporting from an informal meeting (<http://www.aoml.noaa.gov/ocd/gcc/uwpc2/workshops/>) held in Miami in 2002, which were later adopted by the international community at the IOCCP workshop “Ocean Surface  $p\text{CO}_2$ , Data Integration and Database Development” (Tsukuba, 2004). The IOCCP Recommended Format for  $f\text{CO}_2$  Metadata and Data from Underway Systems is on the IOCCP web-site: [www.ioccp.org](http://www.ioccp.org) under Data and Products. This is the format now used by CDIAC for its metadata forms. Wanninkhof reminded the group that 2007 Guide to Best Practices for Ocean  $\text{CO}_2$  Measurement (available at CDIAC: [http://cdiac.ornl.gov/oceans/Handbook\\_2007.html](http://cdiac.ornl.gov/oceans/Handbook_2007.html)) contains a Standard Operating Procedure (SoP) for Determination of  $p\text{CO}_2$  in air that is in equilibrium with a continuous stream of sea water.

Wanninkhof reviewed the WOCE flag designations for Data and Cruise QC flags. The data QC flags are:

- 2 = Good (measurement meets specified accuracy or precision)
- 3 = Questionable (measurement might not meet specified accuracy or precision but probably is useful for some / many purposes)
- 4 = Bad (measurement has no interpretive value and value is not reported)

For surface water  $f\text{CO}_2$  data with a WOCE flag 3, a set of sub-QC flags is used by several groups to indicate the problem with the particular measurement and to provide the user with additional information if the data are to be used, where 1 = outside standard range; 2 = questionable/interpolated sea surface temperature; 3 = questionable equilibrator temperature; 4 = anomalous  $\Delta T$  ( $\pm 1^\circ\text{C}$ ); 5 = questionable SSS; 6 = questionable pressure; 7 = low equilibrator gas flow; 8 = questionable air values (e.g. possible stack gas contamination); 9 = interpolated standard; and 10 = other, see metadata.

#### Discussions:

It was recognized that the vast majority of users of the SOCAT products will never refer to the data flags, but will instead accept the data that this group says are useable. The

system of flags should be as simple as possible. After discussing the proposed structure, the group agreed to the following flag system:

#### I. Cruise Flags:

Category A (11): A good cruise:

- 1) followed approved methods / SoP criteria
- 2) metadata documentation complete
- 3) 2<sup>nd</sup> level QC performed and deemed acceptable

Category B (12): An acceptable cruise:

- 1) followed methods / SoP criteria
- 2) metadata documentation complete
- 3) no 2<sup>nd</sup> level QC could be performed because data were obtained in a region or time where no other data were available for comparison

Category C (13): An acceptable cruise:

- 1) did not follow methods / SoP criteria
- 2) metadata documentation complete
- 3) 2<sup>nd</sup> level QC performed and deemed acceptable

Category D (14): Un-documented data

- 1) inadequate information about methods / procedures
- 2) metadata documentation incomplete

Category F (15): (F for “failure”) Failed QC checks (do not show in data set).

The initial status of data contributed to data set that has not yet undergone 2<sup>nd</sup> level QC would be labeled as “N” (19). These category labels will later be given numerical values but in such a way that they will not be confused with WOCE flags. Suggested numerical values are between brackets.

#### II. Data Flags:

The group agreed to use the WOCE flags: 2 = Good, 3 = Questionable, and 4 = Bad for individual data points. The group decided not to retain the sub-flag system use by some groups in their internal QC. The group agreed that only fCO<sub>2</sub> data flagged as 2 and 3 would be retained in the SOCAT data products, while level 4 data would be eliminated.

### **2.3 Crossover and Other Checks for Underway Data (U. Schuster and A. Olsen)**

Ute Schuster provided an overview of crossover analyses carried out for the *MV Santa Maria* and *MV Falstaff* lines, showing results of comparisons between the ship values and NCEP / NCAR reanalysis products for SST, sea-level pressure, wind speed, and air temperature. Schuster’s results showed that the comparisons involve a lot of work. Other parameters may be used depending on how much time the PI has to put into the analysis.

Are Olsen presented examples of crossover analyses from the *Nuka* and *Skogafoss* lines, and described the automated crossover analyses used in the CARINA project. He noted that much of the work can be automated, and Matlab routines developed for the CARINA

project may facilitate this process. Tests were carried out that showed that the automated scripts work just as well as user-supervised analyses, but expert judgment on the quality of the crossovers remains essential. Olsen suggested that the automated routine used for CARINA may be adapted to manage surface  $f\text{CO}_2$  crossovers, although interpretation of the results will be much more difficult than for inorganic carbon data from below 2000 m due to much larger spatial and temporal variability of  $f\text{CO}_2$  in the surface ocean.

Olsen suggested that another approach for  $f\text{CO}_2$  data comparison would be to establish MLR algorithms for  $f\text{CO}_2$  as a function of temperature, mixed-layer depth (such as from the FOAM UK Met Office reanalysis), and Chlorophyll a (from SeaWiFS and MODIS) and check observations against calculated  $f\text{CO}_2$  ranges for the period and region. He emphasized that this MLR technique is only intended to detect offsets that would require further investigation.

#### Combined discussions on 2<sup>nd</sup> level QC procedures:

- Comparison of atmospheric  $\text{CO}_2$  - It was agreed that a valuable first check would be to compare the available shipboard atmospheric  $\text{CO}_2$  measurements (especially  $x\text{CO}_2$  in dry air) with GlobalView data. The group agreed that these data should be provided as a separate data set for QC purposes, but not included in the surface  $f\text{CO}_2$  data set itself. If the measured atmospheric  $x\text{CO}_2$  is systematically biased, this could mean that calibration is off and that  $f\text{CO}_2$  in water has a similar bias.
- Variations in  $f\text{CO}_2$  normalized to a constant temperature - Another option for data where the  $f\text{CO}_2$  variability is largely temperature driven is to compute  $f\text{CO}_2$  at a constant temperature and look at the seasonal variations in a given area. This would make it possible to use cross-over analysis to identify major offsets and then to do a closer inspection to see if the variability is real or a data problem. This is something that is best done by the regional groups.
- Use Multiple Linear Regressions (MLR) – It may be useful to perform a MLR on all the data in a given area to look at the residuals on a cruise-by-cruise basis to see if anything stands out. The regression could be on temperature or salinity, and this could be done regionally in an automatic way on the already-assembled data set. This would not be particularly useful in data sparse areas, and it would probably be time consuming to prepare the code, although it is not necessary to worry about the precision of the regressions since the goal is simply to identify flyers. MLR will need to be examined closely for each region to determine the applicability to a particular area, the data or data products available, and the amount of work involved.
- The group also agreed that crossover analyses might be useful in certain regions, if much of the work could be automated as was done for CARINA. The Matlab scripts used by CARINA are a good starting point, but they would need to be adapted to surface  $f\text{CO}_2$ . It is likely that the crossover results will have to be interpreted manually.
- The cruise flag will indicate which data have been examined by 2<sup>nd</sup> level QC. Accompanying comments will describe the procedures followed.

Based on discussions of 1<sup>st</sup> and 2<sup>nd</sup> level QC procedures, the group summarized procedures remaining to be carried out:

Additional 1<sup>st</sup> level QC procedures

- Check unrealistic ship speed to find GPS failures.
- Check  $\Delta T = \text{inlet temperature} - \text{equilibrator temperature}$  (size, sign as expected, spikes) (automated), e.g. 2°C as upper limit for flagging. Plot equilibrator temperature versus sea surface temperature; good way to find outliers. Put negative flag on any  $f\text{CO}_2$  for high  $\Delta T$

List of priorities for 2<sup>nd</sup> level QC

- Atmospheric  $\text{CO}_2$  data (especially in open ocean regions, e.g., the Southern Ocean) (co-locate with GlobalView data, gridded/interpolated) within spline fit. (automate) (LAS to co-locate) (within 2-3 ppm), use with care (regional groups to look at differences). Shipboard atmospheric  $x\text{CO}_2$  will be available as a companion file to SOCAT in future versions of the data set.

Suggested regional and experimental tools for 2<sup>nd</sup> level QC

- Check sea surface temperature (SST) (reanalysis products, satellite, CTDs) (automate)
- Look at  $f\text{CO}_2$  at a constant temperature (subtropical, only works in some regions). Plot data as a function of time after correcting for temperature, should be within 40  $\mu\text{atm}$  bandwidth (at least in subtropics of Pacific) (visual, lots of work).
- MLR (multiple linear regression) approaches using SST, salinity, nitrate or mixed layer depth. Plot MLR as function of sea surface temperature and salinity, find outlier cruises. Likely to work in temperature controlled regions, e.g., the subtropics. Note: this approach will not be applicable to all regions.
- Crossovers to help identify potential issues (adopt CARINA procedures). This approach may not work very well for highly variable surface water  $f\text{CO}_2$ .
- Despiking, dangerous, no one keen

ACTION ITEM 2: Develop a separate data set of  $x\text{CO}_2$  in dry air and  $f\text{CO}_2$  in air to compare with long-term atmospheric monitoring stations and GlobalView. This is not a top priority and should be left to future SOCAT versions (Pfeil, Olsen, Hankin, Malczyk).

ACTION ITEM 3: Test the MLR approach in one area and assess this approach. (Regional groups)

## 2.4 Quality-Controlling Non-Carbon Data (B. Pfeil)

Benjamin Pfeil referred to his earlier presentation on the status of the data set and indicated the 1<sup>st</sup> level QC checks that were applied to the non-carbon variables already. Much more remained to be done for the QC of the non-carbon data (see list above), but other commitments have prevented him from dedicating much time to this.

### Discussions:

The group agreed that many of these activities are time consuming. Automating these activities as much as possible is critical. It was suggested that we could make more use of the LAS system, and that the group needed to develop a list of the checks to be done for the non-carbon data. It was suggested that a toolbox should be created and made available such that individual investigators can use it to send clean data to the data center. However, some 1<sup>st</sup> level QC will always need to be done at a higher level (e.g., the data assembly center), since it will ensure a more coherent global data set.

## **2.5 Live-Access Server and Shared Data Management Strategies for Ocean Carbon Observations (S. Hankin and J. Malczyk)**

Steve Hankin provided a demonstration of the LAS for the Takahashi 2008 collection, <http://ferret.pmel.noaa.gov/OCDMS/index.htm> and an initial preview of the SOCAT data version 1.0. He noted that the server was working, although there were still some bugs to track down and the latest version of the data set needed to be entered. He also provided an example of how LAS might be used to compare data from the Takahashi climatology and the SOCAT data set. Of particular interest were the generation of on-the-fly anomaly maps that may be generated by comparing observations to a climatology or other data set. (On the fly means that a smaller data set can be extracted online from a larger data set.) Finally, Hankin provided some thoughts on what would be possible with a shared data set:

1. Support for 2<sup>nd</sup> level QC:
  - Able to download by region – e.g., N. Pacific; regions can overlap;
  - Able to upload flags, altered (recalibrated) values, QC “scripts” that may be applied to other regions;
  - Able to track additions to metadata - who uploaded? when? what was altered?
  - Able to add new fields (e.g., wind shear)
2. “Releases” and/or “tags” (versions)
  - major (external) and minor (internal)
  - each observation can be audit-tracked
  - compare scientific conclusions by version
3. Comparison support
  - between cruises on repeat tracks
  - time-track visualizations and anomalies
  - cross-over points
4. On-the-fly gridding

Given an agreement on gridding techniques, LAS can create gridded fields on the fly:

- Choose the data subset and constraints (e.g., Central Pacific Spring, excluding El Niño years)
- Choose gridding technique and parameter values
- Choose grid resolution

### Discussions:

It was noted that initial comparisons between the Takahashi climatology and the SOCAT data set reveal that Takahashi has significantly more Pacific data. It was not clear if those data are publicly available, but have simply not yet been contributed to CDIAC. LAS may help us to identify data sets missing in SOCAT.

The group needs to decide on definitions for the regions to allow LAS to have a drop-down menu to extract data by region. For the coastal areas, the definition is done by bathymetry (<200 meters is defined as coastal). The LAS and SOCAT data set have the necessary bathymetry data to do this. The group must also decide what to do with cruises that cross from one region to another, or from the open ocean to the coastal zones.

**ACTION ITEM 4:** Decide on regional definitions and how to deal with cruises that cross boundaries (Regional groups, Pfeil, Olsen, Hankin, Malczyk)

## **2.6 Record-Keeping for Secondary QC Procedures (R. Key)**

Bob Key advised the group to record everything, or at least enough so that someone 10 years from now can figure out exactly what was done to the data. This includes

- Origin of the data
- Data sets included, with appropriate cruise names or code numbers.
- Methods used – a formal publication-quality level of detail including methods, software used, etc.
- Detailed results about how the data were treated, including a description by data set with an emphasis on anomalies, tabulated results of all tests with supporting graphics, a consistent sense of the offset by all data contributors, and a consistent offset type (additive or multiplicative) used by all, if applicable.
- Where applicable, robust adjustments applied to data, including a publication quality table (web or manuscript) that includes EXPCODES and internal identifiers for automated application, a special code to indicate data that were not tested (which must be different from the indicator for data that have been tested, but no adjustment recommended), and a “Master” cruise indicator if used.
- Summary Documents – white papers with as much detail as practical, and publication level summary.

### Discussions:

In response to questions about how to deal with errors found in the data set, Key advised that the PI should be contacted and asked to fix it and to resubmit the data. The original data should always be kept unchanged.

## **2.7 Gridding and Interpolation Procedures (C. Sabine and U. Schuster)**

Chris Sabine and Ute Schuster led an open discussion about the basic gridding procedures the group would use. They reminded the group that the goals set out by the 2007 SOCOVV meeting were to develop both a regularly updated 2<sup>nd</sup> level QC data set

and a monthly 1° x 1° gridded data product with no interpolation. In discussions with potential users of the data, this gridded product was agreed to be more useful for more people than an interpolated product, and to remain truer to the original data. It was agreed that the data set would not include CO<sub>2</sub> air-sea fluxes, since PIs will calculate this in different ways.

The binning statistics to the 1° x 1° gridded monthly data product may need some careful consideration to determine whether some weighting scheme should be used to differentiate between data from a single cruise that made measurements every minute versus a more sparse data set that made measurements every hour. It was suggested that some information about the fraction of the month that is covered by the different data sets would also be useful. There needs to be some information for all bins that will allow a user to identify, for example, the number of data points, the number of cruises and the standard deviation of the data.

The group stressed that the 2<sup>nd</sup> level QC data set need to be clearly identified in LAS with respect to version number and level of QC, so that someone using the LAS knows exactly which version of the 2<sup>nd</sup> level QC data set was used.

**ACTION ITEM 5:** Develop a recommended procedure for providing information about binning (e.g., the number of data points, the number of cruises and the standard deviation of the data.) (Sabine, Schuster, Wanninkhof)

## **2.8 Data Sharing Issues (C. Sabine, B. Pfeil)**

Drawing from the status report provided by Benjamin Pfeil, Chris Sabine led a discussion on the data currently included in the SOCAT data set. Some data provided to Takahashi for the climatology has not been approved for public sharing, citing a two-year embargo period that is more-or-less traditional for research programs. Some CarboOcean PIs have adopted this two-year period, while other CarboOcean data are released more quickly.

With this in mind, the SOCAT data set will be approximately 2 years behind for some data (e.g., the data set for 2<sup>nd</sup> level QC to be published in late 2009 will have all available data up to the end of 2007). The group agreed that it would be helpful to set a goal for science publications that would come from using the SOCAT data set, and to work back from that goal in order to determine at what point we will freeze the data for 2<sup>nd</sup> level QC for the first version. While ambitious, the group agreed that the 8<sup>th</sup> International Carbon Dioxide Conference (13-19 September 2009, Jena, Germany) would be the best opportunity. The group agreed that all data to be included in the first-release SOCAT data set should be sent to CDIAC no later than 1 September 2008, and the 2<sup>nd</sup> level QC data set will be frozen on 15 September. The data set for 2<sup>nd</sup> level QC will then be sent to LAS, which can do the gridding on-the-fly to any specified grid size.

The group agreed it would be desirable to continue acquiring new data as they become available after the 1 September 2008 deadline in order to have them ready for the next

version of the data set. It would be best to carry out 1<sup>st</sup> level QC on the data as soon as possible after submission and to contact data contributors about any problems.

Benjamin Pfeil also pointed out that many data sets may be publicly available, but are not yet in CDIAC. A special effort will be made before the 1 September 2008 deadline to identify those cruises and contact the data contributors as soon as possible.

ACTION ITEM 6: Benjamin Pfeil, Maria Hood and Alex Kozyr will identify data contributors whose data are probably publicly available, but not yet at CDIAC.

### **III. REGIONAL GROUP REPORTS**

The regional group chairs established at the 2007 SOCOVV meeting were asked to provide an update on regional group activities, including members of the group, identification of missing data sets in their regions, possible approaches to 2<sup>nd</sup> level QC and scientific synthesis priorities.

#### **3.1 Indian Ocean Group Report (N. Metzl for V.V.S.S. Sarma)**

Group members:

V.V.S.S. Sarma (Chair, NIO, India), A. Murata (JAMSTEC, Japan), C. Sabine (PMEL, NOAA, USA), N. Metzl (Paris, France), C. Goyet (Univ. Perpignan, France), B. Tilbrook (CSIRO, Australia)

Data set identification:

The group has not yet examined the SOCAT version 1.1 data set for missing data.

Scientific synthesis issues and priorities:

The Indian Ocean is very much undersampled compared to the rest of the world ocean (Annex 4). Seasonal sampling has only been carried out for one year in the Arabian Sea (1995), which happened to be an unusually warm year. Good seasonal and inter-annual coverage exists at the French time-series station in the southwestern Indian Ocean. The rest of the region (more than 90%) has been sampled mostly only once or twice. Interannual variability of  $f\text{CO}_2$  in the Indian Ocean is completely unknown. The work of Nicolas Metzl suggests that there is a significant effect of the Indian Ocean Dipole (IOD) on  $\text{CO}_2$  parameters, but the nature and mechanisms of the variability are unknown. Sarma has applied a simple MLR model to understand the influence of the IOD on  $\text{CO}_2$  fluxes and primary productivity (PP) and observed that the fluxes were decreased by about 30% during an IOD year. In the case of Bay of Bengal, nothing much known is at all, including how river runoff affects dissolved inorganic carbon cycling, how biological controls (such as PP and bacterial respiration), change due to river runoff, etc. Hence, little is known about how climatic events and land-driven processes influence different factors controlling  $f\text{CO}_2$  in the Indian Ocean.

Dr. Murata of JAMSTEC reports that the Japanese vessel, *R/V Mirai*, is scheduled to visit the eastern equatorial Indian Ocean every year for servicing or deployment of TRITON buoys, and says that it may be possible to install an underway-surface fCO<sub>2</sub> system on the ship for these cruises. Sarma has proposed a couple of cruises in the Bay of Bengal during the next few years. The Japanese vessel *Hakuo Maru* is also arriving in the Indian Ocean this year, and will make a deep meridional section across the Indian Ocean from 15°N to Antarctica.

Murata and Sabine have observations of several new repeat hydrography sections in the Indian Ocean, and comparisons with the 1995 WOCE cruise sections will allow for an examination of how the Indian Ocean is responding to the anthropogenic forcing.

A new time-series station in the central Bay of Bengal will be operational by the end of this year and is expected to continue for another decade. A coastal time series in the Bay of Bengal has already been operating for the past 9 months. Carbon measurements are planned in the Arabian Sea time series. Therefore, it would be interesting to compare time-series variations of carbon parameters in the northern (Arabian Sea and Bay of Bengal) and southwestern Indian Ocean (French time series).

While we may expect better seasonal and spatial coverage, it will not be possible to create a 1° x 1° grid with the current data coverage.

The following key questions should be addressed in the future to better understand fCO<sub>2</sub> cycling in the Indian Ocean:

1. The highest priority should be given to examine seasonal and inter-annual variability of fCO<sub>2</sub> in different regions such as the Bay of Bengal, the equatorial Indian Ocean and the South Indian Ocean.
2. Efforts should be focused towards examining the influence of Indian Ocean Dipole on surface pCO<sub>2</sub> and the influence of different processes such biological, thermal, and mixing effects, as well as air-sea exchange.
3. More in situ buoys, such as NOAA/PMEL, should be moored in different locations to cover temporal variability.

### **3.2 Pacific Ocean Group Report (R. Feely)**

Group members:

- Equatorial and tropical Pacific Group (30 °S- 30 °N): Dick Feely (Chair).
- South Pacific Group: Bronte Tilbrook (Chair).

Other sub-regions and group members are not yet established.

Data sets identified:

The group has not yet examined the SOCAT version 1.1 data set for missing data.

Scientific synthesis issues and priorities:

For the Pacific Ocean, the goal of surface CO<sub>2</sub> measurement projects is to quantify the daily-to-interannual and decadal variability in air-sea CO<sub>2</sub> fluxes and understand the mechanisms controlling these fluxes. The approach used is to make autonomous surface fCO<sub>2</sub> measurements using research and volunteer observing ships (VOS) to get spatial coverage at seasonal time scales and use a network of surface moorings to get high-frequency temporal resolution. There are currently 7 ships in the Pacific (variable from year to year) with an additional 7 open-ocean moorings and 1 coastal mooring. NOAA, LDEO, JMA, IOS, and CSIRO have been maintaining fCO<sub>2</sub> systems on research and VOS ships in the Pacific Ocean for many years.

### **3.3 Atlantic Ocean Group Report (U. Schuster and N. Lefèvre)**

Group members:

Ute Schuster and Nathalie Lefèvre (Co-Chairs): other group members not yet established.

Data set identification:

The group has not yet examined the SOCAT version 1.1 data set for missing data.

Scientific synthesis issues and priorities:

In the North Atlantic, the priorities are:

- to identify variability on seasonal, annual, and interannual variability in CO<sub>2</sub> air-sea fluxes to a higher precision as done so far, and
- to identify the underlying causes for the variability, including changes in e.g. sea surface temperature, biological activity, circulation changes, and the North Atlantic Oscillation .

The North Atlantic is an ocean basin where a high number of measurements are available now, and continuous measurements on research vessels, VOS, moorings, and buoys is continuing. However, some geographical regions, as the south-eastern tropical region, are undersampled in comparison with others, making basin-wide estimations still uncertain. Where sufficient measurements are available, the MLR technique will be used to identify flyers. Additionally, with a relatively high density of measurements (compared to other ocean basins), the use of cross-over analyses will be explored as a tool for quality control.

In the equatorial and Southern Atlantic, data coverage is still low , and 2<sup>nd</sup> level quality control will be very difficult.

### **3.4 Southern Ocean Group (N. Metzl for B. Tilbrook)**

Group members:

Bronte Tilbrook and Nicolas Metzl (Co-Chairs): other group members not yet established.

Data set identification:

The SOCAT data set version 1.1 (April '08) has been explored for the Southern Ocean (south of 30°S). Not surprisingly, compared to Northern Hemisphere, the S.O. is clearly undersampled and large regions have never been sampled (or the data are not yet in SOCAT, see Annex 4, figures 1 and 2). Most of the cruises in S.O.-SOCAT have been conducted since 1986 (only 11 cruises reported before 1984). During the austral winter season (July-September), SOCAT contains very few cruises, especially south of 50°S. However, it is recognized (Figure 1) that three sectors have been better sampled: Drake Passage (Western Atlantic sector), the sections between Tasmania/New-Zealand and Antarctica (Western Pacific sector, including the Ross Sea) and regions around sub-Antarctic French Islands (Western Indian sector).

QC issues:

For QC issues, five topics have been explored using SOCAT, that may also help in selecting/producing QC1 and QC2 for other regions:

1. Detecting bad or missing data and information in the data base: for example, Julian day not included for all cruises, the number of digits for hydrography incorrect for some cruises, etc.. This should be checked as QC1 for all cruises, not only the S.O. region. The easiest way would be to ask all contributors to check their data (e.g. by selecting name of cruises in SOCAT) and send back comments to B. Pfeil.
2. Check differences between Equilibrium Temperature and SST: this is especially important in cold waters of the Southern Ocean as well as in large SST gradients when crossing frontal zones. In the SOCAT data set we identified some cruises that present very large differences between Equilibrium Temperature and SST (differences between 5°C and 20°C). It has been noted that for some of these cruises, recomputed fCO<sub>2</sub> is not reported; therefore these data should be deleted. Such comparison (plots Teq versus SST) should be performed for all regions.
3. Comparison of atmospheric xCO<sub>2</sub> recorded on board to independent atmospheric xCO<sub>2</sub> data: this is especially interesting in the Southern Ocean as the spatio-temporal variability of atmospheric CO<sub>2</sub> is relatively low in the southern hemisphere. Examples of such comparisons have been presented (south Indian region), which help verify the quality of the standards and data processing as well as long-term analysis. The comparison has been done using atmospheric xCO<sub>2</sub> data from monitoring stations (using the original xCO<sub>2</sub> data, not using the GlobaView 2D (time/latitude) interpolated product). It has been noted that for some cruises such a comparison is difficult, mostly because of pollution by the ship.
4. Identifying intercomparison cruises: For the S.O., four intercomparisons have been identified.
  - Several groups that provided data to SOCAT (S.O. region) were involved in the intercomparison cruises onboard *Meteor* (1996, see section 2.1 in this

report; also Körtzinger et al., 2000). During this cruise, the differences between systems from CSIRO (Australia), MRI (Japan) and IPSL (France) was about +/- 2  $\mu\text{atm}$ .

- Intercomparison conducted during a cruise on the *Aurora Australis* (2003, south of Tasmania) between CSIRO (Australia) and Ulg (Belgium) systems: difference was about +/- 4  $\mu\text{atm}$ .
- Intercomparison conducted during OISO-8 cruise on the *Marion-Dufresne* (Dec-1998, South Indian sector) between systems of IPSL (France) and Ulg (Belgium) : difference about +/- 5  $\mu\text{atm}$
- Intercomparison conducted onboard *L'Astrolabe* (Jan-2008, south of Tasmania) between CSIRO ("old system" Specify type of the old system. Eg. Infrared, McNeil, equilibrator) and the GO-CSIRO ("new instrument". Specify type of this new system. What is the main difference with the old system.): difference 0.2 +/- 2.2  $\mu\text{atm}$ .

These results are in the range (4-5  $\mu\text{atm}$ ) of the uncertainty for most surface  $\text{pCO}_2$  data (see discussion in section II of this report). The good results obtained during the recent comparison in January 2008 with the GO system (now used by many groups) suggest that such intercomparisons should be performed more regularly (when possible) to verify "old systems" and thus help to verify historical data in SOCAT.

5. Identifying Cross-over for intercomparison: this concept is tricky (see report SOCAT-I meeting), and we think it is especially tricky for the S.O. because of high mesoscale variability in frontal zones and in the Seasonal Ice Zone (where large daily changes and small-scale, variations (i.e. few km), are common during austral summer ). Using Cross-over analysis for QC is not recommended for the S.O. region.

Scientific synthesis issues and priorities:

The scientific issues have been separated following temporal scale (here referring mostly to data-based analysis).

#### *Past and recent past activities*

Seasonality: In the sub-Antarctic zone and Polar Front zone, several analyses have been published over the last ten years, and in these sectors, the  $\text{pCO}_2$  seasonality is relatively well known. At higher latitudes (south of the Polar Front) the seasonality, as well as the climatology of  $\text{pCO}_2$ , have been improved only in recent years. This is mainly because of new data collected during the austral winter season (high  $\text{fCO}_2$ ). Although winter data are sparse, a simple interpolation of SOCAT recomputed  $\text{fCO}_2$  data clearly shows that  $\text{fCO}_2$  is higher south of  $50^\circ\text{S}$ . On an annual scale, the summer  $\text{CO}_2$  sink is balanced by the winter source. Recent estimates suggest that the net annual air-sea  $\text{CO}_2$  flux is low (range 0-0.2  $\text{PgC/yr}$  south of  $50^\circ\text{S}$ ) compared to previous data-based and modelling approaches.

Interannual variability: a few studies analyzed the interannual variability in the Southern Ocean. These were mainly based on cruise to cruise comparison (e.g. two periods) and chosen to highlight and understand large pCO<sub>2</sub> anomalies (in summer only and often associated to warm anomalies and biological activity at local/regional scales). A synthesis of 22 cruises (1991-2003) has been recently composed to evaluate the interannual variability of pCO<sub>2</sub> in the subtropical and sub-Antarctic regions south of Tasmania; this analysis suggests that pCO<sub>2</sub> anomalies are likely related to large-scale climate indices (SAM, ENSO).

Decadal variability: detecting the decadal change (or stability) of ocean pCO<sub>2</sub> obviously needs long-term and accurate observations. Only very recently some groups (in Japan, France, USA and New-Zealand) have been able to explore the decadal trends of ocean pCO<sub>2</sub> and compare with the rate observed in the atmosphere. In short, all studies show that ocean pCO<sub>2</sub> increases in the S.O. but the rates are different depending the location and period. The decadal rates deduced from the data range between 10 to 20 µatm/decade in the region 30°S-50°S, and between 18 to 24 µatm/decade south of the Polar Front.

#### *Present and near future activities*

Although the results listed above indicate encouraging progress in the S.O. regions, it is worth recalling that most of these analyses were conducted at local or regional scales. It is expected that the SOCAT data set will help to investigate the spatio-temporal variability of pCO<sub>2</sub> at a basin scale. In this context several issues have been identified:

Re-Investigate the seasonal variability in all S.O. regions, especially at high latitudes including the complex seasonal ice zone where pCO<sub>2</sub> temporal variability (when observed) is very high, both rate (~ order of 1 week) and size of change (> 50 µatm). A better understanding of the processes controlling the seasonal variability in all sectors is also needed to improve biogeochemical parameterization of ocean models.

Understanding the interannual to decadal variability: Recent studies based on atmospheric CO<sub>2</sub> observations as well as ocean modelling suggest that the Southern Ocean CO<sub>2</sub> sink has been saturated for about two decades (or is no longer keeping pace with atmospheric CO<sub>2</sub> increases). This is due to an intensification of the Southern Annular Mode (SAM) which increases ventilation of carbon-rich deep water. Detecting the interannual to decadal change of the S.O. carbon uptake from in-situ observations and at basin scale is challenging, as the data are sparse. The few studies that have documented decadal trends of pCO<sub>2</sub> (see above) show that the rate of increase in surface water pCO<sub>2</sub> is not far from the atmospheric change (given the uncertainties). Both the SOCAT data set and derived accompanying monthly Atlas will help to identify key regions where such a decadal analysis could be performed and compared. In this context, we also expect to separate the anthropogenic signal and natural or climate induced variability. For this, an important new goal would be to create a synthesis of both DIC and TA data that have been also measured in the surface waters during several cruises.

Interpolation and extrapolation methods: As we expect to estimate air-sea CO<sub>2</sub> fluxes at basin-scale, for both seasonal and long-term analysis, interpolation/extrapolation methods need to be investigated in the S.O. where data are sparse in space and time. Unfortunately, simple regression techniques (e.g. pCO<sub>2</sub> versus SST, Chl-a, etc.) are not able to reconstruct pCO<sub>2</sub> fields in the S.O. Indeed, it has been shown that there is no simple relationship with SST (especially at high latitude). Methods based on MLR approaches or neural networks (NN) that used SST, Chl-a, MLD, etc, have been tested in the North Atlantic, but not in the S.O.; these are attractive approaches but cloud cover often limits the use of satellite products in the S.O. (e.g. Chl-a). Therefore, new approaches have to be developed or adapted (e.g. linking pCO<sub>2</sub> data with altimetry).

High winds and changes in the winds: an important source of uncertainty in CO<sub>2</sub> flux estimates from observations is the gas transfer velocity, especially at the high wind speeds of the Southern Ocean. It is expected that recent experiments conducted in the Southern Ocean (e.g. GasEx III, 2008) will offer new results and reduce the uncertainties attached to the gas exchange coefficient. The effects of sea ice dynamics on air-sea CO<sub>2</sub> fluxes are poorly known.

Other topics for the S.O. group: Based on the SOCAT data, it should be possible after selecting specific data to address new questions relevant for the S.O. region: the role of mesoscale (eddies, fronts), the role of sea-ice, comparing naturally fertilized areas (Islands in circumpolar regions, sea-ice regions), identifying large pCO<sub>2</sub> anomalies as a proxy in Climate change, etc.

### **3.5 Coastal Regions (A. Borges and A. Chen)**

Group members:

Alberto Borges and Arthur Chen (Co-Chairs); Potential group members contacted include Abdir Omar, Arthur Chen, Bruno Delille, Burke Hales, Dorothee Bakker, Fiz Fernandez Perez, Francisco Chavez, Helmuth Thomas, Joe Salisbury, Magdalena Santana-Casiano, Minhan Dai, Nick Bates, Nick Hardman-Mountford, Simone Alin, Ute Schuster, Vedula V.S.S. Sarma, Wei-Jun Cai.

Several more potential candidates have been identified during SOCAT-2.

Data set identification:

The group has not yet examined the SOCAT version 1.1 data set for missing data.

Scientific synthesis issues and priorities:

One of the major scientific issues to be addressed by an enhanced global coastal data set is to reconcile opposing views on carbon cycling in the coastal ocean; namely, that continental shelves act as sinks and near-shore ecosystems act as sources of CO<sub>2</sub> to the atmosphere. The direction of the flux changes both spatially and seasonally, and the controls on this are not fully understood. The impact of rivers is of particular interest.

Quality control of coastal data is even more complicated than in open oceanic waters. There are large areas of the world's coastal zones that do not have data. Where data do exist, they are often of poor quality. In many cases, very simple measurement techniques are used, and many scientists working in these areas do not know the difference between  $x\text{CO}_2$  (mixing ratio) and  $f\text{CO}_2$ , and do not record atmospheric pressure or humidity. Mooring data indicate that typical day-to-day variability in  $f\text{CO}_2$  data is up to  $50 \mu\text{atm}$ , and that sub-daily variability can be up to  $25 \mu\text{atm}$  (due to tidal advection of water masses with different chemical signatures). This will make the cross-over approach very difficult or impossible. Atmospheric  $x\text{CO}_2$  cannot help due to continental influence. One way to QC the data is to use quality checked open oceanic data and assume the same quality on the coastal part of the transect. A simple averaging approach without interpolation of data, as decided for the open ocean regions, is also recommended for the coastal region. A  $1^\circ \times 1^\circ$  resolution of grids is considered too coarse, while a  $1/4^\circ \times 1/4^\circ$  resolution of grids is assumed over-ambitious. Hence, a grid resolution of  $1/2^\circ \times 1/2^\circ$  seems a good compromise. LAS should be able to handle  $1^\circ \times 1^\circ$  grids for the open ocean and  $1/2^\circ \times 1/2^\circ$  grids for the coastal ocean. It is also recommended that LAS should be able to filter data for depths 0 m to 200 m to extract the coastal ocean data.

**ACTION ITEM 7.** Regional group chairs will establish their groups (Atlantic, Pacific sub-regions, Southern) and all groups will look through the SOCAT version 1.1 data set to identify data sets that are missing and provide that information to Pfeil. Pfeil, Kozyr, and Hood will contact data contributors to include those data in the SOCAT data set before the 1 September deadline.

#### **IV. SUMMARY AND ACTION ITEM LIST**

The group offered its congratulations and thanks to Benjamin Pfeil and Are Olsen for the extraordinary amount of work they have done to get the SOCAT data set to its present stage. The next steps will require input from the regional groups, including identifying missing data sets and looking at the data to see which 2<sup>nd</sup> level QC checks may be performed in each region (deadline 1 September 2008). The regional groups are asked to carry out 2<sup>nd</sup> level QC after this date, aiming for a first release of the 2<sup>nd</sup> level QCed SOCAT data set and SOCAT gridded product by late 2009. The above will require a significant time commitment on the part of the regional groups, and it is not clear that there is either sufficient time or commitment to do this adequately. Having funds to support working group meetings would be extremely helpful, but is unlikely before September 2008. It was also strongly suggested that the groups should identify new science that can be generated from the data set, both regionally and globally, to encourage scientists in the regional groups to put in effort to the QC procedures, and to work towards strong scientific SOCAT presentations (e.g., at the ICDC-8 Conference in September 2009) and publication.

**ACTION ITEM 1:** Coordinate with OceanSITES on the Platform Names / Codes for time series stations (Pfeil, Olsen, Kozyr, Hood).

ACTION ITEM 2: Develop a separate data set of  $x\text{CO}_2$  in dry air to compare with long-term atmospheric monitoring stations and GlobalView. The  $f\text{CO}_2$  in air and the difference between  $f\text{CO}_2$  in air and water should be added to the main data set eventually, but should be left to future SOCAT versions (Pfeil, Olsen, Hankin, Malczyk).

ACTION ITEM 3: Test the MLR approach in one area to see how difficult this is. (Regional groups)

ACTION ITEM 4: Decide on regional definitions and how to deal with cruises that cross boundaries (Regional groups, Pfeil, Olsen, Hankin, Malczyk)

ACTION ITEM 5: Develop a recommended procedure for providing information about binning (e.g., the number of data points, the number of cruises and the standard deviation of the data.) (Sabine, Schuster, Wanninkhof)

ACTION ITEM 6: Benjamin Pfeil, Maria Hood and Alex Kozyr will identify data contributors whose data are probably open for data sharing but not yet at CDIAC.

ACTION ITEM 7. Regional group chairs will establish their groups (Atlantic, Pacific sub-regions, Southern) and all groups will look through the SOCAT version 1.1 data set to identify data sets that are missing and provide that information to Pfeil. Pfeil, Kozyr, and Hood will contact data contributors to include those data in the SOCAT data set before the 1 September deadline.

ANNEX 1  
Participants List

**Dorothee Bakker**

School of Environmental  
Sciences, University of East  
Anglia, UK  
D.Bakker@uea.ac.uk

**Alex Kozyr**

Carbon Dioxide Information  
Analysis Center, USA  
kozyra@ornl.gov

**Are Olsen**

Bjerknes Centre for Climate  
Research, Norway  
are@gfi.uib.no

**Alberto Borges**

University of Liège, Belgium  
Alberto.Borges@ulg.ac.be

**Arne Koertzinger**

Leibniz-Institut fuer  
Meereswissenschaften  
akoertzinger@ifm-geomar.de

**Benjamin Pfeil**

Bjerknes Centre for Climate  
Research, Norway  
Benjamin.Pfeil@bjerknes.uib.no

**Arthur Chen**

Institute of Marine Geology  
and Chemistry, National Sun  
Yat-sen University, Taiwan  
ctchen@mail.nsysu.edu.tw

**Nathalie Lefèvre**

IPSL/CNRS, France  
nathalie.lefevre@locean-  
ipsl.upmc.fr

**Denis Pierrot**

Rosenstiel School for Marine and  
Atmospheric Science NOAA, USA  
denis.pierrot@noaa.gov

**Richard Feely**

Pacific Marine Environmental  
Laboratory, NOAA, USA  
Richard.A.Feely@noaa.gov

**Andrew Lenton**

IPSL/CNRS, France  
andrew.lenton@locean-  
ipsl.upmc.fr

**Chris Sabine**

Pacific Marine Environmental  
Laboratory, NOAA, USA  
chris.sabine@noaa.gov

**Steve Hankin**

Pacific Marine Environmental  
Laboratory, NOAA, USA  
Steven.C.Hankin@noaa.gov

**Jeremy Malczyk**

Pacific Marine Environmental  
Laboratory, NOAA, USA  
Jeremy.Malczyk@noaa.gov

**Ute Schuster**

School of Environmental Sciences,  
University of East Anglia, UK  
U.Schuster@uea.ac.uk

**Truls Johannessen**

Bjerknes Center for Climate  
Research, Norway  
truls@gfi.uib.no

**Nicolas Metzl**

IPSL/CNRS, France  
Nicolas.Metzl@locean-  
ipsl.upmc.fr

**Rik Wanninkhof**

Atlantic Oceanographic and  
Meteorological Laboratory,  
NOAA, USA  
Rik.Wanninkhof@noaa.gov

**Robert M. Key**

Princeton University, USA  
key@princeton.edu

Unable to attend: VVSS Sarma (NIO, India); Taro Takahashi (LDEO, USA); Bronte Tilbrook (CSIRO, Australia).

ANNEX 2  
Agenda

Monday, 16 June 2008

0900 - 0910	Welcome, Opening, Announcements (Maria)
0910 - 0930	Overview of SOCAT-1 Meeting and decisions for 1 <sup>st</sup> level QC data set; goals for SOCAT-2 (Dorothee)
0930-1030	Overview of actions from 1 <sup>st</sup> level QC and discussions (status of level 1 QC data set; treatment of unrealistic values; input on draft technical document, Matlab files, metadata, public data not to be included, etc.). (Benjamin / Are )
1030-1100	<i>Break</i>
1100-1130	Overview of issues for 2 <sup>nd</sup> level QC data set (need for consistent data set, assessment of uncertainties per cruise, flags, crossovers, etc.) (Benjamin / Are / Dorothee, etc.)
1130-1200	Assessment of uncertainty per cruise / getting data to a certain level of certainty (Chris, Arne)
1200-1230	Flags (Denis/ Rik)
1230-1400	<i>Lunch</i>
1400-1430	Crossover checks (Are, Ute, others with experience)
1430-1500	Quality controlling non-carbon variables (Benjamin)
1500-1530	Tools in the Live Access Server (Steve / Jeremy)
1530-1600	<i>Break</i>
1600-1630	Guidelines for documenting 2 <sup>nd</sup> level QC (Bob)
1630-1700	Gridding and interpolation procedures for gridded data set (Chris, Ute)
1700-1730	Data sharing issues ? What should be included in this 1 <sup>st</sup> version of the data set with 2 <sup>nd</sup> level QC? How should this data set be built upon in the future ? (Chris)
1730-1800	Summary of 2 <sup>nd</sup> level qc issues and open discussion (Dorothee)

Tuesday, June 17, 2008

0930 -1000	Overview of Day 1 decisions (Dorothee)
1000-1030	Regional QC and science issues (each group should identify major science issues, major data QC issues, describe experience with QC (if any) and who will carry out tasks.) (20 minutes presentation / 10 minutes discussion)
1030-1100	Atlantic (Ute and Nathalie)
1100-1130	<i>Break</i>
1130-1200	Pacific (Dick)
1200-1230	Indian (Nic to present report of Sarma)
1230-1400	Southern Ocean (Nic)
1400-1430	<i>Lunch</i>
1430 - 1530	Coastal Regions (Alberto and Arthur)
1530 - 1600	Issues for bringing it all together (Bob, Alex, Benjamin, etc.)
1600-1700	<i>Break</i>
1700-1730	Open discussions and summary decisions on 2 <sup>nd</sup> level QC (Dorothee)
1730	Dates for tasks, public release of 2 <sup>nd</sup> level QC, science meeting follow-up (Maria, Dorothee)
1800 - 2000	Close of meeting.
	Joint Reception SOCAT / CARINA; Miollis Coffee Bar

ANNEX 3  
Summary of SOCAT-1

*Summary of SOCAT-1 meeting, Bremen, 5 December 2007*

I. On the 1st level QC data set:

- Cut-off for data included in this 1<sup>st</sup> version data set on 31/12/2007. Freeze set of 1<sup>st</sup> level QC data, until 2<sup>nd</sup> level QC has been completed.
- Technical issues on the 1<sup>st</sup> level QC data set by Taro and Ray have been resolved.
- 1<sup>st</sup> level QC is not a full data QC. Only outliers in salinity and atmospheric pressure have been removed.
- Parameters in the 1<sup>st</sup> level QC data set: all the measured CO<sub>2</sub> parameters in water and air, equilibrator temperature, sea surface temperature, atmospheric pressure, as well as salinity from the World Ocean Atlas and atmospheric pressure from NCEP reanalysis, as well as fCO<sub>2</sub> calculated in a consistent way from other CO<sub>2</sub> parameters.
- Data will be co-located with bathymetry, especially useful for coastal regions.
- Naming of cruises with EXPO codes (country, ship, year, month, date of sailing).
- The data will be archived at CDIAC and will be made accessible to the data PIs and the regional groups for 2<sup>nd</sup> level QC.
- The data will be put onto a Live Access Server by the PMEL group before the spring meeting. The LAS will catch problems in the data. The LAS will be live during the meeting.
- Advise all data PIs of data calculation routines and a description of the methods in January 2008 with a month for comments. Distribution via an email from IOCCP, link on IOCCP website and an announcement in the IOCCP Newsletter.
- Ask all PIs if they agree that their data are included (Wanninkhof).
- Encourage the PIs to update the metadata (eg. publications).
- Any further issues?

II. Suggestions on the 2nd level QC:

- Regional groups have been established at the Paris meeting. The groups will lead the 2<sup>nd</sup> level QC and the science based on the data synthesis.
- Need for consistent 2<sup>nd</sup> level QC
- How to do this in practical terms?
- Assessment of uncertainty per cruise (Sabine).
- Get the data to a certain level of certainty (Körtzinger).
- Add flags to the data sets. Design a formal table for all flags, such that flags can be assigned in a consistent and transparent manner (Watson). Flags could be based on the accuracy of the analyser, the number of standards etc.
- Crossover concept tricky.
- How critical should one be?
- Identify unrealistic values
- Document everything!
- Use atmospheric xCO<sub>2</sub> for QC
- If regions are identified, use at least 10° overlaps (Key)
- 2<sup>nd</sup> level QC ready by September 2008

III. Meeting of the group leaders on approaches for 2<sup>nd</sup> level QC in May-June 2008:

- Main aim: Discuss approaches for 2<sup>nd</sup> level QC (see also section II above)
- Identify approaches for gridding/ interpolation. LAS may help.
- Identify science issues.

#### IV. Action items:

- 1) Circulate report of the Bremen meeting among participants. Ask absent group leaders for input (Maria Hood, Dorothee Bakker) (December 2007).
- 2) Contact PIs via IOCCP (January 2008) (Maria Hood, Dorothee Bakker, Are Olsen, Benjamin Pfeil).  
Make the items below available to data PIs (Benjamin Pfeil, Are Olsen):
  - Draft technical document with description of methods of 1<sup>st</sup> level QC and calculation of surface water fCO<sub>2</sub>,
  - Matlab routines for 1<sup>st</sup> level QC and calculation of surface water fCO<sub>2</sub>,
  - Original surface CO<sub>2</sub> cruise data files,
  - Metadata,
  - List of cruises with EXPOcodes and original cruise names.
 Ask data PIs within a month:
  - To comment on the draft technical document and the Matlab routines for 1<sup>st</sup> level QC and calculation of fCO<sub>2</sub>.
  - To contact Benjamin Pfeil, if their public data should NOT be included in the data synthesis.
  - To check whether the database has the final version of their data.
  - To check and update the metadata, e.g. publications, information on calibration, information on ice cover.
- 3) Put 1<sup>st</sup> level QC data on to Live Access Server by March 2008 (Chris Sabine).
- 4) Make 1<sup>st</sup> level QC data set available to the group leaders in March 2008 (Benjamin Pfeil, Are Olsen), if possible via the Live Access Server.
- 5) Organize a meeting for the group leaders by June 2008 (Maria Hood, Dorothee Bakker).
- 7) 2<sup>nd</sup> level data QC ready by late 2008.
- 8) Plan a scientific meeting in 1-2 years for all data PIs.
  - E.g. Discussion of interpolation methods. LAS can assist in making a gridded product at a 1° by 1° resolution with finer resolution in coastal regions.
- 9) Prepare scientific papers for the global CO<sub>2</sub> meeting in summer 2009 (Jena).

ANNEX 4  
SOCAT version 1.1 Data Distribution Graphics (B.Pfeil)

**Figure 1. Spatial Overview of the Data in SOCAT v1.1**

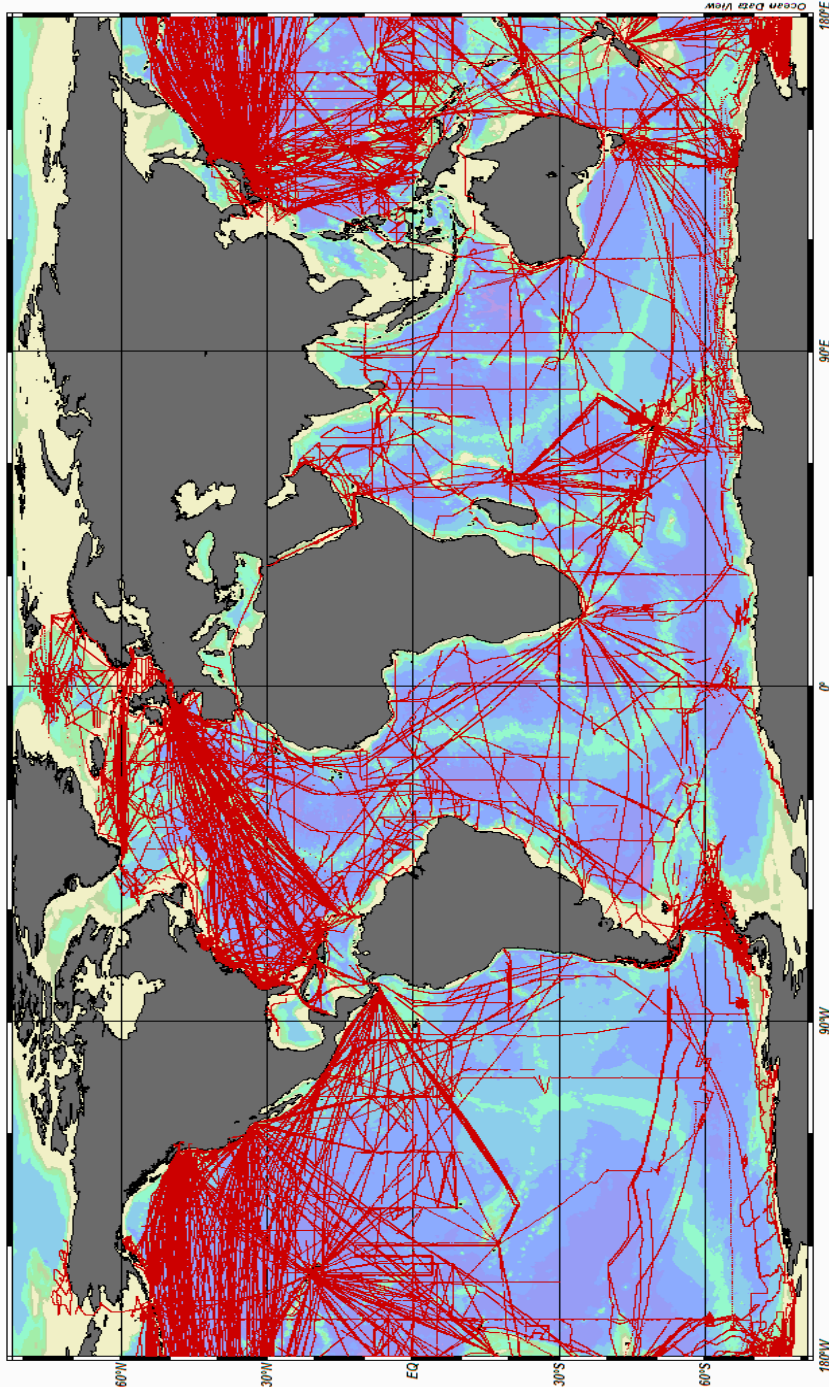


Figure 2. Spatial Distribution of the data in SOCAT v1.1

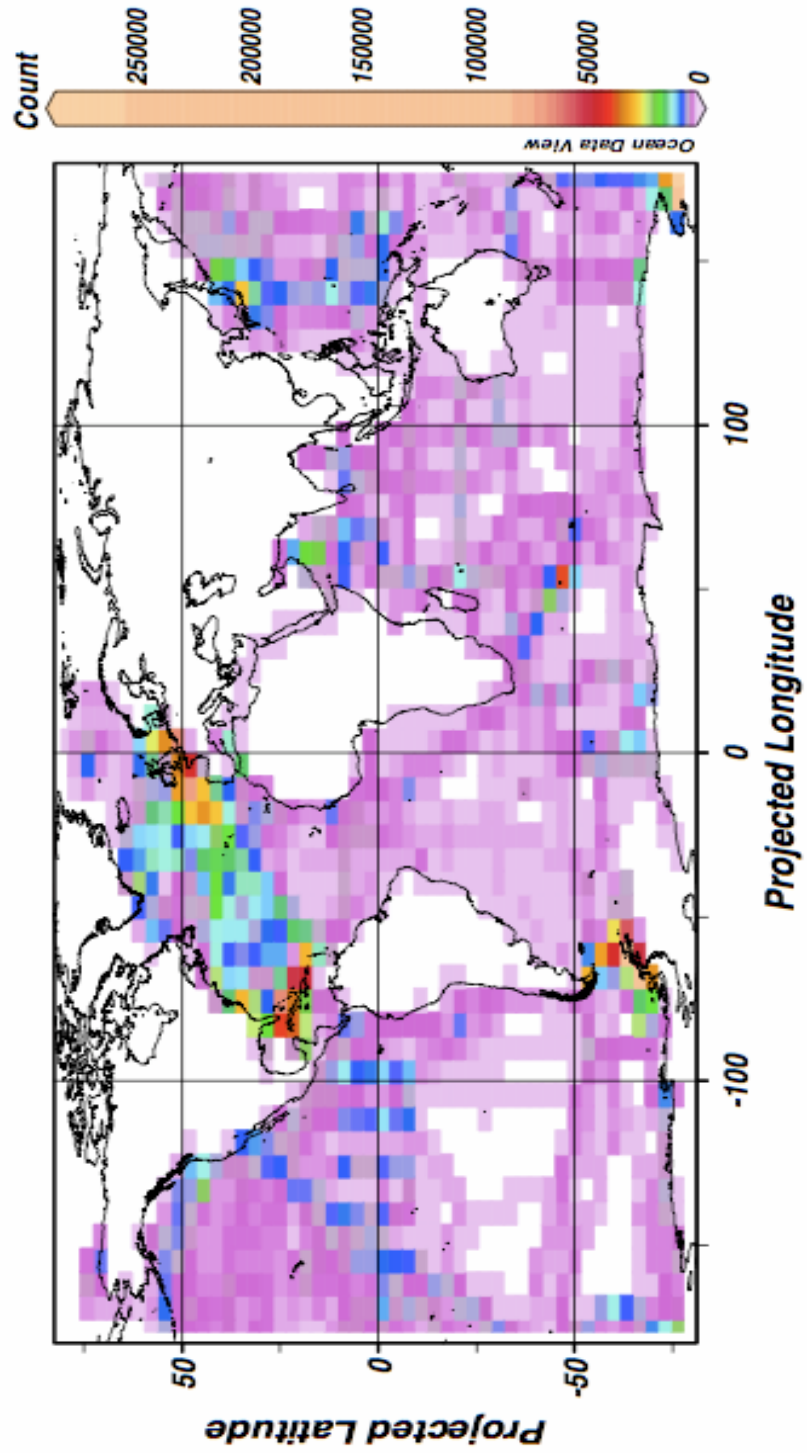
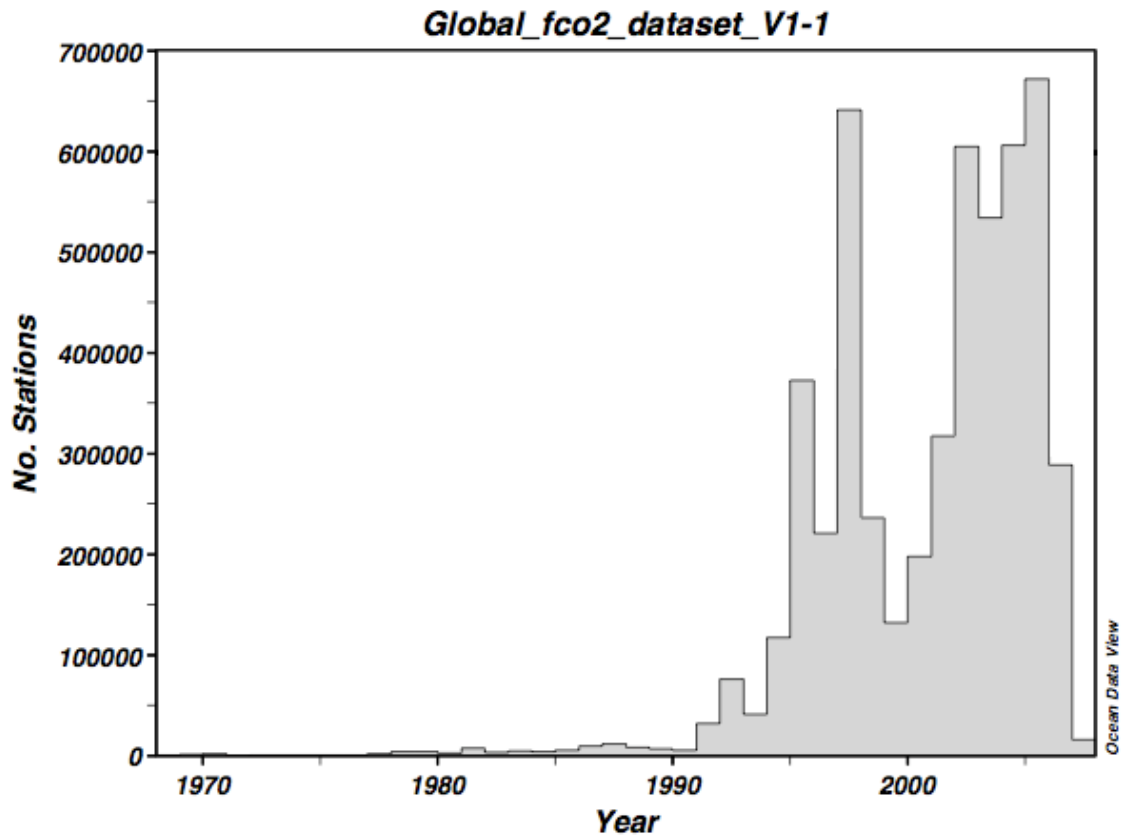


Figure 3. Temporal Distribution of the data in SOCAT v1.1



## Annex 5

Calculations used for the 1<sup>st</sup> level QC data set

### **Uniform Format Data Set for the Surface Ocean CO<sub>2</sub> Atlas (SOCAT)**

Are Olsen and Benjamin Pfeil, Bjerknes Center for Climate Research, Norway.

August 2008

#### **Introduction**

Over the last few decades several million measurements of the surface ocean CO<sub>2</sub> concentration have been made, in particular following the advent of infrared based measurement devices which determines the CO<sub>2</sub> concentration in an air headspace in equilibrium with a continuous stream of sea water. The concentration can be expressed as the mole fraction of CO<sub>2</sub> in the headspace ( $x_{CO_2}$ ), the CO<sub>2</sub> partial pressure ( $p_{CO_2}$ ), and the fugacity of CO<sub>2</sub> ( $f_{CO_2}$ ) in the headspace, which takes into account the non-ideal behavior of CO<sub>2</sub> gas. It is this latter which should be used for gas exchange calculations. Conversion between these can be carried out using a set of standard procedures (DOE, 1994, Dickson et al., 2007).

Unfortunately investigators have reported data differently. Some have reported  $x_{CO_2}$ , some  $p_{CO_2}$ , and some  $f_{CO_2}$  of the surface seawater. Given this, and given that the format of the files have varied, it has always been a time consuming task to handle and use data from publicly available data repositories like CDIAC (Carbon Dioxide Information Analysis Center).

To alleviate this situation a uniform format global surface ocean  $f_{CO_2}$  data set has been developed, as encouraged by both the Surface Ocean – Lower Atmosphere Study (SOLAS) and the International Ocean Carbon Coordination Project (IOCCP) at the workshop on “Ocean Surface  $p_{CO_2}$  Data Integration and Database Development”, Tsukuba 2004. This document briefly describes the work that has been done.

#### **Equations**

To ensure consistency we decided to recompute  $fCO_2$  data whenever possible using a set of standard equations.

Calculations were carried out according to *Recommendation for autonomous underway  $pCO_2$  measuring systems and data reduction routines* by Pierrot et al, 2008, which follows the DOE handbook (DOE, 1994).

Unless otherwise specified, reported  $xCO_2$  data were assumed to be dry mole fractions standardized by each investigator with respect to calibration gas runs at the temperature of equilibration. Calculation of  $CO_2$  partial pressures from these data follows:

$$(pCO_2)_{equT}^{wet} = (xCO_2)_{equT}^{dry} (P_{equ} - pH_2O) \quad (1)$$

where  $(xCO_2)_{equT}^{dry}$  is the  $CO_2$  mole fraction and  $pH_2O$  is the water vapor pressure at the equilibrator temperature.

Water vapor pressure is calculated according to Weiss and Price (1980):

$$pH_2O = \exp(24.4543 - 67.4509(100/T) - 4.8489 \ln(T/100) - 0.000544 S) \quad (2)$$

The correction for difference in intake and equilibrator temperatures was carried out using the empirical relationship derived by Takahashi et al. (1993)

$$(pCO_2)_{SST}^{wet} = (pCO_2)_{equT}^{wet} \exp\{0.0423(SST - equT)\} \quad (3)$$

where SST is the sea surface temperature in the same units as  $equT$ . Note, our approach here differs from that of Pierrot et al. (2008) who suggest that the conversion to intake temperatures are done on the  $fCO_2$  values. We chose to do it on the  $pCO_2$  values since the Takahashi et al. (1993) relationship were derived using  $pCO_2$ . The difference between these two approaches is very very small. If only  $fCO_2$  at the equilibrator temperature was provided, the conversion to in situ temperature was carried out using these.

Although several approaches are available (e.g., Copin-Montégut, 1988; Goyet et al., 1993; Takahashi et al., 1993; Weiss et al., 1982), the one of Takahashi et al. (1993) was preferred as it does not require knowledge of the alkalinity and TCO<sub>2</sub> of the waters and was determined for isochemical conditions, while the others were not.

The conversion of pCO<sub>2</sub> to fCO<sub>2</sub> values is carried out according to:

$$(fCO_2)_{SST}^{wet} = (pCO_2)_{SST}^{wet} \exp \left\{ \frac{\left[ B(CO_2, SST) + 2 \left( 1 - (x_{CO_2})_{equT}^{wet} \right)^2 \delta(CO_2, SST) \right] P_{equ}}{R \times SST} \right\} \quad (4)$$

where  $P_{equ}$  is the pressure (atm) of the equilibrator, and SST is the sea surface temperature (in K). R is 82.0578 cm<sup>3</sup> atm mol<sup>-1</sup> K<sup>-1</sup>. B(CO<sub>2</sub>,T) and δ(CO<sub>2</sub>,T) are the virial coefficients for CO<sub>2</sub> (Weiss, 1974).

B(CO<sub>2</sub>,T) in cm<sup>3</sup> mol<sup>-1</sup> is given by:

$$B(CO_2, SST) = 1636.75 + 12.0408 SST - 3.27957 \cdot 10^{-2} SST^2 + 3.16528 \cdot 10^{-5} SST^3 \quad (5)$$

and δ(CO<sub>2</sub>,T) in cm<sup>3</sup> mol<sup>-1</sup> by:

$$\delta(CO_2, SST) = 57.7 - 0188 SST$$

### Implementation

The sea surface CO<sub>2</sub> concentration data in the files were reported in 11 different ways, and the large majority of the files contained CO<sub>2</sub> data expressed in at least two different manners (e.g. xCO<sub>2</sub> and fCO<sub>2</sub>).

Ideally we would like to have always computed or recomputed  $f\text{CO}_2$  values from dry mole fractions along with reported equilibrator and intake temperatures, equilibrator pressure, and surface salinity using the set of equations given above. However, on many occasions not all of the required data were reported in the data files, and this necessitated the use of different starting points for our calculations and/or the use of data from external sources. In particular, atmospheric pressure and/or salinity data were sometimes missing. When pressure was missing we used 6 hourly sea level pressure data from the NCEP/NCAR reanalysis project (Kalnay et al., 1996). When salinity was not reported we used climatological monthly mean salinity data from the World Ocean Atlas 2005 (Antonov et al., 2005). The salinity has only a minor effect on the calculations as it only influences the  $f\text{CO}_2$  through the water vapor pressure (see eqs 1 and 2).

Table 1 lists the various starting  $\text{CO}_2$  parameters for our recalculations, the additional supplied data, and what external data were required for the calculation. This list also gives the order of preference for our recalculations. That is, if (1) was possible this was used. If (1) was not possible but (2) was, then (2) was used. If neither (1) nor (2) was possible, but (3) was, then this was used and so on. The philosophy behind this scheme was to (a) start out as close to dry  $x\text{CO}_2$  values as possible and (b) to limit use of external data unless absolutely required (i.e. when no in situ  $f\text{CO}_2$  data could be obtained without resorting to WOA salinities or NCEP/NCAR pressures.). If  $f\text{CO}_2$  data were provided, but no  $x\text{CO}_2$  or  $p\text{CO}_2$ , like from for instance a CARIOCA buoy, the  $f\text{CO}_2$  values were retained.

Finally, if either atmospheric pressure or NCEP/NCAR were used, 3 hPa were added to account for the overpressure normally maintained in ships (Takahashi and Sutherland, 2007).

We note that even though 11 different types of surface  $\text{CO}_2$  data were reported, only 5 were used for the recalculations. The others were incorrectly classified or somewhat obscure (for instance  $x\text{CO}_2$  at sea surface temperature at 100% humidity) and could not

be used in either equation 1 or 3 directly. However, no data were lost since all the data sets containing non-usable parameters also had parameters that could be used for recalculations.

Table 1: Reported CO<sub>2</sub> Parameters Used for the Calculations in Order of Preference

Preference <sup>b</sup>	CO <sub>2</sub> parameter	required extra var.	number of occurrences	occurrences in %
1	dry xCO <sub>2</sub> at eq or intake temp <sup>a</sup>		3,144,803	67.19
2	pCO <sub>2</sub> at eq. or intake temp <sup>a</sup>		278,690	5.96
3	fCO <sub>2</sub> at eq or intake temp <sup>a</sup>		573,859	12.26
4	pCO <sub>2</sub> at eq or intake temp but no pressure <sup>a,c</sup>	pressure	617,377	13.19
5	xCO <sub>2</sub> but no pressure and SSS <sup>a</sup>	SSS & pressure	65,060	1.39
	total		4,679,789	

<sup>a</sup>Some xCO<sub>2</sub>/pCO<sub>2</sub>/fCO<sub>2</sub> data were reported at intake temp. For these, the correction to intake temperature was not carried out. It is clear from the data file what data were reported at intake temperature and what data were reported at equilibrator temperature.

<sup>b</sup>Number is also used within the output data file for identifying which reported CO<sub>2</sub> variable was used for calculations. See Table 2.

<sup>c</sup>In case pCO<sub>2</sub> at eq or intake temperature was reported without an accompanying pressure it was assumed that the pCO<sub>2</sub> was reported at 1 atm, i.e. NCEP/NCAR sea level pressure was only used for the conversion to fCO<sub>2</sub>

### Reported data

All scripts and in- and output data have been made available along with this report. Transparency is essential for assuring the best quality data product and we encourage all to evaluate our calculations to identify any errors.

The data file contains all of the reported data, the NCEP/NCAR pressures and the WOA salinities. In addition bottom depth from ETOPO2 (<http://www.ngdc.noaa.gov/mgg/global/global.html>) has been included for identification of shelf and coastal data.

The file also contains an identifier which shows what input parameter was used. It should also indicate whether the original data were provided at equilibrator or intake temperature.

Table 2: Parameters Reported in the Output File:

Abbreviation	Description	unit
intake_depth	water intake depth	m
sst	sea surface temperature	deg C
sss	sea surface salinity	PSU
fCO <sub>2</sub> _rec	recomputed fCO <sub>2</sub>	µatm
eq_t	temperature at equilibration	deg C
atm_press	atmospheric pressure as reported	hPa
equ_press	equilibrator headspace pressure as reported	hPa
Press	pressure used for recomputing (if atmospheric pressure, 3hPa has been added)	hPa
win_dir	wind direction as reported	deg
wind_speed	wind speed as reported	m/s
Jday	Julian day as reported	
ship_speed	ship speed	knot
ship_dir	ship heading direction	deg
Hum	humidity	%
WOA_sss	salinity extracted from WOA 2005	PSU
NCEP_slp	atmospheric pressure extracted from NCEP/NCAR 6 hourly data	hPa
ETOPO_z	Bottom depth from ETOPO2v2c, <a href="http://www.ngdc.noaa.gov/mgg/global/etopo2.htm">http://www.ngdc.noaa.gov/mgg/global/etopo2.htm</a> <u>1</u>	M
XCO <sub>2</sub> _w_sst_wet	xCO <sub>2</sub> water at sea surface temperature in wet air	µmol/mol
XCO <sub>2</sub> _w_eqt_wet	xCO <sub>2</sub> water at equilibrator temperature in wet air	µmol/mol
XCO <sub>2</sub> _w_sst_dry	xCO <sub>2</sub> water at sea surface temperature in dry air	µmol/mol
XCO <sub>2</sub> _w_eqt_dry	xCO <sub>2</sub> water at equilibrator temperature in dry air	µmol/mol

XCO2_w_eqt_dry_n	xCO <sub>2</sub> water at equilibrator temperature in dry air. Duplicate*	μmol/mol
fCO2_insitu_from_XCO2	fCO <sub>2</sub> recomputed from xCO <sub>2</sub> , salinity and either atmospheric pressure or pressure at equilibration provided in the file	μatm
fCO2_insitu_from_XCO2_WOA	fCO <sub>2</sub> recomputed from xCO <sub>2</sub> , salinity data used from WOA 2005	μatm
fCO2_insitu_from_XCO2_NCEP	fCO <sub>2</sub> recomputed from xCO <sub>2</sub> , atmospheric pressure used from NCEP/NCAR	μatm
fCO2_insitu_from_XCO2_WOA_NCEP	fCO <sub>2</sub> recomputed from xCO <sub>2</sub> , salinity used from WOA 2005 and atmospheric pressure used from NCEP/NCAR	μatm
pCO2_theta_w_sst	pCO <sub>2</sub> , water at sea surface temperature in wet air Duplicate*	μatm
pCO2_w_eqt	pCO <sub>2</sub> , water at equilibrator temperature in wet air	μatm
pCO2_w_sst_wet	pCO <sub>2</sub> , water at sea surface temperature in wet air	μatm
pCO2_w_sst_wet_n	pCO <sub>2</sub> , water at sea surface temperature in wet air Duplicate*	μatm
fCO2_insitu_from_pCO2	fCO <sub>2</sub> recomputed from pCO <sub>2</sub> , salinity and either atmospheric pressure or pressure at equilibration provided in the file	μatm
fCO2_insitu_from_pCO2_NCEP	fCO <sub>2</sub> recomputed from pCO <sub>2</sub> , salinity provided atmospheric pressure used from NCEP/NCAR	μatm
fCO2_w_corr20	fCO <sub>2</sub> , water corrected to 20 deg C sea surface temperature in wet air	μatm
fCO2_w_corr25	fCO <sub>2</sub> , water corrected to 25 deg C in wet air	μatm
fCO2_w_eq	fCO <sub>2</sub> , water in wet air at equilibrator temperature	μatm
fCO2_w_sst_wet	fCO <sub>2</sub> , water at sea surface temperature in wet air	μatm
fCO2_insitu_from_fCO2	fCO <sub>2</sub> , recomputed from fCO <sub>2</sub> salinity and either atmospheric pressure or pressure at equilibration	μatm

	provided in the file	
Id_CO2_used	Identifies which reported CO <sub>2</sub> value was used for the calculations (see Table 1 for details)	
station	number of measurements during a cruise	
identifier	Important for later data handling, which will enable easier tracking of changes.	

\* Duplicate: Will be corrected in next version.

## References

- Antonov, J.I., Locarnini, R.A., Boyer, T.P., Mishonov, A.V., Garcia, H.E., 2006. World Ocean Atlas 2005, Volume 2: Salinity. Levitus, S., ed, NOAA Atlas NESDIS 62, U.S. Government Printing Office, Washington, D.C., 182 pp.
- Copin-Montégut, C., 1988. A new formula for the effect of temperature on the partial pressure of CO<sub>2</sub> in seawater. *Marine Chemistry* 25 (1), 29-37.
- DOE, 1994. Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water. Version 2, Dickson, A.G., Goyet, C., eds, ORNL/CDIAC-74.
- Dickson, A.G., Sabine, C.L., Christian, J.R., 2007. Guide to best practices for ocean CO<sub>2</sub> measurements. PICES Special Publication 3, 191pp.
- Goyet, C., Millero, F.J., Poisson, A., Shafer, D.K., 1993. Temperature dependence of CO<sub>2</sub> fugacity in seawater. *Marine Chemistry* 44 (2-4), 205-219.
- Kalnay, E. et al, 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of American Meteorological Society* 77, 437-470.
- Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Lüger, H., Johannessen, T., Olsen, A., Feely R.A., Cosca, C.E., 2008. Recommendations for Autonomous Underway pCO<sub>2</sub> Measuring Systems and Data Reduction Routines. *Deep-Sea Research Part II*. To be published.
- Takahashi, T., Olafsson, J., Goddard, J.G., Chipman, D.W., Sutherland, S.C., 1993. Seasonal-Variation of CO<sub>2</sub> and Nutrients in the High-Latitude Surface Oceans – a

- Comparative-Study. *Global Biogeochemical Cycles* 7 (4), 843-878.
- Weiss, R.F., 1974. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine Chemistry* 2, 203-215.
- Weiss, R.F., Jahnke, R.A., Keeling, C.D., 1982. Seasonal effects of temperature and salinity on the partial pressure of CO<sub>2</sub> in seawater (Pacific Ocean). *Nature* 300 (5892), 511-513.
- Weiss, R.F., Price, B.A., 1980. Nitrous oxide solubility in water and seawater. *Marine Chemistry* 8, 347-359.