

Project Information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	EuroSea
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D4.7
Deliverable title	SOCAT Quality Control (QC) procedures
Description	Manual describing the needs for operational QC
Work Package number	4
Work Package title	Data integration, Assimilation, and Forecasting
Lead beneficiary	University of Bergen
Lead authors	Benjamin Pfeil
Contributors	
Due date	31.10.2022
Submission date	31.12.2022
Comments	Delayed since lead author changed employer in spring 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626.

Table of contents

1. Executive summary.....	1
2. Background.....	1
2.1. Quality control within SOCAT.....	1
2.2. Secondary QC by regional groups.....	2
2.3. Cruise quality flags.....	3
3. Implementing automated QC for the EOVI Inorganic Carbon.....	5
3.1. Introduction to ICOS.....	5
3.2. The marine network within ICOS.....	6
3.3. Improving the TRL by establishing a stable data flow and quality control scheme.....	7
3.4. QuinCe (adopted from Jones et. al. (2022)).....	7
3.5. QC within QuinCe and export to SOCAT.....	8
4. Improving the TRL for SOCAT by implementing automated QC.....	9
4.1. Standardization of instrumentation.....	9
5. Conclusions - Actions needed to reaching a higher TRL.....	10
6. References.....	11

1. Executive summary

The Surface Ocean CO₂ Atlas (SOCAT) is a synthesis of quality-controlled fCO₂ (fugacity of carbon dioxide) values for the global surface oceans and coastal seas with annual updates. SOCAT aims to provide data with the highest possible quality for carbon data – consistent quality control (QC) is essential in achieving this primary goal of SOCAT. Currently there are various steps of quality control, and within this task of EuroSea we aimed to develop an operational implementation of QC as a showcase for data within SOCAT from the European Research Infrastructure Integrated Carbon Observing System. The aim within EuroSea is to increase the Technology Readiness Level (TRL) from 5 (Technology validated in relevant environment) to 7 (system prototype demonstration in operational environment) for relevant ICOS data for direct submission to SOCAT. This was achieved by creating automated quality control into the ICOS state-of-art-software QuinCe, a web-based tool for processing and quality control of data from in situ sensors and underway instruments that is used for first and second level quality control for operational ICOS stations. One important aspect of SOCAT is the assessment of data quality, to ensure that all published data is fit for purpose and manual eyes-on QC is currently essential to lower uncertainties. Currently, this assessment consists of evaluating the metadata of each dataset to ensure that the correct Standard Operational Procedures (SOPs) have been followed during data collection, that the system setup is correct, instruments are calibrated and in addition examining data to ensure they are of good quality. SOCAT consists of three steps of QC: 1.) QC while data is being ingested; 2.) Eyes-on QC by regional experts and 3.) QC for the entire dataset defining the uncertainty based upon the submitted metadata and within this task it has been shown that certain parts of this QC process can be automated while other levels bear challenges if a higher level of TRL is aimed for.

2. Background

SOCAT¹, the Surface Ocean CO₂ ATlas, is a synthesis activity for quality-controlled surface ocean fCO₂ measurements by more than 100 contributors from around the globe. It enables quantification of the ocean carbon sink, ocean acidification and the evaluation of ocean biogeochemical models. As such, it is used in the Global Carbon Budget (e.g., Le Queré et al, doi: [10.5194/essd-10-2141-2018](https://doi.org/10.5194/essd-10-2141-2018)). Version 1 of SOCAT was released in 2011 and due to the implementation of the SOCAT Automation Upload Dashboard hosted by NOAA PMEL (Pacific Marine Environmental Laboratory) annual releases have been realised since version 4 in 2016.

2.1. Quality control within SOCAT

Surface measured fCO₂ data submitted to SOCAT have been initially QCed by their respective data providers prior to submission also known as primary QC. During submission using the SOCAT Data Upload Dashboard data first goes through a series of automated QC checks described in Table 1. They are applied to the timestamp, position and value of the variables used to calculate fCO₂. The “flag of 4” mentioned refers to the WOCE QC flags, meaning bad data, and is given to the fCO₂ values.

¹ <https://socat.info>

Table 1: Automatic QC checks performed for SOCAT

Parameter	Unit	Criteria	Action
Time	–	Duplicate times	Artificial seconds added; flag of 4 if < 50 duplicate times in data set.
Sampling depth, water	m	< –20 or > 20	flag of 4
Salinity	–	< 0 or > 50	flag of 4
Sea surface temperature	°C	< –8 or > 50	flag of 4
Equilibrator temperature	°C	< –10 or > 45	flag of 4
Atmospheric pressure	mbar	< 800 or > 1200	flag of 4
Equilibrator pressure	mbar	< 800 or > 1200	flag of 4
$x\text{CO}_2$, $p\text{CO}_2$, $f\text{CO}_2$ water	$\mu\text{mol mol}^{-1}$ or μatm	< 0 or > 10 000	flag of 4
$x\text{CO}_2$, $p\text{CO}_2$, $f\text{CO}_2$ air	$\mu\text{mol mol}^{-1}$ or μatm	< 0 or > 10 000	flag of 4
$\Delta x\text{CO}_2$, $\Delta p\text{CO}_2$, $\Delta f\text{CO}_2$	$\mu\text{mol mol}^{-1}$ or μatm	< –10 000 or > 10 000	flag of 4
$x\text{H}_2\text{O}$ equilibration	mmol mol^{-1}	< 0 or > 200	flag of 4
WOCE flag, from PI	–	< 1 or > 9	flag of 4
Air temperature	°C	< –35 or > 60	flag of 4
Relative humidity	%	< 0 or > 100	flag of 4
Specific humidity	–	< 0 or > 40	flag of 4
Wind direction	°	< 0 or > 360	flag of 4
Wind speed	m s^{-1}	< 0 or > 50	flag of 4
Ship direction	°	< 0 or > 360	flag of 4
Ship speed, from PI	km h^{-1}	< 0 or > 100	flag of 4
Ship speed, calculated	km h^{-1}	> 720	flag of 4 for following point

Since SOCAT is a global data product, the criteria are ample to allow for a wide range of environmental conditions. Timestamps in non-chronological order or an excessive number of duplicated times or an excessive number of flag 4 data points are causes for the dataset to be sent back to the provider.

2.2. Secondary QC by regional groups

The next step in QC is manual secondary QC, performed by scientists specialized in the Essential Ocean Variable Inorganic Carbon, divided in groups of regional expertise² lead by the SOCAT global group. The $f\text{CO}_2$ data are visually inspected using the SOCAT Live Access Server hosted by NOAA PMEL and WOCE flags of good (2)/questionable (3)/bad (4) data are assigned just to the respective $f\text{CO}_2$ data while just data with WOCE flag 2 are included in the final product. The metadata provided is reviewed for completeness, and additional information is requested if missing. Temperature and salinity are not explicitly QCed but e.g., bad temperature values have a direct impact on the $f\text{CO}_2$ data and reflect a WOCE flag 4. Figure 1 shows the distribution of percentage of WOCE flags 2, 3 and 4 per month within SOCATv2019 when the EuroSea project started.

In SOCATv2019, the percentage of good data over the total (Figure 1) is mostly over 90%, with some particular exceptions, especially since the mid-1990s. The low percentage of “bad data” is likely due to the provider removing data points that are known to be erroneous (e.g., the instruments are on, but the mooring is not in the water yet; known periods of instrument malfunction, issues with the calibration gases, etc.).

² <https://www.socat.info/index.php/regional-groups/>

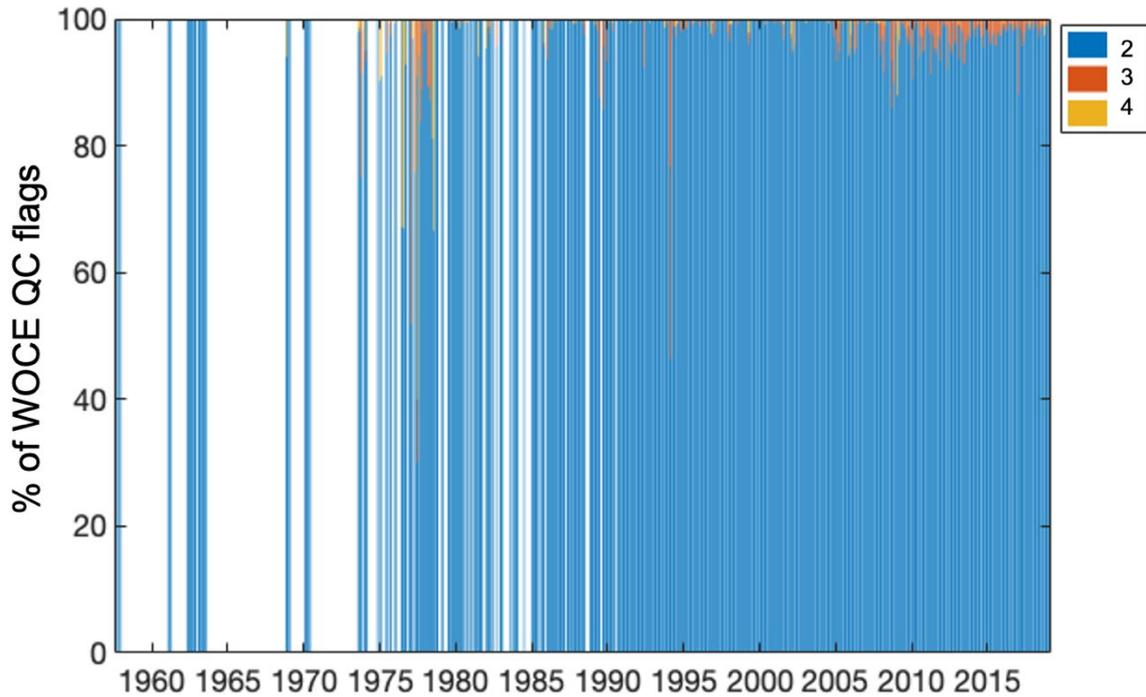


Figure 1: Distribution of data points QC flags per month in SOCAT

2.3. Cruise quality flags

Part of the manual QC procedure by the regional groups is the assignment of cruise quality flags A-E (Pfeil et al. (2013), Bakker et al. (2016)). These are indicators of the estimated accuracy of $f\text{CO}_2$ in the datasets, from $\leq 2 \mu\text{atm}$ for A-B to $\leq 10 \mu\text{atm}$. The criteria for assigning the dataset flag include whether the method follows standard operating procedures (e. g. accuracy limits for individual sensors, standard gas, etc), the existence of high-quality crossovers and metadata completeness. Table 2 shows the criteria set by SOCAT for assigning the cruise quality flags. It is obvious that a detailed inspection of the provided metadata is essential in order to be able to assign this flag but the eyes-on QC on the data also has an impact.

Currently metadata is submitted in various ways – as a pdf, cruise report or using a template provided by the SOCAT group. Metadata is not machine-readable but machine readability is essential in order to perform and automate this step of QC preferably using the Extensible Markup Language (XML) which is a markup language and file format for storing, transmitting, and reconstructing arbitrary data. The vision of a standardised and machine-readable metadata entry for SOCAT has been existing since 2016 but has not been realised for various reasons mostly regarding to the lack of funding and its overall complexity.

One aim of SOCAT was to align this effort with the automation of the metadata reporting scheme for the Sustainable Development Goal 14.3.1 Indicator: Average marine acidity (pH) measured at agreed suite of representative sampling stations where the methodology and metadata template are partly based upon the information requested by the SOCAT template. The machine readability of the SDG 14.3.1 metadata has not been realised so far due to the lack of funding and the pandemic but certain steps are ongoing and currently processes are undergoing to make the controlled vocabularies by NERC/BODC which are the backbone for SeaDataNet and EMODnet fit for purpose for the EOVI Inorganic Carbon.

Table 2: Data set quality control flags for SOCAT version 3 and later. All criteria need to be met for assigning a flag of A to E (Source: Bakker et al. (2016)).

Flag	Criteria ^a
A	(1) Accuracy of calculated $f\text{CO}_2\text{rec}$ (at SST ^b) is better than 2 μatm . (2) A high-quality cross-over ^{c,d} with another data set (also flagged A or B) is available. (3) Followed approved methods/SOP ^e criteria. (4) Metadata documentation complete. (5) Data set QC was deemed acceptable.
B	(1) Accuracy of calculated $f\text{CO}_2\text{rec}$ (at SST) is better than 2 μatm . (2) Followed approved methods/SOP criteria. (3) Metadata documentation complete. (4) Data set QC was deemed acceptable.
C	(1) Accuracy of calculated $f\text{CO}_2\text{rec}$ (at SST) is better than 5 μatm . (2) Did not follow approved methods/SOP criteria. (3) Metadata documentation complete. (4) Data set QC was deemed acceptable.
D	(1) Accuracy of calculated $f\text{CO}_2\text{rec}$ (at SST) is better than 5 μatm . (2) Did or did not follow approved methods/SOP criteria. (3) Metadata documentation incomplete. (4) Data set QC was deemed acceptable.
E	(Primarily for alternative sensors) (1) Accuracy of calculated $f\text{CO}_2\text{rec}$ (at SST) is better than 10 μatm . (2) Did not follow approved methods/SOP criteria. (3) Metadata documentation complete. (4) Data set QC was deemed acceptable.
S (Suspend)	(1) More information is needed for data set before flag can be assigned (2) Data set QC revealed non-acceptable data and (3) Data are being updated (part or the entire data set).
X (Exclude)	The data set duplicates another data set in SOCAT.
N (New)	Data submitted to SOCAT that has not undergone independent data set quality control.
U (Updated)	Data re-submitted to SOCAT following updates by the data provider. Will be quality controlled as if new.
Q	A data set with conflicting flags, usually different flags in different regions.

^aThe accuracy takes precedent over the criteria that follow.

^bSST or sea surface temperature.

^cA high-quality cross-over is defined in version 3, as a cross-over between two data sets with a maximum cross-over equivalent distance of 80 km, a maximum difference in sea surface temperature of 0.3°C and a maximum $f\text{CO}_2\text{rec}$ difference of 5 μatm . Inconclusive cross-overs, defined as having a temperature difference greater than 0.3°C or a $f\text{CO}_2\text{rec}$ difference exceeding 5 μatm , will not have a flag A.

^dA cross-over is defined as a distance of less than 80 km. This criterion combines distance and time as $(dx^2 + (30 dt)^{20.5}) \leq 80$ km. One day of separation in time is equivalent (heuristically) to 30 km of separation in space.

^eSOP or Standard Operating Procedure following Dickson et al. (2007).

Figure 2 illustrates the distribution of cruise flags within SOCAT over more than six decades where one can see the improvements in technology and metadata reporting since the start of SOCAT. Within SOCAT metadata was rescued and completed in many cases during the manual QC process resulting in improved cruise quality flags. The time series of the distribution of cruise quality flags (Figure 2) shows two different phenomena: a general increase in accuracy of the data, due to methodological improvements and completeness of enriched metadata, and since the mid-2000s, an increase of “E” data. These can be a reflection of the recent, wider variety of CO₂ sensors available, whose accuracy has not been independently determined yet, but provide useful data.

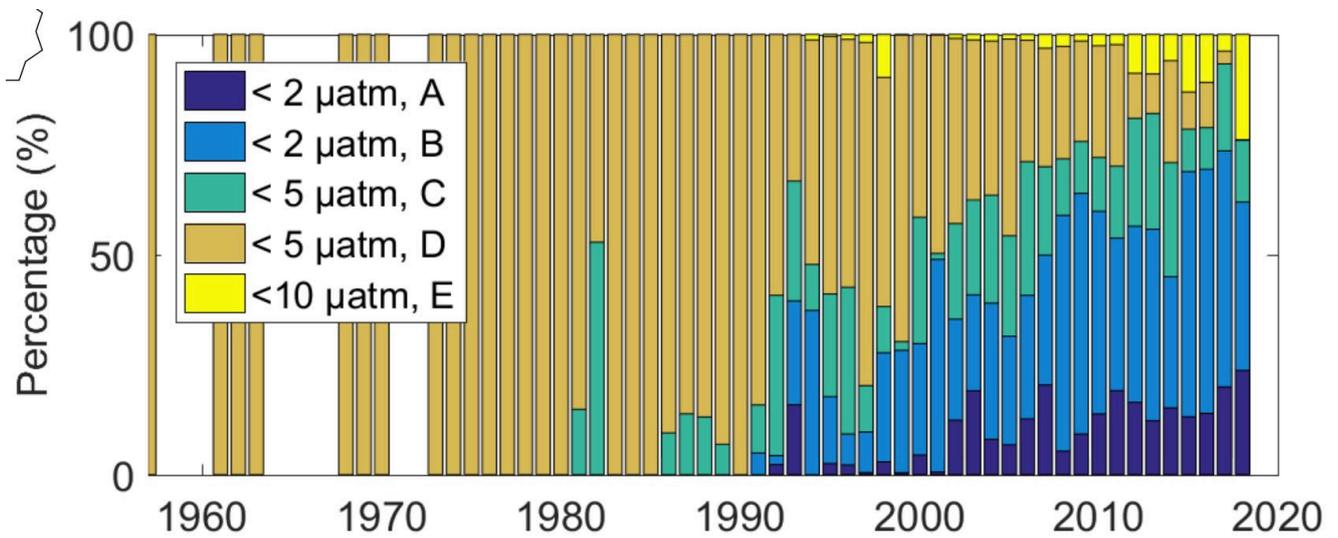


Figure 2: Distribution of cruise flags (accuracy of fCO_2) in SOCATv2019

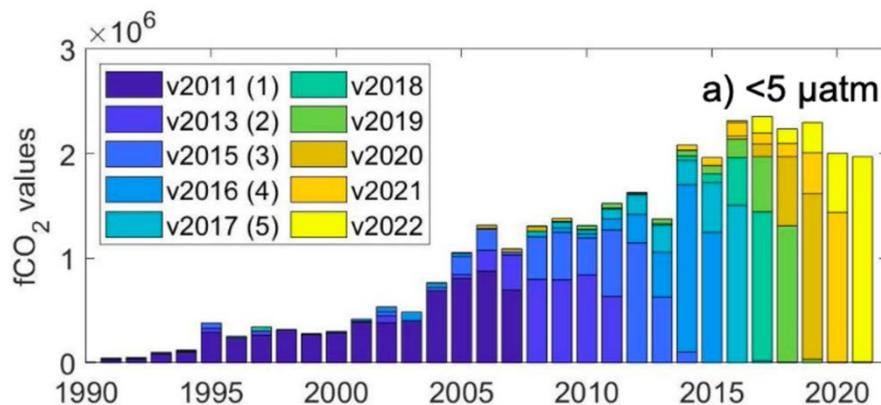


Figure 3: Number of surface ocean fCO_2 values with an estimated accuracy of a) $< 5 \mu atm$ and b) $5 -10 \mu atm$ for each year by SOCAT version (Source: Dorothee Bakker, University of East Anglia)

Time wise, SOCAT shows a very clear increase in the number of observations with time (Figure 3), specially starting in the mid-1990s. In each version there seems to be a slight lag of 1-2 years between the most recent data and the peak in number of observations.

3. Implementing automated QC for the EOVI Inorganic Carbon

3.1. Introduction to ICOS

The Integrated Carbon Observing System (ICOS) is a European Research Infrastructure that has been developed by a community of European scientists to study the carbon cycle in the framework of Climate Change including the influence of human activity and the changing climate on the balance of greenhouse gases of Europe and its surroundings. From a series of partly overlapping national and European projects this has developed into a full blown so called Landmark European Research Infrastructure Consortium (ERIC, established in 2015). ICOS ERIC is now the legal representation for the whole of ICOS Research Infrastructure. ICOS ERIC now has 14 member countries and encompasses more than 150 observation stations distributed over (mostly) Europe and organised in three domains: Atmosphere, Ecosystem and Ocean.

ICOS provides long-term, high-quality observations that follow (and cooperatively set) the global standards for the best possible quality data on the atmospheric composition for greenhouse gases (GHG), greenhouse gas exchange fluxes measured by eddy covariance and CO₂ partial pressure at water surfaces. All measurement methods follow published common specifications and protocols. The data is quality controlled and processed at dedicated central Thematic Centres, one for each domain, using open and published processing chains. All data from raw data up to the final quality controlled (averaged) data is openly accessible from the ICOS Carbon Portal with minimal delays.

As ICOS is committed to provide all data and methods in an open and transparent way as free data, a dedicated system was set up to secure the long-term archiving and availability of the data together with the descriptive metadata that belongs to the data and is needed to find, identify, understand and properly use the data, also in the far future, following the FAIR data principles. An added requirement is that the full data lifecycle should be completely reproducible to enable full trust in the observations and the derived data products.

3.2. The marine network within ICOS

The marine network is based on instrumented “Ships of Opportunity (SOOP)” and fixed time-series stations like moorings. The SOOP are usually commercial ships operating regularly repeated routes, e.g., ferry routes on European shelf and marginal seas, and cargo vessels on open marine routes but research vessels are also part of the network. For the fixed time-series, observations are recorded by means of moorings. These platforms need visits of well-equipped research vessels preferentially 4 to 12 times a year.

The ships and fixed stations are equipped with a varying suite of automated instrumentation to measure atmospheric and surface ocean pCO₂, sea surface temperature, salinity and related variables. On SOOP lines, measurements are repeated along the same transects at intervals of days to months; they only cover the marine surface.

In contrast to atmospheric measurements where the World Meteorological Organization (WMO) has certain requirements for a common standardisation of instrumentation, to provide automated QC and to lower the overall uncertainties, a common setup of instrumentation is not existing for the marine part of ICOS. The overall performance of a station is evaluated during the ICOS labelling process (for more information see³) but a variety of instrumentation measuring pCO₂ is used making data flow, metadata and QC unique for most stations. Detailed assistance by a technician or scientists at the station level is essential to get high quality data during the QC process.

Within the Ocean Thematic Centre (OTC) of ICOS, EuroSea partner UiB is responsible for handling the marine data aiming to establish an operational data flow and standardized QC. The first aim was to establish a standardized QC system for carbon data aiming to increase the TRL for the Essential Ocean Variable Inorganic Carbon. The design criteria of this system was to integrate as much as possible the operational (legacy) database systems within the field – SOCAT. With the aim to preserve the investments in the robust and proven QA/QC and database systems and combining these with a newly established state-of-the-art QC software called QuinCe which aims to automate and standardize first level QC within ICOS and enabling station PIs for perform additional QC.

³ <https://otc.icos-cp.eu/labelling>

3.3. Improving the TRL by establishing a stable data flow and quality control scheme

As mentioned earlier within the marine part of ICOS, a scheme for data flow was established aiming to document every step in order to allow reproducibility, making the data flow operational and to increase the Technology Readiness Level (TRL) from 5 (Technology validated in relevant environment) to 7 (system prototype demonstration in operational environment) for ICOS data.

The data flow covers the entire data life cycle from raw (L0) to Near-Real-Time (L1) and final quality controlled and calibrated data (L2) of the data life cycle and the submission to external data products. Note that the definition of data levels and QC varies from the SOCAT QC as described in chapter 2.1 - the ICOS levels L0 to L2 would be defined as primary QCed using SOCAT terminology. Figure 4 shows the data flow within the marine part of ICOS illustrating the complexity, involvement of various entities and indicates that often eyes-on QC by the PI is essential for Quality Assurance (QA) and QC in order to reduce the uncertainty making data fit for purpose for climate change science.

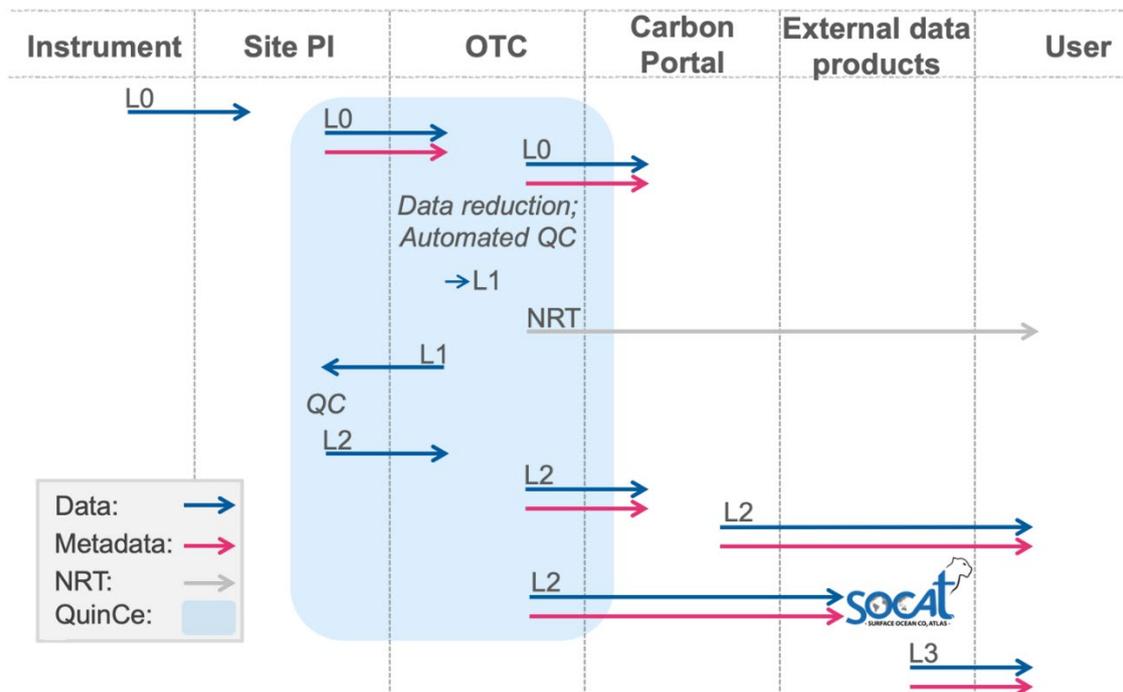


Figure 4: ICOS data life cycle as an example for an operational environment for QC of marine carbon data.

Figure 4 also indicates where the ICOS state-of-the-art QC tool QuinCe is being used to control and to document the data flow within ICOS including an automated submission to SOCAT where ICOS data gets an external QC (L3 in ICOS terms).

3.4. QuinCe (adopted from Jones et. al. (2022))

QuinCe is a web-based tool being developed by the ICOS OTC to process surface ocean carbon dioxide and other measurements, providing a standardised data processing system based on published best practices using community-approved algorithms. It accepts data in any common text format, automatically applies calibrations, and runs a suite of automatic quality control routines to find common data issues. QuinCe provides comprehensive mapping and plotting tools for full scientific quality control, with all changes recorded throughout. Fully quality-controlled data sets can be automatically submitted to data centres (e.g.

Copernicus), synthesis projects (e.g. SOCAT) and ICOS RI. A fully automated processing pipeline allows retrieval, processing, automatic quality control and publication for those stations capable of supplying near real-time data.

QuinCe addresses one crucial aspect that is currently missing in SOCAT - data provenance. Data that comes off the sensors requires processing to calculate the final published value, i.e. data reduction (e.g. Dickson 2007, Pierrot et al. 2009), and the operators (researchers, engineers, technicians) are required to do a certain level of initial quality control to identify and eventually remove bad data points. However, none of these processes are controlled or documented making SOCAT QC more challenging. For the most part the scientists process the data using code written in-house which is neither published nor audited. Aside from the duplication of development and maintenance effort across the community that this entails, there is no way to know of any faults in the implementations or other bugs within this process. Similarly, any decisions regarding which data points are retained, flagged or adjusted are not necessarily well documented. Finally, the raw data files on which the processing is performed are rarely archived in a publicly accessible manner. This leaves a large gap in the data's provenance record, which hinders the reliability and transparency of the processed data (Jones et al., 2022).

3.5. QC within QuinCe and export to SOCAT

QuinCe performs automatic QC and data reduction on each sensor independently of the others. QC routines are written as separate modules that take in a set of values from a single sensor, along with variables suitable for expected values from that sensor, allowing for maximal code reuse. Once data reduction is complete QuinCe performs a round of automatic QC on the calculated values to detect any issues that cannot be found through examining direct sensor values alone. For example, equilibrator-based pCO₂ systems need to adjust the CO₂ values to account for the difference between the temperature in the equilibrator and at the intake (ΔT). If ΔT is larger than a certain threshold this often indicates an issue somewhere in the instrument. QuinCe can detect this once data reduction is complete and flag the relevant sensor values so the user can check the cause. Part of the automated QC within QuinCe is based upon the automated QC used in SOCAT and described in Table 1. Each QC routine sets flags on values according to the criteria it is given. QuinCe uses the same QC flags as being used for SOCAT: Good (2), Questionable (3) or Bad (4).

Once QuinCe has finished all automated quality control (L1) and calculations, the station PI or technician can perform manual quality control (L2). A dedicated QC page provides access to all data through two plot windows and a table of all data. QC flags set by the automatic QC routines are highlighted. It is clear that automatic QC is not 100% trustworthy, with both false negatives and false positives. Therefore, it is inevitable that the operators (researchers, engineers, technicians) must examine these and determine whether they accept the flags provided by the automatic QC, or override them.

Once the user has completed manual QC and assessed all the flags from the automatic QC they can export the L2 dataset to predefined export formats for a number of possible uses, such as for automated import into the ICOS Carbon Portal and SOCAT.

QuinCe does not maintain full metadata records of the instruments and platforms whose data it processes - this information is managed by the metadata database controlled by the ICOS Carbon Portal which can export machine-readable metadata but as highlighted in section 2.1 cannot be handled automatically by SOCAT at this stage.

4. Improving the TRL for SOCAT by implementing automated QC

The example of ICOS illustrates that it is possible to increase the Technology Readiness Level (TRL) from 5 (Technology validated in relevant environment) to 7 (system prototype demonstration in operational environment) for ICOS data including export all data and metadata to SOCAT. Note that the RI ICOS is the first step for establishing an operational data flow for the EOVI Inorganic Carbon. Major efforts in the field of data management have been achieved in ICOS which is unfortunately not often the case for other European data in the field.

QuinCe heavily mimics the initial SOCAT QC but in order to get the best possible quality that is needed to make data fit for purpose for SOCAT additional manual QC is needed in order to quality assure and quality control data. QuinCe or the SOCAT Upload Dashboard could be optimized using machine learning to automate most of the manual QC but severe investments would be needed.

Currently detailed knowledge of the various systems being used and scientific knowledge of the various regions is essential to keep the uncertainties as low as possible during the assignment of the cruise quality flags.

One major lack of implementing automated QC is the lack of standardization of metadata which is essential for assigning the cruise quality flags which define the uncertainty and are the backbone of SOCAT enabling the user to choose the data based upon the to be expected uncertainty.

In order to automate this process of QC, metadata would need to use controlled vocabularies e.g. for every instrument allowing to determine the uncertainty of each instrument. In addition, metadata needs to be machine-readable which is currently not the case and outside the scope of EuroSea to implement this feature.

4.1. Standardization of instrumentation

In 2021 ICOS performed a pCO₂ instrument inter-comparison exercise. The unpublished results by ICOS OTC were provided by Tobias Steinhoff from NORCE towards this deliverable.

Some of the reasons for performing an inter-comparison exercise were the usage of different instrumentation within ICOS, the evaluation of new sensors and the availability of new platforms (e.g. Saildrones, sailing boats) with the aim to evaluate the performance and overall uncertainty of the various systems used. The results have a direct impact on the assignment of the SOCAT cruise quality control flags since they directly reflect the respective uncertainty (A,B – accuracy of <2 µatm pCO₂; C,D – accuracy <5 µatm pCO₂; and F – accuracy <10 µatm pCO₂). In total 29 instruments were divided into three groups for underway systems e.g. used on SOOPs and RVs; surface instruments used on buoys and submersible instruments used on moorings. The performance of the systems was tested using 6 reference gases and QuinCe was used to the data reduction and QC of the data for most systems allowing a comparable data treatment. Figure 5 shows preliminary results that show a systematic deviation of certain instruments. Other preliminary results indicate that data needs to be well documented in a transparent and comprehensible way, that more calibration of other sensors e.g., temperature is essential and that around 50% of all instruments already within ±5 µatm.

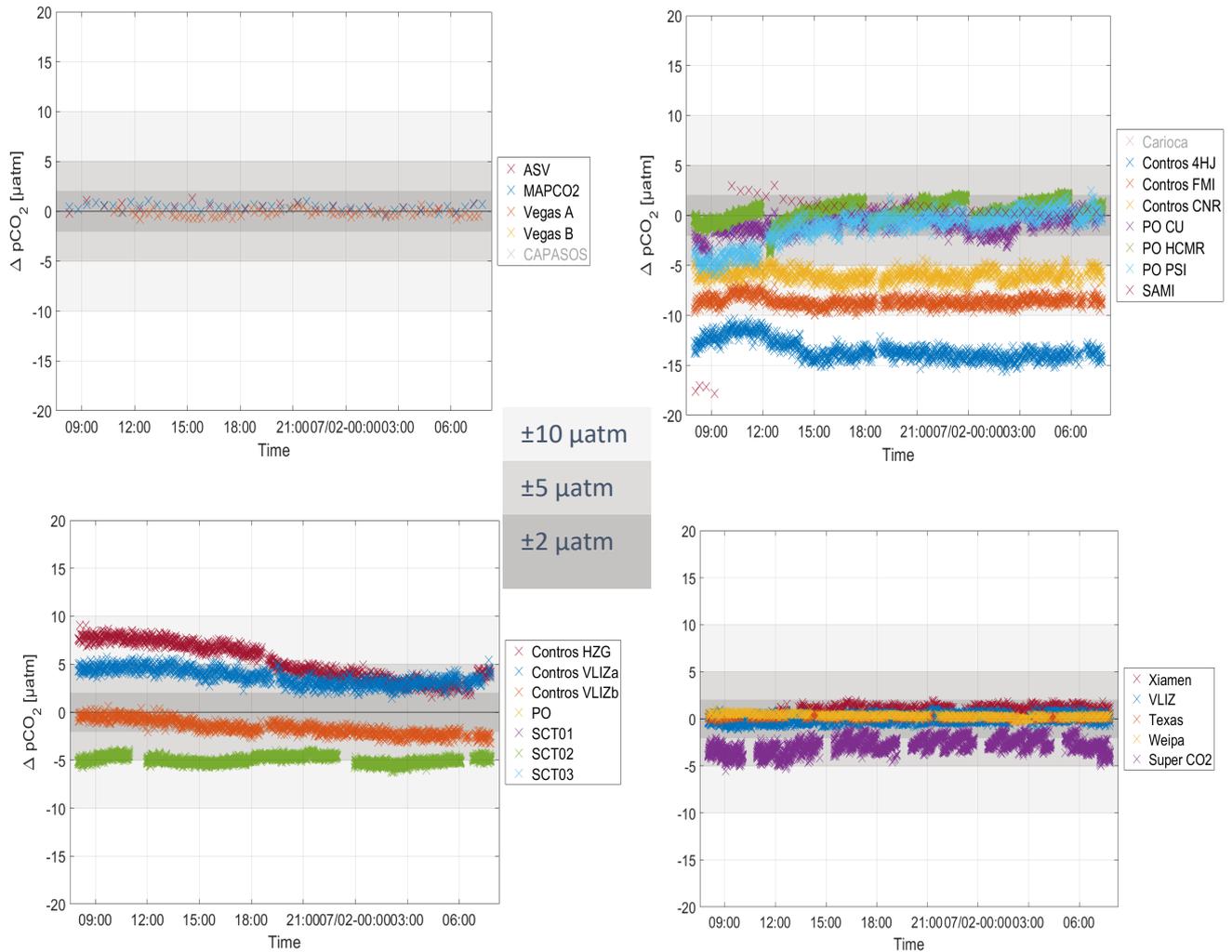


Figure 5: Example of the results from the ICOS inter-comparison exercise showing the deviation of $\Delta p\text{CO}_2$ of the various instruments measuring CO_2 . The gray scale shows the respective uncertainty limits for cruise quality flags with SOCAT.

5. Conclusions - Actions needed to reaching a higher TRL

Certain actions are needed in order to reach a higher level of TRL for automated QC within SOCAT:

- Standardization of instruments: The current usage of a fleet of instruments measuring $p\text{CO}_2$ is a hinder to establish an automated QC that will make data fit for purpose.
- Metadata reporting: metadata needs to be standardized with controlled vocabularies and made machine readable. Activities within the SDG14.3.1 activity to extend the BODC/NERC.
- Major investments in the field of data management efforts: it should be recognized that the lack of instrumentation needs more resources in order to make these data fit for purpose for high end data products.
- Need for automation: currently the SOCAT process is a major effort for the SOCAT regional groups where these tasks are often performed on a voluntary basis which is not a sustainable and efficient

usage of resources. Automation could help to make QC more efficient and use the expert's handling tasks that cannot be automated.

- Primary QC: the lack of standardization for QC for data entering SOCAT needs to be standardized and documented in order to be assured that the overall dataset has the needed quality for SOCAT.

6. References

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., ... Xu, S. (2016). A multi-decade record of high-quality CO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT). *Earth System Science Data*, 8(2), 383–413. doi:10.5194/essd-8-383-2016

Dickson, A. G., Sabine, C. L., & Christian, J. R. (Eds.). (2007). *Guide to Best Practices for Ocean CO₂ Measurements* (IOCCP Repo, σ. 191). ISBN 1897176074. <http://hdl.handle.net/11329/249>

Lauvset, S., Currie, K., Metzl, N., Nakaoka, S.-I., Bakker, D. C. E., Sullivan, K., ... Olsen, A. (2018). SOCAT Quality Control Cookbook. Available from SOCAT website: https://www.socat.info/wp-content/uploads/2020/12/2018_SOCAT_QC_Cookbook_final_with-v2021-contact-details.pdf

Jones, S. et al (2022). QuinCe: A web-based tool for processing and quality control of data from in situ sensors and underway instruments. To be submitted

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneeth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregón, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, *Earth Syst. Sci. Data*, 10, 2141–2194, <https://doi.org/10.5194/essd-10-2141-2018>, 2018.

Lauvset, S., Currie, K., Metzl, N., Nakaoka, S.-I., Bakker, D. C. E., Sullivan, K., ... Olsen, A. (2018). SOCAT Quality Control Cookbook. Available from SOCAT website: https://www.socat.info/wp-content/uploads/2020/12/2018_SOCAT_QC_Cookbook_final_with-v2021-contact-details.pdf