

Scientific Data

<https://doi.org/10.1038/s41597-026-07532-5>

Article in Press

Recalculating the Surface Ocean CO₂ Atlas (SOCAT) to a sea surface temperature climate data record

Received: 15 November 2025

Accepted: 27 May 2026

Cite this article as: Ford, D.J., Shutler, J.D., Holding, T. *et al.* Recalculating the Surface Ocean CO₂ Atlas (SOCAT) to a sea surface temperature climate data record. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-07532-5>

Daniel J. Ford, Jamie D. Shutler, Thomas Holding, Richard P. Sims, Ian Ashton, David K. Woolf, Lonneke Goddijn-Murphy, Dorothee C. E. Bakker, Christopher Sabine, Christopher J. Merchant, Owen Embury, Yuanxu Dong & Rik Wanninkhof

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Recalculating the Surface Ocean CO₂ Atlas (SOCAT) to a sea surface temperature climate data record

Daniel J. Ford ¹, Jamie D. Shutler ¹, Thomas Holding ^{1,12}, Richard P. Sims ^{1,13}, Ian Ashton ², David K. Woolf ³, Lonneke Goddijn-Murphy ⁴, Dorothee C. E. Bakker ⁵, Christopher Sabine ⁶, Christopher J. Merchant ^{7,8}, Owen Embury ^{7,8}, Yuanxu Dong ^{9,10} and Rik Wanninkhof ¹¹

¹ Centre for Geography and Environmental Sciences (CGES), University of Exeter, Penryn, UK

² College of Engineering Mathematics and Physical Sciences, University of Exeter, Penryn, UK

³ International Centre for Island Technology, Heriot-Watt University, Stromness, Orkney, UK

⁴ Environmental Research Institute, University of the Highlands and Islands, Thurso, UK

⁵ Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, UK

⁶ Department of Oceanography, University of Hawai'i at Mānoa, 1000 Pope Rd., Honolulu, Hawaii 96822, United States

⁷ Department of Meteorology, University of Reading, Reading, UK

⁸ National Centre for Earth Observation, University of Reading, Reading, UK

⁹ Marine Biogeochemistry Research Division, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel 24148, Germany

¹⁰ Institute of Environmental Physics, Heidelberg University, Im Neuenheimer Feld 229, 69120 Heidelberg, Germany

¹¹ NOAA/OAR Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida 33149, United States

¹² Now at: BirthRites Lise Meitner Research Group, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

¹³ Now at: Government of Jersey Marine Resources, Natural Environment, Howard Davis Farm, La Route de la Trinitie, Trinity JE3 5JP, Jersey

Corresponding author: Daniel J. Ford (d.ford@exeter.ac.uk)

Abstract

The Surface Ocean CO₂ Atlas (SOCAT) is a global scientific community effort to collate and provide additional quality control and standardisation for surface ocean carbon dioxide (CO₂) data. Each year the international marine carbon community submit any new measurements collected on research vessels, ships of opportunity, moorings, uncrewed surface vehicles and sailing yachts for inclusion in the annual update of the SOCAT database. The data synthesis effort, which published its first data product in 2011, includes a variety of systems, sampling strategies, maintenance cycles and instrument calibrations. Each in-water CO₂ gas measurement is paired, and linked, with a sea surface temperature (SST) measurement. However, the differences in measurement systems means that data pairs from different platforms are representative of differing depths in the ocean, whilst SST measurements can suffer from warming within the observation platform. These complexities can limit the accuracy and precision of any atmosphere-ocean CO₂ assessments that use the SOCAT products. Here the SOCATv2025 database with an estimated uncertainty in the fugacity of CO₂ in seawater ($f\text{CO}_2(\text{sw})$) of less than 5 μatm is recalculated to a reference temperature at a consistent depth of 0.2 m using the European Space Agency (ESA) Climate Change Initiative (CCI) SST climate data record. This recalculation process of the $f\text{CO}_2$ values does not assume isochemical conditions and so temperature driven carbonate speciation is captured. The data pairing is maintained so the resulting dataset is well suited for the analysis of atmosphere-ocean CO₂ exchange.

The synthesis cruise data and gridded data products, that include both the original and recalculated data, are provided and consistency with the original SOCAT data products and format is confirmed. The importance of robustly accounting for the observed warm bias is demonstrated as removing this signal by recalculation to a climate data record temperature shows a $\sim 0.4 \text{ Pg C yr}^{-1}$ ($\sim 12 \%$) increase in the 2024 ocean CO_2 sink (3.4 Pg C yr^{-1}). These recalculated data products are needed for annual carbon assessments therefore these will be routinely provided each year following each annual SOCAT dataset release.

Background and Summary

The ever-increasing anthropogenic carbon dioxide (CO_2) emissions force the global oceans to absorb more CO_2 . This annual absorption, or sink, is increasing with time and has reached $3.2 \pm 0.4 \text{ Pg C yr}^{-1}$ during the period 2015-2024¹. Consequently the ocean has sequestered $\sim 25\%$ of anthropogenic emissions during this decade¹. Our ability to routinely quantify the annual ocean CO_2 sink comes from: (1) global biogeochemical models and (2) statistical multi-parameter regression and machine learning approaches of in situ observations of the fugacity of CO_2 in seawater ($f\text{CO}_{2(\text{sw})}$). These assessments require in situ observations which, via a large voluntary community effort, are annually collected, quality controlled and compiled into the Surface Ocean CO_2 Atlas (SOCAT)². The latest release, SOCATv2025³ contains 41.4 million $f\text{CO}_{2(\text{sw})}$ observations, each with an estimated uncertainty of less than $5 \mu\text{atm}$ spanning 1957 to 2024, with most of the data between 1990 and 2024 (99.7 %). These data provide the basis for the observation-based assessments of $f\text{CO}_{2(\text{sw})}$ and air-sea CO_2 fluxes and are also used for the evaluation of ocean biogeochemical models. The SOCAT dataset is therefore a pivotal dataset for the ocean carbon and climate communities.

Air-sea CO_2 exchange, or flux, occurs across the air-water interface and the rate is largely controlled within the top $\sim 0.02 \text{ m}$ of the ocean^{4,5}. But due to the physical limitations and differences in sampling techniques across the ships and other platforms, the $f\text{CO}_{2(\text{sw})}$ observations within SOCAT are not representative of the top 0.02 m , and the depth can vary depending on the ship or observational platform. Although SOCAT reports observation depth, this depth does not always reflect the true sampling depth as ballasting, cargo and sea state, for example, can vary the true sampling depth. Consequently the sampling depths within SOCAT are variable (with the average and variance unknown) but will not be representative of the depth directly relevant for air-sea CO_2 exchange. Alongside these depth-related issues, a warm bias has also been identified within the SOCAT observations^{6,7}. This warm bias stems from a multitude of temperature related issues that can affect the seawater temperature measurement (reported as the coincident SST) in SOCAT including: (1) the inadequate calibration of SST sensors (2) uncorrected warming of the seawater within the ship after the temperature observations are taken (i.e., the SST measurement is generally taken near the intake, whereas the CO_2 measurement is taken some time later after the water has travelled within the ship).

The presence of the warm bias is well understood, but its magnitude varies depending upon the method used to detect it. Goddijn-Murphy et al. ⁶ suggested the global mean bias was on order ~ 0.09 °C between 1991 and 2007 when comparing monthly 1° (i.e. 1° latitude by 1° longitude grid) binned SOCAT temperature observations to a satellite temperature dataset that was designed to represent the SST representative of air-sea CO₂ exchange. Dong et al. ⁷ indicated the global mean bias was on order ~ 0.02 °C between 1982 and 2022, when comparing a subset of the monthly 1° binned SOCAT temperature observations to coincident drifting buoys measuring SST (sampling depth ~ 0.2 m). Dong et al. ⁷ also showed a latitudinal variability to the magnitude of the warm bias, varying from 0.04 °C at the equator to 0.14 °C at 25°N to -0.05 °C at 35°N and 0.15 °C at 50°N (and a similar variation in the Southern Hemisphere). The $f\text{CO}_2(\text{sw})$ in the SOCAT dataset can be recalculated or “corrected” to a new temperature dataset allowing for the removal of the warm bias. Careful choice and use of a SST climate data record which is stable with respect to independent observations through time for this recalculation, provides data representative of a constant depth that is directly relevant for air-sea CO₂ fluxes ⁴. The recalculated $f\text{CO}_2(\text{sw})$ can be used to quantify the impact of this seemingly small warm bias. Watson et al. ⁸ suggested this warm bias would increase the global ocean CO₂ sink by 0.4 Pg C yr^{-1} using the Goddijn-Murphy et al. ⁶ approach to recalculate the SOCAT data, Dong et al. ⁷ later revised this to an increase of 0.2 Pg C yr^{-1} .

The previous assessments of the warm bias, however, have caveats. Both compare monthly 1° binned SOCAT observations to monthly 1° reference data, which inherently introduces both a spatial and temporal mismatch between the differing observations. For example, Goddijn-Murphy et al. ⁶ use of satellite based observations could be affected by a cool bias within the satellite observations with respect to SST drifter observations ^{9,10}, that would overestimate the warm bias within the SOCAT temperature observations. Meanwhile, the Dong et al. ⁷ approach is limited in the ability to capture any drift in this warm bias through time due to the small number of SST drifters coincident to SOCAT observations. These example caveats indicate that an approach that uses the strengths of both observational types will likely produce a more robust evaluation of, and correction for, the warm bias.

Since 2019, a team at the University of Exeter have followed the methods of Goddijn-Murphy et al. ⁶, and produced a recalculated SOCAT dataset that addresses this warm bias. In early versions, the SOCAT observations were recalculated to the Optimum Interpolated SST data record (OI-SSTv2.0) ^{11–13}. After 2021, the recalculated dataset ¹⁴ was updated to use OI-SSTv2.1, where the v2.1 SST data addressed an observed cool bias (0.14 °C) in the original OI-SSTv2.0 temperature record ⁹. Since 2022, a recalculated version of the SOCAT data using the European Space Agency (ESA) Climate Change Initiative (CCI) SST v2.1 ¹⁵ climate data record has been provided alongside the OI-SST referenced version ^{16,17}.

In this paper and the associated dataset, we consolidate recent efforts to improve the recalculated SOCAT observation dataset, through different lines of evidence. Firstly, the SOCAT observations are recalculated to a daily reference satellite SST dataset at the native spatial resolution of the SST data record from the platform, instead of monthly 1° data. This step reduces the spatial and temporal mismatch between the observational types. Secondly, the reference temperature dataset is updated to the most recent ESA CCI-SSTv3.0^{10,18} which provides daily ~ 5 km satellite based SST observations at a fixed time of 10:30 h (local solar time) and a depth of 0.2 m. Thirdly, the CCI-SSTv3.0 are corrected for a cool bias ($\sim 0.04^\circ\text{C}$) that is characterised to exist within the CCI-SST reference satellite SST dataset^{19,20} following recommendations in Dong et al.⁷. Finally, the temperature relationship used to transform the $f\text{CO}_2(\text{sw})$ data to be representative of the CCI-SST temperature is updated (from the previously used quadratic Takahashi et al.²¹ relationship) to the more recent advance recommended in Humphreys²² which does not assume isochemical conditions.

Methods

SOCATv2025 individual cruise synthesis file

The SOCATv2025 dataset was retrieved as a synthesis text file containing individual cruises with an estimated $f\text{CO}_2(\text{sw})$ uncertainty of less than $5\ \mu\text{atm}$ (containing data Flag as A, B, C or D) from the National Centers for Environmental Information (NCEI)³ covering the period 1957 to 2024. The synthesis text file contains the individual $f\text{CO}_2(\text{sw})$ and coincident SST observations at the seawater intake along with important ancillary information, including the coincident equilibrator temperature and pressure, sea surface salinity (SSS) and air pressure.

We also retrieved the SOCATv2025 Flag E data, which contains individual cruise observations with an estimated $f\text{CO}_2(\text{sw})$ uncertainty of less than $10\ \mu\text{atm}$, but greater than $5\ \mu\text{atm}$. The methods and technical validation will focus on the main SOCATv2025 dataset (i.e observations with estimated $f\text{CO}_2(\text{sw})$ uncertainty less than $5\ \mu\text{atm}$), but the recalculation process has also been conducted for this Flag E dataset and are included within the dataset record.

The CCI-SST climate data record and cool bias correction

The ESA CCI-SSTv3.0 daily Level 4 analysis¹⁰ is a satellite based SST climate data record that is constructed from both infrared and microwave satellite radiometers. Compared to the previous version of ESA CCI-SST (v2.1), version 3.0 contains improvements to the temporal stability (with fewer timeseries spikes), and the

handling of SST retrievals under dust aerosol conditions¹⁰. The Level 4 analysis uses 22 available radiometer observations that have been consistently recalibrated, and bias adjusted and combines these with an interpolation analysis to produce a daily gap filled dataset at a spatial resolution of 0.05° (~5 km at the equator). These SST have been adjusted to a time of 10:30 h (local time) at a depth of 0.2 m. The time-of-day standardisation avoids aliasing the varying diurnal-cycle sampling from the changing constellation of satellites into the climate record which would manifest as a time-varying SST bias; the CCI-SST record is unique in removing this issue. The interpolation techniques used to produce spatially complete Level 4 products do result in a level of data smoothing which varies spatially, especially when filling across larger gaps. In the case of CCI-SST the Level 4 feature resolution is typically coarser than 15 km. Alongside each SST observation, an SST uncertainty that incorporates the underlying satellite observation and interpolation uncertainties is provided which is representative of the standard deviation of a normal distribution. This approach of characterising the uncertainties has been verified against in situ observations¹⁰. The CCI-SSTv3.0 record was retrieved from Centre for Environmental Data Analysis (CEDA)¹⁸, and covers the period 1980 to 2024.

The CCI-SST analysis does not ingest in situ SST observations from drifters (which are also representative of ~0.2 m depth) or any other platforms. This independence allows in situ observations to be routinely used to objectively assess the accuracy and precision of the CCI-SST analysis. The CCI-SST validation report²⁰ indicates that, compared to in situ SST drifters, the CCI-SST has a relatively stable global cool bias of ~0.04 °C (median) through time. The same validation report showed a small latitudinal variation in the cool bias with a median bias near 0 °C in the subtropics (35 °S to 20 °S and 20 °N to 35 °N), and a larger cool bias ~0.1 °C in the high latitude northern hemisphere (above 55 °N)²⁰. The stability of the CCI-SST record was also independently confirmed by a climate data record intercomparison¹⁹, which also used in situ SST drifter observations. Here, this work is attempting to provide a dataset most representative of the near-surface conditions, therefore before input into the recalculation process this cool bias was removed from the CCI-SST data assuming a globally fixed bias (i.e., 0.04 °C was added to all CCI-SST values). An alternative choice of SST climate data record was the OI-SSTv2.1⁹ but due to temporal stability issues this was less suitable as the reference dataset. This choice and issue are discussed further in the final section of the technical validation.

Matching SOCAT observations to daily reference data

The GitHub version controlled software that followed the method within Goddijn-Murphy et al.⁶ was used as the starting point. Within Goddijn-Murphy et al.⁶ the satellite reference SST data were linearly interpolated to the locations of the SOCAT observations, where the reference dataset was monthly 1 ° satellite observations. This software was updated to allow for different spatial resolution reference data to be provided and the ability to

flexibly provide the time resolution of the reference data (from daily to calendar month). The daily reference SST and uncertainty data were linearly interpolated in space to the SOCAT observations, providing both SST and SST uncertainty values from the reference data representative for each individual SOCAT observation.

Recalculation of $f\text{CO}_2$ (sw) to the reference temperature

The recalculation of the $f\text{CO}_2$ (sw) observations involves transforming the $f\text{CO}_2$ (sw) at the reported SOCAT SST, back to the equilibrator temperature (i.e. back to the original measurement) using the Takahashi et al.²¹ temperature correction used by SOCAT. The $f\text{CO}_2$ (sw) consistent with the reference SST climate data record is then calculated using the Humphreys²² temperature correction. This process simply reverses part of the original calculation performed by SOCAT (using the metadata), before then inserting the SST reference data and repeating the forward calculation. The process and theory are explained in detail within Goddijn-Murphy et al.⁶ therefore here we focus on the main process, and updates from the previous work. For the transformation of $f\text{CO}_2$ (sw) between different temperatures we use PyCO2SYS (v1.8.3.3)^{23,24} because these tools are traceable and they enable the use of both the Takahashi et al.²¹ and updated Humphreys²² temperature sensitivities.

For the first step, the $f\text{CO}_2$ (sw) reported at SOCAT SST were converted back to the equilibrator temperature using the same linear temperature sensitivity²¹ (with respect to the natural logarithm $f\text{CO}_2$ (sw) and temperature) used within the SOCAT processing^{2,25}. If the air pressure and equilibrator pressure are recorded these were provided to PyCO2SYS. If the equilibrator pressure was not provided (17 % of the observations), we use the air pressure +3 hPa due to the overpressure generally maintained in ships²⁶, which is consistent with the SOCAT processing²⁵. If the equilibrator temperature was not reported (12 % of the observations), the reverse calculation step was omitted and the SOCAT SST was considered the equilibration temperature (e.g. a membrane-based system, in close proximity to and at the same temperature as the water inlet).

The second step transforms the $f\text{CO}_2$ (sw) within the equilibrator or near the membrane to the reference SST data (i.e., the climate data record). In this step the $f\text{CO}_2$ (sw) was transformed using the updated temperature sensitivity described in Humphreys²², based on equations governing the marine carbonate system and the van 't Hoff equation. The updated temperature sensitivity (of Humphreys²²) generally indicates a smaller sensitivity at lower temperatures and a larger temperature sensitivity at higher temperatures in comparison to the quadratic Takahashi et al.²¹ approach. The differences occur due to the Humphreys²² approach using the OceanSODA complete carbonate system dataset²⁷ to carry out synthetic Takahashi et al.²¹ experiments at each location (1 ° spatial resolution) to retrieve the b_h term required by the van 't Hoff equation. The b_h was then parameterised as a function of SST, SSS and $f\text{CO}_2$ (sw), where the $f\text{CO}_2$ (sw) acts as a proxy for the dissolved inorganic carbon (DIC) to total alkalinity (TA) ratio. Therefore the approach does not assume an isochemical change²⁸. The assumption of

isochemical conditions (i.e. using the Takahashi et al. approach) would have a small impact for small temperature adjustments (i.e. $<1\text{ }^{\circ}\text{C}$)^{22,28}. However, the impact will increase for $f\text{CO}_2(\text{sw})$ observations that require larger temperature adjustments (i.e. $>1\text{ }^{\circ}\text{C}$) or occur in waters with DIC/TA ratios that diverge from those used to estimate the Takahashi et al.²¹ approach. Thus the use of the updated temperature sensitivity allows the recalculation process to consider more complex interactions and improve the recalculated $f\text{CO}_2(\text{sw})$ ^{22,28}.

It is noted that combinations of SST, SSS and $f\text{CO}_2(\text{sw})$ that do not occur regularly in the global oceans (as assessed using the OceanSODA data) may lead to more unconstrained b_h , and therefore temperature sensitivities, which could be improved with regional fits of the coefficients used in Humphreys²². We do not see this as a limitation for our analysis on the global scale because this detail is only relevant for $\sim 0.1\%$ of the SOCAT database or $\sim 40,000$ data points (these are the data within SOCAT database that have low salinity of less than 20 and high $f\text{CO}_2(\text{sw})$ conditions greater than $700\text{ }\mu\text{atm}$ and typically occur in coastal, estuarine conditions).

Equivalent partial pressure of CO_2 ($p\text{CO}_2(\text{sw})$) is also provided by converting each $f\text{CO}_2(\text{sw})$ to $p\text{CO}_2(\text{sw})$ at the reference temperature using PyCO2SYS and providing the air pressure, so that both variants of the CO_2 measurements are provided. The SOCAT $f\text{CO}_2(\text{sw})$ observations that are made prior to 1980 (i.e., before the start of the CCI-SST climate data record) are not recalculated as the reference SST cannot be provided. Therefore $\sim 50,000$ observations between 1957 and 1979 do not undergo the recalculation process but remain present in the synthesis files.

Finally, we provide the uncertainty on the recalculated $f\text{CO}_2(\text{sw})$ and $p\text{CO}_2(\text{sw})$ which considers the uncertainties in the SOCAT SST and atmospheric pressure from the SOCAT protocols²⁵, and the reference temperature uncertainty. These were propagated through the recalculation process using PyCO2SYS.

Gridding into a monthly 1 ° synthesis file

The original SOCAT and recalculated observations were then gridded at monthly 1 ° resolution following the SOCAT gridding methodology²⁹ into both unweighted and cruise-weighted variants of the observations. The unweighted gridding considers all observations within a monthly 1 ° grid cell of equal weighting. The cruise-weighted gridding follows the SOCAT developed method²⁹ and considers that not all individual cruises sample at the same frequency (i.e. some sample every minute, whereas others could be every 20 minutes), therefore an unweighted monthly 1 ° grid cell would be weighted towards the value of those from the highest frequency sampling. The cruise-weighted gridding therefore considers each individual cruise within a monthly 1 ° grid cell first, and then averages the cruise means for each monthly 1 ° grid cell. Results from both gridding procedures are generated (i.e., unweighted, and cruise-weighted), and the reference SST are also gridded following the same

procedures. Additionally, we provide equivalent gridded recalculated $p\text{CO}_2(\text{sw})$ that some users may use (Table 1 summarises the netCDF variable name for these data).

As a consistency check to confirm the gridding method and precision, values within the gridded dataset are cross-checked against the original SOCAT gridded dataset. The unweighted SOCAT $f\text{CO}_2(\text{sw})$ showed a mean absolute difference of $0.04 \mu\text{atm}$ and for the cruise-weighted $f\text{CO}_2(\text{sw})$ a difference of $0.22 \mu\text{atm}$ ($N = 384,729$). Within the unweighted data, ~ 1200 monthly 1° regions ($\sim 0.3\%$) have an absolute difference greater than $1 \mu\text{atm}$, which occur in locations where SOCAT $f\text{CO}_2(\text{sw})$ observations could not be matched to the satellite reference data and therefore were not included in the gridding. The cruise-weighted data has $\sim 10,000$ of the monthly 1° regions ($\sim 2\%$) with a absolute difference greater than $1 \mu\text{atm}$, which occur in the same regions as the unweighted data.

Data record

The recalculated SOCATv2025 dataset is available at Ford et al.³⁰. The dataset contains four files. The first file (a .tsv file) contains the individual observations from the original SOCAT dataset (i.e observations with an uncertainty less than $5 \mu\text{atm}$) which has been appended with six new columns: the daily reference satellite SST ($T_{\text{reference}}$), the uncertainty on the daily reference satellite SST ($T_{\text{reference_uncertainty}}$), the $f\text{CO}_2(\text{sw})$ (recalculated_ $f\text{CO}_2$) and $p\text{CO}_2(\text{sw})$ at the reference temperature (recalculated_ $p\text{CO}_2$) and the uncertainty on the recalculated $f\text{CO}_2(\text{sw})$ (recalculated_ $f\text{CO}_2_{\text{uncertainty}}$) and $p\text{CO}_2(\text{sw})$ (recalculated_ $p\text{CO}_2_{\text{uncertainty}}$).

The second file (a netCDF file) contains the monthly 1° gridded dataset containing the original SOCAT gridded observations, with corresponding fields for the recalculated observations at the reference SST and associated reference SST observation as both cruise-weighted and unweighted variants (variable names are summarised in Table 1). We refer the reader to the netCDF metadata for further information on the specific fields, which includes minimum, maximum and standard deviation for each recalculated variables. The netCDF format is consistent with the original SOCAT gridded netCDF dataset.

The third and fourth files contains the SOCATv2025 Flag E dataset (i.e. observations with an uncertainty less than $10 \mu\text{atm}$ but greater than $5 \mu\text{atm}$) and are provided as a .tsv file for the individual observations that make up the Flag E data, and the monthly 1° gridded dataset in netCDF format. These files have the same format as described for the first two datasets.

The recalculated SOCAT dataset will continue to be produced annually for all future releases of SOCAT as a community best effort and as funding allows. The repository information and datasets will be given on the

FluxEngine^{31,32} GitHub repository which also contains the underlying analysis software and links to the previous versions of the recalculated data (<https://github.com/oceanflux-ghg/FluxEngine>).

Table 1: Summary of the netCDF variables within the gridded recalculated SOCAT dataset

Variable	NetCDF Variable Name	Units
Reference temperature unweighted	sst_reference_unweighted	°C
Reference temperature cruise-weighted	sst_reference_weighted	°C
Recalculated $f\text{CO}_2$ (sw) unweighted	fco2_recalculated_ave_unwtd	μatm
Recalculated $f\text{CO}_2$ (sw) cruise-weighted	fco2_recalculated_ave_weighted	μatm
Recalculated $p\text{CO}_2$ (sw) unweighted	pco2_recalculated_ave_unwtd	μatm
Recalculated $p\text{CO}_2$ (sw) cruise-weighted	pco2_recalculated_ave_weighted	μatm

Technical Validation

Assessing the warm bias in SOCAT SST

As previously mentioned, the assessment of the warm bias within the SOCAT SST dataset was previously based on monthly 1 ° reference temperature data^{13,14,16}. This inherently introduced a spatial and temporal mismatch between the SOCAT SST and reference temperature observations, and this mismatch could lead to an over- or underestimation of the warming bias. By using daily full spatial resolution reference temperature data the mismatch in time and space is reduced. This is apparent from the narrowing of the histogram when using the daily full resolution data in comparison to the monthly equivalent (Figure 1). Using the daily CCI-SST data indicated a significant global mean of the SOCAT minus reference SST difference of 0.11 °C (Two-tailed t-test; $t(41292327) = 1115$; $p < .01$), compared to the equivalent analysis using the monthly 1 ° reference which identified a 0.12 °C mean bias (Figure 1) across the period 1980 to 2024. This indicates that the warm bias within the SOCAT data is slightly overestimated by ~ 0.01 °C if it is characterised using monthly 1 ° SST data. The CCI-SST reference used to analyse the data within Figure 1 has been first corrected for the cool bias with respect to SST drifters (see earlier methods). Therefore the warm bias would be increased by another 0.04 °C if this cool bias had not been removed from the CCI-SST reference dataset (i.e., the daily mean bias would appear to be 0.15 °C instead of 0.11 °C).

To assess the temporal stability of the SOCAT warm bias, the mean warm bias was extracted for each year between 1980 and 2024 (Figure 2a). All years, except 1989 and 1992, indicated significant differences between the SOCAT and reference SST (Two tailed t-test; $p < .01$). In the 1980s, a large warm bias of the order ~ 0.3 °C was identified but the number of SOCAT SST observations in this period was lower than later years ($\sim 60,000$

between 1980 to 1990). From 1990, the number of SOCAT observations increases and the mean bias stabilises at around ~ 0.01 °C until 2000 where it reduced to near 0 °C. After 2000 the mean bias shows an increase from 0 °C in 2000 to ~ 0.1 °C in 2024, with a notably elevated bias in the 2003 to 2005 period. The analysis therefore indicates that the warm bias is not a temporally stable feature and appears to be increasing with time since ~ 2000 . This analysis also reinforces that using monthly 1 ° reference SST data generally overestimates the warm bias in the SOCAT dataset, further supporting the use of the daily SST dataset.

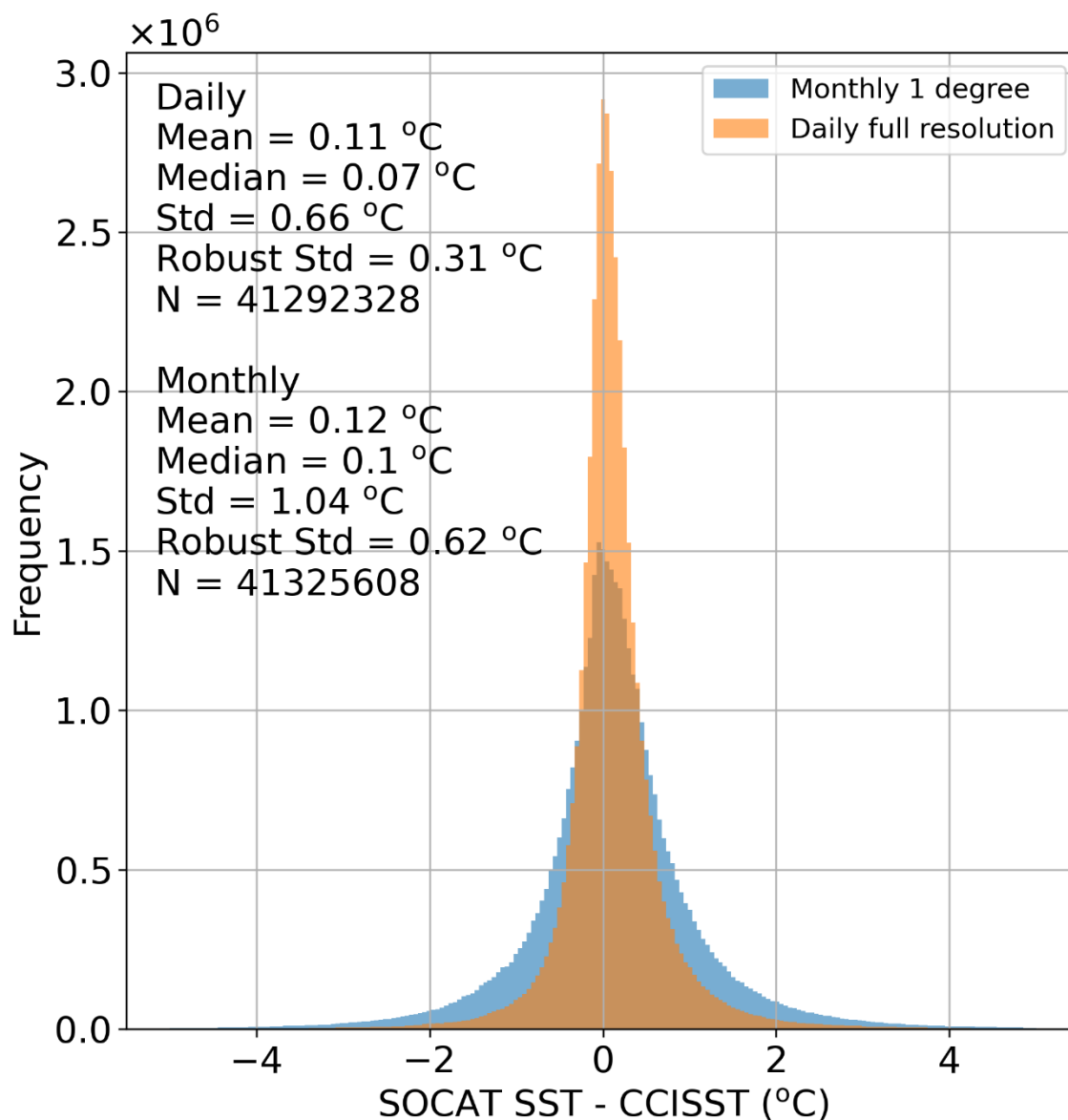


Figure 1: Histogram of the SOCAT SST minus the bias corrected CCI-SST plotted in 0.05 °C bins for the monthly 1 °, and daily full resolution reference SST data. In plot statistics are mean bias (Mean), median bias (Median), standard deviation (Std), robust standard deviation (Robust Std) and number of observations (N).

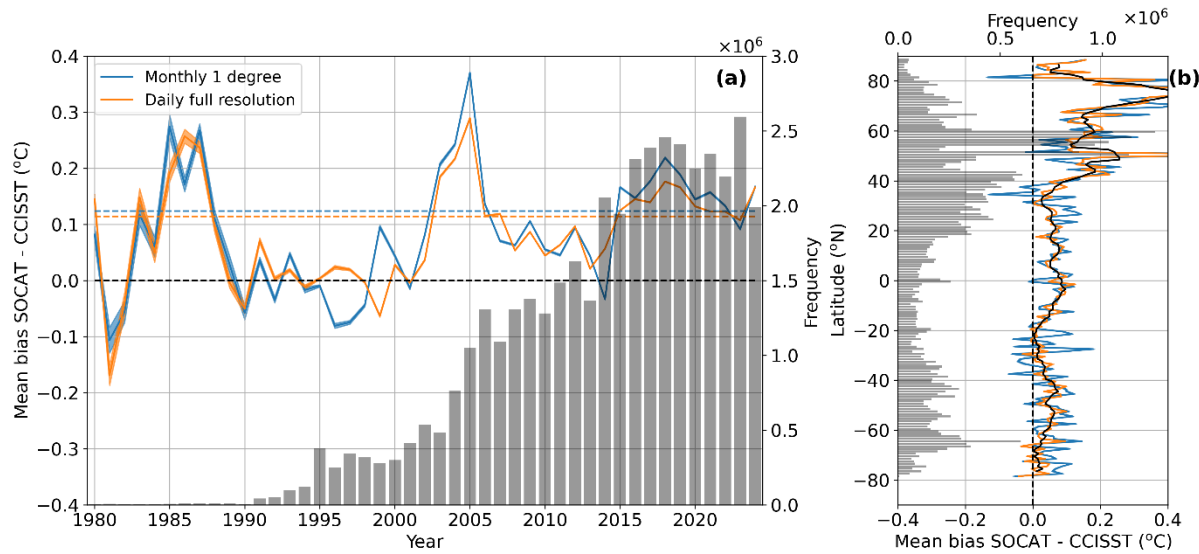


Figure 2: (a) Mean annual SOCAT SST minus the bias corrected CCI-SST between 1980 and 2024 (left y-axis). Blue line indicates the monthly 1 ° data, and orange line the daily full spatial resolution CCI-SST observations. Coloured banding represents the standard error on the calculated bias for the monthly 1 ° (blue) and daily full spatial resolution (orange). Dashed coloured lines indicate the global mean values from Figure 1. Dashed black line indicates a bias of 0. Grey bars indicate the number of observations for each year (right y-axis). (b) Mean SOCAT SST minus bias corrected CCI-SST for 1 ° latitude bins considering all observations. Coloured lines are the same as in (a). Black solid line indicates a 5 ° moving mean of the daily full resolution (orange solid line).

Dong et al.⁷ indicated that there was a latitudinal variability to the SST warm bias within the monthly 1 ° gridded SOCAT SST data. Within our analysis, the warm bias shows a consistent latitudinal variability (Figure 2b) except our analysis shows a stronger warm bias within the high latitude North Hemisphere.

The variability of the warm bias was explored regionally for the recalculated dataset.. Figure 3 shows the mean bias timeseries for the eight Regional Carbon Cycle Assessment and Processes (RECCAP2) regions (and additional regions including the Baltic Sea). These timeseries highlight some of the spatial variability and how regionally there are different temporal changes in the warm bias. These regional differences will reflect the individual systems and platforms used to collect the SOCAT SST data within each region and indicates which regions are likely driving the global mean warm bias. Within the North Atlantic and North Pacific Oceans the warm bias appears to increase from 2000 until present, with the North Atlantic generally having a larger bias than the North Pacific Ocean (Figure 3a, d). The Southern Ocean tends to show a mean bias of ~0 °C for the period 2000 to 2010 around°, before rapidly increasing from 2015 to a maxima in the most recent years (Figure 3j). The South Pacific Ocean generally showed a small mean bias around ~0.04 °C with fluctuations between years (Figure 3g). The Indian Ocean, Arctic Ocean and Mediterranean Sea have limited data, and therefore the mean bias is likely sensitive to variations in the data density (Figure 3 c, f, k). There is a tendency for the Arctic Ocean to have large warm bias (Figure 3c), and the Indian Ocean (Figure 3k) to have a cool bias in the 1990 that increases to a warm bias by 2023. In the global analysis (Figure 2) there was a notable increase in the warm bias

in 2003 to 2005 period, which can be attributed to the Baltic Sea (Figure 3f). These results indicate that there are geographical variabilities in the warm bias, but also temporal variability within each region (Figure 3), which is different from the global analysis (Figure 2). We note that the CCI-SST was corrected for a cool bias, that was assumed to be temporally and spatially homogenous (i.e., the fixed bias correction of 0.04 K covered within the methods), which may impact the regional analyses of the warm bias. However, Figure 3 shows that many regions have a mean SOCAT SST to CCI-SST bias larger than the cool bias correction of 0.04K, which are temporally varying suggesting that the fixed bias correction is not the source.

The colocation between SOCAT SST and the CCI-SST reference SST allows for the recalculation of the SOCAT $f\text{CO}_2(\text{sw})$. Equally the analysis can aid in identifying and understanding the underlying datapoints that contribute to the warm bias within the SOCAT dataset. For each SOCAT SST observation the SOCAT minus CCI-SST difference can be considered alongside the uncertainty information provided with the CCI-SST to indicate the significance of a particular bias. For example, Figure 4 shows the individual SOCAT observations where this difference falls outside two times the CCI-SST uncertainty, which would indicate a ~95% confidence that the difference was significant, for two time periods. These difference results are binned into five year periods, which broadly indicates that about 10% of the SOCAT observations get flagged within the analysis, and individual cruises with particularly large differences are apparent (e.g., some with differences that are greater than 2°C; Figure 4b). This analysis could therefore provide another level of SOCAT quality control and would allow submitters to the SOCAT dataset to return and re-check their data (i.e., towards reducing these warm biases within the SOCAT dataset or identifying the possible causes). Equally, some biases may be real and can be ignored within the SOCAT quality control, for example within the vicinity of the Gulf Stream boundaries where small changes in sampling times between the in situ SOCAT data and the daily SST reference will likely, and correctly, lead to differences due to the rapidly shifting temperature boundaries and moving eddies (i.e Figure 4c showing large positive and negative bias geographical on the same cruise line in the Gulf Stream region).

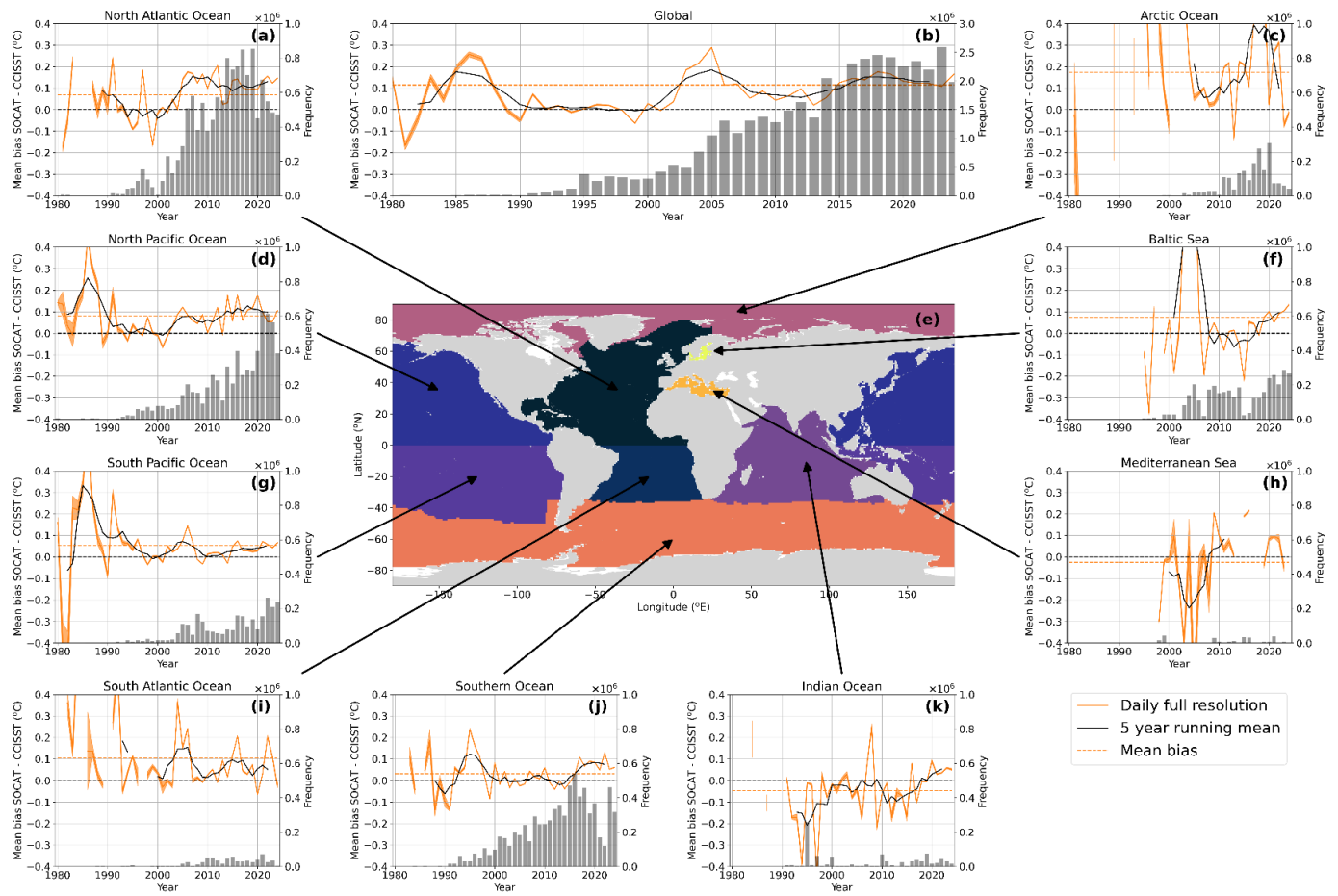


Figure 3: (a) Mean annual SOCAT SST minus the bias corrected CCI-SST between 1980 and 2024, for North Atlantic Ocean (left y-axis). Solid orange line indicates the daily full spatial resolution CCI-SST observations. Dashed coloured line indicate the mean bias for the region. Dashed black line indicates a bias of 0. Solid black line indicates a five year moving average (mean) of the daily full resolution mean bias. Grey bars indicate the number of observations for each year (right y-axis). Note y-axis limits in (b) are different from the remaining subplots. (b) same as (a) for all the data (as in Figure 2). (c) same as (a) for the Arctic Ocean. (d) same as (a) for the North Pacific Ocean. (e) RECCAP region definitions. (f) same as (a) for Baltic Sea. (g) same as (a) for South Pacific Ocean. (h) same as (a) for Mediterranean Sea. (i) same as (a) for South Atlantic Ocean. (j) same as (a) for Southern Ocean. (k) same as (a) for Indian Ocean. Basemap in (e) from Natural Earth v4.0.0 (<https://www.naturalearthdata.com/>).

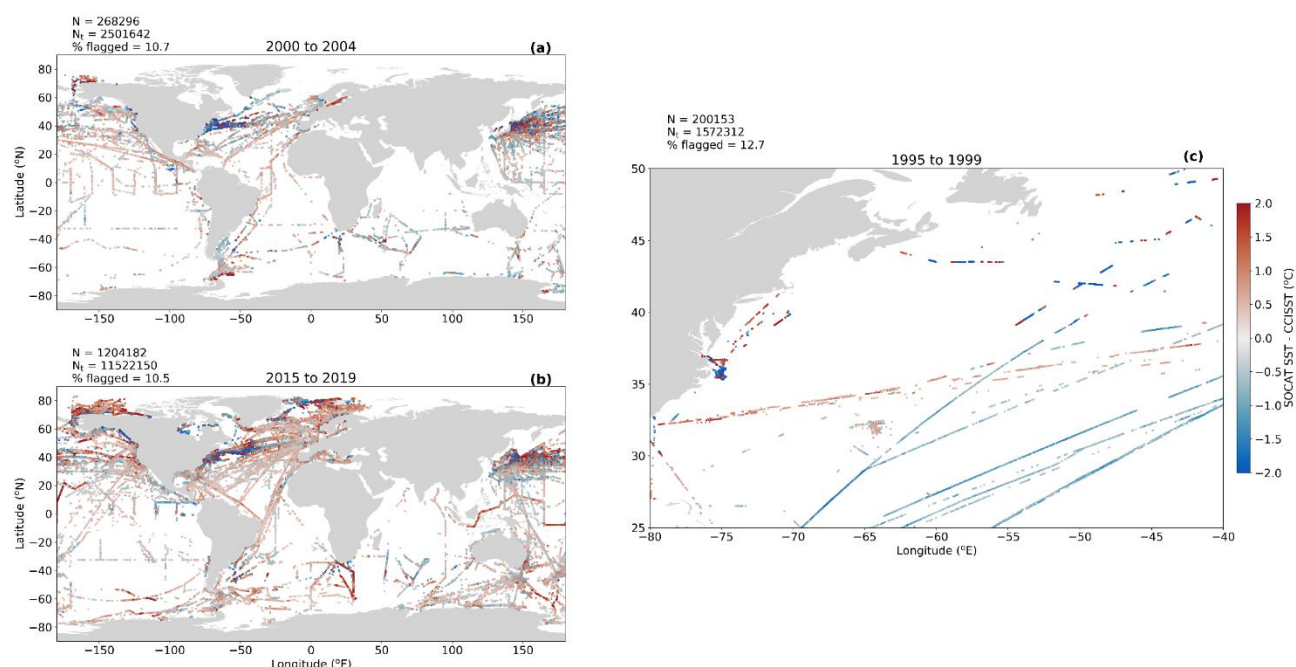


Figure 4: Global map of individual cruises where the SOCAT SST minus CCI-SST difference was greater than two times the CCI-SST uncertainty ($\sim 95\%$ confidence), for cruises in the time period above each plot. Statistics above plot are number of flagged data (N), number of total SOCAT observations (N_t) and percentage of data flagged ($\%$ flagged). Each subplot indicates a different period from (a) 2000 to 2004 and (b) 2015 to 2019. (c) shows a subset region over the Gulf Stream for the period 1995 to 1999. Basemap from Natural Earth v4.0.0 (<https://www.naturalearthdata.com/>).

Impact of updating the temperature sensitivity on the recalculated $f\text{CO}_2$ (sw)

The choice of the temperature sensitivity approach used to recalculate the SOCAT reported $f\text{CO}_2$ (sw) to the satellite reference temperature will affect the resulting recalculated $f\text{CO}_2$ (sw). Comparing the updated temperature sensitivity of Humphreys²² to the previously used Takahashi et al.²¹ we find small differences in the recalculated $f\text{CO}_2$ (sw) (Figure 5). On the global scale, a median difference of $-0.02\ \mu\text{atm}$ indicates that the Humphreys²² temperature sensitivity produces slightly lower $f\text{CO}_2$ (sw) (Figure 5a). This global difference is due to the Humphreys²² approach having a lower sensitivity in cold waters with higher DIC/TA ratios²⁸, and a higher sensitivity in warmer waters when compared to Takahashi et al.²¹. This feature is apparent when the differences were split regionally, into the Northern Hemisphere (Figure 5b), Tropics (Figure 5c) and Southern Hemisphere (Figure 5d). For example, the Northern and Southern Hemispheres there is a tendency for the Humphreys²² method to produce lower $f\text{CO}_2$ (sw) shown by negative median biases of $-0.05\ \mu\text{atm}$ and $-0.13\ \mu\text{atm}$, respectively. These regions largely cover the cooler waters, where the temperature sensitivity was smaller. The reciprocal is apparent in the Tropics (positive median bias) where the temperature sensitivity is larger than Takahashi et al.²¹. In some coastal locations, with low salinity and high $f\text{CO}_2$ (sw), the Humphreys²² temperature sensitivities can lead to larger $f\text{CO}_2$ (sw) differences. These are highlighted by the elevated standard deviations shown particularly in the Northern Hemisphere (Figure 5b). Although these $f\text{CO}_2$ (sw) differences can be relatively large, they affect only $<0.1\%$ of the SOCAT data analysed (or $\sim 40,000$ observations).

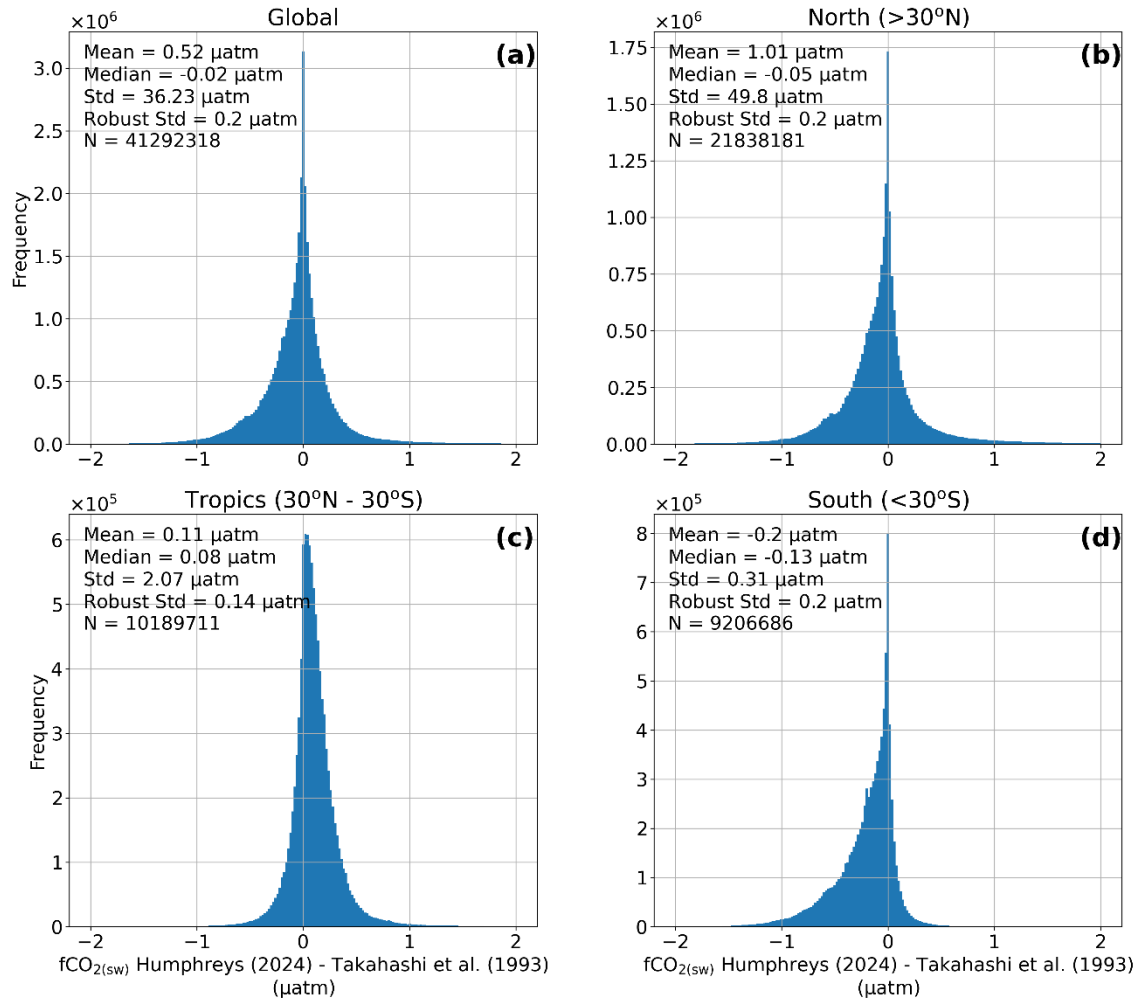


Figure 5: (a) Histogram of the global difference in the recalculated $f\text{CO}_2(\text{sw})$ using the Humphreys²² and Takahashi et al.²¹ temperature sensitivities with the daily bias corrected CCI-SST. In plot statistics are mean bias (Mean), median bias (Median), standard deviation (Std), robust standard deviation (Robust Std) and number of observations (N). (b) same as (a) but for the North Hemisphere (>30°N). (c) same as (a) for the Tropics (30°N to 30°S). (d) same as (a) for the Southern Hemisphere (<30°S).

Reasoning for using the CCI-SST climate data record instead of OI-SST data record

Some previous versions of the recalculated SOCAT datasets have included those recalculated using OI-SSTv2.1 or CCI-SSTv2.1^{14,16}. The temporal stability and consistency of the reference SST data is important when assessing and correcting the warm bias within SOCAT dataset, as any drift or changes in the SST reference will affect the correction of the warm bias and therefore the resultant global ocean CO_2 sink values. The OI-SSTv2.1 data have an inconsistency through time due to an updated methodology applied after 2015, as described in Huang et al.⁹, where data prior (1981 to 2014) follow the previous methodology (v2.0). An inconsistency in the generation of the SST reference data could manifest into a bias change with respect to SST drifters. The climate

record intercomparison¹⁹ and an intercomparison in Yang et al.³³ showed that for the period 1981 to 2010 the OI-SST was cold by about 0.1°C with respect to the in situ SST drifter observations, before transitioning towards being consistent to the SST drifters in 2022. This drift was apparent within the annual mean biases retrieved for OI-SSTv2.1, where larger and more variable warm biases between 1981 to 2010 were observed (Figure 6) compared to CCI-SSTv3.0 (Figure 2).

The OI-SST provides daily SST, but these have a reduced spatial resolution of 0.25 °s (i.e., ~25 km at the equator) compared to the CCI-SST daily data which are ~5 km. If the same analysis conducted for the daily CCI-SST data (Figure 1) is applied to the daily OI-SST data, there is a narrowing of the histogram (Figure 7). But the narrowing is smaller than that achieved when using the daily CCI-SST data. This is particularly clear when evaluating the standard deviation (σ) and robust standard deviation (σ_r ; median absolute deviation scaled by 1.4826 so that it is consistent with the standard deviation for a normal distribution) of each distribution (Figure 1; Table 2, where $\sigma=0.66$ for CCI-SST versus $\sigma=0.80$ for OI-SST, and $\sigma_r=0.31$ for CCI-SST versus $\sigma_r=0.48$ for OI-SST). This finding is consistent with the increased spatial resolution of the CCI-SST (i.e., 5 km with CCI-SST versus the 25 km of OI-SST) further reducing the spatial mismatch between the SOCAT observations and the reference SST dataset used to recalculate them. The reduction in the spatial mismatch will improve the representativeness of the recalculated SOCAT observations when considering individual cruise observations, or smaller regional subsets, further improvement in the applicability of these recalculated data.

It should be noted that the gap filling techniques used by both the OI-SST and CCI-SST methodologies will smooth out features, such that the feature resolution will typically be coarser than the stated gridded resolution except when there is a high density of satellite observations^{33,34}. The increased spatial mismatches due to the lower spatial resolution of the OI-SST data are consistent with the slightly larger mean bias that were identified in comparison to that identified by using the CCI-SST daily 5 km equivalent (Figure 1; Table 2). The greater narrowing of the histogram for the CCI-SST (Figure 1; Table 2) supports the conclusion that the higher spatial resolution of the CCI-SST is more appropriate for use with the SOCAT observations and enhances their use for studies in smaller regions or those that use individual cruises.

Table 2: Summary statistics from the comparison of SOCAT SST minus the satellite reference SST at both monthly 1 ° and the daily native resolution of the reference SST dataset (CCI-SSTv3.0 = ~5km; OI-SSTv2.1 = ~25km). The statistics are standard deviation (Std, σ) and robust standard deviation (Robust Std, σ_r).

Dataset	Monthly 1 °				Daily full resolution			
	Mean (°C)	Median (°C)	Std, σ (°C)	Robust Std, σ_r (°C)	Mean (°C)	Median (°C)	Std, σ (°C)	Robust Std, σ_r (°C)
CCI-SSTv3.0	0.12	0.1	1.04	0.62	0.11	0.07	0.66	0.31
OI-SSTv2.1	0.11	0.1	1.07	0.67	0.11	0.1	0.80	0.48

Overall the reduced temporal stability of the OI-SSTv2.1 throughout the 1981 to 2024 period and the higher spatial resolution of the CCI-SST which results in a closer temporal and spatial match to the SOCAT dataset observations, makes the CCI-SSTv3.0 more suitable for assessing and removing the warm bias within the SOCAT dataset.

Although we show that the CCI-SST is more suitable for assessing the warm bias within SOCAT, we must highlight that this is a particularly demanding application for any of these SST climate data records (i.e., requiring an accuracy and stability on the order of 0.01 °C). These climate data records are under constant development^{9,10}, and this work highlights the importance of the ongoing work to revisit satellite-era data for the purpose of creating more accurate and stable climate data records. °

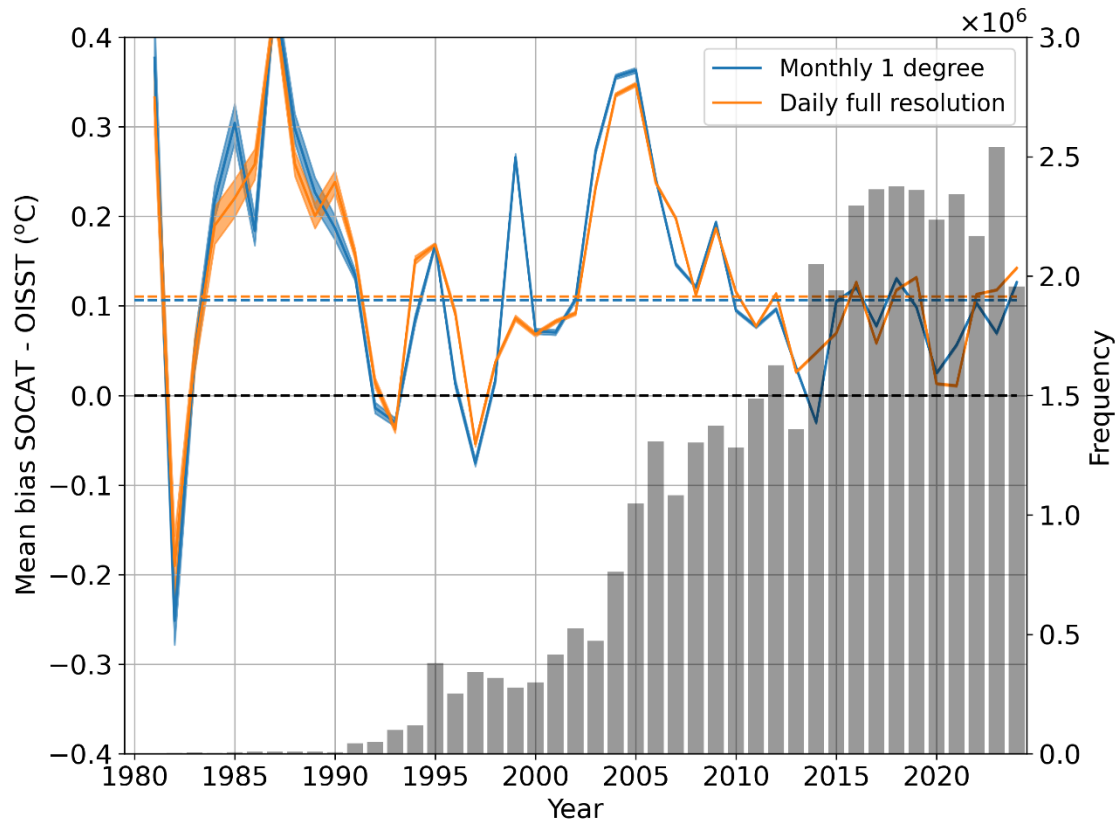


Figure 6: Mean annual SOCAT SST minus the OI-SSTv2.1 between 1981 and 2024 (left y-axis). Blue line indicates the monthly 1 ° data, and orange line the daily 0.25 ° spatial resolution OI-SST observations. Dashed coloured lines indicate the global mean values from Figure 7. Dashed black line indicates a bias of 0. Grey bars indicate the number of observations for each year (right y-axis).

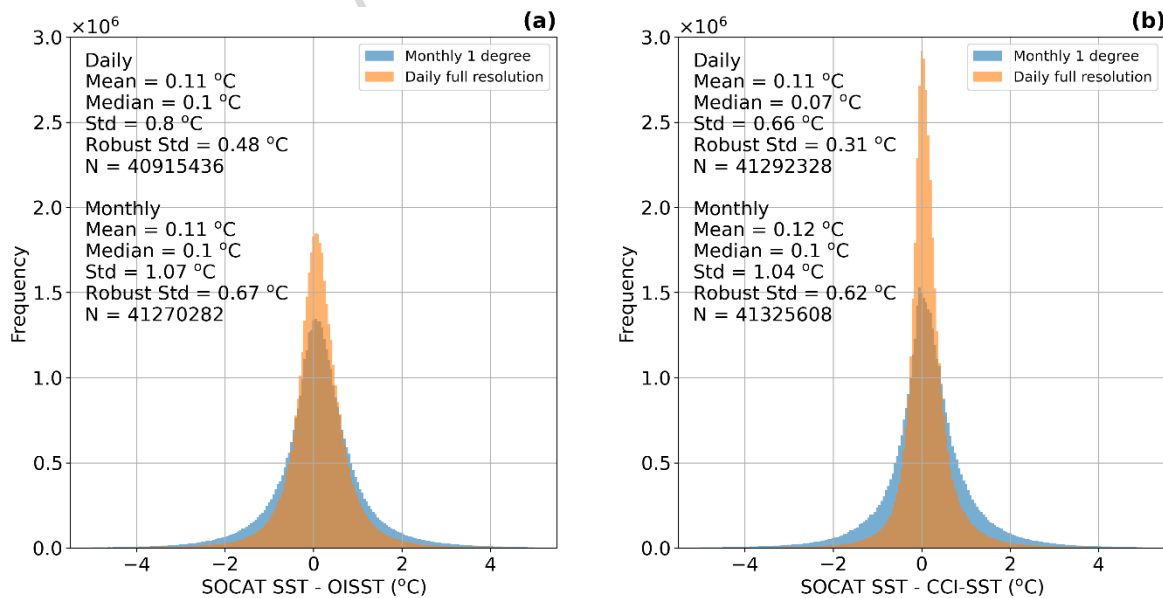


Figure 7: (a) Histogram of the SOCAT SST minus the bias corrected OI-SSTv2.1 plotted in 0.05 °C bins for the monthly 1 °, and daily 0.25 ° resolution reference SST data. In plot statistics are mean bias (Mean), median bias (Median), standard deviation (Std), robust standard deviation (Robust Std) and number of observations (N). (b) same as (a) but for the CCI-SSTv3.0 as plotted in Figure 1.

Demonstrating the impact of the recalculation on the estimate of the global ocean CO₂ sink

The recalculation of the SOCAT $f\text{CO}_2(\text{sw})$ to a reference SST dataset corrects for the warm bias present within SOCAT, which if left uncorrected, leads to a modified air-sea CO₂ flux global ocean CO₂ sink^{7,8,35–37}. Here we use the University of Exeter Feed Forward Neural Network with uncertainties for interpolating the $f\text{CO}_2(\text{sw})$ data and calculating the air-sea gas fluxes and net sink (UExp-FNN-U)^{8,35} to provide a single example of the impact of removing the warm bias (Figure 8). The UExp-FNN-U methodology uses the self-organising map feed forward neural network (SOM-FNN) two step approach to estimate $f\text{CO}_2(\text{sw})$. The SOM phase geographically clusters the global oceans into 16 regions with a similar seasonality in $f\text{CO}_2(\text{sw})$. For each cluster region, an FNN ensemble is trained to estimate the $f\text{CO}_2(\text{sw})$ within each region from a set of predictor variables. The approach follows similar principles to some other $f\text{CO}_2(\text{sw})$ interpolation methodologies (e.g.^{38,39}), but the different prediction approaches (e.g. XG-Boost, decision trees, multi linear regression) and methodological structures could show different sensitivities, so this analysis is provided as an exemplar. Within the UExp-FNN-U methodology, the air-sea CO₂ fluxes were computed following the bulk methods in Ford et al.³⁵ but with the cool salty skin correction omitted.

The UExp-FNN-U methodology was first run using the original SOCAT $f\text{CO}_2(\text{sw})$ observations (i.e. the non-recalculated version that contain the warm bias) to set the ‘baseline’, and this run is consistent with the majority of the current $f\text{CO}_2(\text{sw})$ data-products used within the annual Global Carbon Budget assessments¹. The UExp-FNN-U method was then run with the recalculated $f\text{CO}_2(\text{sw})$ data that were recalculated using the daily bias corrected CCI-SST with the Humphrey’s temperature sensitivity (i.e., the dataset presented and described within this paper). The differences between the two analyses (Figure 8) are generally consistent with the identified mean warm bias’s (i.e., those shown in Figure 2). Around the year 2000 the recalculated data showed a $\sim 0.05 \text{ Pg C yr}^{-1}$ decrease in the ocean CO₂ sink. After the year 2000, the difference grows to an increase in the ocean CO₂ sink of $\sim 0.4 \text{ Pg C yr}^{-1}$ in 2024. These results show that if the original SOCAT dataset is used, the ocean CO₂ uptake would likely be underestimated from 2010 onwards due to the increasing warm bias. This underlines the importance of recalculating the SOCAT data to a temporally stable climate data record before using them in any ocean carbon sink assessments.

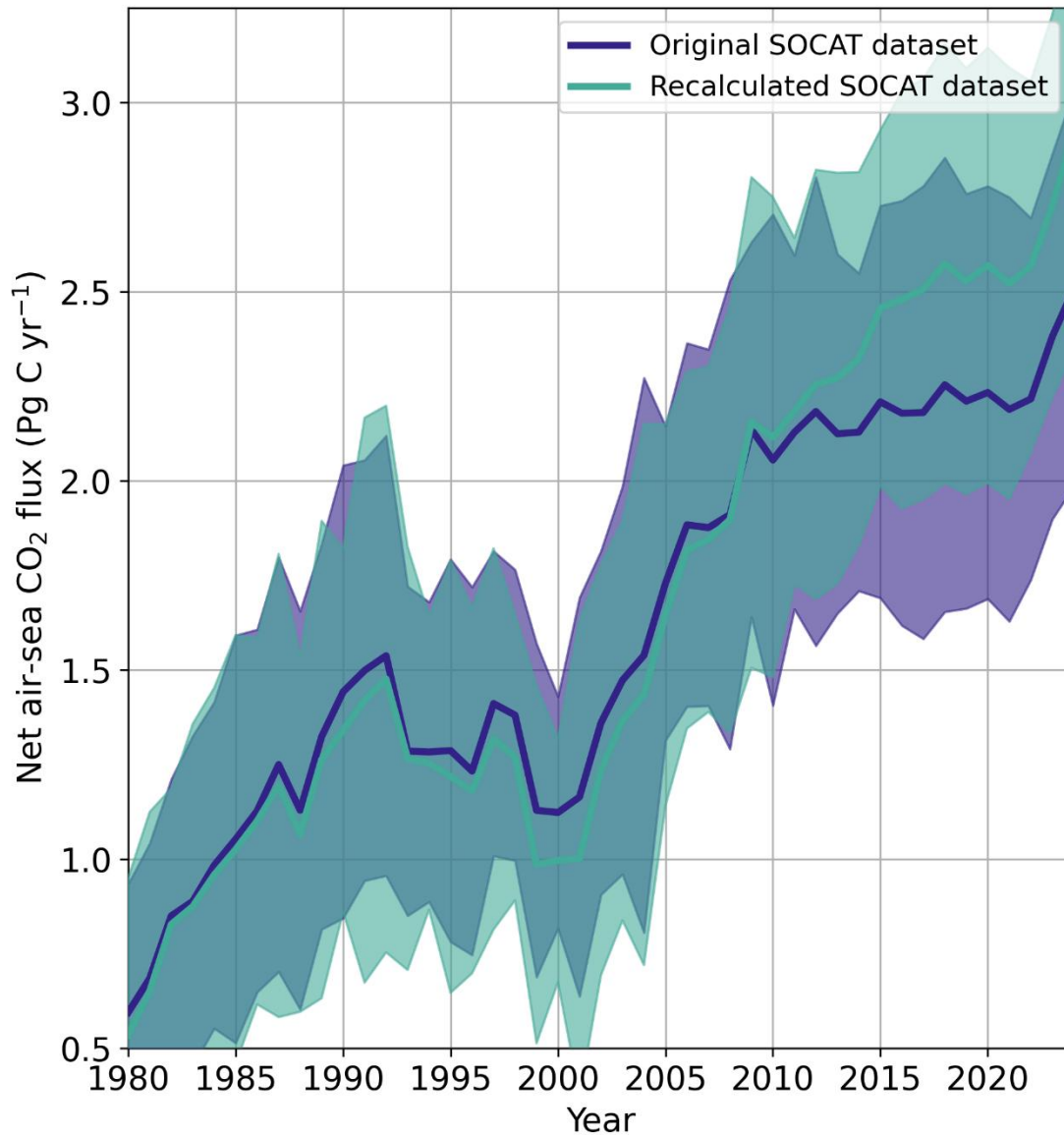


Figure 8: Global ocean net air-sea CO₂ flux, or ocean CO₂ sink using the UExP-FNN-U methodology. The ‘Original SOCAT dataset’ line indicates the UExP-FNN-U run with the original SOCAT dataset, and the ‘Recalculated SOCAT dataset’ line uses the recalculated SOCAT dataset as described within this paper. Positive net air-sea CO₂ fluxes indicate CO₂ uptake by the oceans. The ocean CO₂ sink values presented do not include the fixed riverine flux adjustment as included within the Global Carbon Budget assessments. Shaded region indicates the 1 σ uncertainty on the global ocean CO₂ sink due only to the $f\text{CO}_2(\text{sw})$ interpolation as described in Ford et al.³⁵, as this is the component that would be directly affected by changing the original SOCAT to the recalculated SOCAT dataset.

Data Availability

The recalculated SOCATv2025 dataset is available at Ford et al.³⁰. The recalculated SOCAT dataset will continue to be produced and the repository information and datasets will be given on the FluxEngine^{31,32} GitHub repository (<https://github.com/oceanflux-ghg/FluxEngine>).

Code Availability

The FluxEngine Python toolbox (v4.1.0)^{31,32} contains the updated software implementation of the recalculation methodology. The main driver script for this is ‘fe_reanalyse_fco2_driver.py’, which can be run from the command line.

Acknowledgements

This work was funded by the European Space Agency under the project Ocean Carbon for Climate (OC4C; 3-18399/24/I-NB). The Surface Ocean CO₂ Atlas (SOCAT) is an international effort, endorsed by the SCOR Infrastructural Project International Ocean Carbon Coordination Project (IOCCP) and the Surface Ocean Lower Atmosphere Study (SOLAS), to deliver a uniformly quality-controlled surface ocean CO₂ database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT. We thank two anonymous reviewers for their comments which have improved the manuscript and underlying dataset. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Funding

This work was funded by the European Space Agency under the project Ocean Carbon for Climate (OC4C; 3-18399/24/I-NB).

References

1. Friedlingstein, P. *et al.* Global Carbon Budget 2025. Preprint at <https://doi.org/10.5194/essd-2025-659> (2025).
2. Bakker, D. C. E. *et al.* A multi-decade record of high-quality fCO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT). *Earth System Science Data* **8**, 383–413 (2016).
3. Bakker, D. C. E. *et al.* Surface Ocean CO₂ Atlas Database Version 2025 (SOCATv2025) (NCEI Accession 0304549). NOAA National Centers for Environmental Information <https://doi.org/10.25921/648F-FV35> (2025).

4. Woolf, D. K., Land, P. E., Shutler, J. D., Goddijn-Murphy, L. M. & Donlon, C. J. On the calculation of air-sea fluxes of CO₂ in the presence of temperature and salinity gradients. *Journal of Geophysical Research: Oceans* **121**, 1229–1248 (2016).
5. McGillis, W. R. & Wanninkhof, R. Aqueous CO₂ gradients for air – sea flux estimates. *Marine Chemistry* **98**, 100–108 (2006).
6. Goddijn-Murphy, L. M., Woolf, D. K., Land, P. E., Shutler, J. D. & Donlon, C. The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO₂ fugacity in support of air-sea gas flux studies. *Ocean Science* **11**, 519–541 (2015).
7. Dong, Y. *et al.* Update on the Temperature Corrections of Global Air-Sea CO₂ Flux Estimates. *Global Biogeochemical Cycles* **36**, (2022).
8. Watson, A. J. *et al.* Revised estimates of ocean-atmosphere CO₂ flux are consistent with ocean carbon inventory. *Nature Communications* **11**, 1–6 (2020).
9. Huang, B. *et al.* Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate* **34**, 2923–2939 (2021).
10. Embury, O. *et al.* Satellite-based time-series of sea-surface temperature since 1980 for climate applications. *Sci Data* **11**, 326 (2024).
11. Holding, T., Ashton, I. G. C. & Shutler, J. D. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2019. 1 GBytes PANGAEA <https://doi.org/10.1594/PANGAEA.905316> (2019).
12. Holding, T., Ashton, I. G. C. & Shutler, J. D. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2020. 1.4 GBytes ICOS ERIC - Carbon Portal <https://doi.org/10.18160/VMT4-4563> (2020).
13. Shutler, J. D., Holding, T., Sims, R. P. & Ashton, I. G. C. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2021. 1.4 GBytes PANGAEA <https://doi.org/10.1594/PANGAEA.939233> (2021).

14. Sims, R. P., Ford, D. J., Shutler, J. D., Ashton, I. & Holding, T. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2022. Zenodo
<https://doi.org/10.5281/ZENODO.8228586> (2023).
15. Merchant, C. J. *et al.* Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci Data* **6**, 223 (2019).
16. Ford, D. J., Sims, R. P., Shutler, J. D., Ashton, I. & Holding, T. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2023. Zenodo
<https://doi.org/10.5281/ZENODO.8229316> (2023).
17. Ford, D. J., Shutler, J. D., Ashton, I., Sims, R. P. & Holding, T. Reanalysed (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2024 (v1.1). Zenodo
<https://doi.org/10.5281/ZENODO.13284017> (2024).
18. Good, S. A. & Embury, O. ESA Sea Surface Temperature Climate Change Initiative (SST_cci): Level 4 Analysis product, version 3.0. 15341 Files, 246711441999 B NERC EDS Centre for Environmental Data Analysis [dataset] <https://doi.org/10.5285/4A9654136A7148E39B7FEB56F8BB02D2> (2024).
19. Atkinson, C. *et al.* ESA CCI Phase 3 Sea Surface Temperature (SST): Climate Assessment Report D5.1 v1.1. https://climate.esa.int/documents/2370/SST_CCI_D5.1_CAR_v1.1-signed.pdf (2023).
20. Embury, O. SST CCI Product Validation and Intercomparison Report.
https://climate.esa.int/documents/2369/SST_CCI_D4.1_PVIR_v2.1-signed.pdf (2023).
21. Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W. & Sutherland, S. C. Seasonal variation of CO₂ and nutrients in the high-latitude surface oceans: A comparative study. *Global Biogeochemical Cycles* **7**, 843–878 (1993).
22. Humphreys, M. P. Temperature effect on seawater fCO₂ revisited: theoretical basis, uncertainty analysis and implications for parameterising carbonic acid equilibrium constants. *Ocean Sci.* **20**, 1325–1350 (2024).
23. Humphreys, M. P., Lewis, E. R., Sharp, J. D. & Pierrot, D. PyCO2SYS v1.8: marine carbonate system calculations in Python. *Geosci. Model Dev.* **15**, 15–43 (2022).

24. Humphreys, M. P. *et al.* PyCO2SYS: marine carbonate system calculations in Python. Zenodo
<https://doi.org/10.5281/ZENODO.3744275> (2024).
25. Pfeil, B. *et al.* A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT). *Earth System Science Data* **5**, 125–143 (2013).
26. Takahashi, T. *et al.* Climatological mean and decadal change in surface ocean pCO₂, and net sea-air CO₂ flux over the global oceans. *Deep-Sea Research Part II: Topical Studies in Oceanography* **56**, 554–577 (2009).
27. Gregor, L. & Gruber, N. OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth Syst. Sci. Data* **13**, 777–808 (2021).
28. Wanninkhof, R., Pierrot, D., Sullivan, K., Mears, P. & Barbero, L. Comparison of discrete and underway CO₂ measurements: Inferences on the temperature dependence of the fugacity of CO₂ in seawater. *Marine Chemistry* **247**, 104178 (2022).
29. Sabine, C. L. *et al.* Surface Ocean CO₂ Atlas (SOCAT) gridded data products. *Earth System Science Data* **5**, 145–153 (2013).
30. Ford, D. J., Shutler, J. D., Ashton, I., Sims, R. P. & Holding, T. Recalculated (depth and temperature consistent) Surface Ocean CO₂ Atlas (SOCAT) version 2025. Zenodo
<https://doi.org/10.5281/ZENODO.15656802> (2025).
31. Holding, T. *et al.* The FluxEngine air–sea gas flux toolbox: simplified interface and extensions for in situ analyses and multiple sparingly soluble gases. *Ocean Science* **15**, 1707–1728 (2019).
32. Shutler, J. D. *et al.* FluxEngine: A flexible processing system for calculating atmosphere-ocean carbon dioxide gas fluxes and climatologies. *Journal of Atmospheric and Oceanic Technology* **33**, 741–756 (2016).
33. Yang, C. *et al.* Sea Surface Temperature Intercomparison in the Framework of the Copernicus Climate Change Service (C3S). *Journal of Climate* **34**, 5257–5283 (2021).
34. Reynolds, R. W. *et al.* Objective Determination of Feature Resolution in Two Sea Surface Temperature Analyses. *Journal of Climate* **26**, 2514–2533 (2013).
35. Ford, D. J. *et al.* A Comprehensive Analysis of Air-Sea CO₂ Flux Uncertainties Constructed From Surface Ocean Data Products. *Global Biogeochemical Cycles* **38**, e2024GB008188 (2024).

36. Shutler, J. D. *et al.* Satellites will address critical science priorities for quantifying ocean carbon. *Frontiers in Ecology and the Environment* **18**, 27–35 (2020).
37. Woolf, D. K. *et al.* Key Uncertainties in the Recent Air-Sea Flux of CO₂. *Global Biogeochemical Cycles* **33**, 1548–1563 (2019).
38. Chau, T.-T.-T., Gehlen, M., Metzl, N. & Chevallier, F. CMEMS-LSCE: a global, 0.25°, monthly reconstruction of the surface ocean carbonate system. *Earth Syst. Sci. Data* **16**, 121–160 (2024).
39. Bennington, V., Gloege, L. & McKinley, G. A. Variability in the Global Ocean Carbon Sink From 1959 to 2020 by Correcting Models With Observations. *Geophysical Research Letters* **49**, (2022).