ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

C3S2 311 Lot 1: Access to a comprehensive archive of historical surface observations, with support to data rescue

Eighth version Marine User Guide

Issued by: NUIM / Peter Thorne
Date: 27/11/2024
Ref: C3S2_D311_Lot1.2.3.2.202308_Eighth version Marine_User_Guide
Official reference number service contract: 2021/C3S2_311_Lot1_NUIM/C2

This document has been produced in the context of the Copernicus Climate Change Service (C3S).
The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of C3S on behalf of the European Union (Delegation Agreement signed on 11/11/2014). All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose.
The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission and the European Centre for Medium-Range Weather Forecasts has no liability in respect of this document, which is merely representing the authors view.

Contributors

NATIONAL OCEANOGRAPHY CENTRE (NOC)

1. Stavroula Biri[1]
2. Richard Cornes
3. Elizabeth Kent
4. Thomas Cropper

DEUTSCHER WETTERDIENST (DWD)

1. Andreas Wernecke[2]
2. Axel Andersson
3. Tina Leiding
4. Martina Schulz

NATIONAL UNIVERSITY OF IRELAND MAYNOOTH (NUIM)

1. Peter Thorne
2. Corinne Voces

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION (NCEI)

1. Huai-min Zhang

The service would like to acknowledge in kind contributions by NOC and NOAA NCEI and their valuable and continuing input and support to the scientific and technical development, improvement and production of the data sets released through this service.

[1] SB left the NOC in Spring 2023
[2] AW left DWD in June 2023

Contents

Executive Summary

The C3S2 311 Lot 1 GLAMOD (Global Land and Marine Observations Database) service is concerned with the provision of globally available land and marine surface meteorological records. The service includes inventorying of, and brokering access to, data sources, their harmonization (via conversion to a Common Data Model, merging, and quality assurance) and their provision via the Copernicus Climate Change Service Data Store (CDS).

This marine user guide describes all relevant aspects of the marine data service necessary for a user to access and work with the marine data appropriately and with confidence. This document does not constitute a technical service document which instead is available separately. There is also an accompanying land user guide that describes the land data and its processing.

This is a living document which shall be subject to regular revision to reflect the status of the service at any given point in time. Releases of this document will always accompany new data releases. On an exceptional basis, as warranted, additional releases may occur to clarify issues raised by users or document important changes between releases such as any modification to the modalities of data access. Feedback on the adequacy and completeness of the document is welcomed at any time. Feedback should be provided via the C3S helpdesk facility to allow C3S tracking of user input. All such feedback shall be passed on in full to the service team for due consideration.

Former versions are archived and available upon request. The version history is given below:

Version	**Release Date**	**Release notes**
0.0	31/08/17	Initial version consisting of section outlines and description of initial archiving material
1.0	14/12/17	Updates to reflect the status of the test release
2.0	17/12/18	Update to reflect status of beta release
3.0	14/10/19	Update for first data release
4.0	21/07/20	Update after second data release
5.0	29/03/21	Update after third data release
6.0	03/08/21	Update after fourth data release
7.0	12/06/23	Update after fifth data release
8.0	30/08/23	Update after sixth data release

Quick start

1 Quick start

1.1 Overview

The C3S2 311 Lot 1 - Global Land and Marine Observations Database (GLAMOD) service provides access to in situ surface observations from various climate archives in a common data model. This document describes the marine component of the service, providing access to surface marine meteorological observations. This component is currently based on Release 3.0 and Release 3.0.2 of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman et al. [2017], Liu et al. [2022]) and has been enhanced by:

- advanced quality control tests;
- enhanced duplicate detection and flagging;
- association of instrumental metadata with the observations.

Additional sources, such as from data recovery and digitisation, will be included in future releases.

The observations in ICOADS come from a range of observing platforms, before 1980 these are primarily ships but with an increasing number of measurements made by sensors installed on moored and drifting buoys after this date. The ship observations are clustered over the shipping routes prevalent at the time of observation. Coverage by drifting buoys tends to be more dispersed but with lower sampling over regions of ocean upwelling. Moored buoy data, with the exception of the tropical arrays, tend to be more coastal and concentrated around North America and Europe.

Each meteorological report in ICOADS contains observations of multiple essential climate variables (ECVs, e.g. Bojinski et al. [2014]) made at the same time and location, for example coincident measurements of the air temperature, humidity, wind speed and direction, sea surface temperature and sea level pressure. From the mid 2000s the record is dominated by reports from drifting buoys but these only report a small subset of the ECVs included in this service, typically only sea level pressure or sea surface temperature. From 2015 on, this release does not contain drifting buoy data from the ICOADS Release 3.0.2. A consolidated drifting buoy data record based on additional data sources such as the C-RAID projects drifting buoy dataset (Zunino Rodriguez [2023]) will be included in future releases.

The reporting frequency ranges from hourly or more frequent in the recent period to daily observations prior to ~1860. In the intervening period observing frequency ranged from three times per day based on the watches on board ships to 3-hourly or 6-hourly observations.

Several sources of metadata have been used, for the period after ~1982 metadata from WMO-No. 47, the "List of Selected, Supplementary and Auxilary Ships" (e.g. Kent et al. [2007]) has been merged with the ship observations. Some ships observations prior to 1982 have also been merged with the WMO-No. 47 metadata but this is hampered by a lack of ship callsigns within the data before this date. Additional metadata has been extracted from the ICOADS Supplemental Data (see Freeman et al. [2017]) and documentation available on the ICOADS website.

Apart from the summary information presented in this document, a detailed overview of the data contents for each of the input sources is available in the Marine Data Inventory (C3S2_D311_Lot1.2.3.1.202308_Updated Inventory for Marine datasets_v3).

1.2 Essential Climate Variables and observing methods

Table 1 lists the Essential Climate Variables included in the current release. Table 1 also summarises the principle observing methods and changes over time for those ECVs. Early observations (< 1950s) were typically recorded in whole units, for example whole degrees Celsius or Fahrenheit depending on national practice. In some cases the data were recorded to higher precisions but truncated or rounded when digitised (e.g. Chan et al. [2019]) or shared operationally. This rounding for operational data streams persisted until the 1980s (e.g. Willett et al. [2008]). Later observations were typically recorded and shared with precisions of tenths or higher. As part of the processing to create ICOADS (and earlier datasets) the observations were converted to standardised units (°C for temperatures, hPa for pressure and m/s for speeds) and uniform resolution, typically to one or two decimal places. These have been further processed as part of the GLAMOD service and fully converted to S.I. units. Where recorded, the originally reported value and units of the observations can be traced back to the original input data. There are known issues with the marine meteorological observations, these are discussed further in Section 2. No corrections or adjustments have been applied as part of this service, instead the user is directed to the literature identified within Section 2.

1.3 Coverage

The temporal coverage of the data available in the current release spans the period 1851 - 2022, Figure 1 shows the coverage per ECV. Before the Brussels International Maritime Conference of 1853 the most common ECVs to be reported were: air temperature; wind speed / force and direction; and sea level pressure, with typically $\sim$100 - 200 observations per month. A small number of sea surface temperature and humidity observations are available (< 100). Following the Brussels conference there is a large increase in number of observations to O(10,000) per month for all parameters excluding dew point temperature. Dew point observations only reach 10,000 per month on a routine basis around the time of the First World War. For all parameters, the number of observations increases over time to around 100,000 per month in the 1960s and remains at this level until the 1990s. From the mid 1990's there is a large increase in the number of sea level pressure and sea surface temperature observations from drifting buoys. From 2015 on, drifting buoys are not included in this release, resulting in a drop in the number of sea surface temperature and sea level observations.

Figure 2 shows the total number of reports and total number of months per 1x1 degree grid cell, however over the period 1851 - 2022 there have been major changes to the spatial sampling. Early in the record there were typically a few hundred 1x1 degree grid cells with data per month (see Figure 1). These tended to be concentrated in the Atlantic Ocean and around the southern capes to the Indian and Pacific Oceans. With the opening of the Suez and then Panama canals the major shipping routes changed, with less shipping transiting the capes and a greater proportion transiting the Mediterranean and Caribbean seas to reach the Indian and Pacific oceans.

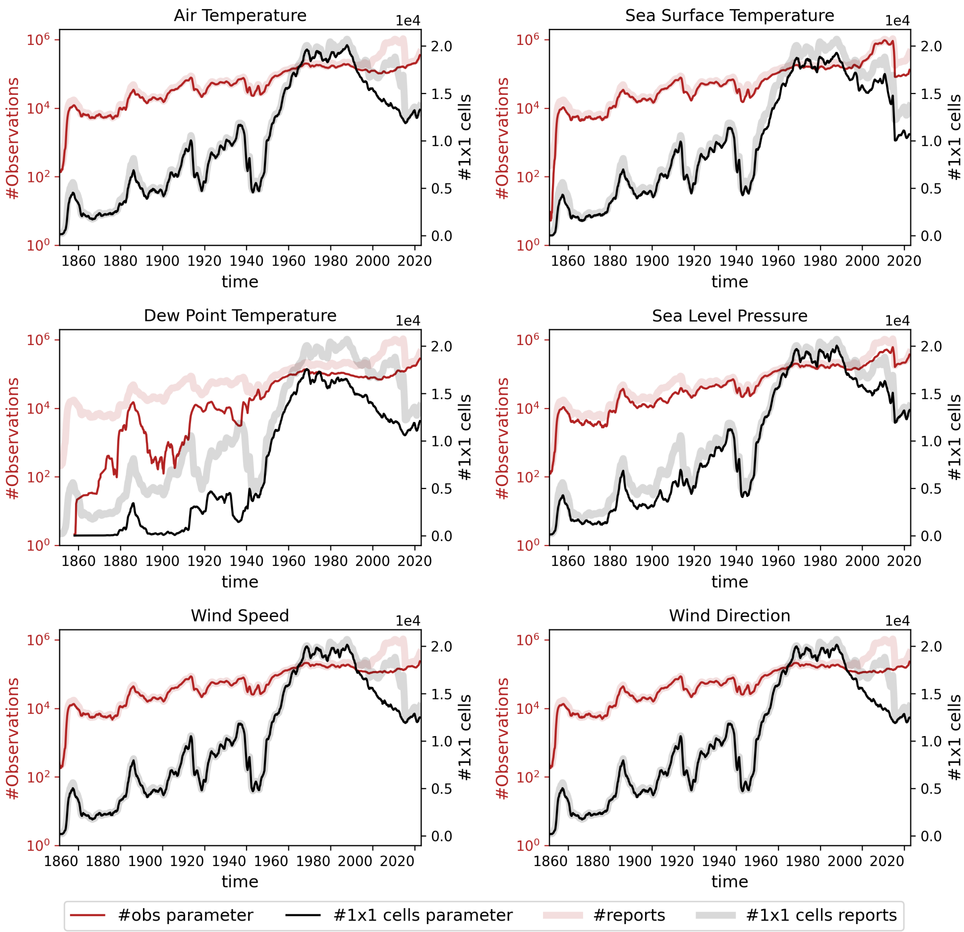

Figure 1: Availability of ECVs listed in Table 1 by time. Both the number of observations (red) and number of grid cells (black) with data are shown. Note log scale for the number of observations.

Spatial coverage reached a peak between the 1960s and 1990s but declined sharply after 1990. The number of grid cells dropped from around 20,000 to around 10,000 - 12,000 by 2020 for the majority of parameters. The contrast between the stable number of observations and decreasing spatial coverage (contrasting red and black lines in Figure 1) is due to the change in reporting frequency from 6 hourly to 3 or 1 hourly and greater over the past 30 years, increasing the temporal sampling at the cost of spatial sampling. Sea level pressure and sea surface temperature experience a much smaller marked decline in spatial coverage due to the contribution of drifting buoys.

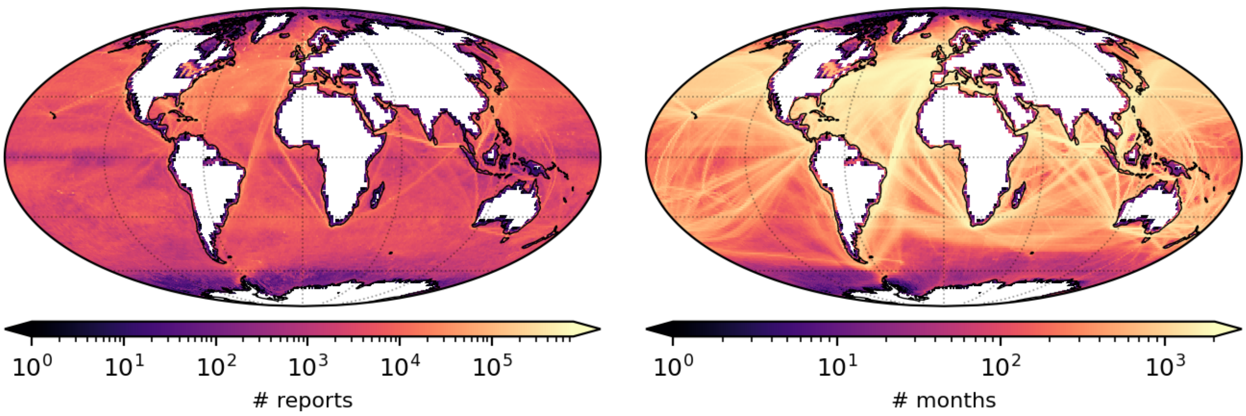

Figure 2: Spatio-temporal distribution of marine data holdings 1851 - 2022. Left: number of reports per grid cell; right: number of months with at least 1 report. Grid size is 1x1 degrees lat/lon. Note logarithmic colour scale. Only reports passing the report level QC checks flag have been included.

1.4 Data access

The data can be downloaded and accessed via the Copernicus Climate Change Service (C3S) Climate Data Store (CDS):

`https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-surface-marine?tab=form`.

The page has 3 tabs, the Overview tab presents an overview of the marine data and provides details on the data description, main and related variables available, contact email and links to licence/data policy statements. The Documentation tab provides links to PDF's of the Marine Product User Guide, Marine data inventories, Common Data Model specifications and a link to the Data Deposit Server webpage. The Download Data tab provides 6 sections that need to be completed in order to download data. The first selects one or more variables (Figure 3) from the available ECVs (Table 1). The second selects the level of quality control to apply, selecting either those that have passed all stages of quality control or all observations (Figure 4). The next three select the time period by selecting the year (Figure 5), month (Figure 6) and day (Figure 7). The final selection selects the region, specifying either all data (global) or selecting observations within a bounding box (Figure 8).

Data can additionally be downloaded using the CDS Python API. Selecting the `Show API request` button reveals the API commands to use for the given selection. Example code can be found in Figure 28. More information can be found at `https://cds.climate.copernicus.eu/api-how-to`

Table 1: Essential Climate Variables included in the GLAMOD service

Variable	Description
Air temperature	Air temperature at observation height, ranging from 4 - 5 m in the 19th century to over 30 m for the modern period. Typically measured using liquid in glass thermometers (mercury or alcohol) but with a transition to electronic sensors over the past 30 years.
Dew point temperature	Dew point temperature at observation height, usually the same as for air temperature. Typically calculated from wet and dry bulb thermometers housed in marine screens or whirling psychrometers. Over the past 30 years there has been a transition to electronic sensors measuring relative humidity.
Sea surface temperature	Water temperature near the sea surface measured using a variety of methods. This temperature does not represent the skin sea surface temperature. Depending on the method, the temperature is measured in depths from a few centimeters to several meters. Typically bucket measurements pre 1950 but with a transition to engine intake measurements beginning around 1900. Dominated by drifting buoy observations after $\sim$2003.
Sea level pressure	Atmospheric pressure reduced to sea level and corrected for temperature where appropriate. Methods range from mercury barometers in the 19th Century, through aneroid barometers to electronic sensors. It should be noted that the sea level reduction and temperature corrections are applied at the time of observation.
Wind speed	Wind speed at observation height. As with air temperature and dew point temperature, the observation height has varied from $<$ 10 m in the early record to more than 30 m for recent years. Methods include visual wind speed estimation for the early record transitioning to mechanical sensors (cup/propellor anemometers) and electronic sensors (sonic anemometers). Instrumental observations corrected for ship speed and course (not applicable for visual estimation).
Wind Direction	See wind speed.

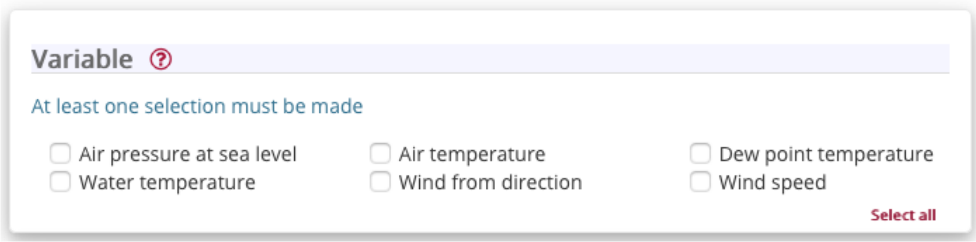

Figure 3: Selection box for variable selection on the Global Land And Marine Observations Database download page on the CDS. It should be noted that ”water temperature” is used in the data model to include river, lake and sea surface temperature.

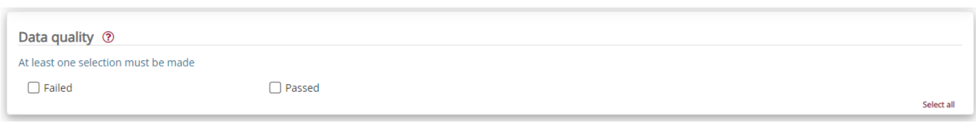

Figure 4: Selection box for quality status on the Global Land And Marine Observations Database download page on the CDS.

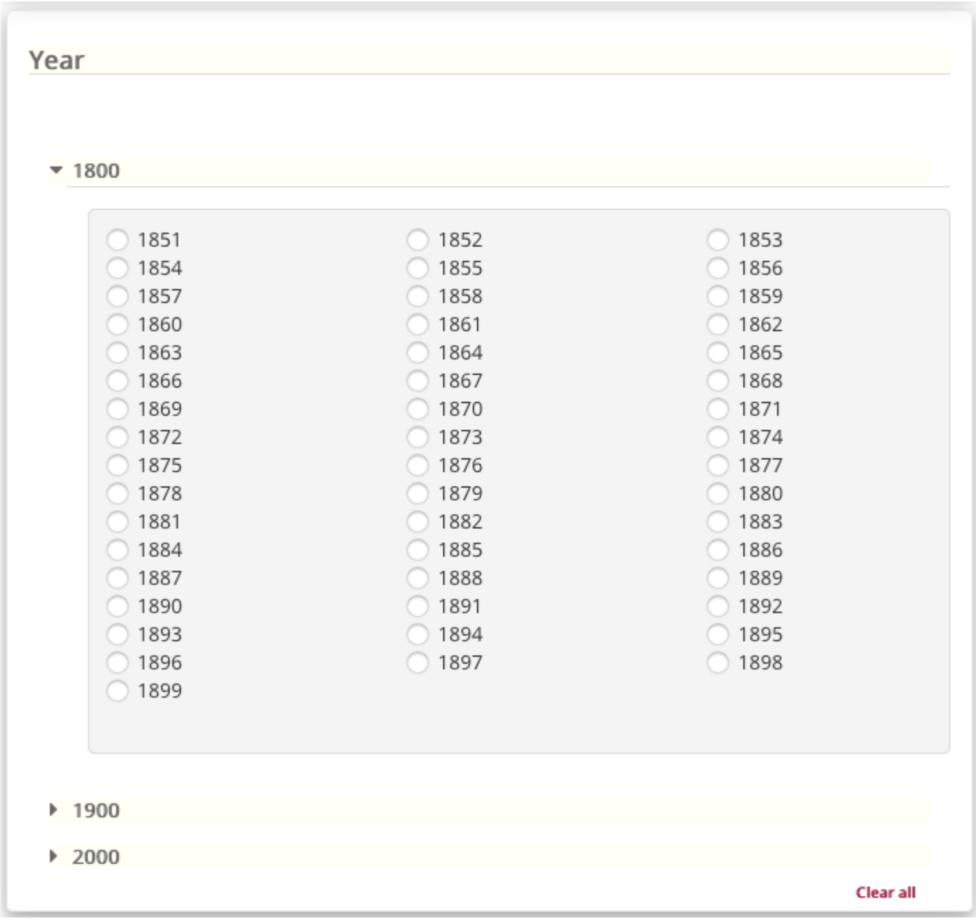

Figure 5: Selection box for selecting year on the Global Land And Marine Observations Database download page on the CDS.

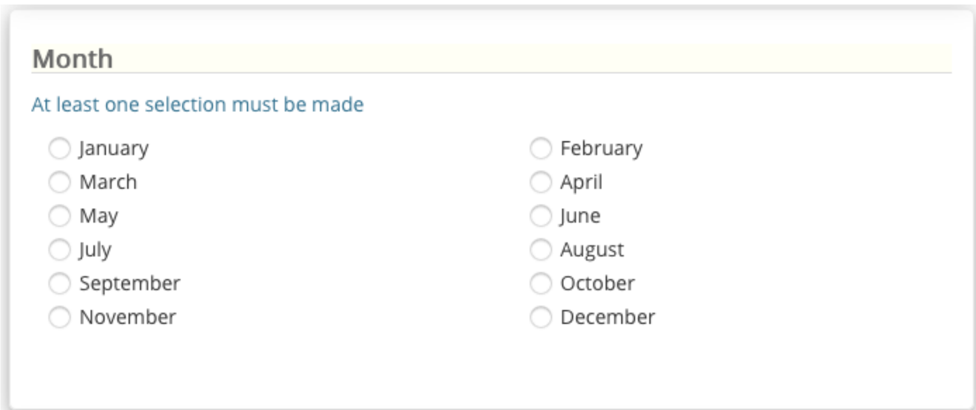

Figure 6: Selection box for selecting month on the Global Land And Marine Observations Database download page on the CDS.

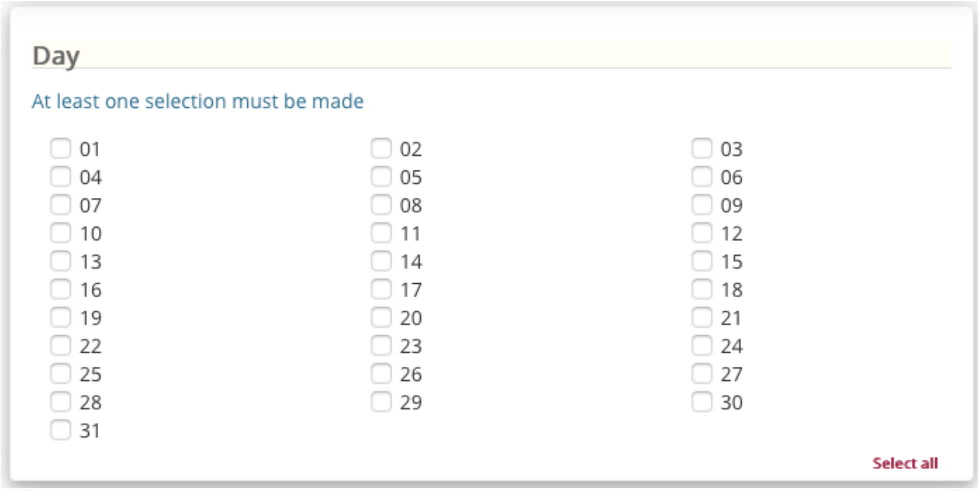

Figure 7: Selection box for selecting day on the Global Land And Marine Observations Database download page on the CDS.

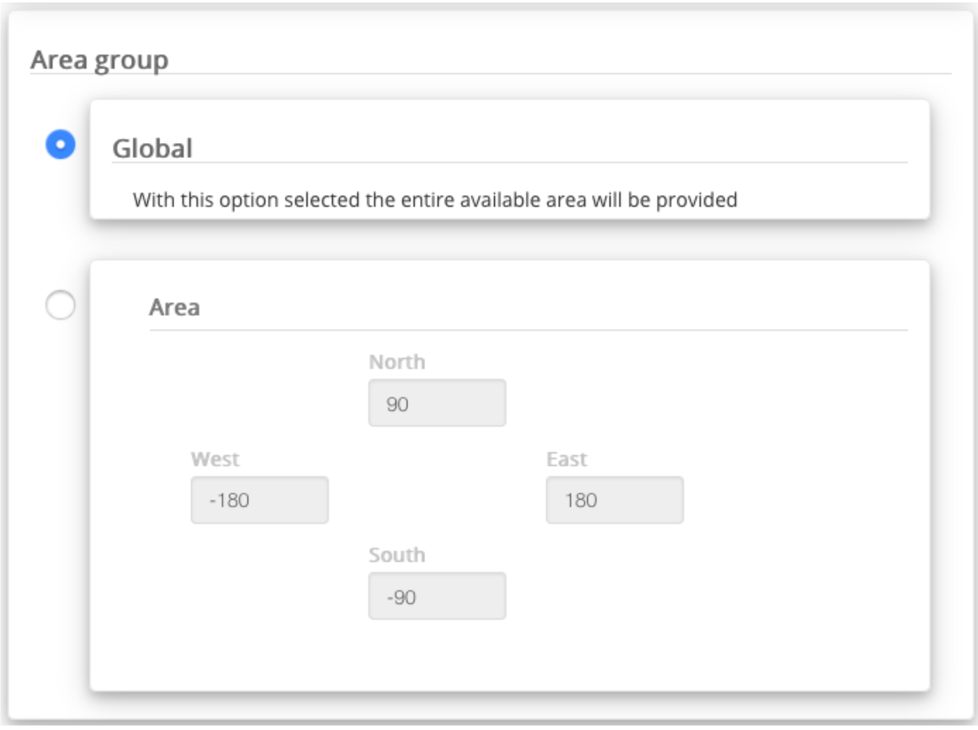

Figure 8: Selection box for selecting area on the Global Land And Marine Observations Database download page on the CDS.

```
import cdsapi
c = cdsapi.Client()
c.retrieve(
    'in-situ-observatins-surface-marine',
    {
        'format': 'csv-obs.zip',
        'day': '07',
        'month': '01',
        'year': '1898',
        'data_quality': 'quality_controlled',
        'variable': 'water_temperature',
    },
    'download.csv-obs.zip')
```

Figure 9: Example API request to download data from the Global Land And Marine Observations Database.

1.5 Data model

Data from the Climate Data Store are returned in delimited text files (.csv), with 19 columns per row and one observed value per row. The columns are listed in table 2 together with a brief description. The date and time of the observation is given by the `date_time` column, with the value returned as a string of the format `YYYY-MM-DD hh:mm:ss`. All data are returned in the UTC time zone. The location of the observation is given by the `latitude` and `longitude` columns in the WGS84 / EPSG4326 coordinate reference system. The variable being reported / observed is given by the `observed_variable` column, the units by the `units` column and the observed value by the `observation_value` column. The remaining columns provide further contextual information for the observations, such as platform type and station identifiers etc.

Table 2: Data model.

Field	**Kind**	**Description**
observation_id	character	Unique identifier for observation. For marine observations this has been set as <dataset>-<version>-<uid>-<field>, e.g. ICOADS-30-0NBVN1-AT for an air temperature observation with UID 0NBVN1 from ICOADS release 3.
report_type	character	Subdaily / hourly data.
date_time	timestamp with time zone (YYYY-MM-DD hh:mm:ss+0Z)	Timestamp (date/time) of observation.
date_time_meaning	character	date_time represents the end of the time period over which the observation was made.

Table 2 – continued from previous page

Field	Kind	Description
latitude	numeric	Latitude of the observation (degrees North), range -90 to 90.
longitude	numeric	Longitude of the observation (degrees East), range -180 to 180.
observation_height_above_station_surface	numeric	Height of the sensor above local ground or sea level in metres. Positive values for observations above the surface. For visual observations, height of the observing platform.
observed_variable	character	The variable being observed / measured.
units	character	Units of the observed variable.
observation_value	numeric	The value of the observed variable.
value_significance	character	The significance of the observed variable, e.g. instantaneous value or mean over indicated period.
observation_duration	character	The time period over which the observation was made.
platform_type	character	The type of structure upon which the sensors are mounted, e.g. ship, drifting buoy, tower etc.
station_type	character	Type of station, e.g. sea station, land station
primary_station_id	character	The primary station identifier, e.g. ship call-sign, buoy number, WIGOS station identifier.
station_name	character	Name of the station or site.
quality_flag	integer	Flag indicating the quality of the observation.
data_policy_licence	character	Usage permissions / licence for the data.
source_id	character	Identifier used to link the observation to the original source of the data. For marine data this links to the version of ICOADS, source and deck identifier and which ICOADS monthly file the record is from. Combined with the observation ID this can be used to trace an observation back to the original source record.

Detailed description

2 Data and metadata sources

2.1 Observations

The International Comprehensive Ocean-Atmosphere Data Set

Weather observations have routinely been made at sea since at least the 1650s, with descriptive observations of the prevailing weather conditions and early instrumental measurements included in the general ships logbooks as part of the daily entries. Following the Brussels Conference of 1853, and recognition of the importance of the weather observations to international trade and safe navigation, standardised meteorological logbooks and observing instructions were developed and made available to ships captains. In return for following the instructions and returning the completed logbooks the captains were provided with the latest sailing directions and weather charts, thereby collectively gaining benefit from the collecting and sharing of meteorological data. This international coordination, standardisation and data sharing has continued to the present day, currently overseen by the World Meteorological Organisation (WMO) and Global Ocean Observing System (GOOS) (e.g. see Smith et al. [2019]). Observations from other platforms, such as moored and drifting buoys, are similarly coordinated internationally.

Through the international coordination and data sharing, national weather services independently developed global archives of marine meteorological and oceanographic observations. Many of these archives contained overlapping data. For example, the US Navy Marine Atlases developed after the Second World War contained data from both the US archives and data from, inter alia, the UK, German and Netherlands archives that had been shared internationally. This resulted in increased numbers of observations but at the expense of having to perform duplicate detection and elimination. In recognition of the importance of historic data to understanding climate variability, building on the earlier Marine Atlases, the US National Oceanic and Atmospheric Administration (NOAA) developed the Comprehensive Ocean - Atmosphere Data Set (COADS) in the early 1980s. The first version was published in 1985 (e.g. see Woodruff et al. [1987]), consisting of both monthly summary statistics for selected ECVs and the raw weather reports from ships and other vessels used as input to the monthly summaries. Each weather report contained coincident measurements of multiple ECVs (typically air temperature and humidity, wind speed and direction, sea level pressure and sea surface temperature) together with visual estimates of the sea state, wind force (when not measured), cloud cover and weather. This first version formed the largest and most comprehensive archive of marine meteorological observations publicly available at the time.

Building on the first version, COADS continued to be developed with coverage extended to present day through major updates and reprocessing and through incremental near real time updates. As part of this development, and the evolving observing system, COADS has expanded from primarily ship based observations to surface meteorological measurements from most marine platforms shared internationally. Examples include measurements from moored and drifting buoys, coastal stations and offshore platforms, all shared internationally in real time over the WMO Global Telecommunication System (GTS) and in delayed mode through Global Data Assembly Centres. The development has also included blending COADS with other national archives and delayed mode data sources, such as the UK Met Office marine data bank. In recognition of the importance of the international contributions

COADS was renamed as the International Comprehensive Ocean - Atmosphere Data Set (ICOADS) in 2002 (e.g. Worley et al. [2005]). The current major release of ICOADS, release 3.0, was published in 2017 (Freeman et al. [2017]), releasing observations up to the end of 2014. Near realtime updates for the period after 2014 are available from ICOADS Release 3.0.2 (Liu et al. [2022]). The ICOADS near realtime updates contain regular releases of marine observations, updated within the first 10 days of the most current month. The near realtime observations are based on blended marine observations in the TAC and BUFR formats from NOAA's National Centers for Environmental Information (NCEI) GTS collections.

As noted above, due to the combination of different sources over many decades, including from different national archives, and the long tradition of data sharing in marine meteorology there are many duplicates in the raw data. The publicly released version of ICOADS has attempted to remove or combine duplicates when detected, but this has not always been optimal due to choices made over the history of ICOADS. For example, some duplicates are missed due to one archive recording the latitude / longitude in whole degrees and another archive storing the same location information but recording the grid box centres (i.e. with a 0.5° trailing digit). These decisions have been revisited as part of this service, with the available ICOADS source files (known as "total" files within the NCEI/ICOADS archives) reprocessed and the data flagged appropriately. This processing is described in Section 3. The original source of the data is recorded in ICOADS with the data accessions indexed by "deck" (DCK) and "source" (SID) identifiers (Woodruff et al. [1987], Freeman et al. [2017]). A detailed summary of the data accessions is given in the marine inventory (C3S2_D311_Lot1.2.3.1.202308_Updated Inventory for Marine datasets_v3) available from the CDS website.

2.2 Selection of data for inclusion in Service

As part of the development of this service the input sources and decks contributing observations to ICOADS have been prioritised and reprocessed based on the volume of observations, spatial / temporal coverage and availability of ECVs.

Release 1 focussed on the post World War II period (1951 - 2010) and included observations from drifting buoys (sea surface temperature and sea level pressure) and Voluntary Observing Ships (VOS) (air temperature, dew point temperature, wind speed and direction, sea level pressure and sea surface temperature). As part of the first release VOS observations were merged with instrumental metadata from WMO Publication Number 47 (e.g. Kent et al. [2007]) to provide information on observing heights, instrument types / observing methods and information on the vessels contributing the observations.

Release 2 sought to extend the record backwards to 1851, including all available ship observations prior to 1951.

Release 3 updated the record to the end of 2014.

Release 4 updated the record to the end of 2020 using the observations from ICOADS Release 3.0.1, including a full reprocessing of the data from the previous releases. As part of this reprocessing additional metadata has been recovered for some of the earlier records and the observations for the period.

Release 5 updated the record to the end of 2021 using the observations from ICOADS Release 3.0.2, adding near realtime data updates including decoded BUFR messages from 2015 onwards.

Release 6, the most recent at the time, updated the record to the end of 2022 using the observations from ICOADS Release 3.0.2, migrated the data processing to the Irish Center for High-end Computing and fixed a bug where the additional metadata was not added in the past. The metadata has been added to the previous releases as a post-processing step.

Drifting buoy data from the ICOADS Release 3.0.2 are not included in this release. A consolidated drifting buoy data record based on additional data sources such as the C-RAID Drifting buoys (Zunino Rodriguez [2023]) will be processed and included in future releases.

The marine inventory (C3S2_D311_Lot1.2.3.1.202308_Updated Inventory for Marine datasets_v3, available from the CDS website) gives a detailed overview of the contents from each data source used in this release.

Figure 10 shows the monthly number of reports available in the current release. Only those passing the full quality control check are shown, in total 390,335,546 reports from 248 sources have been processed. Of these 282,367,426 are unique and pass the report quality control checks (see Section 3), the remainder either fail the duplicate check or the aggregate report level QC check (see Section 3.6). Figure 11 shows the spatial sampling by total number of reports and by number of months. The total number of reports is dominated by the recent decades and the increase in buoy sampling (see Figure 12). The contribution to the ship based reports can be seen in the number of months available, with the ship tracks clearly visible.

Several notable sources and platform types have been excluded. Namely, fixed stations (mobile installations and rigs, coastal stations and moored buoys), meteorological observations from research vessels, and meteorological observations from oceanographic programmes and datasets such as the World Ocean Database (WOD) and the Global Ocean Surface Underway Data (GOSUD). Users requiring access to these data are recommended to access data directly from the respective datasets (e.g. WOD, GOSUD, SAMOS).

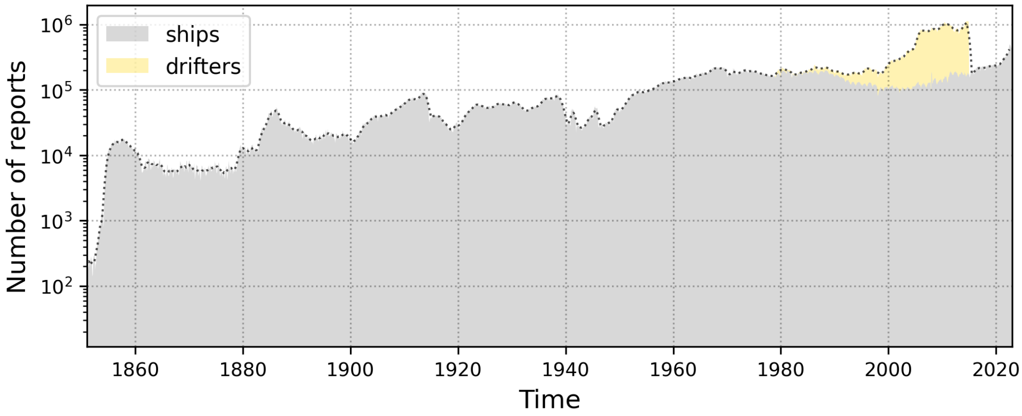

Figure 10: Monthly number of marine in situ reports. The stacked areas indicate the monthly number of ship reports (grey) and drifting buoy reports (yellow). Only reports that have passed all the quality control checks are included. Time series have been smoothed using a 12-month running mean filter. Note the logarithmic scale in the y axis.

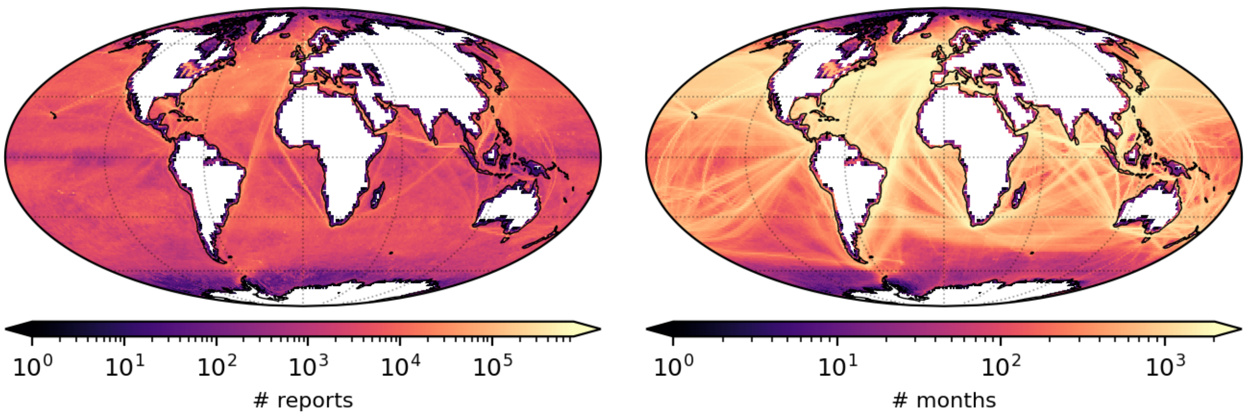

Figure 11: Spatio-temporal distribution of marine data holdings 1851 - 2022. Left: number of reports per grid cell; right: number of months with at least 1 report. Grid size is 1x1. Note logarithmic colour scale. Only reports passing the report level QC checks flag have been included.

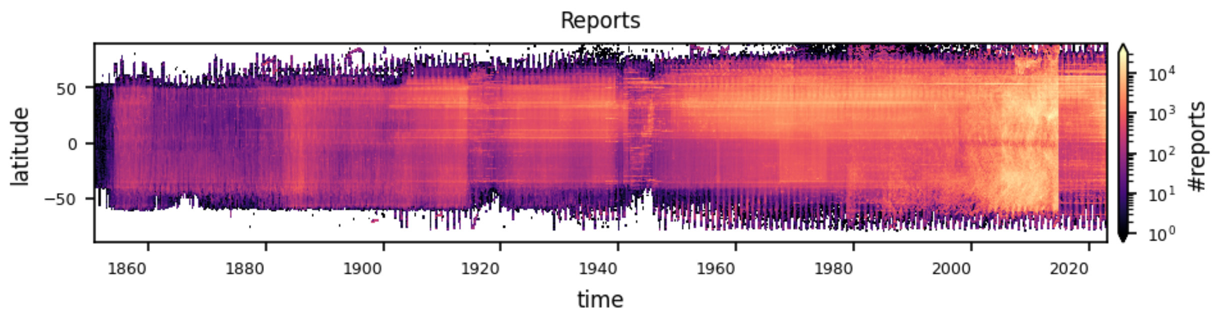

Figure 12: Spatio-temporal distribution of marine reports: latitude-time (Hovmøller) plots of number of reports. Data is binned and aggregated in 1° latitude × 1 month boxes. Only reports passing the report level QC checks have been included.

2.3 Metadata

WMO Publication Number 47 and other metadata

WMO Publication No. 47 (https://www.ocean-ops.org/share/SOT/PUB47/), the International List of Selected, Supplementary and Auxiliary Ships, contains metadata on the ships, and observing methods used by those ships contributing to the WMO Voluntary Observing Ships Programme. A summary of the information available in WMO Publication No. 47 can be found in Kent et al. [2007]. This includes information on the ship dimensions and types, type(s) of instruments used, location of the sensors on board the ships and height of the sensors above the sea surface. A subset of the metadata from WMO Publication No. 47 has been included in ICOADS since Release 2.5 based on a basic match between the reports in ICOADS and the WMO Publication 47 edition coincident with the observation (Kent et al. [2007], Woodruff et al. [2011]).

As part of C3S2 311 Lot 1 the WMO Publication 47 metadata has been reprocessed and selected fields, including ship names, types and instrument heights, have been included in this release. During the reprocessing duplicate records have been combined, reducing the number of records from around 875 000 to 255 000 over the period 1956 - 2020. Validity dates have been added, with the first record for a given ship extended by 1 year prior to its first appearance and the final record extended by 5 years. This allows for a period of time between a ship being recruited and it first appearing in WMO Publication No. 47 and a period where a ship continues reporting but is no longer registered by the original recruiting country. Limited quality control has been applied to the data, with correction of typographical errors and standardisation of entries. Similarly, heights have been corrected for known errors, such as incorrect packing of numeric data into alpha-numeric representation (e.g. i5 keyed instead of 15). In addition to WMO Publication No. 47, metadata is present in the ICOADS supplemental records for a small number of observations, including information such as ship type, barometer height, routes etc. Some information has been extracted from the ICOADS supplemental records and included in this service.

Figure 13 shows the fraction of ship observations passing quality control for which the sea level pressure observing height is known (red line). The blue shaded area indicates the 25th - 75th percentiles of the observing heights and the blue dashed line the median height. The earliest heights available are from the late 1870s for a period of $\sim$15 years, with a significant proportion of the observations made during this period at a known height of around 5 m. There is then a large gap in the metadata until the 1960s where WMO Publication No. 47 becomes available and call signs begin to be used in the observational record allowing the metadata to be associated with the observations. However, the values before the 1970s should be used with caution, only a small fraction of the observations have been associated with the metadata, there is a large inter-quartile range and the median value is higher than expected. This is likely due to a number of uncorrected heights in feet above sea level rather than meters above sea level but further work is required to fully understand the metadata. Between 1970 and 2020 the median observation height increases from $\sim$15 m to $\sim$25 m.

The decrease in the proportion of observations with a known height from $\sim$2014 is due to a decrease in the availability of call signs in the observational data due to call sign masking and use of generic call signs (such as SHIP) in response to security and commercial concerns of the participating Voluntary Observing Ships. Metadata for the air temperature, humidity and wind speed observations does not become available until the 1960s although the observing heights are expected to be similar to that for pressure. From the 1970s onwards the change in heights (not shown) shows a similar trend to sea level pressure, with the temperature and humidity heights increasing from ∼15 m to ∼27 m. The increase in wind speed height is greater (not shown), increasing from ∼15 m in the early 1970s to almost 35 m by 2020.

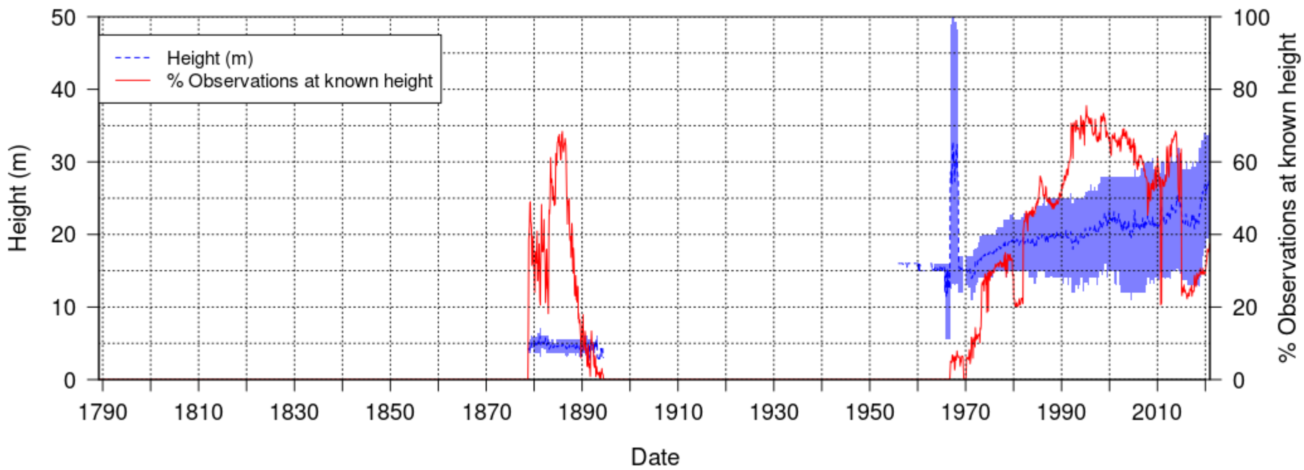

Figure 13: Median sea level pressure observing height (blue) and % of observations where the height is known (red). Also shown is the inter-quartile range of the observing height (blue shading).

2.4 Observing methods, coverage and known issues

Air temperature and humidity

Figure 14 shows climatological values of the mean air temperature and dew point temperature over the period 1851 - 2022. Figure 15 shows the number of months with at least 1 observation on a 1x1 grid. The clustering over the major shipping routes is clearly visible. The air temperature and humidity measurements made on board ships have traditionally been made using liquid (mercury or alcohol) in glass thermometers (wet and dry bulb) sheltered from direct solar radiation, rain and sea spray and housed either in a fixed shelter or in a hand-held instrument. However, due to the need for the thermometers to be accessible the location of the thermometers is often not ideal, with the shelters sometimes located in poorly exposed locations and with inadequate ventilation. This can lead to biases in humidity measurements due to inadequate ventilation of the wet bulb thermometers (e.g. Berry and Kent [2011]; Willett et al. [2008]). Similarly, daytime air temperature measurements can contain biases due to the warming of the ship superstructure by solar radiation, in turn warming the air and giving biased air temperature measurements (Rayner et al. [2003]). Bias adjustments have been developed to account for these effects (e.g. Berry et al. [2004]) but have not been applied as part of this service. With the exception of the Second World War (e.g. see Cornes et al. [2020]; Kent et al.

[2013]; Rayner et al. [2003]) pervasive biases have not been identified in the night time air temperature measurements or artificially ventilated wet bulb thermometers measurements during the period 1900 – 2020. However, biases have been identified in the 19th Century and earlier data due to the use of thermometers located in the captains cabin (e.g. Chenoweth [2000]).

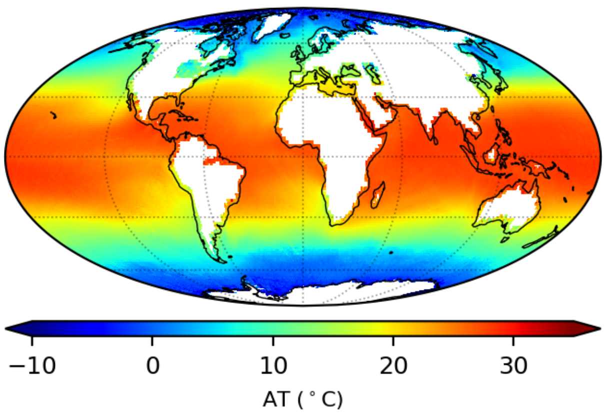

Figure 14: Mean air temperature (left) and dew point temperature (right) over the period 1851 - 2022. All observations passing quality control have been averaged to give monthly mean values. These have then been averaged to give the long term mean.

Another factor that could lead to biases are systematic changes in the height at which the observations are made. The average observation height on merchant ships has has changed substantially over the period, ranging from ∼4 - 5 m in the late 1800s, increasing to ∼30 m in the late 2000s (Figure 13, Kent et al. [2007], Kent et al. [2013]). Adjustments to a fixed reference height are required to avoid inhomogeneities in the record and leading to underestimation of trends in the data. Adjustments are typically made based on Monin-Obukhov similarity theory and the approximation of a constant flux layer near the surface (e.g. Businger et al. [1971]). Observing heights are included with the observations where known but adjustments have not been made as part of this service.

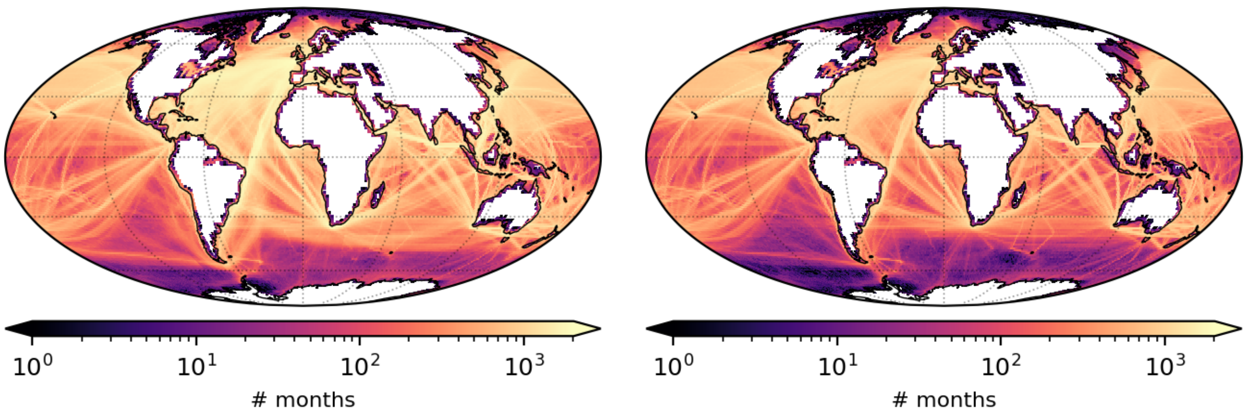

Figure 15: Number of months with at least one observation of air temperature (left) or humidity (right) over the period 1851 - 2022 on a 1x1 grid. Only observations passing quality control have been included.

Sea surface temperature

Figure 16 shows climatological values of the mean sea surface temperature over the period 1851 - 2022 (left) and the number of months with at least 1 observation on a 1x1 grid (right). The clustering over the major shipping routes is clearly visible but with better sampling than for air temperature and humidity away from the shipping lanes due to the sampling by drifting buoys. Sea surface temperature observations have been historically made using a wide variety of methods, ranging from wooden, canvas and rubber buckets to infrared radiometers. The most common methods are the bucket based and engine intake methods, both of which suffer from biases (Kent et al. [2017]). Due to evaporative and conductive cooling the water samples in the buckets can be biased low compared to the true sea surface temperature. Conversely, due to heat from the engine room, engine intake measurements can be biased warm. Bias adjustments have been developed and implemented when constructing gridded datasets from the observations (Kennedy et al. [2011]). The quality of SST observations has been widely discussed (e.g. Kennedy [2014], Kent et al. [2017]). As with the other parameters, no adjustments have been made to the observations to account for known biases as part of this service.

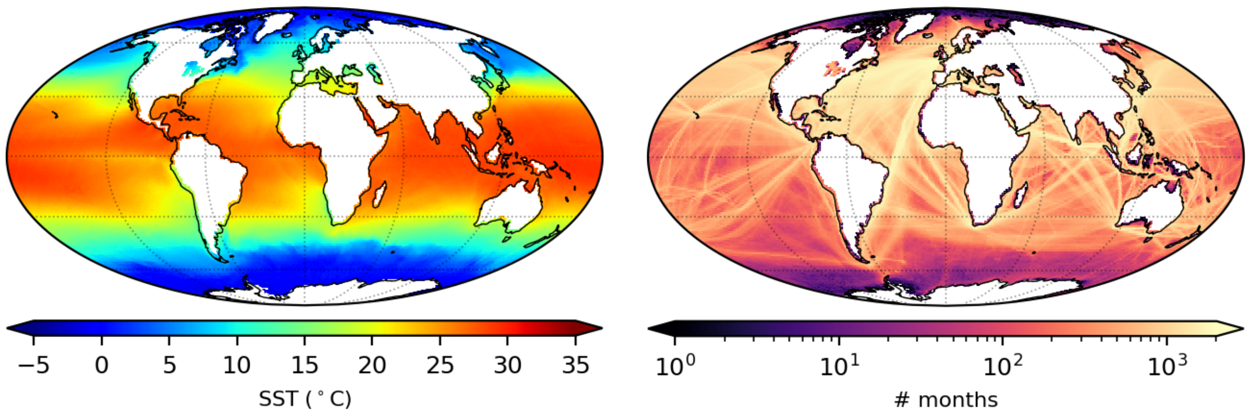

Figure 16: Mean sea surface temperature (left) and number of months with at least one observation (right) over the period 1851 - 2022. All observations passing quality control have been averaged to give monthly mean values. These have then been averaged to give the long term mean.

Wind speed and direction

Figure 17 shows climatological values of the mean wind speed and direction over the period 1851 - 2022. Figure 18 shows the number of months with at least 1 observation on a 1x1 grid. As with the other parameters, the clustering over the major shipping routes is clearly visible. Wind speed observations made on board ships were historically made by estimating the wind force and recording the estimate using the Beaufort scale. This provides estimates of the upper, lower and mid-point wind speeds for each value on the scale at a reference height of 10 m. When Beaufort scale wind estimates have been converted to a speed the mid-point has typically been used. More recently, the visually estimated wind speeds have been estimated as a speed (e.g. knots or m/s) using the Beaufort scale as a guide.

Through comparison with instrumental measurements and co-located observations the original Beaufort scale has been shown to be biased and corrections proposed (Kent and Taylor [1997]). Over the past several decades there has been an increasing move to using anemometers to observe and report the wind speed over the oceans. These measurements can contain biases due to the impact of flow distortion on the wind speed (Moat et al. [2005]). As with the air temperature and humidity, the typical height of wind speed measurement has changed with time (e.g. Thomas et al. [2008]) and the measured values require adjustment to a common reference height. Again, no adjustments have been made as part of this service.

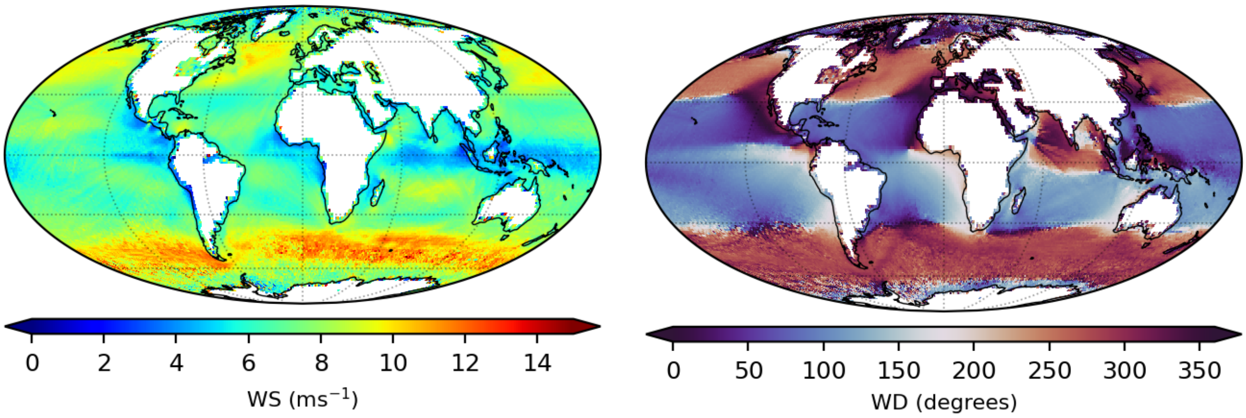

Figure 17: Mean wind speed (left) and direction (right) over the period 1851 - 2022. All observations passing quality control have been averaged to give monthly mean values. These have then been averaged to give the long term mean.

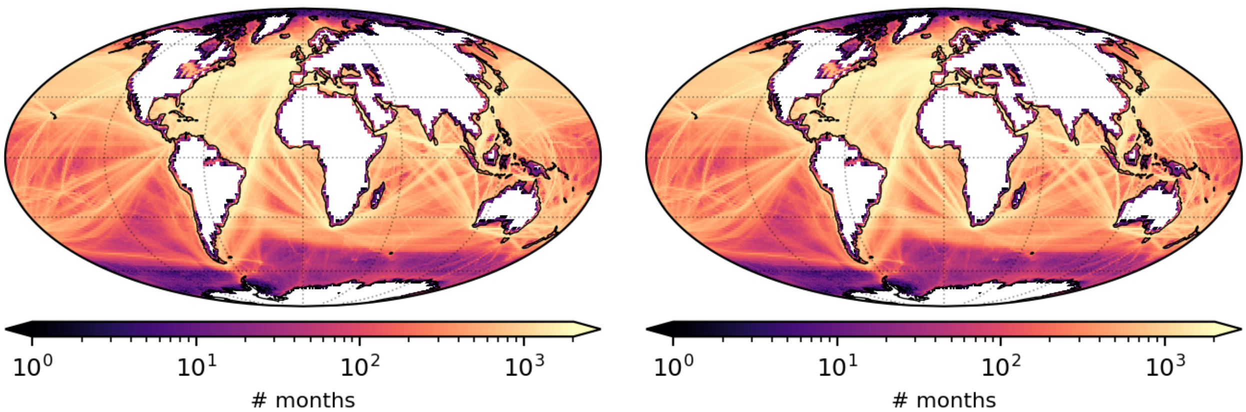

Figure 18: Number of months with at least one observation of wind speed (left) or wind direction (right) over the period 1851 - 2022 on a 1x1 grid. Only observations passing quality control have been included.

Sea level pressure

Figure 19 shows climatological values of the mean sea level pressure over the period 1851 - 2022 (left) and the number of months with at least 1 observation on a 1x1 grid (right). The clustering over the major shipping routes is clearly visible and, as with the sea surface temperature, better sampling away from the shipping lanes due to the sampling by drifting buoys can be seen.

Early pressure observations were typically made using mercury barometers, with a transition to marine aneroid barometers in the mid 20th century and to electronic instruments more recently. The different sensors each have corrections that need to be applied but these are typically applied either at the time of reporting or automatically prior to the values being read in the case of the electronic sensors. The mercury and aneroid barometers require correction for temperature and height above sea level, with an additional adjustment for gravity required for the mercury barometers. These corrections are believed to have been applied in the ICOADS data. When looking at long term means / climatological values the observations also need to be corrected for diurnal and semi-diurnal oscillations (e.g. Ansell et al. 2006). These additional adjustments have not been applied as part of this service.

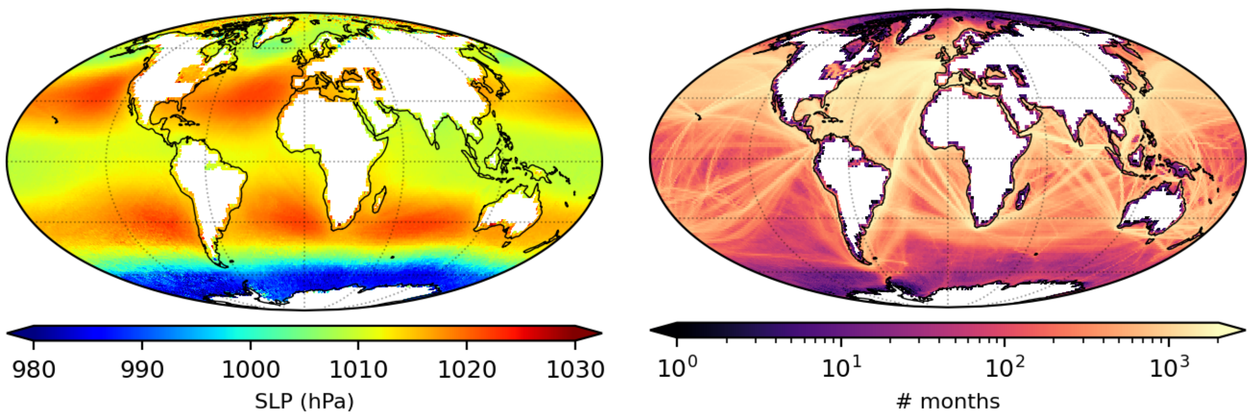

Figure 19: Mean sea level pressure over the period 1851 - 2022. All observations passing quality control have been averaged to give monthly mean values. These have then been averaged to give the long term mean.

3 Quality control and processing

3.1 Verification of input data and conversion to a common data model

The source data for ICOADS is stored in the fixed width International Maritime Meteorological Archive (IMMA) format (see `https://icoads.noaa.gov/doc.html`). This format consists of one weather report per record with a core data section and optional attachments containing additional information. Table 3 summarises the different attachments, those in italic are typically reported with each report regardless of source. The presence of the other attachments depends on both the type of platform/station making the weather report and the source of the data. For example, the IMMT attachment will only be present for ship observations reporting in delayed mode.

Prior to converting to the common data model, used by the C3S2 service for both land and marine data, the individual ICOADS records are checked against the specification of the IMMA format. Each element in IMMA has a defined data type and valid range. Similarly, valid enumerated (or coded) values are specified via a code table. Any record with any element falling outside the specified valid range, with the exception of missing data, or with any invalid coded value is considered invalid. Similarly, reports with an invalid date, time or location are excluded from further processing. No QC flags are set. After validation, the IMMA records are converted to the common data model (see Section 5) and the quality control described below applied.

Table 3: Sections of an IMMA record (see `https://icoads.noaa.gov/doc.html` for a full description)

Attachment name	Abbreviation	Description
Core	Core	ICOADS core record containing commonly observed parameters, date and time
ICOADS attachment	ICOADS	Attachment containing ship/ platform identification, additional observed parameters and ICOADS QC flags
IMMT-5 / FM 13 attachment	IMMT	Attachment containing additional information reported by Voluntary Observing Ships in either the International Marine Meteorological Tape (IMMT) or WMO FM-13 formats.
Model QC attachment	Mod-QC	Attachment containing NWP model data collocated to the location the observations.
Ship metadata attachment	Meta-VOS	Attachment containing selected metadata from WMO-No. 47 merged with ship observations in ICOADS.
Near surface oceanographic data attachment	NOcn	Attachment containing surface ocean biogeochemical measurements
Edited cloud report	ECR	Attachment containing corrected / edited cloud observations following Hahn and Warren (1999). This is only available from 1950 onwards.
Reanalysis QC / feedback attachment	Rean-QC	Attachment providing the facility to include observation feedback information from reanalysis models
IVAD attachment	IVAD	Attachment to store data from the ICOADS Value Added Database Project
Error attachment	Error	Attachment designed to support the correction of erroneous IMMA elements.
Unique report ID attachment	UIDA	Attachment with unique ID assigned to each report
Supplemental data attachment	Suppl.	Attachment to store additional data not representable in other attachments and to store the weather reports in the orignal format / units where available. The format of the Suppl. attachment is source and deck dependent.

3.2 Station/platform identification (ship only)

The station/platform identification within ICOADS is stored in the ID field and over the lifetime of the marine meteorological observations and archives a wide variety of formats have been used to identify individual ships. For example, in the early punchcard decks from the 1960s and earlier, ships were assigned a numeric identifier allocated in alphabetical order each year with the ship names written on the back of the punchcard. Lists matching the ship numbers to ship names were created and archived. However, over time and with the conversion from punchcards to magnetic tape and then to digital formats the link between ships names and numbers has been lost, with only the ship number remaining. For some decks the ship number has subsequently been dropped where it was thought to have no value without the link to the ship name. In other cases the logbook number was used to identify the digitised data, sometimes with a page number appended. More recently, ship names or abbreviated ship names have been recorded directly in the digitised data when logbooks have been keyed. For realtime data sources we typically have the ITU callsign or a masked/generic identifier. Overall, there are many sources of error and ambiguity in the station/platform identification field. These issues are discussed further in Carella et al. [2017] and we follow the method of Carella et al. [2017] to link observations from the same ship together. As part of this process the ID field has been corrected for known errors and new ids assigned where the ID field is missing, a generic ID assigned or set to a logbook and page number.

3.3 Duplicate detection and flagging (ship only)

Duplicates exist in ICOADS due to the way that the dataset has been created over many decades with data from many different sources, often with overlapping observations, merged. For example, many national climate archives such as those in the US, UK, Germany and the Netherlands were based on the same source data with the ship logbooks digitised once and then shared between nations. This process has been repeated several times resulting in many duplicates in the underpinning data. For recent decades multiple real time GTS sources have been ingested as this has been shown to increase the number of unique reports but at the expense of increased duplicates. Similarly, when the higher quality delayed mode versions of the real time data are ingested the real time versions need to be identified, flagged and replaced. As part of the processing to create ICOADS the duplicate detection and elimination software (DUPELIM) is run, with only those reports believed to be unique released into the final versions (e.g. see Freeman et al. [2017], Woodruff et al. [2011])[3]. The processing for C3S2 uses the “total” ICOADS files, which include observations identified as duplicates that have been excluded from the “final” data.

The ICOADS DUPELIM software relies heavily on the date, time, location and station / platform identification. All of these fields may contain errors due to a number of reasons. For example, an observation may be mis-assigned to the wrong day or an observation put in the wrong location due to the miscoding of the quadrant of the globe within the data records when the data were originally keyed. In other cases, the data have been modified or stored differently in the different archives. For example, the location of an observation was often recorded as the one degree latitude/longitude grid box

[3] additional information is also available from https://icoads.noaa.gov/dupelim.html

the observation was located in. On conversion to a latitude/longitude pair the location could be given either as the edge of the grid box or the grid box centre, resulting in different locations for the same observation. Conversion between units and the enumeration/translation between code tables can also introduce minor differences. As such, due to those differences not all duplicates have been identified by the DUPELIM software and unidentified duplicates exist in the ICOADS final files.

Within the C3S2 service an additional level of duplicate detection has been performed making allowances for these known issues. In many cases the results are similar to that of DUPELIM but with an increase of up to 5 - 10 % in the number of duplicates identified in some periods. The method and results are fully described in Kent et al. [2019], Table 4 lists the categorisations used.

Table 4: Duplicate flags and meanings used within the common data model (see Section)

Flag value	Meaning	Comments
0	Unique observation, no known duplicates	
1	Best duplicate	
2	Duplicate	This has been used to identify reports thought to be duplicates but where no selection was made, e.g. because there were unique variables present in each report.
3	Worst duplicate	
4	Unchecked	Reports have not been through the duplicate identification process.

3.4 Drifting buoy identification and duplicate flagging

Due to the different data management practices and data streams for the drifting buoy data compared to ships many of the issues present in the ship data are avoided. The prevalence of mis-coding errors in the date, time, location and buoy ID fields is thought to be negligible. As such, within the C3S2 service we make use of the ICOADS intermediate reject flag (IRF) to identify the best duplicates to keep. Drifting buoy reports with IRF set to `retain` have their report quality flag within the common data model set to 0 (pass). All other buoy reports have their report quality flag set to 1 (failed). In all cases, the duplicate flag is set to 4 (unchecked).

3.5 Advanced quality control

Once the observations have been through the initial format checks, station / platform identification and duplicate flagging three additional stages of quality control are performed. These are based on a scheme developed by the UK Met Office and described in Kennedy et al. [2019] and repeated here for completeness. Prior to proceeding to the advanced quality control some pre-screening is performed. For a ship based report to proceed to the advanced quality control stage it needs to have been flagged as either the best duplicate or a unique report (section 3.3). For a buoy based report to proceed it needs to have been flagged for retention by the ICOADS IRF flag and initially marked as passing the report quality check (section 3.4). All other observations have their report and observation quality flags set to unchecked. No additional quality control data will be generated for these reports.

Report level quality control

The report level quality control checks are similar to those performed as part of the verification of input data (section 3.1). Three different flags are set based on the validity of the date, time and location of a report. 1 indicates a fail and 0 a pass. In addition to these checks the station/platform identifier is checked against a list of those stations known to produce low quality data (see Table 5). Observations with a station/platform identifier matching one on the list or from a source or location known to be of low quality have their black list flag set to 1. All others are set to zero. The flags are:

- `Time check`: Check whether time of observation is valid ($0 \leq$ hour ≤ 24). 1 returned for invalid or missing times, 0 otherwise.
- `Date check`: Check whether date is valid and in the range 1850 to 2024. 1 returned for invalid dates or dates outside the specified range, 0 otherwise.
- `Location check`: Check whether location (latitude and longitude) is valid. 1 returned for latitude outside the range 90°S to 90°N or for longitude outside the range 180°W to 360°E. 1 also returned for missing data, 0 returned in all other cases.
- `Landlocked check`: Check whether the record is located on land according to the ICOADS 'LZ' landlocked flag.
- `Blacklist`: Flag set to 1 for any observation matching a condition listed in table 5, 0 otherwise.

Table 5: Conditions for blacklisting

Condition	**Reason / details**
Report at 0°N, 0°E	Often when the location is missing the latitude and longitude are set to 0°.
ICOADS platform type 13	Data from C-MAN coastal station. Data not representative of the open ocean.
Data from the ICOADS SEAS deck (deck 874)	There was an unrecoverable error in the encoding of data from this source into IMMA.
Mis-calibrated buoys	Data from buoys 52521, 53522, 53566-68, 53571, 53578, 53580, 53582, 53591-96, 53599, 53600-09, 53901, 53902 are known to have calibration errors during the period November 2005 to January 2006.
Mis-positioned MORMET data (deck 732)	Observations from the Russian Marine Meteorological Dataset (MORMET) are known to be mis-positioned in certain regions and time periods. These are listed in Table 6.

Table 6: Exclusion regions and periods the ICOADS MORMET deck (deck 732)

Region	Bounds				Years (inclusive)
	W	S	E	N	
1	-175	40	-170	55	1958 - 1971
2	-165	40	-160	60	1958 - 1971
3	-145	40	-140	50	1958 - 1964, 1968 - 1971
4	-140	30	-135	40	1958 - 1959, 1969 - 1974
5	-140	50	-130	55	1958 - 1964, 1967 - 1971
6	-70	35	-60	40	1958 - 1961, 1963 - 1971
7	-50	45	-40	50	1969, 1971 - 1974
8	5	70	10	80	1969 - 1974
9	0	-10	10	0	1960, 1966 - 1972
10	-30	-25	-25	-20	1965, 1969, 1972 - 1974
11	-60	-50	-55	-45	1972 - 1974
12	75	-20	80	-15	1962 - 1965
13	50	-30	60	-20	1962 - 1965, 1969, 1971 - 1973
14	30	-40	40	-30	1958 - 1971
15	20	60	25	65	1958 - 1963, 1965 - 1970
16	0	-40	10	-30	1962 - 1965, 1972 - 1974
17	-135	30	-130	40	1972 - 1974

Observation level quality control

- `No value check`: checks for the presence of missing data. Returns 1 for missing data, 0 otherwise.
- `No normal check`: checks for the availability of a climatological value at the date and position of the observation. Returns 1 if no climatology is available, 0 otherwise.
- `Climatology check`: checks the magnitude of anomalies against a predefined range. Returns 1 for values outside the range, 0 otherwise,
- `Freezing value check`: Sea surface temperature only, checks whether the temperature is below the approximate freezing point of water (-1.8°C). Set to 1 for temperatures below -1.8°C, 0 otherwise.
- `Super-saturation check`: Dew point temperature only, checks whether the dew point temperature is greater than the air temperature. Returns 1 if dew point temperature greater than air temperature, 0 otherwise.

Along track / voyage level quality control

Only stations/platforms with a non-generic identifier are subject to the along track quality control. The checks are performed on 1 month of data at a time but include the data for the preceding and following months for a number of the checks. All checks are applied to time ordered data, full details are provided in Kennedy et al. [2019].

- `Speed check`: Checks the speed required to get to the observation location forwards and backwards in time. Returns 1 if check fails, 0 otherwise.
- `Distance from estimated location`: Calculates the expected position of the current observation by extrapolating forward and backward in time from the last and next observation and reported speed and course. If the distance between both the forward and backward extrapolated location and reported location is greater than threshold return 1, 0 otherwise.
- `Direction consistency`: Calculates the course between location of current and prior observation and compares to reported course. If difference between reported course and calculated course > 60° returns 1, 0 otherwise.
- `Speed consistency`: Calculates the speed required to travel between the location of the current observation and prior observation and compares to reported speed. Returns 1 if difference in speed > 10 knots, 0 otherwise.
- `Extreme speed check`: If the calculated speed is greater than 40 knots return 1, 0 otherwise.
- `Mid point check`: Compares the reported position to the mid point of the next and previous observations taking into account time difference. If distance between reported location and mid point greater than 150 nm return 1, 0 otherwise.
- `Combined track check`: Returns 1 if an observation fails the mid point check, speed check and at least one other check. Returns 0 otherwise.
- `Repeated value check`: Checks whether a significant proportion of observations along a track have a repeated value. Returns 1 if more than 70% of observations in a three month period have the same value.
- `Rounded value check`: The round flag is set to 1 for all whole-number observations of DPT, when the fraction of whole-number observations exceeds 70% of all observations. All other observations have their round flag set to 0.
- `Repeated super-saturation check`: Dew point temperature only, checks whether the wet bulb reservoir has dried out resulting in repeated super-saturated values. If in a series of observations spanning more than 72 hours, the DPT and AT were equal for four consecutive observations or more, their repsat flags were set to 1.
- `Fewsome check`: Returns 1 if 3 or fewer observations from an ID, 0 otherwise.
- `IQUAM track check`: Track check as described by Xu and Ignatov [2014] but only using the 5 observations immediately preceding and following the observation being checked.
- `IQUAM spike check`: Spike check as described by Xu and Ignatov [2014] but only using the 5 observations immediately preceding and following the observation being checked.

Buddy check

The final check applied to the ship observations is the buddy check. All observations of a given parameter passing the previously described checks are converted to anomalies by subtraction of the climatological value and averaged onto a 1 by 1 degree 5 day grid. The grid box standard deviation (σ)is also calculated at this stage. For each grid box (target grid box) the neighbouring area is searched for grid boxes with data according to the search parameters listed in Table 7. If no grid boxes are found the search area is increased in size, doubling the spatial search radius. If still no grid boxes are found the temporal search radius is then doubled. Once neighbouring grid boxes have been identified the arithmetic mean anomaly of all the neighbouring grid boxes is calculated and an acceptable anomaly range determined based on the number of neighbouring grid boxes(μ). The anomaly for each observation in the target grid box is compared to the acceptable range, any observation outside of the range has its buddy flag set to 1 (fail) and 0 otherwise.

- `Buddy check`: Anomaly for observation compared to mean anomaly for neighbouring grid boxes. 1 returned if anomaly outside of specified range (Table 7), 0 otherwise.

Table 7: Limits for buddy checks

Search area	Number of neighbouring grid boxes	Range
± 2 pentads (10 days) and ± $\sim$111 km	n >100	μ± 2.5 σ
	15 <n <100	μ± 3.0 σ
	5 <n <15	μ± 3.5 σ
	0 <n <5	μ± 4.0 σ
± 2 pentads (10 days) and ± $\sim$222 km	n >0	μ± 4.0 σ
± 4 pentads (20 days) and ± $\sim$222 km	n >0	μ± 4.0 σ
	n = 0	NA (± Inf)

Additional drifting buoy quality control

The UK Met Office QC suite includes an additional set of checks for assessing observations from drifting buoys. These are based on an updated version of the checks described in Atkinson et al. [2013]. Only buoy observations passing previous quality checks are assessed. Checks are made to complete drifting buoy records (i.e. the checks are at along track/voyage level) for the period 1985-2014. Sea Surface Temperature (SST) checks are made relative to a satellite-based analysis that has been matched to report locations as a reference. The checks are:

- `Aground check`: Uses position reports to assess whether a buoy has run aground and is out of water. Returns 1 for reports deemed aground, 0 otherwise.
- `Picked-up check`: Uses position reports to assess whether a buoy has been picked up by a boat and is out of water. Returns 1 for reports deemed picked-up, 0 otherwise.

- `SST tail-check`: Compares buoy SSTs to a reference to look for poor quality data at the start of a buoy record (e.g. mis-calibrated at deployment) and the end of a buoy record (e.g. sensor failure). Returns 1 for reports where a start or end 'tail' is found, 0 otherwise.
- `SST bias and noise check`: Compares buoy SSTs to a reference to assess whether a buoy record is unacceptably biased or noisy as a whole. Returns 1 for all record reports if failing the check, 0 otherwise.
- `SST short record check`: For buoy records with a small number of observations, compares SSTs to a reference to assess whether too many buoy reports are in gross error. Returns 1 for all record reports if failing the check, 0 otherwise.

3.6 Report and Observation Quality Flags

To aid the user in selecting data the flags described above have been aggregated into in two simple flags. The first indicates the overall quality of the weather report. The second indicates the quality of the individual observations. Both the report quality and observation quality flags use the same encoding (see Table 8). If a QC flag is raised for any of the observation level checks described above the overall observation quality is marked as failing. Similarly, if a flag is raised for any of the report level or track level checks the report quality is marked as failing. Additionally, if a report passes all of the relevant report level checks but has no passing observations then it is marked as failing.

3.7 Impact of quality control

Figure 20 shows the outcome of the duplicate flagging. Prior to $\sim$1950 the majority of data observations are unique but with a small number flagged as either a best duplicate or worst duplicate. It was not possible to differentiate between best and worst duplicates for a small number of reports. In these cases, both observations are marked as duplicates with no qualification as best or worst. The impact of widespread data sharing can be seen in the 1950s and early 1960s with a significant number of best and worst duplicates. Between 1965 and the late 1970s the number of duplicates decreases

Table 8: Quality flag

Quality flag	Description
0	Passed
1	Failed
2	Not checked
3	Missing
4	Observed value updated and changed (manual correction)
5	Observed value updated and changed (automatic correction)

again. After 1980 there is an increasing number of unchecked reports due to the rise in drifting buoy numbers, with a sharp increase around 2000. During this period the drifting buoy network underwent a significant expansion. The rise in duplicates during this period is due to the collation of multiple real time data sources and delayed mode data into ICOADS.

Figure 21 shows the impact of the overall report level QC on the data available through the current data release. The unchecked reports in the 1950s to 1965 and after ∼1980 are due to the reports already being flagged as a duplicate (worst duplicate or duplicate) or flagged by the ICOADS IRF flag as a duplicate drifting buoy report. Typically, an extra 5 - 10% of reports are removed by the overall report quality flag.

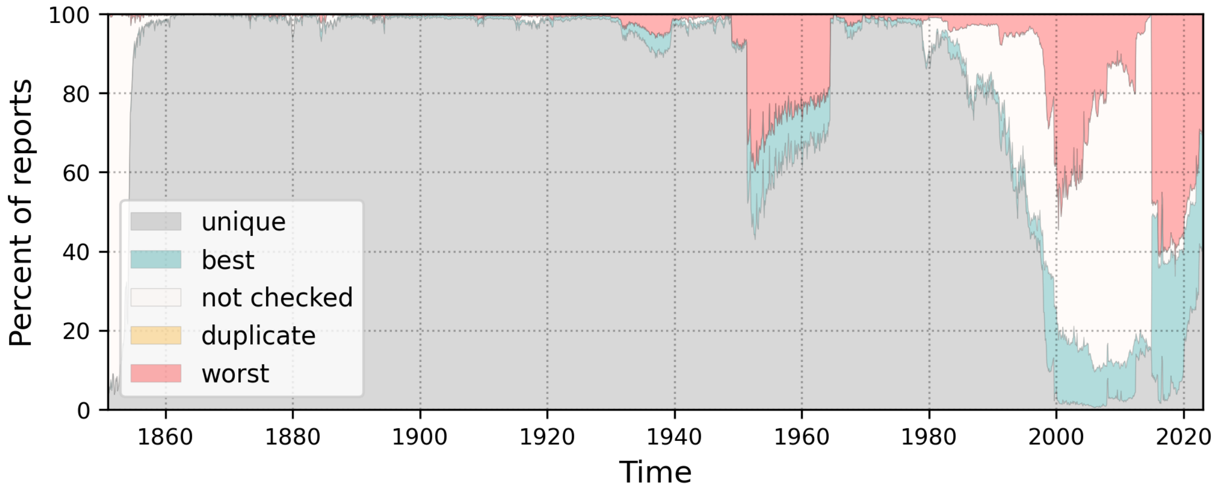

Figure 20: Percentage of reports flagged as unique, best duplicate, duplicate, worst duplicate or unchecked in the current data release. Reports from drifting buoys are unchecked at this stage.

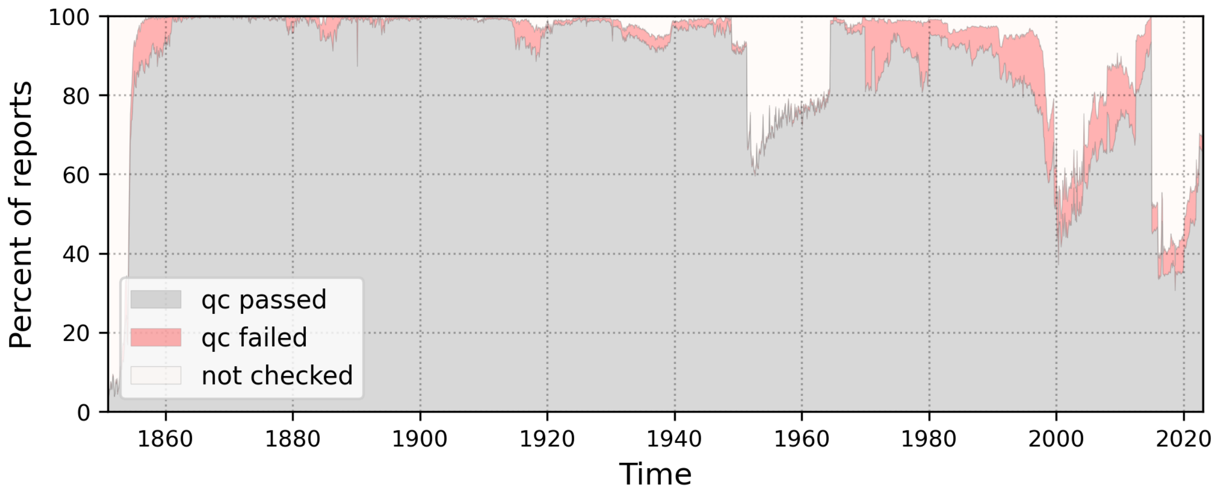

Figure 21: Percentage of reports flagged as unchecked, passing or failing the overall report quality check in the current data release.

4 Data access

The data can be downloaded and accessed via the Copernicus Climate Change Service (C3S) Climate Data Store (CDS):

`https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-observations-surface-marine?tab=form.`

The page has 3 tabs, the Overview tab presents an overview of the marine data and provides details on the data description, main and related variables available, contact email and links to licence/data policy statements. The Documentation tab provides links to PDF's of the Marine Product User Guide, Marine data inventories, Common Data Model specifications and a link to the Data Deposit Server webpage. The Download Data tab provides 6 sections that need to be completed in order to download data. The first selects one or more variables (Figure 22) from the available ECVs (Table 1). The second selects the level of quality control to apply, selecting either those that have passed all stages of quality control or all observations (Figure 23). The next three select the time period by selecting the year (Figure 24), month (Figure 25) and day (Figure 26). The final selection selects the region, specifying either select all data (global) or observation within a bounding box (Figure 27).

Data can additionally be downloaded using the CDS Python API. Selecting the `Show API request` button reveals the API commands to use for the given selection (Figure 28). More information can be found at:

`https://cds.climate.copernicus.eu/api-how-to`

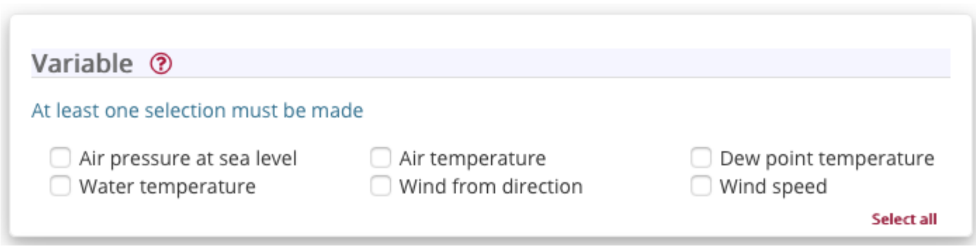

Figure 22: Selection box for variable selection on the Global Land And Marine Observations Database download page on the CDS.

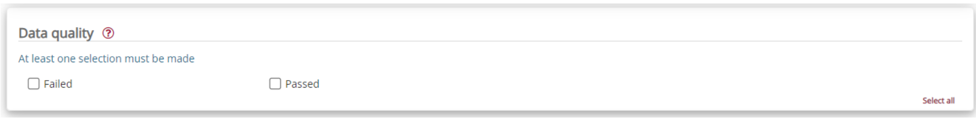

Figure 23: Selection box for quality status on the Global Land And Marine Observations Database download page on the CDS.

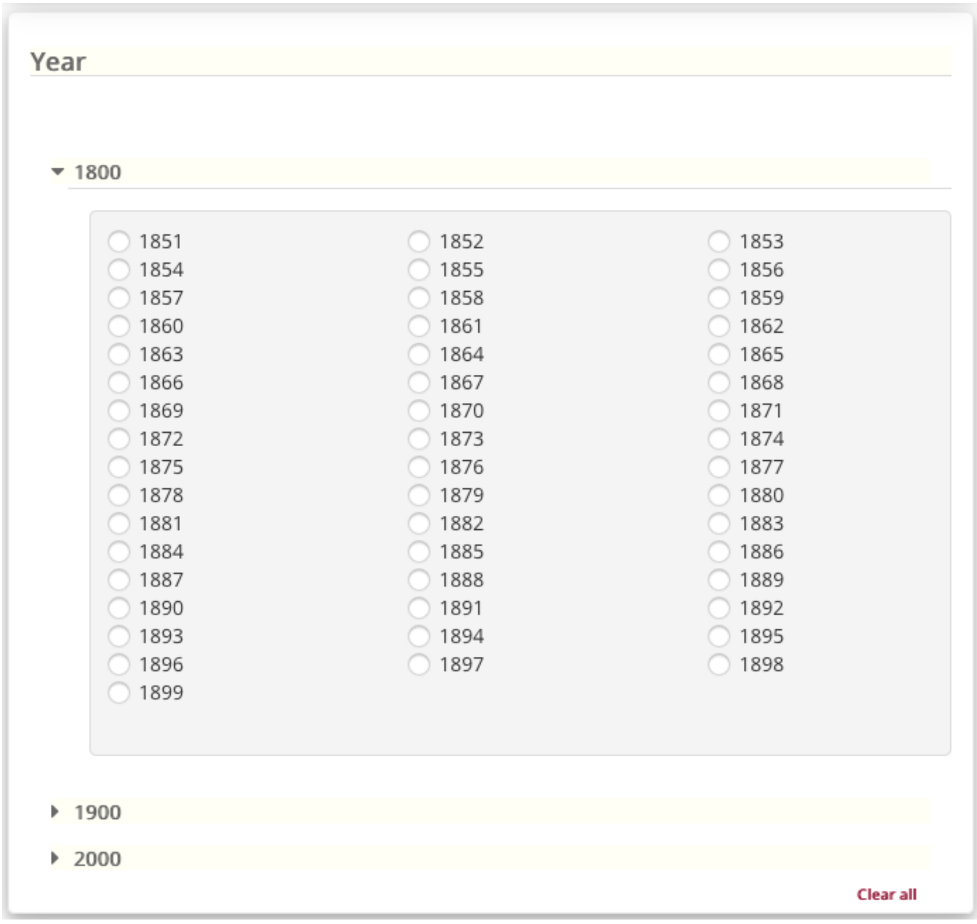

Figure 24: Selection box for selecting year on the Global Land And Marine Observations Database download page on the CDS.

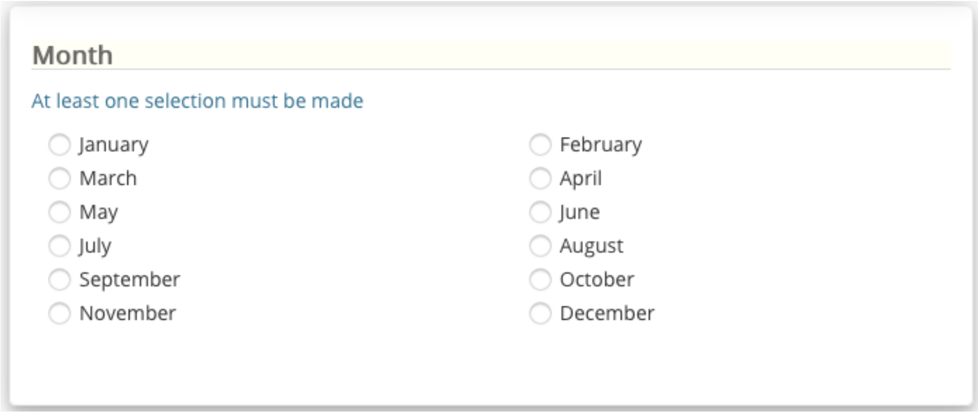

Figure 25: Selection box for selecting month on the Global Land And Marine Observations Database download page on the CDS.

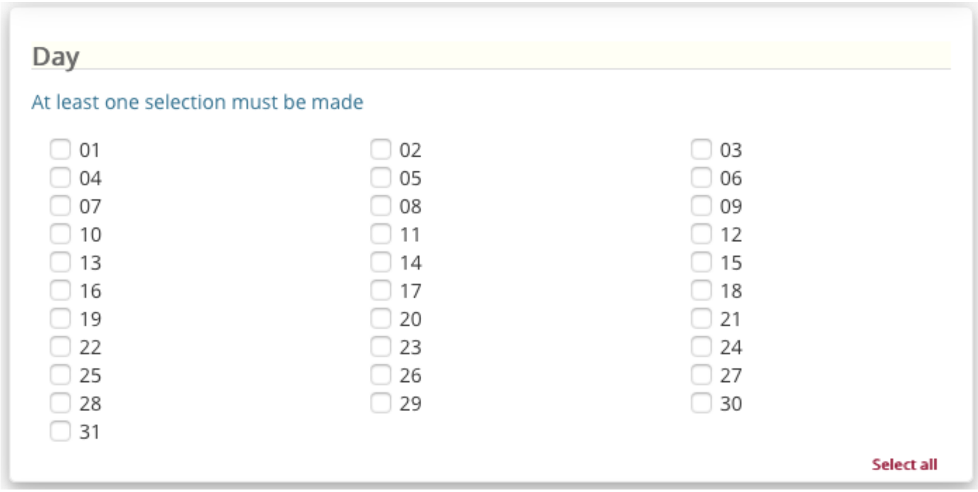

Figure 26: Selection box for selecting day on the Global Land And Marine Observations Database download page on the CDS.

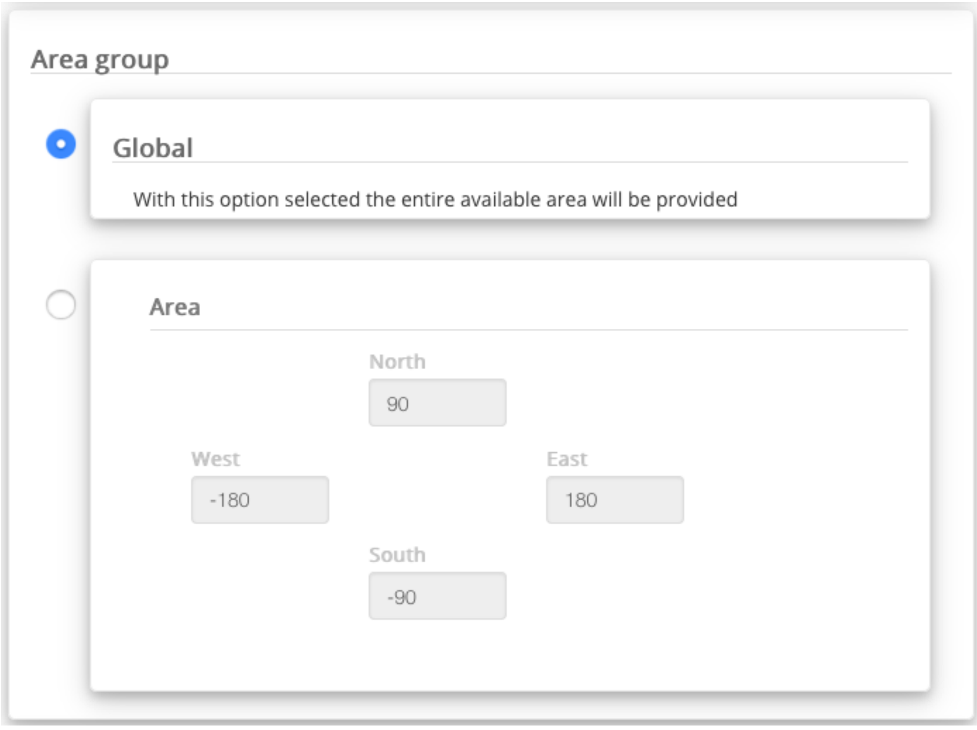

Figure 27: Selection box for selecting area on the Global Land And Marine Observations Database download page on the CDS.

```
import cdsapi
c = cdsapi.Client()
c.retrieve(
    'in-situ-observatins-surface-marine',
    {
        'format': 'csv-obs.zip',
        'day': '07',
        'month': '01',
        'year': '1898',
        'data_quality': 'quality_controlled',
        'variable': 'water_temperature',
    },
    'download.csv-obs.zip')
```

Figure 28: Example API request to download data from the Global Land And Marine Observations Database.

5 Data model and formats

Data from the Climate Data Store are returned in delimited text files (.csv), with 19 columns per row and one observed value per row. The columns are listed in Table 9 together with a brief description. The date and time of the observation is given by the `date_time` column, with the value returned as a string of the format `YYYY-MM-DD hh:mm:ss`. All data are returned in the UTC time zone. The location of the observation is given by the `latitude` and `longitude` columns in the WGS84 / EPSG4326 coordinate reference system. The variable being reported / observed is given by the `observed_variable` column, the units by the `units` column and the observed value by the `observation_value` column. The remaining columns provide further contextual information for the observations, such as platform type and station identifiers etc.

Table 9: Data model.

Field	**Kind**	**Description**
observation_id	character	Unique identifier for observation. For marine observations this has been set as <dataset>-<version>-<uid>-<field>, e.g. ICOADS-30-0NBVN1-AT for an air temperature observation with UID 0NBVN1 from ICOADS release 3.
report_type	character	Subdaily / hourly data.
date_time	timestamp with time zone (YYYY-MM-DD hh:mm:ss+0Z)	Timestamp (date/time) of observation.
date_time_meaning	character	date_time represents the end of the time period over which the observation was made.
latitude	numeric	Latitude of the observation (degrees North), range -90 to 90.
longitude	numeric	Longitude of the observation (degrees East), range -180 to 180.
observation_height_above_station_surface	numeric	Height of the sensor above local ground or sea level in metres. Positive values for observations above the surface. For visual observations, height of the observing platform.
observed_variable	character	The variable being observed / measured.
units	character	Units of the observed variable.
observation_value	numeric	The value of the observed variable.
value_significance	character	The significance of the observed variable, e.g. instantaneous value or mean over indicated period.
observation_duration	character	The time period over which the observation was made.

Table 9 – continued from previous page

Field	Kind	Description
platform_type	character	The type of structure upon which the sensors are mounted, e.g. ship, drifting buoy, tower etc.
station_type	character	Type of station, e.g. sea station, land station
primary_station_id	character	The primary station identifier, e.g. ship call-sign, buoy number, WIGOS station identifier.
station_name	character	Name of the station or site.
quality_flag	integer	Flag indicating the quality of the observation.
data_policy_licence	character	Usage permissions / licence for the data.
source_id	character	Identifier used to link the observation to the original source of the data. For marine data this links to the version of ICOADS, source and deck identifier and which ICOADS monthly file the record is from. Combined with the observation ID this can be used to trace an observation back to the original source record.

Table 10: Meaning of platform_type variable. Note: Settings 0-4 are derived from the "OSV or Ship Indicator" in NCDC (1968); Settings 0-1 are very poorly documented and probably should be regarded as equivalent to ship data (value=5). As used by ICOADS.

Flag Value	Meaning
0	US Navy or "deck" log, or unknown
1	Merchant ship or foreign military
2	Ocean station vessel—off station or station proximity unknown
3	Ocean station vessel—on station
4	Lightship
5	Ship
6	Moored buoy
7	Drifting buoy
8	Ice buoy [note: currently unused]
9	Ice station (manned, including ships overwintering in ice)
10	Oceanographic station data (bottle and low-resolution CTD/XCTD data)
11	Mechanical/digital/micro bathythermograph (MBT)
12	Expendable bathythermograph (XBT)
13	Coastal-Marine Automated Network (C-MAN) (NDBC operated)
14	Other coastal/island station
15	Fixed (or mobile) ocean platform (plat, rig)
16	Tide gauge
17	High-resolution Conductivity-Temp.-Depth (CTD)/Expendable CTD (XCTD)
18	Profiling float
19	Undulating oceanographic recorder
20	Autonomous pinneped bathythermograph
21	Glider

6 Data licensing and permissions

The marine data builds upon several decades of effort under ICOADS. Since its outset ICOADS has been served under a fully open licence by NCAR[4] and NOAA[5] . Any business built upon ICOADS, e.g. a derivation of the data to provide guidance to marine shipping companies, is perfectly acceptable and encouraged. Further, there are no limits to re-use and third party data sharing. Source decks which did not meet these restrictions were not ingested into ICOADS. There are no 'hidden' holdings of more restricted IPR available within ICOADS. There will, however, be marine data (e.g. from research vessels, military, coastguard, bio-monitoring) that have not been prioritised for accession for a wide variety of reasons (including resources, quality, volume, range of parameters, complexity and perhaps uncertain IPR).

References

C. P. Atkinson, N. A. Rayner, J. Roberts-Jones, and R. O. Smith. Assessing the quality of sea surface temperature observations from drifting buoys and ships on a platform-by-platform basis. *Journal of Geophysical Research-Oceans*, 118(7):3507–3529, Jul 2013. doi: {10.1002/jgrc.20257}.

D. Berry, E. Kent, and P. Taylor. An analytical model of heating errors in marine air temperatures from ships. *Journal of Atmospheric And Oceanic Technology*, 21(8):1198–1215, Aug 2004. ISSN 0739-0572. doi: {10.1175/1520-0426(2004)021<1198:Aamohe>2.0.CO;2}.

D. I. Berry and E. C. Kent. Air-Sea fluxes from Icoads: the construction of a new gridded dataset with uncertainty estimates. *International Journal of Climatology*, 31(7, SI):987–1001, Jun 15 2011. ISSN 0899-8418. doi: {10.1002/joc.2059}. 3rd Jcomm Workshop on Advances in Marine Climatology (Climar-Iii), Gdynia, Poland, May, 2008.

S. Bojinski, M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp. The Concept of Essential Climate Variables in Support of Climate Research, Applications, And Policy. *Bulletin of The American Meteorological Society*, 95(9):1431–1443, Sep 2014. ISSN 0003-0007. doi: {10.1175/Bams-D-13-00047.1}.

J. Businger, J. Wyngaard, Y. Izumi, and E. Bradley. Flux-Profile Relationships in Atmospheric Surface Layer. *Journal of The Atmospheric Sciences*, 28(2):181–&, 1971. ISSN 0022-4928. doi: {10.1175/1520-0469(1971)028<0181:Fprita>2.0.CO;2}.

G. Carella, E. C. Kent, and D. I. Berry. A probabilistic approach to ship voyage reconstruction in Icoads. *International Journal of Climatology*, 37(5):2233–2247, Apr 2017. ISSN 0899-8418. doi: {10.1002/

[4] https://www2.ucar.edu/terms-of-use

[5] https://www.nodc.noaa.gov/about/datapolicy.html

joc.4492}. 4th Jcomm Workshop on Advances in Marine Climatology (Climar), Asheville, NC, Jun 09-12, 2014.

D. Chan, E. C. Kent, D. I. Berry, and P. Huybers. Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, 571(7765):393+, Jul 18 2019. ISSN 0028-0836. doi: {10.1038/s41586-019-1349-2}.

M. Chenoweth. A new methodology for homogenization of 19th century marine air temperature data. *Journal of Geophysical Research-Atmospheres*, 105(D23):29145–29154, Dec 16 2000. ISSN 2169-897X. doi: {10.1029/2000JD900050}.

R. C. Cornes, E. C. Kent, D. I. Berry, and J. J. Kennedy. Classnmat: A global night marine air temperature data set, 1880-2019. *Geoscience Data Journal*, 7(2):170–184, Nov 2020. ISSN 2049-6060. doi: {10.1002/gdj3.100}.

E. Freeman, S. D. Woodruff, S. J. Worley, S. J. Lubker, E. C. Kent, W. E. Angel, D. I. Berry, P. Brohan, R. Eastman, L. Gates, W. Gloeden, Z. Ji, J. Lawrimore, N. A. Rayner, G. Rosenhagen, and S. R. Smith. Icoads Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5):2211–2232, Apr 2017. ISSN 0899-8418. doi: {10.1002/joc.4775}. 4th Jcomm Workshop on Advances in Marine Climatology (Climar), Asheville, NC, Jun 09-12, 2014.

J. J. Kennedy. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, 52(1):1–32, Mar 2014. doi: {10.1002/2013RG000434}.

J. J. Kennedy, N. A. Rayner, C. P. Atkinson, and R. E. Killick. An ensemble data set of sea surface temperature change from 1850: The met office hadley centre hadsst.4.0.0.0 data set. *Journal of Geophysical Research: Atmospheres*, 124(14):7719–7763, 2019. doi: https://doi.org/10.1029/2018JD029867. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029867`.

J. J. Kennedy, N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research-Atmospheres*, 116, Jul 22 2011. ISSN 2169-897X. doi: {10.1029/2010JD015220}.

E. Kent and P. Taylor. Choice of a Beaufort equivalent scale. *Journal of Atmospheric And Oceanic Technology*, 14(2):228–242, Apr 1997. ISSN 0739-0572. doi: {10.1175/1520-0426(1997)014<0228:Coabes>2.0.CO;2}.

E. C. Kent, D. I. Berry, I. P. González, R. Cornes, and J. Kennedy. Documentation for marine duplicate identification and linking of platform identifiers. Technical report, National Oceanography Centre, 2019.

E. C. Kent, S. D. Woodruff, and D. I. Berry. Metadata from Wmo publication no. 47 and an assessment of voluntary observing ship observation heights in Icoads. *Journal of Atmospheric And Oceanic Technology*, 24(2):214–234, Feb 2007. ISSN 0739-0572. doi: {10.1175/JTECH1949.1}.

E. C. Kent, N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker. Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *Journal of Geophysical Research-Atmospheres*, 118(3):1281–1298, Feb 16 2013. ISSN 2169-897X. doi: {10.1002/jgrd.50152}.

E. C. Kent, J. J. Kennedy, T. M. Smith, S. Hirahara, B. Huang, A. Kaplan, D. E. Parker, C. P. Atkinson, D. I. Berry, G. Carella, Y. Fukuda, M. Ishii, P. D. Jones, F. Lindgren, C. J. Merchant, S. Morak-Bozzo, N. A. Rayner, V. Venema, S. Yasui, and H.-M. Zhang. A Call For New Approaches to Quantifying Biases in Observations of Sea Surface Temperature. *Bulletin of The American Meteorological Society*, 98(8): 1601–1616, Aug 2017. ISSN 0003-0007. doi: {10.1175/Bams-D-15-00251.1}.

C. Liu, E. Freeman, E. C. Kent, D. I. Berry, S. J. Worley, S. R. Smith, B. Huang, H. min Zhang, T. Cram, Z. Ji, M. Ouellet, I. Gaboury, F. Oliva, A. Andersson, W. E. Angel, A. R. Sallis, and A. Adeyeye. Blending TAC and BUFR marine in situ data for ICOADS near-real-time release 3.0.2. *Journal of Atmospheric and Oceanic Technology*, Aug. 2022. doi: 10.1175/jtech-d-21-0182.1. URL `https://doi.org/10.1175/jtech-d-21-0182.1`.

B. Moat, M. Yelland, R. Pascal, and A. Molland. An overview of the airflow distortion at anemometer sites on ships. *International Journal of Climatology*, 25(7):997–1006, Jun 15 2005. ISSN 0899-8418. doi: {10.1002/joc.1177}. 2nd Workshop on Advances in Marine Climatology (Climar-2), Brussels, Belgium, Nov, 2003.

N. Rayner, D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research-Atmospheres*, 108(D14), Jul 17 2003. ISSN 2169-897X. doi: {10.1029/2002JD002670}.

S. R. Smith, G. Alory, A. Andersson, W. Asher, A. Baker, D. Berry, I, K. Drushka, D. Figurskey, E. Freeman, P. Holthus, T. Jickells, H. Kleta, E. C. Kent, N. Kolodziejczyk, M. Kramp, Z. Loh, P. Poli, U. Schuster, E. Steventon, S. Swart, O. Tarasova, L. P. de la Villeon, and N. Vinogradova-Shiffer. Ship-Based Contributions to Global Ocean, Weather, and Climate Observing Systems. *Frontiers in Marine Science*, 6, Aug 2 2019. doi: {10.3389/fmars.2019.00434}.

B. R. Thomas, E. C. Kent, V. R. Swail, and D. I. Berry. Trends in ship wind speeds adjusted for observation method and height. *International Journal of Climatology*, 28(6):747–763, May 2008. ISSN 0899-8418. doi: {10.1002/joc.1570}.

K. M. Willett, P. D. Jones, N. P. Gillett, and P. W. Thorne. Recent Changes in Surface Humidity: Development of the HadCruh Dataset. *Journal of Climate*, 21(20):5364–5383, Oct 2008. ISSN 0894-8755. doi: {10.1175/2008Jcli2274.1}.

S. Woodruff, R. Slutz, R. Jenne, and P. Steurer. A Comprehensive Ocean-Atmosphere Data Set. *Bulletin of The American Meteorological Society*, 68(10):1239–1250, Oct 1987. ISSN 0003-0007. doi: {10.1175/1520-0477(1987)068<1239:Acoads>2.0.CO;2}.

S. D. Woodruff, S. J. Worley, S. J. Lubker, Z. Ji, J. E. Freeman, D. I. Berry, P. Brohan, E. C. Kent, R. W. Reynolds, S. R. Smith, and C. Wilkinson. Icoads Release 2.5: extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, 31(7, SI):951–967, Jun 15 2011. ISSN 0899-8418. doi: {10.1002/joc.2103}. 3rd Jcomm Workshop on Advances in Marine Climatology (Climar-Iii), Gdynia, Poland, May, 2008.

S. Worley, S. Woodruff, R. Reynolds, S. Lubker, and N. Lott. Icoads release 2.1 data and products. *International Journal of Climatology*, 25(7):823–842, Jun 15 2005. ISSN 0899-8418. doi: {10.1002/joc.1166}. 2nd Workshop on Advances in Marine Climatology (Climar-2), Brussels, Belgium, Nov, 2003.

F. Xu and A. Ignatov. In situ NOAASST Quality Monitor (iQuam). *Journal of Atmospheric And Oceanic Technology*, 31(1):164–180, Jan 2014. ISSN 0739-0572. doi: {10.1175/JTECH-D-13-00121.1}.

P. Zunino Rodriguez. C-raid improve the access to historical drifter data: Copernicus reprocessing of argos and iridium drifters (c-raid), 2023. URL `https://doi.org/10.17882/77184`.

Copernicus Climate Change Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-climate.eu

climate.copernicus.eu copernicus.eu ecmwf.int