

Customer	:	ESRIN	Document Ref	:	SST_CCI-SIRDS-UOL-001
Contract No	:	4000109848/13/I-NB	Issue Date	:	05 June 2017
WP No	:	32	Issue	:	1

Project : SST-CCI- Phase-II

Title : SIRDS Reference Data

Abstract : This document contains details of the reference data part of the SIRDS

Author : G. Corlett. **Checked by** : Merchant

Chris Atkinson, MOHC; Gary
Corlett, UoL

Chris Merchant
Science Leader, UoR

Accepted : _____

Craig Donlon
ESA Technical Officer

Distribution : SST CCI team members
ESA (Craig Donlon)

EUROPEAN SPACE AGENCY CONTRACT REPORT

The work described in this report was done under ESA contract.
Responsibility for the contents resides in the author or organisation
that prepared it.

AMENDMENT RECORD

This document shall be amended by releasing a new edition of the document in its entirety. The Amendment Record Sheet below records the history and issue status of this document.

AMENDMENT RECORD SHEET

ISSUE	DATE	REASON FOR CHANGE
A	26 May 2017	Initial draft
1	05 June 2017	Issued to ESA after internal review

TABLE OF CONTENTS

1. INTRODUCTION.....	4
1.1 Purpose and Scope.....	4
1.2 Structure of the Document	4
2. DATASET CREATION.....	5
2.1 Introduction.....	5
3. NETCDF FILE FORMAT.....	9
3.1 Collections and Sub-Collections.....	10
4. QUALITY CONTROL SCHEMES	12
5. CONTENT OF REFERENCE DATASET	14

1. INTRODUCTION

The SST CCI project is part of the ESA Climate Change Initiative (CCI), which aims to produce and validate sea surface temperature (SST) essential climate variable (ECV) data products. As part of the validation activities an SST CCI Independent Reference Data Set (SIRDS) has been assembled. The SIRDS is comprised of two components:

- Reference data
- Match-up data

The SIRDS reference data are quasi-independent of satellite SST products for the period 1981-2016 for use in validation, verification and other SST CCI SST product performance assessments.

Match-up data are specified in a separate document. This will be available once the final match-up processing has been completed.

1.1 Purpose and Scope

This document provides details of the content and generation of the reference data for the SST CCI SIRDS.

1.2 Structure of the Document

After this introductory section, the document is divided into a number of major sections. Section 2 provides details of how the SIRDS reference data were assembled. Section 3 gives details of the file format and the quality control methods applied are summarised in Section 4. Finally, in Section 5 a list of files comprising the SIRDS reference data is provided.

2. DATASET CREATION

Version 1.0 of the final reference dataset was extracted from HadIOD.1.2.0.0 using r322 of the /sstcci_refdata code in the hadiod FCM repository: <http://fcm1/projects/HadOBS/browser/hadiod>

2.1 Introduction

HadIOD is the Met Office Hadley Centre Integrated Ocean Database, which brings together ocean temperature and salinity observations made by surface and sub-surface-profiling instruments. The database includes quality flags and, where possible, assigns bias corrections and uncertainty estimates for different instrument types. The database contains observations of the global oceans from 1850-present, with latest monthly updates made approximately 1.5 months after the end of the target month.

The primary data sources for HadIOD.1.2.0.0 are the International Comprehensive Ocean-Atmosphere Dataset (ICOADS) version 2.5.1 for surface-only data, and the Met Office Hadley Centre Ensembles dataset version 4.2.0 (EN4) for sub-surface profiling data. Delayed mode Global Tropical Moored Buoy Array (GTMBA) data taken from NOAA PMEL are also included in the database. In November 2016, surface buoy data (drifting and moored) were absent from the ICOADS Real Time data stream and so GTS buoy data from the Met Office's MetDB were used in HadIOD as a substitute.

For further information about these sources see:

- ICOADS - <http://icoads.noaa.gov/>
- EN4 - <http://www.metoffice.gov.uk/hadobs/en4/>
- PMEL GTMBA - <http://www.pmel.noaa.gov/tao/>

A paper describing HadIOD.1.0.0.0 has been published in JGR-Oceans. HadIOD.1.2.0.0 is very similar in design to HadIOD.1.0.0.0 but with some updates to the source data used (e.g. to the latest version of EN4) and metadata assignment (such as QC flags and bias corrections):

- *Atkinson, C.P., N.A. Rayner, J.J. Kennedy, S.A. Good, 2014: An Integrated Database of Ocean Temperature and Salinity Observations. JGR-Oceans, 119, 7139-7163, doi:10.1002/2014JC010053.*

Reference data files are in NetCDF format. Data cover 1978-present. Each NetCDF file contains monthly SST data from a particular instrument type. Instruments selected for the reference dataset are drifters, voluntary observing ships (ship), GTMBA, mooring data (excluding GTMBA), bottles, Conductivity-Temperature-Depth casts (CTDs), Mechanical Bathymeters (MBTs), expendable bathythermographs (XBTs), Argo floats and animal data (these are data from instrumented marine animals, e.g. elephant seals). Files may not be available for each instrument type each month because of changes in the usage of different instrument types over time. Figure 1 shows plots of the number of observations in the files per month for each instrument type. Note that the large drop in the number of mooring (excluding GTMBA) observations in November 2016 is due to a change of data source (from ICOADS to MetDB).

Files are named SSTCCI2_refdata_INSTTYPE_YYYYMM.nc, where YYYY=YEAR, MM=MONTH and INSTTYPE=instrument type (e.g. Argo, drifter)

GTMBA data are a blend of delayed mode data taken from NOAA PMEL supplemented by ICOADS real time data. ICOADS data are sub-sampled to hourly. NOAA PMEL data are provided at native sampling frequency (mainly 10-minutely and hourly). The NOAA PMEL data were downloaded on 21/09/16. The ICOADS contribution to the GTMBA dataset increases in the years leading up to this date (see Figure 1).

For observations made by sub-surface profiling instruments (bottle, CTD, MBT, XBT, Argo and animal), the shallowest temperature observation passing quality control in the depth range 4-6m is provided in the files. This depth range was selected as the shallowest range common to all profiling instrument types; XBTs do not reach thermal equilibrium with seawater until around 4 dbar depth whilst Argo float CTD sensors stop pumping as they approach the surface to avoid sensor degradation. For SST observations from surface-only instruments (drifter, ship, GTMBA, mooring) all available data are included in the files (whether or not they pass QC).

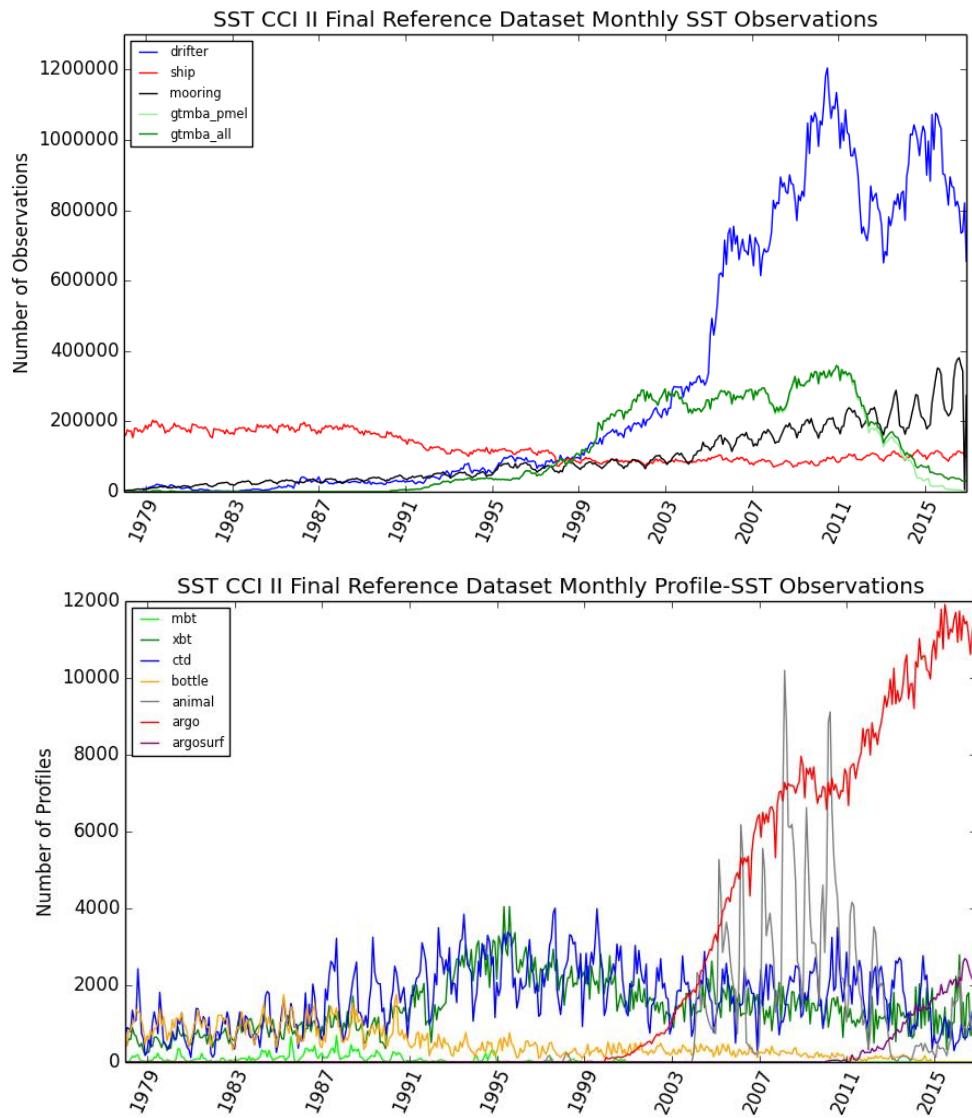


Figure 1. Monthly counts of number of SST observations in ESA SST CCI Phase II final reference dataset files. Period shown is January 1978–December 2016, coverage is extended monthly. (Top) number of observations from surface-based instruments (these are not filtered for observations passing QC); for GTMBA, PMEL-only (gtmbo_pmel) and PMEL+ICOADS (gtmbo_all) are shown. (Bottom) number of profiles from sub-surface profiling instruments (the shallowest observation passing QC in the depth range 4–6m for each profile is included in the file i.e. one observation per profile, except surface Argo where all observations in the depth range 0–10m are included). Note the difference in scale between the two plots.

In addition to the Argo data described above, a further set of ‘near-surface’ Argo data files was extracted. This contains near-surface data from Solo2 and S2A type Argo floats. These floats have been modified to continue pumped primary sampling up to 1 dbar, providing measurements

of vertical temperature structure close to the surface. The near-surface Argo files include data for all depths ranging from 0-10m for profiles that sample shallower than 4m. Observations from the same profile can be identified using the PROF_ID variable.

From December 2007 onwards, for security reasons most ship callsigns in ICOADS v2.5.1 are masked using the ID 'MASKSTID'. This mainly affects ICOADS Real Time monthly updates (2008-onwards) which are based on GTS data. During HadIOD production, instances of 'MASKSTID' are unmasked where possible using data kept by the Met Office, however these unmasked IDs cannot be released externally. For the ship reference data, IDs of form 'MKnnnnnnn' (where nnnnnnn is a 7-digit zero-padded integer) have been mapped to these unmasked IDs for release. These are stored in the PLAT_ID variable and can be used to track observations from the same ship. Note that some of these IDs may map to generic IDs used by multiple ships, e.g. 'SHIP'. In particular 'MK0000016' and 'MK0000172' cannot be used for tracking as they correspond to the IDs 'MASKSTID' and 'SHIP' respectively. This ID scheme is only applied to ship reference data from December 2007 onwards, prior to this original ICOADS callsigns are provided in the PLAT_ID variable.

All data are provided in a common file format (see 'NetCDF File Format' below). The content of the QC variables varies dependent on instrument type (see 'QC Schemes' below). Note that depths for drifter, ship and mooring observations are estimated (i.e. an estimated depth is assigned because none is available from ICOADS).

HadIOD.1.2.0.0 assigns bias corrections and uncertainties to observations using the following error model:

True SST =

observed SST + instrument-type-bias + individual-instrument-bias + random-measurement-error

'instrument-type-bias' is a bias that arises at the level of instrument type (e.g. ship bucket, XBT); a correction for this effect is included in the variable SST_TYPE_CORR. 'individual-instrument-bias' is a bias in a particular instrument (e.g. a particular float); a correction for this effect is stored in the variable SST_PLAT_CORR. SST_TYPE_CORR_UNC and SST_PLAT_CORR_UNC give an estimate of the uncertainty in the value of a correction. After applying the corrections, these uncertainties will be correlated amongst observations of the same instrument type or from the same platform respectively. Correlated uncertainties must be propagated appropriately if aggregating groups of observations (see also the examples below).

Bias corrections and their uncertainties are not provided for all instrument types. At present SST_TYPE_CORR corrections are provided for ship, XBT and MBT instrument types. SST_TYPE_CORR_UNC is only provided for ships. For MBTs and XBTs not all observations are assigned corrections and only those observations whose metadata allows a correction to be applied have been included in the files. It is recommended that users make use of the corrections where provided. Note that the depth correction (DEPTH_CORR) and SST_TYPE_CORR should be applied in tandem. This is because the XBT and MBT correction schemes correct for both thermal and depth biases and the ship SST corrections adjust relative to a reference depth.

At present, SST_PLAT_CORR is not provided for any instrument type but SST_PLAT_CORR_UNC is specified for observations from surface only instruments (i.e. drifters, ships, moorings and GTMBA). This means we are not yet able to provide bias corrections for individual platforms, but for some instrument types we are able to estimate the range in which the individual platform bias is likely to lie. An estimate of random measurement uncertainty (SST_RAND_UNC) is provided at minimum for all instrument types. This is uncorrelated from one observation to the next. All uncertainties are given as 1 standard deviation.

Some examples of handling observation correction and uncertainty information for different instrument types are given below. For further information and a more detailed example please see the HadIOD paper.

- **Argo.** No corrections or their associated uncertainties are provided for Argo data. Random measurement uncertainty is 0.002°C. Total observation uncertainty, which combines all available uncertainty terms for an observation in quadrature, is therefore 0.002°C. This observation uncertainty is uncorrelated amongst Argo observations.
- **Drifter.** No corrections are provided for drifter data. SST_PLAT_CORR_UNC is 0.29°C and random measurement uncertainty is 0.26°C. Combining uncertainty in quadrature gives a total observation uncertainty of 0.39°C in drifter observations. If aggregating drifter observations, the random component of uncertainty is uncorrelated, but SST_PLAT_CORR_UNC is correlated amongst observations from the same platform (identifiable using PLAT_ID).
- **XBT.** SST_TYPE_CORR and DEPTH_CORR should be added to the observations. No correction uncertainties are provided. Random measurement uncertainty is 0.15°C. Total observation uncertainty, which combines all available uncertainty terms for an observation in quadrature, is therefore 0.15°C. This observation uncertainty is uncorrelated amongst corrected XBT observations
- **Ship.** SST_TYPE_CORR and DEPTH_CORR should be added to the observations. SST_TYPE_CORR_UNC, SST_PLAT_CORR_UNC and random measurement uncertainty are provided. SST_PLAT_CORR_UNC is 0.71°C, random measurement uncertainty is 0.74°C and SST_TYPE_CORR_UNC varies in value. Combining these uncertainties in quadrature gives the total observation uncertainty in corrected ship observations. If aggregating corrected ship observations, the random component of uncertainty is uncorrelated, SST_PLAT_CORR_UNC is correlated amongst observations from the same platform (identifiable using PLAT_ID) and SST_TYPE_CORR_UNC is correlated amongst the wider ship population. The correlation of SST_TYPE_CORR_UNC is not straightforward to express and so HadIOD includes multiple realisations of SST_TYPE_CORR which can be used interchangeably to explore this correlation. Please see the HadIOD paper for more information and contact us if interested in using these ensembles.

3. NETCDF FILE FORMAT

Variables:

- OB_ID = sequential numeric observation id applied per month (an ob is uniquely identified using a combination of the ob_id, year and month)
- COLLECTION = a numeric id used to identify the source dataset an observation was taken from (see section below)
- SUBCOL1 = code to identify sub-collections within collections (see section below)
- SUBCOL2 = code to identify sub-collections within collections (see section below)
- PROF_ID = sequential numeric profile id applied per month (a profile is uniquely identified using a combination of the prof_id, year and month; all obs from the same profile receive the same id, obs from surface-only instruments are treated as a profile with a single observation)
- PLAT_ID = a unique identifier for each platform
- LONGITUDE = longitude of ob in -180W to 180E format
- LATITUDE = latitude of ob
- YEAR = year of ob
- MONTH = month of ob
- DAY = day of ob
- HOUR = hour of ob
- MINUTE = minute of ob
- SECOND = second of ob
- DEPTH = depth of ob below surface (metres; this value is inferred for ICOADS obs e.g. 0.2m for drifters)
- DEPTH_CORR = depth correction (metres; added to depth to correct for biases, not prescribed for all ob types)
- SST = observed sea surface temperature (DegC)
- SST_TYPE_CORR = SST correction (instrument type) (DegC, added to SST to correct for biases associated with a particular instrument type, e.g. SST measurements made by ship bucket, not prescribed for all ob types)
- SST_TYPE_CORR_UNC = uncertainty in SST correction (instrument type) (DegC, 1 std.dev.)
- SST_PLAT_CORR = SST correction (individual instrument) (DegC, added to SST to correct for biases associated with a particular instrument, e.g. a particular float, not prescribed for all ob types)
- SST_PLAT_CORR_UNC = uncertainty in SST correction (individual instrument) (DegC, 1 std.dev.)

- SST_RAND_UNC = SST random measurement uncertainty (DegC, 1 std.dev.)
- SST_COMB_UNC = total observation uncertainty (DegC, 1 std.dev.; produced by combining all available uncertainty terms in quadrature; where SST_TYPE_CORR and SST_PLAT_CORR are specified these must be applied for SST_COMB_UNC to be valid)
- QC1 = observation quality flag 1 (see section below)
- QC2 = observation quality flag 2 (see section below)

Attributes:

- Description = description of file content including final reference data version number
- Source = version of HadIOD from which data was extracted and correction scheme selected for bias corrections
- Usage = file specific guidance for using the data (e.g. how bias corrections should be applied)
- History = file creation date

3.1 Collections and Sub-Collections

The content of the SUBCOL1 and SUBCOL2 variables varies dependent on the value of the COLLECTION variable. This is detailed in the table below.

SOURCE	COLLECTION	SUBCOL1	SUBCOL2
ICOADS	1	ICOADS Deck ID*	ICOADS Source ID*
EN4	2	1=WOD13** 2=Argo GDAC 3=GTSPP	N/A
MetDB	3	N/A	N/A
PMEL GTMBA	5	1=TAO/TRITON 2=PIRATA 3=RAMA	PMEL GTMBA SST source code***

* See Table All and AllI in Woodruff, S.D. et al. (2011), ICOADS Release 2.5., Int.J.Climatol., 31:951-967, doi:10.1002/joc.2103

** EN4 blends observations from multiple sources, the three main sources are used in the reference dataset: World Ocean Database 2013 (WOD13), Argo Global Data Assembly Centre (Argo GDAC), Global Temperature and Salinity Profile Programme (GTSPP).

*** PMEL SST source codes

0 - No sensor, no data

1 - Real time (telemetered mode)

2 - Derived from real time

- 3 - Temporally interpolated from real time
- 4 - Source code inactive at present
- 5 - Recovered from instrument RAM (delayed mode)
- 6 - Derived from RAM
- 7 - Temporally interpolated from RAM

4. QUALITY CONTROL SCHEMES

The content of the quality control variables QC1 and QC2 varies dependent on instrument type. This is detailed in the table below.

It is recommended that users omit observations that fail QC. In the case of PMEL GTMBA data, users should filter based on which QC values seem most appropriate (see footnotes below).

Type	QC1	QC2
drifter, ship	Basic Met Office Hadley Centre QC*, pass=0, fail=1	SST tracking QC**, pass=0, fail=1
mooring	Basic Met Office Hadley Centre QC*, pass=0, fail=1	N/A
CTD, XBT, MBT, bottle, animal, Argo, Argo-surface	Basic EN4 QC***, pass=0, fail>=1	N/A
GTMBA^* (COLLECTION=1 or 3)	Basic Met Office Hadley Centre QC*, pass=0, fail=1	N/A
GTMBA^* (COLLECTION=5)	PMEL SST flag^**	Position Flag^***

* Met Office Hadley Centre QC is used for ICOADS observations and comprises basic sanity checks, position track check, climatology check and buddy check.

** SSTs from individual drifters and ships are tracked against a satellite-based reference and persistently poor quality observations are flagged. The coverage of the SST tracking QC is 1986-2015 only. This is an updated version of the tracking QC checks described in: Atkinson, C.P. et al. (2013), Assessing the quality of SST observations from drifting buoys and ships on a platform by platform basis, JGR-Oceans, 118:1-23, doi:10.1002/jgrc.20257.

*** Profile observations have been QC'd by the EN4 system which has many elements (see <http://www.metoffice.gov.uk/hadobs/en4/>). Observations are flagged if they have failed any of the EN4 checks. For profile observations, only observations passing QC are included in the files, with the exception of Argo-surface which includes observations both passing and failing QC.

^* GTMBA files contain a blend of data from ICOADS (COLLECTION ID=1) and PMEL (COLLECTION ID=5). In November 2016, MetDB data (COLLECTION ID=3) are used as a substitute for ICOADS data.

^** PMEL SST flags

0 - datum missing

1 - highest quality; pre/post-deployment calibrations agree to within sensor specifications. In most cases only pre-deployment calibrations have been applied

2 - default quality; pre-deployment calibrations applied. Default value for sensors presently deployed and for sensors which were either not recovered or not calibratable when recovered.

3 - adjusted data; pre/post calibrations differ, or original data do not agree with other data sources (e.g., other in situ data or climatology), or original data are noisy. Data have been adjusted in an attempt to reduce the error.

4 - lower quality; pre/post calibrations differ, or data do not agree with other data sources (e.g. other in situ data or climatology), or data are noisy. Data could not be confidently adjusted to correct for error.

5 - sensor or tube failed

^*** Position flags for PMEL data

0 - datum missing (nominal position added)

1 - highest quality; Pre/post-deployment calibrations agree to within sensor specifications. In most cases only pre-deployment calibrations have been applied.

2 - default quality; Pre-deployment calibrations applied. Default value for sensors presently deployed and for sensors which were either not recovered or not calibratable when recovered.

3 - adjusted data; Pre/post calibrations differ, or original data do not agree with other data sources (e.g., other in situ data or climatology), or original data are noisy. Data have been adjusted in an attempt to reduce the error.

4 - lower quality; Pre/post calibrations differ, or data do not agree with other data sources (e.g., other in situ data or climatology), or data are noisy. Data could not be confidently adjusted to correct for error.

5 - sensor or tube failed

8 - has moved from deployed location

9 - daily drift speed exceeds .1 kts (0.05 m/s)

10+ - missing value(s) in-filled by linear interpolation of neighbouring values (gaps up to and including 10 days); flag value is created by taking the lowest quality indicator of the neighbouring flag values used for interpolation and adding 10 (e.g. if the two neighbouring flag values are 2 [default quality] and 9 [drift speed exceeds .1 kts] then the flag for interpolated values will be 9 + 10 = 19.

5. CONTENT OF REFERENCE DATASET

The SST_CCI SIRDS reference data is comprised of the following files:

