INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **37**: 2211–2232 (2017)
Published online 27 June 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4775

ICOADS Release 3.0: a major update to the historical marine climate record

Eric Freeman,[a,b]* Scott D. Woodruff,[a,c] Steven J. Worley,[d] Sandra J. Lubker,[e] Elizabeth C. Kent,[f] William E. Angel,[a] David I. Berry,[f] Philip Brohan,[g] Ryan Eastman,[h] Lydia Gates,[i] Wolfgang Gloeden,[i] Zaihua Ji,[d] Jay Lawrimore,[a] Nick A. Rayner,[g] Gudrun Rosenhagen[i] and Shawn R. Smith[j]

[a] *NOAA, National Centers for Environmental Information, Asheville, NC, USA*
[b] *STG. Inc., Asheville, NC, USA*
[c] *Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA*
[d] *Computational Information and Systems Laboratory, National Center for Atmospheric Research, Boulder, CO, USA*
[e] *NOAA, Earth System Research Laboratory, Boulder, CO, USA*
[f] *Marine Physics and Ocean Climate, National Oceanography Centre, Southampton, UK*
[g] *Hadley Centre, Met Office, Exeter, UK*
[h] *Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA*
[i] *Deutscher Wetterdienst, Hamburg, Germany*
[j] *Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA*

ABSTRACT: We highlight improvements to the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) in the latest Release 3.0 (R3.0; covering 1662–2014). ICOADS is the most widely used freely available collection of surface marine observations, providing data for the construction of gridded analyses of sea surface temperature, estimates of air–sea interaction and other meteorological variables. ICOADS observations are assimilated into all major atmospheric, oceanic and coupled reanalyses, further widening its impact. R3.0 therefore includes changes designed to enable effective exchange of information describing data quality between ICOADS, reanalysis centres, data set developers, scientists and the public. These user-driven innovations include the assignment of a unique identifier (*UID*) to each marine report – to enable tracing of observations, linking with reports and improved data sharing. Other revisions and extensions of the ICOADS' International Maritime Meteorological Archive common data format incorporate new near-surface oceanographic data elements and cloud parameters. Many new input data sources have been assembled, and updates and improvements to existing data sources, or removal of erroneous data, made. Coupled with enhanced 'preliminary' monthly data and product extensions past 2014, R3.0 provides improved support of climate assessment and monitoring, reanalyses and near-real-time applications.

KEY WORDS marine meteorological data; ship data; buoy data; sea-surface temperature; sea-level pressure; humidity; metadata; ocean

Received 1 June 2015; Revised 5 April 2016; Accepted 9 April 2016

1. Introduction

The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is the most extensive freely available archive of global surface marine data, with over 455 million individual marine reports – each containing the observations and metadata reported from a given ship, buoy, coastal platform, or oceanographic instrument. ICOADS provides a common access point for these marine reports and established the open access and version control now recognized as essential attributes to the production of verifiable climate products (Thorne *et al.*, 2011).

ICOADS spans the period from the late 17th Century to near present with near-real-time (NRT) updates. The overall objectives are (Woodruff *et al.*, 2011) to:

1. collect and preserve as much original surface marine data as possible;
2. treat each observation systematically, converting units and coding schemes to a uniform set;
3. preserve data source identification and measurement metadata with each record;
4. provide data in a common format, identify duplicate reports, and apply basic quality control (QC);
5. freely distribute the data and products.

Release 1 in 1985 of the Comprehensive Ocean-Atmosphere Data Set (COADS; Slutz *et al.*, 1985) provided individual reports and basic gridded monthly summary products covering the period 1854–1979 (Woodruff *et al.*, 1987). Recognizing international contributions over the years, the name was changed in 2002 to ICOADS (WMO, 2002). ICOADS is now formally managed via an international partnership with eight

*Correspondence to: E. Freeman, NOAA/NCEI/STG, 151 Patton Avenue, Asheville, NC 28801, USA. E-mail: Eric.Freeman@noaa.gov

Table 1. ICOADS Release history (update of Table AI in Woodruff *et al.*, 2011).

Release name	Issuance year	References	Resultant period	Subsequent updates and extensions
Release 1	1985	Slutz *et al.*, 1985; Woodruff *et al.*, 1987	1854–1979	1980–1991
Release 1a	1993	Woodruff *et al.*, 1993	1980–1992	1992–1993 1990–1995 1980–1997
Release 1b	1996		1950–1979	1970
Release 1c	2001		1784–1949	(none)
Release 2.0	2002	Woodruff *et al.*, 1998; Woodruff *et al.*, 2003	1784–1997	(none)
Release 2.1	2003	Worley *et al.*, 2005; Woodruff *et al.*, 2005	1784–1997	1998–2002
Release 2.2	2005		1784–2004	(none)
Release 2.3	2006		1784–2004	2005
Release 2.4	2007		1784–2004	2005–05/2007
Release 2.5	2009	Woodruff *et al.*, 2011	1662–2007	Extended monthly (≤5 days after data month)
Release 2.5.1	2013		1662–2007	
Release 2.5.2	2014		2008–08/2014	Extended to 12/2015
Release 3.0	2016	Present paper	1662–2014	Extended monthly (≤5 days after data month)

signatories from the United States, United Kingdom (UK), and Germany. Release 2.1 (Worley *et al.*, 2005) was notable for the introduction of NRT extensions, and Release 2.5 (R2.5) achieved substantial increases in historical coverage through inclusion of many newly recovered and digitized data sources (Woodruff *et al.*, 2011). Table 1 summarizes the release history.

ICOADS contains a wide range of near-surface meteorological and oceanographic parameters stored uniformly in the International Maritime Meteorological Archive (IMMA) data format (currently version 1; Smith *et al.*, 2016). IMMA allows observations to be described by metadata fields and also provides for archiving of the observations in their original (input) format. Preservation of original format data allows for later improvements, e.g. extracting additional variables when the IMMA format is extended, or allowing for correction of any translation errors. ICOADS contains observations of many Global Climate Observing System (GCOS) Essential Climate Variables (ECVs) for both atmospheric and oceanic domains (GCOS, 2010; Bojinski *et al.*, 2014). Available ECVs include air temperature (*AT*), sea-surface temperature (*SST*), wind speed and direction, humidity, sea-level pressure (*SLP*), cloud cover, sea state, and sea ice. ICOADS also contains descriptive information, such as coded weather information, and types and amounts of cloud at different levels in the atmosphere.

The data and products from ICOADS have been used by thousands of users, including universities and government agencies, research institutes, and the general public. Data set developers form a key user group who typically derive gridded data products from the ICOADS observations, vastly increasing its impact. Derived gridded products cover *SST* (e.g. Berry and Kent, 2011; Kennedy *et al.*, 2011a, 2011b; Hirahara *et al.*, 2014; Huang *et al.*, 2015; Liu *et al.*, 2015), wind speed (Berry and Kent, 2011; Tokinaga and Xie, 2011), humidity (Berry and Kent, 2011), *AT* (Berry and Kent, 2011; Kent *et al.*, 2013), *SLP* (Kaplan *et al.*, 2000; Allan and Ansell, 2006), cloud types and amounts (Hahn and Warren, 2009), waves (Gulev *et al.*, 2003; Gulev and Grigorieva, 2006), and air–sea fluxes (da Silva *et al.*, 1994; Josey *et al.*, 1999; Berry and Kent, 2009, 2011; Smith *et al.*, 2011). Such products are vital for climate monitoring and assessment studies such as the Intergovernmental Panel on Climate Change Assessments (e.g. Hartmann *et al.*, 2013; Rhein *et al.*, 2013) and the annual State of the Climate report (e.g. Blunden, 2014).

ICOADS observations are assimilated into reanalysis models (e.g. Saha *et al.*, 2010; Compo *et al.*, 2011; Dee *et al.*, 2011; Rienecker *et al.*, 2011; Kobayashi *et al.*, 2015) and ICOADS-derived gridded *SST* products are used as a boundary condition for atmosphere-only reanalyses. Further applications include calibration and validation of satellite data (e.g. Jackson and Wick, 2010; Roberts *et al.*, 2010; Bentamy *et al.*, 2013; Kinzel *et al.*, 2016) and assessment of climate model output (Flato *et al.*, 2013).

Alongside the IMMA observations, ICOADS regularly updates a suite of basic gridded monthly summary products, in which ten statistics (such as the mean and median) are calculated for each of 22 observed and derived variables, using 2° latitude × 2° longitude boxes back to 1800, and 1° × 1° boxes since 1960 (Worley *et al.*, 2005). These products are valuable for monitoring research (e.g. Wolter and Timlin, 2011) and ecosystem studies (e.g. Martinho *et al.*, 2013; Villegas-Hernández *et al.*, 2015). ICOADS summary products provide a 'quick look' resource based on unadjusted data without interpolation and statistical gap-filling.

Two international workshops series have promoted continuity and coordination for ICOADS:

Since 1999, four World Meteorological Organization (WMO)–Intergovernmental Oceanographic Commission of UNESCO (IOC) Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) Workshops on Advances in Marine Climatology (CLIMAR) have been held, most recently CLIMAR-IV in Asheville, North Carolina, USA (JCOMM, 2015); and, alternating approximately biennially with CLIMAR, Workshops on Advances in the Use of Historical Marine Climate Data (MARCDAT), the latest MARCDAT-III in Frascati, Italy (JCOMM, 2011). These meetings have guided development and maintained links with the international user community. In 2014, the ICOADS program was expanded into international partnership to include research and operational organizations in the UK and Germany (http://icoads.noaa.gov/partners.html, accessed 1 January 2016), to leverage expertise and resources to complete Release 3.0 (R3.0) and enhance future ICOADS releases.

In this paper, we highlight improvements to ICOADS in R3.0. Section 2 provides a description of the data management within ICOADS. Section 3 describes data sources new to R3.0 made available through digitization activities, external archives and additional GTS sources. Section 4 describes extensions to the IMMA data format and Section 5 discusses some of the data issues identified during R3.0 processing. Section 6 outlines future plans and Section 7 contains concluding remarks.

2. Data composition and management

This section provides an overview of the composition of ICOADS in four categories. The first consists of historical data from Release 1 of COADS (Section 2.1). Many of these data were digitized onto computer punched card decks (hence the 'deck' nomenclature still used in ICOADS) and often are limited in terms of available parameters and metadata. The second category is new historical data acquisitions from data recovery and digitization projects (Section 2.2). Much of the ICOADS data for recent decades derives from GTS data streams (Section 2.3). The final data source is periodic merger of delayed mode (DM) archives, e.g. from the JCOMM Global Collecting Centres (GCC) or from reprocessing of research archives (Section 2.4). Lastly, Section 2.5 describes how the ICOADS processing brings together data from all these sources.

2.1. Legacy data: historical archives

The source data for COADS Release 1 are detailed by Woodruff *et al.* (1987) and derived from ships' logbook observations digitized onto punched cards and magnetic tape, extracted from atlas compilations or transmitted using wireless telegraphy and later the GTS. Bilateral exchange agreements with many nations substantially extended the US marine archive (Woodruff *et al.*, 2003, 2005). Release 1 (Slutz *et al.*, 1985; Woodruff *et al.*, 1987) brought together observations from many different national and international climate archives into a single repository with a common format. Buoy data were also included, though amounting to less than 1% of the total.

The quality and completeness of observations in the present ICOADS are strongly related to the source of the observations, the formats used to store the data, and any conversions used. The deck (*DCK*) and source identifier (*SID*) fields provide information on the provenance of each report. For some historical decks, platform identifiers (*ID*s; e.g. ship names or callsigns) are missing; other decks did not include all data elements originally reported. In some cases undocumented data adjustments or format conversions may have been applied. Over time ICOADS has tried to replace early data sources compromised by historical storage in inadequate formats or poor curation (Woodruff *et al.*, 2003). Because the archives from different countries contained much common data, the identification of duplicate reports is vitally important (Section 2.5) and extremely challenging due to variable and often undocumented historical data management practices.

2.2. New historical data from recovery and digitization

ICOADS works with several historical data digitization efforts, including: RECovery of Logbooks And International Marine data (RECLAIM; Wilkinson *et al.*, 2011); The International Atmospheric Circulation Reconstructions over the Earth (ACRE) Initiative (Allan *et al.*, 2011); and 'oldWeather' (http://oldweather.org, accessed 1 January 2016). Records from data sparse regions or periods are given priority. Recently digitized data are typically as complete as feasible, including preservation of accompanying metadata, leading to higher quality.

Digitization projects have helped to shape and expand ICOADS since its inception. Small or large, each set of newly digitized data makes a valuable contribution. Large, multinational efforts such as the Climatological Database for the World's Oceans (CLIWOC) (García-Herrera *et al.*, 2005) provided data, concentrated during 1750–1850, when temporal and spatial coverage remain sparse. However, these data were largely non-instrumental, i.e. mostly once-a-day observations of winds (speed and direction) and weather. A new source of digitized data for R3.0, from English East India Company (EEIC) logbooks, improves coverage during this early period. These voyages were selected from the much larger quantity of non-instrumental 17th–19th century EEIC data also recorded (Farrington *et al.*, 1999), because they included instrumental observations of *AT* and pressure, with occasional reports of *SST*.

This transition from non-instrumental to instrumental observations foreshadows the landmark 1853 Brussels Maritime Conference (Maury, 1854), which established a globally coordinated effort to report in a standardized manner and provide as much data as possible.

2.3. Real-time data from the GTS

Observations circulated in NRT over the GTS in support of Numerical Weather Prediction (NWP) are an important constituent of ICOADS. GTS observations used include reports from ships in the WMO Voluntary Observing Ships (VOS) Scheme, from instruments on moored and drifting buoys, and from coastal or other fixed stations such as tide gauges. Access to GTS data is typically via a National Meteorological and Hydrological Service (NMHS). Section 3.3 describes new and improved data from coastal stations and tide gauge installations from the GTS available in R3.0.

Exactly which reports are retrieved from the GTS depends on how and when the system is interrogated. GTS archives extracted by different operational centres therefore have different contents. Each centre will take a different approach to combining reports that may have been split during transmission, and may apply adjustments to reports. Up through R2.5, ICOADS used GTS data receipts primarily from NOAA's National Centers for Environmental Prediction (NCEP; 1980–forward), and from US Navy (1966–1973) and Air Force (1973–1997) sources.

For R3.0, GTS receipts from the NOAA National Centers for Environmental Information (NCEI) are blended with those from NCEP, resulting in about 5% additional observations. Also, since late 2007, VOS ship identification information has been largely missing from the NCEP stream, hampering the assignment of metadata to the observations and ship-by-ship QC (Woodruff *et al.*, 2011). Adding the NCEI data stream has enabled unique ship identifier information (either callsigns or masked ship identifiers) to be recovered for about 70% of VOS observations. This new blended product now forms the NRT extension for ICOADS, normally published within five days of the end of the preceding calendar month.

2.4. DM data from international exchange and research archives

Data are received in DM from several global data centres, such as the JCOMM GCC for VOS data. DM observations are typically higher quality than the GTS data and may have additional QC, higher numerical precision or additional parameters. Most of the DM data from the GCC will replace data previously included in ICOADS as NRT data, and the duplicate elimination procedure (Section 2.5) identifies NRT reports to be replaced.

Surface or near-surface data extracted from a variety of updated research archives have been incorporated into R3.0, including (see also Section 3.2): near-surface observations selected from the World Ocean Database 2013 (WOD; Boyer *et al.*, 2013); data from the Global Ocean Surface Underway Data (GOSUD) Project; buoy measurements from the Global Tropical Moored Buoy Array (GTMBA; McPhaden *et al.*, 1998, 2009; Bourlès *et al.*, 2008) and from the Canadian Oceanography and Scientific Data (OSD) archive; and a subset of near-surface measurements from research vessels (RVs) in the Shipboard Automated Meteorological and Oceanographic System (SAMOS) initiative.

2.5. Combining the data streams to generate ICOADS

Each Release of ICOADS has incorporated new sources of data (Woodruff *et al.*, 1998, 2003, 2011; Worley *et al.*, 2005). Each new source must be translated into the IMMA format, which may require conversion from historical or non-standard units (e.g. knots or Beaufort Force to m s^{-1}, °F to °C) or between different representations of the same environmental parameter (e.g. from wet- and dry-bulb to dew point temperature, or wind components to wind speed and direction). At the same time, however, the original input data are preserved in the supplemental (*Suppl*) attachment to IMMA (see Section 4).

Following conversion to IMMA format (Section 4) the major task of assembling the final data set for public release remains. This processing is referred to generally as 'dupelim' (duplicate elimination) and includes a complex set of rules specially tailored for each release intended to identify duplicate and poor quality reports, which are not included in the official release. Each report is first processed through a series of QC steps, including the identification and flagging of observations that are statistical outliers compared to a climatology (known as 'trimming'; Wolter, 1997), followed by two separate preconditioning steps to: (1) exclude reports (e.g. known problematic or highly redundant data sources), and (2) to edit known erroneous fields, or compute some missing fields, e.g. dew point temperature, required for some monthly summary variables. Observational metadata from WMO (1955) Publication No. 47 are associated, where possible, with individual reports (Kent *et al.*, 2007) and used to populate the ship metadata (*Meta-vos*) attachment (see Section 5.3).

For R2.5, 'intermediate' (pre-dupelim) data sets were produced containing duplicates and other suspect reports (e.g. landlocked), flagged as to retention status, and were also made available. For R3.0, we have refined this approach to separately flag intermediate data *versus* 'rejects' of clearly erroneous, or otherwise problematic, data in a 'total' file (not provided to users except by special request). This approach ensures preservation of all observations, both unique and duplicates. The total data set is then processed to remove landlocked, duplicate, and other reject reports, to generate the final output data set.

The dupelim procedure is complicated as many of the decks have been reformatted and modified over the years (particularly the legacy data, Section 2.1) predating their more careful data management under ICOADS. This means that reports originating from the same observation may differ in terms of even basic information such as position or time, as well as the available parameter or metadata fields, or the values stored in these fields. Much of the complexity of dupelim therefore arises from the specification of the tolerances required to identify as many duplicates as possible without excluding similar but unique reports and imperfections in this process will impact the final release and any data products derived from it. The inclusion of any new data source can be a complex and time-consuming task, often revealing problems with the new data sources, or with data that have previously been ingested.

3. New data sources and resultant data coverage in Release 3.0

3.1. New historical data sources from digitization

The RECLAIM project identified and prioritized historical data archives for digitization, conversion to IMMA and inclusion in ICOADS (Wilkinson *et al.*, 2011). Some of these have now been digitized, through volunteer crowdsourcing initiatives (oldWeather and Weather Detective), by data scientists or through national digitization programs, e.g. the Deutscher Wetterdienst (DWD) KLIDADIGI and HISTOR projects (Kaspar *et al.*, 2015) and former NOAA Climate Database Modernization Program (CDMP). Each new source is listed in Table 2, with its temporal and spatial coverage also illustrated in Figure 1.

The new data sources provide valuable additional observations in the sparsely observed period before about 1880. The earliest observations, newly digitized by ACRE, are from voyages of exploration and early scientific expeditions (deck 246, Table 3, Figure 1(a)), often from sparsely observed regions. The selected logbooks from EEIC (deck 248, Figure 1(b)) covering 1789–1834 form one of the earliest systematic collections of marine observations, improving coverage in the Atlantic and Indian Oceans, and providing information about global events such as the Tambora eruption of April 1815 (Brohan *et al.*, 2012). Further increasing sampling in this early period are newly digitized observations from other early data sources including the German Maury Collection (1845–1868; Braun, 2000) digitized in 2006 by CDMP (deck 721, Figure 1(c)), and from merchant sailing ships (1868–1907) digitized by the DWD HISTOR project (deck 720, Figure 1(d), http://www.dwd.de/EN/ourservices/meteorological_logbooks/metschiffsjournale.html, accessed 1 January 2016).

Most digitization initiatives focus on particularly data sparse regions and periods. The very early data are always valuable, but the periods of the two World Wars also have sparse coverage and high latitude regions are poorly observed throughout the record. To address this, weather records were transcribed by the oldWeather crowdsourcing project (http://oldweather.org, accessed 1 January 2016) providing more than 185K new reports mostly covering the North Pacific and Arctic oceans; including observations from well-known ships such as the *USRC Bear* and the *USS Jeannette* (deck 710, Figure 1(e)). The Australian Broadcasting Corporation and the University of South Queensland Weather Detective digitization project (http://www.weatherdetective.net.au/, accessed 1 January 2016) provided more than 37K new reports from abstract logs collected by the Queensland climatologist Clement Wragge (deck 711, Figure 1(f)). Environment Canada provided data from the *Fram* providing unique information on the historical climate of the Canadian Arctic (deck 734, Figure 1(g)). A second oldWeather project saw more than 16K volunteer contributors digitize Royal Navy deck logs around WW1 (1912–1925, deck 249, Figure 1(h)).

More general digitization efforts have provided data from German light vessels (deck 720, Figure 1(i)), US Lightships (deck 703, Figure 1(j)), the US Navy (for the periods 1951–1964 and 2001–2012, deck 117, Figure 1(k)) and for the period 1968–1993 from the Chinese/Global Oceanographic Data Archaeology and Rescue (GODAR) project rich in reports from the Western Pacific and China Seas, but also including global voyages (deck 781, Figure 1(l)). The Australian Bureau of Meteorology (BoM) provided a digitized collection of Australian Navy *SST* observations from 1974 to 1977 (deck 750, Figure 1(m)) and lastly some recent US Navy reports were digitized by CDMP (deck 708, Figure 1(n)).

Overall these newly digitized reports contribute many observations, and provide valuable additional coverage in the early instrumental period and during two world wars.

3.2. Data from external archives

Marine observations are often collected and assembled into separate archives, especially for the modern period. Data from external archives form an essential ingredient of ICOADS. New reports are typically added to these archives, and QC'd, on a continuing basis, thus sometimes also improving data that have been utilized previously in ICOADS. For R3.0, ICOADS has acquired many new reports from these sources and have translated the various formats to the common IMMA format. These sources are listed in Table 4. Temporal and spatial coverage for each of these sources are illustrated in Figure 2.

The World Ocean Database (deck 780, Figure 2(a)) is a collection of quality controlled ocean profile and surface oceanographic data (Boyer *et al.*, 2013) beginning in the early 1800s and updated through 2014. New sources of data have been included in WOD 2013 from digitization projects, additional data from recent real-time (RT) and DM streams have been compared and duplicates removed, and standardized QC completed on all data.

Data from the GOSUD project (deck 782, Figure 2(c)), provided by NCEI, complement the WOD data with RT and DM near-surface observations of sea temperature and salinity from RVs and other ship types. The GOSUD archive is new to ICOADS, and both WOD 2013 and GOSUD provide oceanographic information made available in the new Near-Surface-Oceanographic attachment (*Nocn*; see Section 4.2.1).

While GOSUD archives just near-surface sea temperature and salinity, the SAMOS initiative (deck 740, Figure 2(e)) has been collecting a much broader suite of underway ocean and surface meteorological measurements from US and some international RVs. The SAMOS data undergo QC at FSU and both original and quality-processed observations are archived at NCEI (Smith *et al.*, 2009). Hourly 'super-obs' are created by averaging the data from the 10-min period prior to the top of the hour for inclusion in ICOADS. The choice to subset the SAMOS data was made so that the records in R3.0 are more consistent with manual VOS reports, which are typically observed within the 10 min prior to the reporting hour. R3.0 extends earlier FSU-processed World Ocean Circulation Experiment (WOCE) RV data in ICOADS

Table 2. Major new historical data sources blended into R3.0.

Source	Deck/Source ID	Contributor	Period	Region	Number of reports	Information
Expeditionary data from ACRE	*DCK* = 246 *SID* = 167-168	ACRE	1699–1930	Global	104K	Digitized from printed or manuscript forms. Wide variations in format mean that standardization of these data is extremely time consuming
English East India Company Logbooks	*DCK* = 248 *SID* = 148	British Library, CDMP, ACRE	1789–1834	Global	247K	Approx. 1K logbooks containing early instrumental data (*AT* and *SLP*) to give once-daily reports. Sub-daily wind and weather observations were imaged but not digitized
German Maury Collection	*DCK* = 721 *SID* = 152	DWD, CDMP	1845–1868	Global	539K	From a German collection of Maury's logbooks (Braun, 2000). Comparison with the similar US Maury Collection (Woodruff *et al.*, 2005) showed apparent biases in *SLP* observations (Wallbrink *et al.*, 2009)
DWD: HISTOR Merchant Sailing Ships	*DCK* = 720 *SID* = 160	DWD	1868–1907	Global	167K	Data digitized from 542 journals and quality checked with the DWD High Quality Checking Software (HQCS, Section 6)
US Arctic Logbooks	*DCK* = 710 *SID* = 166	oldWeather	1870–1946	Arctic and North Pacific	173K	Logbooks from the US National Archives, concentrating on ships travelling to the Arctic Ocean. Project is ongoing
Australian abstract logs (Wragge collection)	*DCK* = 711 *SID* = 171	Weather Detective	1889–1899	Pacific and Indian Oceans	37K	Abstract logs from the Queensland climatologist Clement Wragge
Second *Fram* Arctic Expedition	*DCK* = 734 *SID* = 128	Environment Canada	1898–1902	Arctic	1.5K	Data from the Second Norwegian Arctic expedition in the *Fram* 1898–1902 (Mohn, 1907), including observations made 3–6 times per day at sea
World War I (WW1) UK Royal Navy Logbooks	*DCK* = 249 *SID* = 165	oldWeather	1912–1925	Global	1.1M	Logbooks from the UK Archives collection (series ADM 53). 350K pages were transcribed
DWD: German Light Vessels	*DCK* = 720 *SID* = 161	DWD	1929–1988	North and Baltic Seas	934K	Data from 16 German light vessels in the North Sea and Baltic Sea. Quality checked using HQCS
US Lightships	*DCK* = 703 *SID* = 144	Woods Hole Oceanographic Institute (WHOI) and CDMP	1931–1980	Atlantic, US East Coast	203K	Reports from 14 US East Coast lightships (Flint, 1994; WHOI, 2002). A subset of the full US Lightship collection
US Navy Hourlies	*DCK* = 117 *SID* = 140–142	US Navy, NCEI	1951–1964	Global	2.9M (est. unique)	An extension of US Navy Marine data available in deck 110 (AWS-WB, 1964) but from card decks with a different original format (AWS-WB, 1958) (http://icoads.noaa.gov/deck117.html, accessed 1 January 2016)
Chinese/Global Oceanographic Data Archaeology and Rescue (GODAR) ships	*DCK* = 781 *SID* = 143	GODAR	1968–1993	Global	414K	Surface marine ship data digitized in China (http://icoads.noaa.gov/godar.html, accessed 1 January 2016)
Australian Navy SST	*DCK* = 750 *SID* = 156	Australian Bureau of Meteorology (BoM)	1974–1977	Global	4.4K	41 cruises comprising *SST* data collected, with insulated buckets and mercury-in-glass thermometers, by Australian Navy ships in a wide region surrounding Australia
US Navy METAR reports	*DCK* = 708 *SID* = 109	US Navy, NCEI, CDMP	2001–2012	Global	401K	DM US Navy Hourly non-classified Aerodrome meteorological reports recorded on paper forms. METAR format (FM 15–XV; WMO, 2011) reports are surface weather observations related to aviation e.g. on US Navy aircraft carriers

Sub-collections within a deck (*DCK*) are indicated by multiple Source ID (*SID*) listings. Images of the original logbooks from some of these collections are archived at NCEI and available, upon request, through the Environmental Document Access and Display System, Version 2 (EV2) application (https://www.ncdc.noaa.gov/EdadsV2/, accessed 1 January 2016).

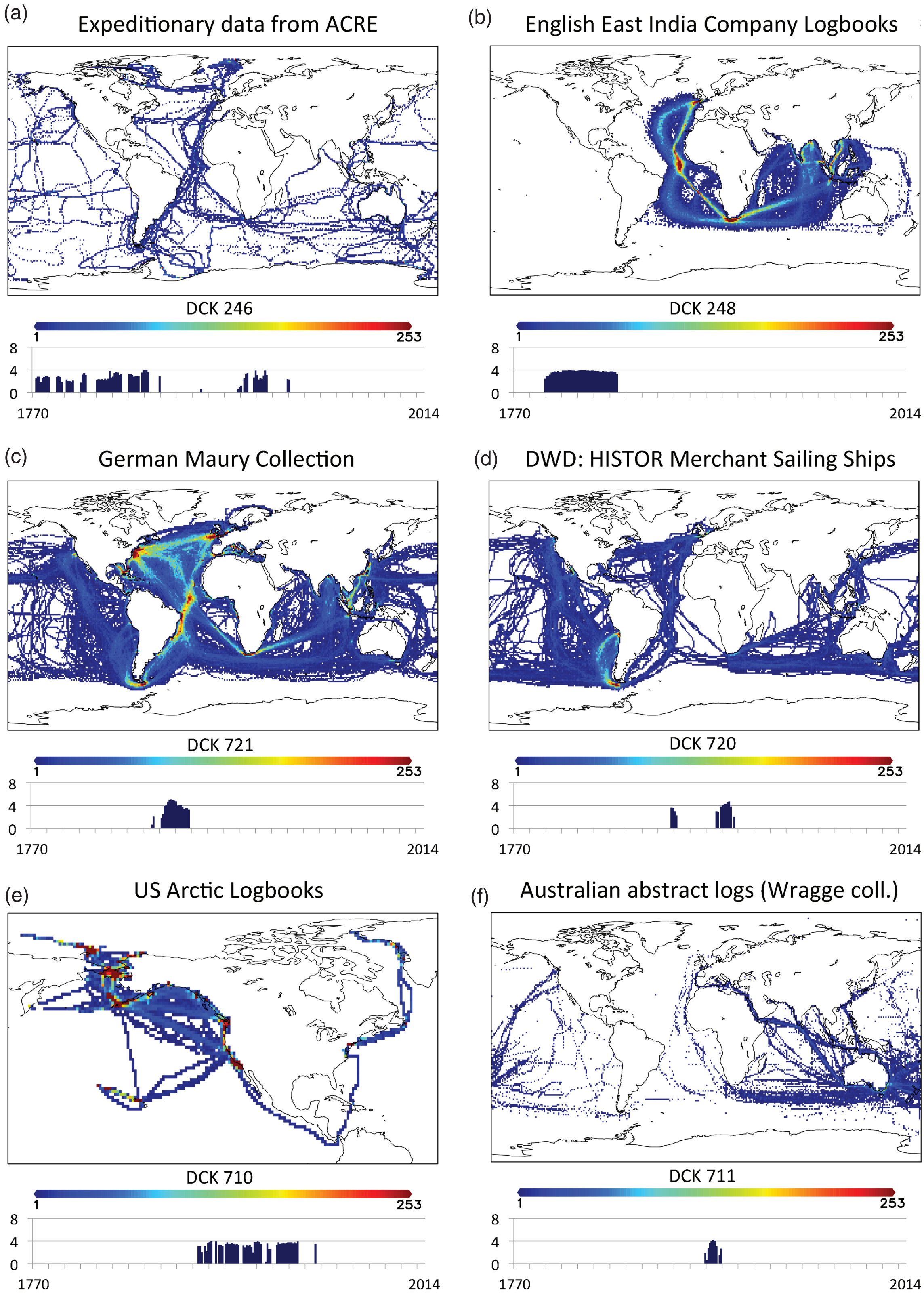

Figure 1. Coverage patterns for new sources of recovered and digitized data showing spatial density of total reports per 1° box (top) and time series showing temporal coverage by year (bottom) (and excluding "landlocked" reports using a 2° × 2° land-sea-mask; see also discussion in Section 6). The time-series vertical axis units are $\log_{10}$ and each tick mark on the horizontal axis represents a 10-year span. More information on these sources can be found in the Section 3 text, and in Tables 2 and 3. (a) Expeditionary data from ACRE (240 reports from 1699 to 1700 are not shown in the time series); (b) English East India Company Logbooks; (c) German Maury Collection; (d) DWD: HISTOR Merchant Sailing Ships; (e) US Arctic Logbooks; (f) Australian abstract logs (Wragge collection); (g) Second *Fram* Arctic Expedition; (h) World War I (WW1) UK Royal Navy Logbooks; (i) DWD: German Light Vessels; (j) US Lightships; (k) US Navy Hourlies; (l) Chinese/Global Oceanographic Data Archaeology and Rescue (GODAR) ships; (m) Australian Navy *SST*; and (n) US Navy METAR reports.

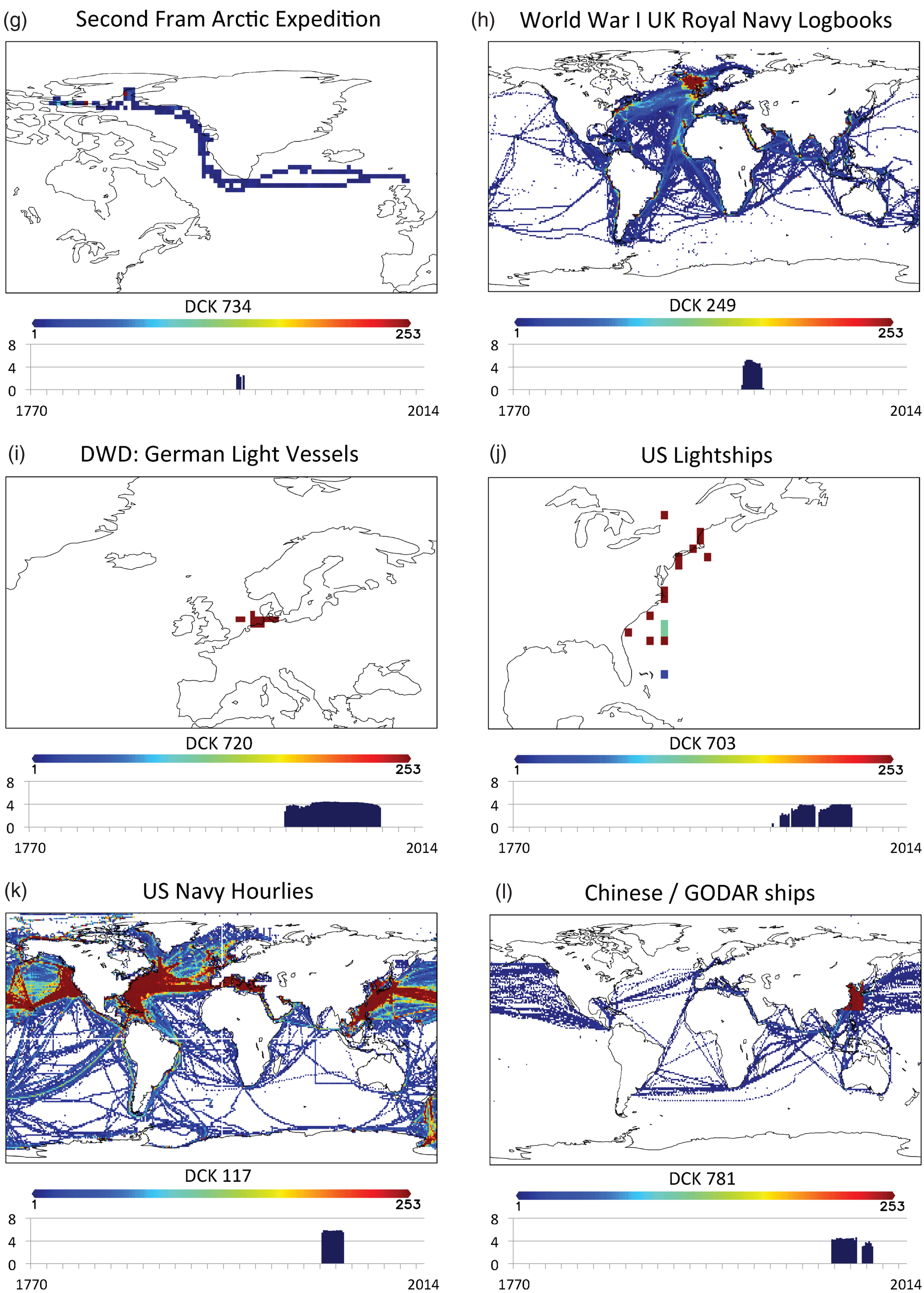

Figure 1. continued

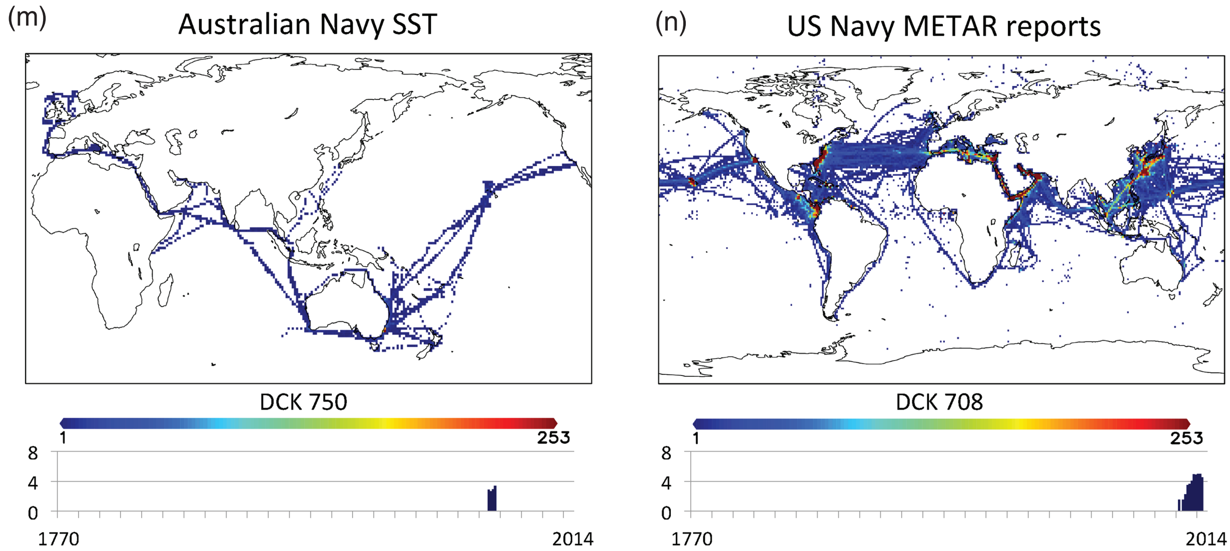

Figure 1. continued

Table 3. Details on specific data sets from ACRE that have been translated for R3.0.

Expedition (and) or commander	Voyage	Ship(s)	Period of record	Number of reports
Edmund Halley	South Atlantic	*Paramour*	1699	247
James Cook	Southern Ocean	*Resolution & Adventure*	1772–1775	401
James Cook	Pacific	*Discovery & Resolution*	1776–1779	2978
La Perouse	Circumnavigation	*Boussole*	1785–1788	857
First Fleet	England to Australia	*Sirius*	1787–1788	281
George Vancouver	Pacific	*Discovery*	1791–1795	686
William Scoresby	Greenland Sea	*Resolution & Esk*	1807–1822	2063
	England to Ceylon	*Prince of Orange*	1816	150
John Franklin	Greenland Sea	*Dorothea*	1818	3743
John Ross	Canadian Arctic	*Isabella*	1818	1670
William Parry	Canadian Arctic	*Fury & Hecla*	1819–1825	14 800
Cyril Laplace	South Pacific	*La Favorite*	1830–1832	5058
Robert Fitzroy	Circumnavigation	*Beagle*	1831–1836	3478
U.S. Exploring Expedition (Charles Wilkes)	Pacific Ocean	*Vincennes*	1838–1842	33 367
Berlin Emigration Society	Hamburg to Adelaide	*Princess Louise*	1849	688
William Penny	Bafffin-Bay and Barrow Straits	*Sophia*	1850–1851	4576
Leigh Smith	Arctic	*Sampson* and *Eclipse*	1872	37
Scottish National Antarctic Expedition (William Speirs Bruce)	Weddel Sea	*Scotia*	1902–1904	4734
International TransAntarctic Expedition (Ernest Shackleton)	Southern Ocean	*Endurance, Aurora, James Caird, Emma & Yelcho*	1914–1916	9200

with observations from 2005 to 2014, yielding approximately 752K new hourly SAMOS reports including population of the new *Nocn* attachment where possible.

The GTMBA is a multi-national effort comprising the Tropical Atmosphere Ocean Project (TAO)/Triangle Trans-Ocean Buoy Network (TRITON) array in the Pacific (McPhaden *et al.*, 1998), Prediction and Research Moored Array in the Atlantic (PIRATA) in the Atlantic (Bourlès *et al.*, 2008), and Research Moored Array for African–Asian–Australian Monsoon Analysis and

Table 4. Major new or improved data sources from external archives blended into ICOADS Release 3.0. Sub-collections within a deck (*DCK*) are indicated by multiple Source ID (*SID*) listings.

Source	Deck/Source ID	Contributor	Period	Region	Number of reports	Information
World Ocean Database (WOD) 2013	DCK = 780 *SID* = 149	NCEI	1770–2014	Global	14.3M*	A collection of quality-controlled ocean profile data, beginning in the 1800's, extending to the present day. Replaces WOD 2005 used in ICOADS R2.5
Global Tropical Moored Buoy Array (GTMBA)	*DCK* = 146 *SID* = 169, 170	NOAA/PMEL, UK NOC	1977–2014	Tropical	13.9M	Multi-national effort comprising the TAO/TRITON array in the Pacific, PIRATA in the Atlantic, and RAMA in the Indian Ocean; updated with new and reprocessed observations and observational metadata (McPhaden *et al.*, 1998; McPhaden *et al.*, 2009; Bourlès *et al.*, 2008)
Global Ocean Surface Underway Data (GOSUD)	*DCK* = 782 *SID* = 159, 162–164	NCEI	1980–2014	Global	2.6M*	IODE/IOC Program collecting near-surface temperature and salinity from underway research and commercial vessels
JCOMM Global Collecting Centres (GCC) archive	*DCK* = *926* *SID* = *112*	JCOMM Global Collecting Centres (GCC)	1982–2014	Global	14.2M	Global delayed mode Voluntary Observing Ship (VOS) data provided by the JCOMM Global Collecting Centres (GCC). Higher quality reports replace associated real-time (RT) GTS reports, where possible, with new data for 2008–2014
Shipboard Automated Meteorological and Oceanographic System (SAMOS)	*DCK* = 740 *SID* = 131	FSU COAPS	2005–2014	Global	750K	Underway marine observations from research vessels (Smith *et al.*, 2009) subsampled to hourly super-observations. Additional parameters often available in supplemental data (e.g. solar and longwave radiation)
Canadian DFO/OSD drifting (and moored) buoy data	*DCK* = 714 *SID* = 063	Environment Canada	New data for 2008–2014	Global	79.6M	Updated archive of QC'd drifting (plus some moored) buoy GTS receipts, including new data for 2008–2014

Numbers of input reports denoted by '*' are counts after reduction to hourly reports.

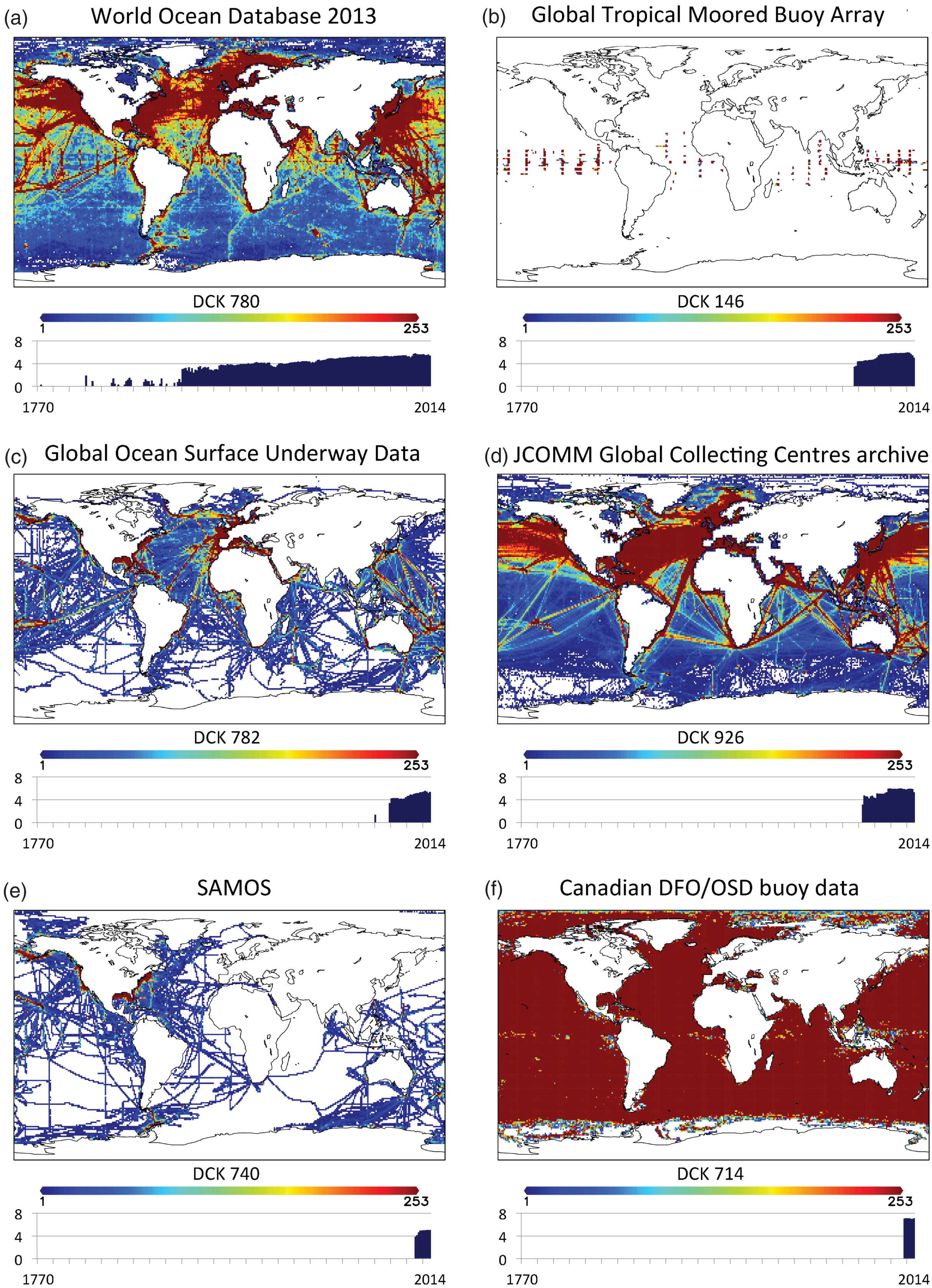

Figure 2. Coverage patterns for new sources of modern data showing spatial density of total reports per 1° box (top) and time series showing temporal coverage by year (bottom) (and excluding "landlocked" reports using a 2° × 2° land-sea-mask; see also discussion in Section 6). The time-series vertical axis units are $\log_{10}$ and each tick mark on the horizontal axis represents a 10-year span. More information on these sources can be found in the Section 3 text, and in Table 4: (a) WOD 2013; (b) GTMBA; (c) GOSUD; (d) JCOMM GCC archive; (e) SAMOS; and (f) Canadian DFO/OSD drifting (and moored) buoy data.

Prediction (RAMA) in the Indian Ocean (McPhaden *et al.*, 2009). ICOADS included some tropical buoy data since Release 1, and subsequent releases (Worley *et al.*, 2005; Woodruff *et al.*, 2011) incorporated DM archive data from NOAA's Pacific Marine Environmental Laboratory (PMEL). For R3.0, the GTMBA holdings have been refreshed again with PMEL data (deck 146; Figure 2(b)), including the population of the new *Nocn* attachment with *SST* and salinity observations and the resolution of problems in R2.5 with unadjusted pressure measurements (Section 5.1) and some erroneous wind directions (Section 5.2).

The Canadian OSD group within the Department of Fisheries and Oceans (DFO) – Canada, manages a continuously updated global archive of drifting buoy data. Acting as the Responsible National Oceanographic Data Centre (RNODC) for Drifting Buoys, OSD/DFO has periodically provided ICOADS with consolidated and QC'd buoy data (deck 714; Figure 2(f)), constructed from more frequent raw GTS reports (Ref http://icoads.noaa.gov/news_fig1.html, accessed 1 January 2016), with this latest extension for R3.0 covering 2008–2014.

The JCOMM GCC distribute global DM data collected on a quarterly basis from VOS paper forms and electronic logbooks (e-logbooks). While the number of Contributing Members (CMs) has decreased in recent years, currently around fifteen CMs, the GCC archive has continued to expand in recent years due to automated weather stations on VOS ships and the use of e-logbooks which help to facilitate the exchange of the archive data. The updated archive replaces many of the RT GTS reports providing additionally observed parameters and a standardized minimum QC applied to all GCC DM reports (deck 926, Figure 2(d)).

3.3. New data sources from the GTS

Coastal data sources also provide a wealth of information for R3.0. Meteorological observations from the US tide-gauge station network derived from the GTS are a new data source. The sea-level measurements are not included, but associated meteorological and *SST* observations have been added to ICOADS. A complication in processing these data arises because location information is not reported in RT. Instead, it has been taken from an historical station metadata archive constructed by NOAA's National Data Buoy Center (NDBC) (http://www.ndbc.noaa.gov/metadata/stationmetadata.xml, accessed 1 January 2016).

Since the early 1980s NDBC has been operating its Coastal-Marine Automated Network (C-MAN), close to 50 stations typically on lighthouses, beaches, near shore islands, and on offshore platforms. Position information for the C-MAN network is also available through the NDBC metadata.

3.4. Composition and data coverage of Release 3.0

Release 3.0 improved ICOADS by increasing the number of available marine reports, starting in the late 1700s, making significant gains into the modern period, and further extending the DM period of record from 2007 in R2.5 to 2014 in R3.0. New and updated data sources described in Section 3 were incorporated into R3.0 and are illustrated in Figure 3. There are notable gains from new digitizations, i.e. around 1855–1865 from the German Maury collection (deck 721), as well as around the World War 1 period, 1912–1925. Additionally, Figure 3 shows gains from reprocessed and updated archives, i.e. from the latest version of WOD (WOD 2013; deck 780) and updated GTMBA buoy archives (deck 146) from PMEL. GTS receipts from NCEI (decks 992–995) and NCEP (decks 792–795, 796) and RV reports from SAMOS (deck 131) also increased report counts, and spatial coverage, in the modern period.

Percentage of ocean coverage has increased with R3.0 (Figure 4). The EEIC collection (Figure 1(b)) provided significant gains in observed *AT* and *SLP* in R3.0 during the early 19th century, and may be the earliest systematic collection of instrumental marine observations in existence. EEIC has provided much needed early observations in the Atlantic and Indian Ocean basins with some noticeable increases also in the South Pacific. Gains in early time periods are critical for extending reanalysis products and gridded analyses farther back in time and may help to constrain proxy reconstructions of past climate (Brohan *et al.*, 2012).

There are also significant gains in coverage during the period 1853–1868, a nearly 5% increase in ocean area sampled in *SST*, *AT*, *SLP* and wind, contributed mostly from the German Maury collection (deck 721) and some from WOD (deck 780). The gains were most notable in the Eastern Pacific and Atlantic Oceans (Figures 1(c) and 2(a)). A nearly 2% increase in ocean coverage during WW1 is from the OldWeather UK Navy logbooks (deck 249, Figure 1(h)), with noteworthy gains in the North Atlantic Ocean.

Starting around the year 2000 there are notable increases in coverage in the modern period due in part to high-frequency reporting systems such as automated ship reports and moored and drifting buoy archives described in Section 3.2. Near-surface ocean temperatures from the global network of Argo profiling floats (plus from other oceanographic instruments), incorporated in R3.0 via WOD 2013, appear to account for a large fraction of the increases in SST coverage from 2005 to 2014. Small decreases in sampled ocean area are also observed in Figure 4. These decreases are associated with removal of suspected erroneous MORMET reports from 1958–74 (deck 732; see Section 5.2), and of erroneous AT values (retaining other elements of the report) from WOD 2013 (deck 780; SID 149).

R3.0 represents progress towards the reduction of uncertainty in products derived from ICOADS. Although digitization activities typically target data sources for which pressure observations are available, efforts are made to digitize the full multivariate record and retain the associated metadata. The benefits of this approach are clear in the early instrumental record, R2.5 contained very few observations of *AT* or *SLP* prior to the 1830s. Although coverage in R3.0 remains low in this period, at around 2%, these

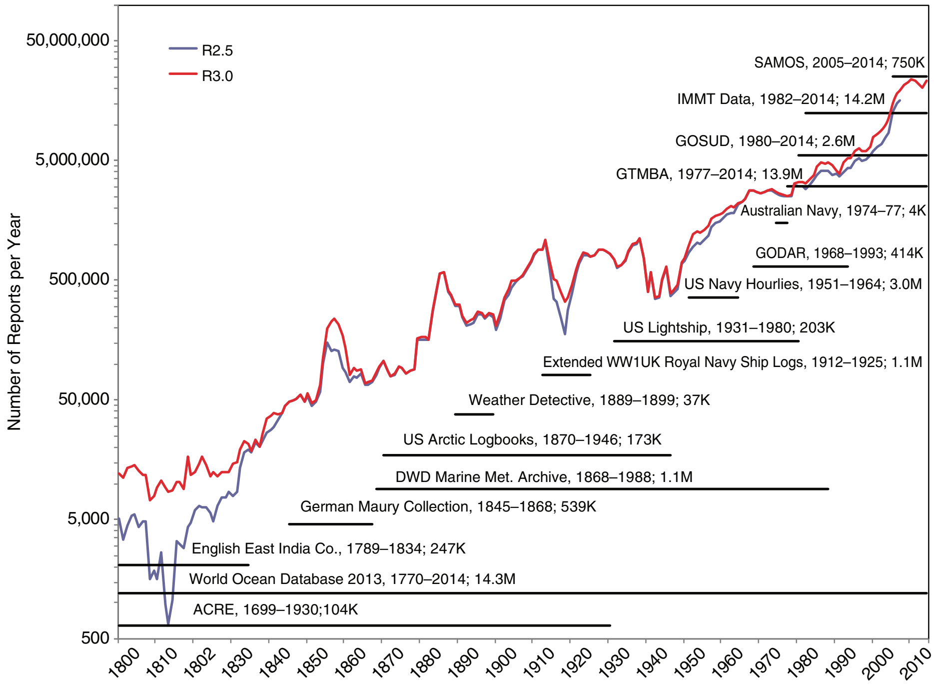

Figure 3. Major historical digitized and external archive data sources added to ICOADS Release 3.0, for 1800–2014. Horizontal black lines illustrate the time range of the original data sources. The annual numbers of reports are plotted as curves (logarithmic scale used on the vertical axis), blue for the previous R2.5, and red for R3.0. Data coverage prior to 1800 is sparse, and that following 2007 continues to grow annually.

new data should provide valuable input or comparison for reanalysis.

The estimation of air–sea fluxes from ICOADS is particularly sensitive to the available sampling and to uncertainties in the observations (Gulev *et al.*, 2007). R3.0 represents progress towards the reduction of uncertainty in products derived from ICOADS through improved coverage (Figure 4) and improved sampling (Figure 3). Figure 4 shows the coverage of reports containing total cloud cover (*N*) needed to estimate solar radiation, and in the evaporation parameter (*G*) shows the coverage for those reports containing variables needed to estimate either latent or sensible heat flux from mean parameters (Berry and Kent, 2009). There are noticeable increases in ocean coverage for clouds in the 1850s and 1860s and around 1900. The surface flux coverage illustrated by *G* is reduced by the lack of availability of humidity measurements in many reports with consistent coverage only from the 1910s.

Coverage has decreased for most variables since the late 1980s (Figure 4), the exception being *SST* (Berry and Kent, 2016; personal communication). The availability of higher quality and more complete observations retaining metadata, as is typical for R3.0 acquisitions, is a further important step towards producing higher quality products from ICOADS both for *SST* (Kennedy, 2014) and for other variables including estimates of air–sea exchange (Gulev *et al.*, 2007).

4. IMMA format updates

The IMMA format (Woodruff, 2007) is used to store and provide ICOADS observational data to users, and also to permanently archive the data and metadata in a technologically stable and readily exchanged form. The format is ASCII-based, containing a *Core* section including date, time, location, and identification information along with commonly reported meteorological variables and associated metadata. A simple fixed-length encoding scheme is used, with floating-point numerical values scaled to integers, e.g. 19.23 stored as 1923, and non-numeric data, such as QC flags and metadata, stored as 1- or 2-character coded values that are then decoded via look-up tables. Following the *Core*, a variety of attachments (attm) is available to hold additional data elements or metadata not represented in the *Core*. A new version of IMMA (IMMA1; Smith *et al.*, 2016) has been developed for R3.0. This introduces several new features, summarized in Table 5, and described in the following sections.

4.1. Unique identifiers and the *UID* attachment (*Uida*)

New to R3.0 and IMMA1 is the addition of the unique report-level identifier (*UID*) and *UID* attachment, i.e. the *Uida* attm. These features assist with record tracking and are assigned by ICOADS. The *UID*s are merely numbers and are not directly related to callsigns or other platform *ID*s. The *UID* allows information such as flags for erroneous values to be linked to each individual report and tracked for investigation and correction where needed. *UID*s also provide the mechanism to link feedback information from products created using ICOADS as input data and from ICOADS Value-Added Database (IVAD) submissions. For example, the *Rean-qc* attm (Section 4.2.5) will be populated with Reanalysis feedback through use of the *UID*s. *UID*s assigned to R2.5.1 and

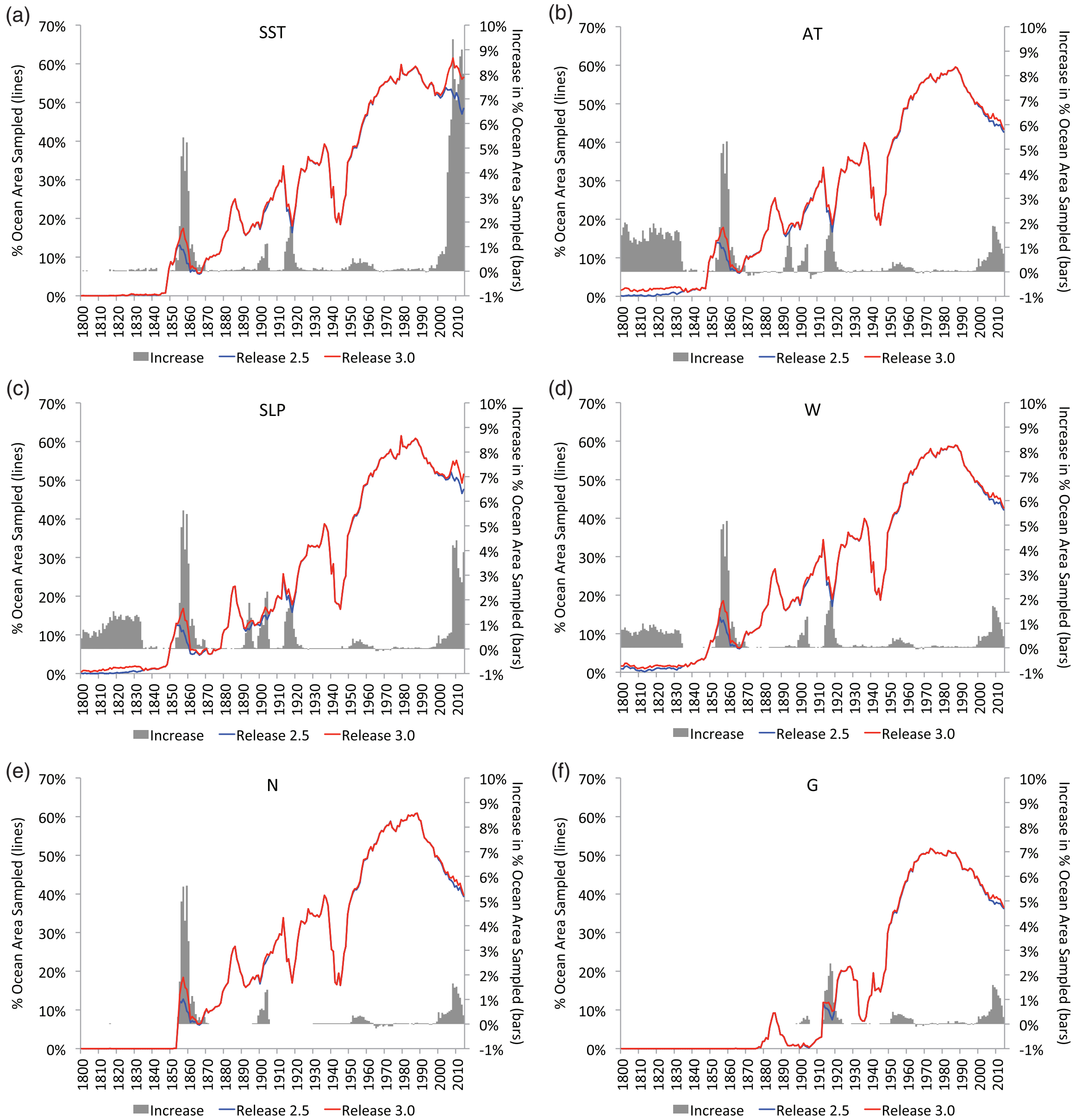

Figure 4. Annual percentage of global ocean and coastal area sampled for (a) *SST* for R3.0 (red curve) compared to R2.5 (blue curve, left axis; extended by NCEP NRT data for 2008–2014). Annual percentage increase (or decrease) in global ocean and coastal area sampled for R3.0, compared to the extended R2.5 (bars, right axis). Remaining panels: similarly for (b) *AT*, (c) *SLP*, (d) wind speed (*W*), (e) total cloud cover (*N*) – and (f) evaporation parameter (*G*), which is computed from *SST*, *SLP*, *AT*, *W*, and dew point temperature, and thus illustrates the extent to which surface fluxes can be computed from the individual observations.

R2.5.2 will be carried forward in R3.0 for continuity and provenance.

4.2. New data and metadata attachments (attms)

*4.2.1. Near-surface oceanographic data attm (*Nocn*)*

Past ICOADS releases have contained limited near-surface oceanographic data, most significantly *SST*. IMMA1 includes a Near-surface oceanographic (*Nocn*) attachment. *Nocn* can hold temperature, salinity, oxygen, phosphate, silicate, nitrate, pH, total chlorophyll, alkalinity, partial pressure of carbon dioxide, and dissolved inorganic carbon and associated sample depths, closest to the surface and less than 10 m. R3.0 contains *Nocn* populated from the WOD 2013, GOSUD, GTMBA, and SAMOS archives (Section 3.2).

*4.2.2. Edited cloud report attm (*Ecr*)*

Visual reports of the amount and type of cloud over the oceans contain biases and inconsistencies (e.g. Hahn *et al.*, 1995; Hahn and Warren, 2009; Eastman *et al.*, 2011). The full ship cloud report allows the reporting of cloud amounts for total and low cloud, cloud types for low, medium, and high clouds and information about the cloud base height. Using software provided by the University of Washington, the *Ecr* attm processing checks cloud and weather reports for internal consistency. If the report is found to be inconsistent it is corrected where possible,

Table 5. Summary of IMMA1 components and related information.

Component	Abbreviation	Reference information	Comment	Designation	Length (char.)
Core	*Core*	Table C0	Unchanged	Main	108
ICOADS attm	*Icoads*	Table C1	Unchanged	Main	65
IMMT-5/FM 13 attm	*Immt*	Table C5	Replaces C2	Main	94
Model QC attm	*Mod-qc*	Table C6	Replaces C3	Main	68
Ship metadata attm	*Meta-vos*	Table C7	Replaces C4	Main	58
Near-surface oceano. attm	*Nocn*	Table C8	New	Main	102
Edited cloud report attm	*Ecr*	Table C9	New	Subsidiary	32
Reanalyses QC/feedback attm	*Rean-qc*	Table C95	New	Subsidiary	61
IVAD attm	*Ivad*	Table C96	New	Subsidiary	53
Error attm	*Error*	Table C97	New	Subsidiary	32
Unique report ID attm	*Uida*	Table C98	New	Main, links to subsidiary	15
Supplemental data attm	*Suppl*	Table C99	Unchanged	Subsidiary	Variable

Components new to IMMA1 are in bold. Updated versions of deprecated tables from the initial version of IMMA, IMMA0 (Woodruff 2007), are noted in the comment column. The designation column indicates whether the component is treated as a part of the main ICOADS report or as a subsidiary record. Reference information links to Smith *et al.* (2016).

or else flagged as invalid. Where the report indicates that the sky is obscured, the reason for this is deduced where possible (e.g. at night with no moonlight, fog or heavy precipitation) or else the report is flagged as invalid. Codes for mid and high clouds are extended from the standard WMO codes to identify fog, cumulonimbus, altostratus or nimbostratus. An estimate of non-overlapping cloud cover is calculated and information on illuminance by the sun or moon is also included. The edited cloud reports are provided in the *Ecr* attm, and cloud parameters as reported remain in the IMMA *Core*.

4.2.3. *ICOADS-value added database attm (*Ivad*)*

The *Ivad* attachment in IMMA1 supports the linking of adjustments, uncertainty estimates, and alternate QC to individual ICOADS records (JCOMM, 2015). Adjusted data values are stored in the *Ivad* attm, while the original unadjusted data will remain in the original *Core* or other attm field locations. The *Ivad* attm supports up to three optional uncertainty values. An author reference code provides users with a link to documentation describing the adjustments.

Ivad attms are being developed by experts in the marine climate community and two prototype adjustments will be associated with R3.0 data shortly after the completion of the release: visually estimated (Beaufort) winds following Lindau (1995) and *AT* adjusted for ship heating following Berry *et al.* (2004). Multiple *Ivad* attms may be associated with a single ICOADS report (e.g. both a wind and *AT* adjustment) or even with a single field in a report (e.g. two different adjustments for *AT*). Through the *Ivad* attm, together with the dedicated database management system infrastructure at the US National Center for Atmospheric Research (NCAR), ICOADS has established the infrastructure in R3.0 to support future expert-developed information linked to ICOADS records.

4.2.4. *Error attm (*Error*)*

To track erroneous values in ICOADS, an *Error* attm has been created. *Error* attms are expected to be created by outside providers as a way to report and track erroneous values. For straightforward errors, e.g. VOS callsign errors, corrected values will replace the erroneous values in the *Core* or other attm field location. The associated erroneous, uncorrected value is stored in the *Error* attm and tracked by the *UID* in the *Uida* attm.

4.2.5. *Reanalysis QC/feedback attm (*Rean-qc*)*

Reanalysis data assimilation systematically evaluates the ingested observational data by comparing the model-driven first guess (approximation) to the incoming *in situ* observations. These comparisons along with pre-ingest data bias adjustments and model actions taken (e.g. rejection of the observation) are known as feedback records. The IMMA1 format includes a *Rean-qc* attachment that identifies if a particular data value was used or rejected, has numerical fields for the model-collocated first guess and analysis values, the bias corrected observational value, and fields that capture the analysis project, data provider, and a code that points to a reference document describing the reanalysis effort.

Implementation of the *Rean-qc* feature in R3.0 is similar to that for the *Ivad* and *Error* attms (Sections 4.2.3–4.2.4). *Rean-qc* records are received from the reanalysis centre and added to the matching ICOADS report, linked via the *UID*. Feedback data from the European Centre for Medium Range Weather Forecasts (ECMWF) Atmospheric Reanalysis of the 20th century (ERA-20C) are expected to be the first *Rean-qc* contributions to ICOADS (Poli *et al.*, 2015) and will become available shortly after completion of R3.0.

4.3. Switch to multi-record ‘linked-report’ approach with subsidiary records

In previous releases and the initial version of the IMMA format, IMMA0 (Woodruff, 2007), each report consisted of a single, variable-length physical record (i.e. line) comprising the *Core* data and zero or more attachments. However, in response to IVAD (JCOMM, 2015) requirements, plus increasing needs to more systematically interchange data and metadata with the user community, IMMA1 (Smith *et al.*, 2016) has been modified to allow multiple records per report, i.e. each such 'linked report' consists of a Main record and optional Subsidiary records all linked by *UID* (see also Table 5), e.g.:

Main IMMA record: *Core* + *Icoads* + *Immt*
+*Mod-qc* + *Meta-vos* + *Nocn* + *Uida* + *Suppl*
Subsidiary
IMMA record: *Uida* + *Rean-qc*
+*Rean-qc* + *Rean-qc* … + *Rean-qc*
Subsidiary
IMMA record: *Uida*
+*Ivad* + *Ivad* + *Ivad* … + *Ivad*
Subsidiary
IMMA record: *Uida* + *Error*
+*Error* + *Error* … + *Error*

This implementation is backward compatible and allows information to be provided in a file containing only Subsidiary records, allowing users to help enhance future versions of ICOADS.

5. Data issues identified

5.1. *SLP* biases

Early *SLP* values from the US Maury Collection, and possibly also in the newly included German Maury Collection, appear to be biased compared to other sources (Wallbrink *et al.*, 2009). While the reasons for these biases remain unclear, for R3.0 differently adjusted (i.e. for temperature and/or for gravity) versions of *SLP* for the German Maury Collection have been generated and will be offered in IMMA format as separate 'Auxiliary' data sets, to allow those interested to examine the adjustment effects.

For R3.0, some additional *SLP* biases have been identified, but only partially addressed, arising from atmospheric pressures reported at the height of the barometer, rather than adjusted to sea level. Inclusion of station pressure data in ICOADS is problematic since the IMMA format presently allocates space only for *SLP*, but station pressures from some data collections have nevertheless been stored in the *SLP* field, e.g. for some RV and moored buoy collections. The additional complication exists that as yet adjustment practices for VOS (and other platform types) have not been fully standardized by WMO (e.g. some RVs also report a value adjusted to sea level, but no standard practice has been followed).

While contemporary VOS data at least should all be adjusted to *SLP*, the remaining scope of possible biases is not fully understood – but improvements have been made for R3.0. In the case of SAMOS RV data, for example, pressures at barometer height were adjusted to sea level, provided sufficient metadata were available. Similarly, discussions with PMEL indicated that the pressure reports from the GTMBA were at barometer height, rather than adjusted to *SLP*. For the new GTMBA archive (deck 146; Section 3.2), the adjustment was made following WMO No. 8 (WMO, 2010) using known barometer heights. (Only very limited amounts of GTMBA data have been carried over from previous ICOADS releases, and while increasing those pressure values by 0.3 hPa as an approximate adjustment appears desirable, this did not prove feasible for R3.0.)

5.2. Identification of problematic data sources

Following completion of R2.5, it was discovered that reports from the US e-logbook software package, Shipboard Environmental Acquisition System (SEAS) versions prior to v9.1 were erroneous: there were unexplained differences between RT GTS and DM observations. These SEAS data are considered unreliable and hence removed from R3.0.

Kennedy *et al.* (2011b) and Minobe and Maeda (2005) documented an approach of excluding data from deck 732, the Russian Marine Meteorological Data Set (MORMET, 7.9M reports output from R2.5). All data reported in prescribed regions and periods in Kennedy *et al.* (2011b) were excluded from R3.0 (339K total excluded) as many were found to be mislocated. The excluded reports are retained in the total file, and approximately 7.83M reports remain in the final R3.0 data set.

For R2.5, data collected by the Institut für Meereskunde at the University of Kiel were provided by Canada's OSD and included as ICOADS deck 715. Until 1990 typically the only available parameter was ocean current (Krauss, 1996), stored in the *Suppl* attm. From 1990, many reports also contained *SST* but these contained large amounts of constant or other suspicious values. For R3.0, it was decided to exclude all deck 715 data from the final data set, but retain the reports in the total file.

Comparisons of past ICOADS holdings of tropical data from PMEL (decks 143–145) with the new GTMBA archive (deck 146) made it apparent that daily average deck 145 wind directions in previous ICOADS releases were in the oceanographic convention (the direction *to* which the wind blows) rather than the ICOADS standard meteorological convention (the direction *from* which the wind blows). This was resolved for R3.0 by removal of all ICOADS reports from deck 145 (1979–1991). While these tropical Pacific data from buoys and some 'flat' island stations remain in the total file, their daily average nature also renders them inhomogeneous with respect to most other contemporary ICOADS data. The older versions of GTMBA data, in contrast, were largely replaced in R3.0 with higher temporal resolution and quality data from deck 146 (Section 3.2).

5.3. Other data considerations

Since late 2007, many VOS began masking their callsigns – either to the anonymous 'SHIP' or to JCOMM approved unique 'mask' identifiers (generally designated on a national basis) – for security and commercial reasons

Table 6. ID patterns in NCEP GTS and NCEI GTS input and final blended output for January 2013.

	REAL (%)	MASK (%)	MASKSTID (%)	SHIP (%)	Unknown (%)
NCEP	1.0	0.2	98.7	0.1	0.0
NCDC	61.9	22.8	0.0	15.1	0.2
Blend	70.1	26.1	1.7	2.0	0.2

Totals are lower after blending due to removal of previously undetected duplicates of Canadian VOS.

(see also Section 2.3). For recent GTS data, R2.5 relied primarily on the NCEP GTS stream in which most callsign information was absent. In contrast, R3.0 also ingests the NCEI GTS stream, which does contain many genuine callsigns and the other uniquely assigned mask identifiers (Table 6). Moreover, inclusion of DM data from the GCC adds additional unique ship identifier information.

For R3.0, UK NOC has again joined WMO Pub. 47 VOS platform/instrument metadata with VOS reports, provided the real callsigns were reported, as was done for R2.5. For future releases, in contrast, NCEI is planning to operationalize this task, which should facilitate access (by NOAA as an NMHS) to securely held JCOMM lists of mask-*versus*-real callsigns. This will permit the association of Pub. 47 metadata with reports containing the unique mask identifiers, further enhancing availability of the *Meta-vos* attm.

Thus far, NCEP has not decoded reports of sea ice (concentration, stage of development, etc.) from the GTS SHIP code, but any originally reported GTS fields should be available in the *Suppl* attm, since NCEP prudently preserves, in their BUFR format, the original SHIP message(s) used to construct each BUFR report. This remains the case for R3.0. However, ice elements in the NCEI GTS are being translated, and to some extent have replaced the near-duplicate NCEP reports missing the ice elements, e.g. during the NRT merging process (Section 2.3).

In translating WOD 2013 into IMMA format, it was recognized that wind direction measurements were in a mixture of oceanographic and meteorological conventions (similar to the situation for some tropical Pacific buoy/island data, as discussed in Section 5.2). Because of this complication and the inability to easily distinguish which convention was used, plus because these wind data generally become available for ICOADS via other sources, all winds from WOD 2013 have been removed from R3.0. Future efforts will be needed to discern the wind direction measurement methods in WOD in order to be able to utilize them fully in future ICOADS releases.

6. Priorities for future development of ICOADS

The trimming approach to climatological QC used through R2.5 was developed for Release 1 (Slutz *et al.*, 1985) and has known limitations (Wolter, 1997). For R3.0, the focus on recovering observations in data sparse regions meant that many new observations were in regions previously unobserved where the QC medians and ranges had not been defined (see Figure 5). It was also apparent that the ranges used were underestimated when observations were sparse, further adding to the number of observations that may be erroneously removed by trimming. An important further QC-related complication is that ICOADS still mainly uses a coarse $2° \times 2°$ land-sea-mask for screening, and thus some of newly rescued historical expeditionary data, e.g. which traverse remote near-coastal regions, will be difficult to access for most users (note: both the $2° \times 2°$ mask, and a $1° \times 1°$ mask used to a more limited degree, are described on: http://icoads.noaa.gov/mask.html, accessed 17 March 2016).

While all data, including untrimmed observations and marine reports flagged as over land, are available in the total IMMA dataset, the final user product omits such coarsely landlocked data. Moreover, most users are presently encouraged to use an ‘enhanced’ (rather than ‘untrimmed’) IMMA product, and thus do not have ready access to observations over areas without trimming limits, or that have been trimmed incorrectly. In addition, downstream users of the enhanced (and other) monthly summary products are also heavily impacted by the results of the outdated trimming, land-sea-mask, and other legacy QC processing.

As an urgent priority following R3.0, the currently used trimming limits are therefore being renovated through a combination of more extensive gap-filling and extrapolation with the application of a lower limit for the range based on estimates of the random uncertainty in ship observations from Kent and Berry (2005). Other improvements to QC are expected to incorporate checks developed by DWD’s Higher-level Quality Control Software (HQCS), including improvements to the land-sea-mask, tracking of reports from individual ships, and climatological and spatial consistency checks.

The inclusion of new historical data sources remains a priority, and ICOADS will continue to work with international digitization initiatives and archives to rescue and translate newly available data and metadata into the IMMA format. Many documents remain in national archives and personal collections that have yet to be inventoried and will provide important new data to further flesh out weak or non-existent areas of temporal/spatial coverage in ICOADS. While it is a significant task to complete a project from paper (or digital images) to keyed records, and then ultimately to translate and blend these records into ICOADS – this critical work needs to continue for the foreseeable future. The vast bulk of extant ship observations reported prior to the 1853 Brussels Conference probably are non-instrumental – like CLIWOC data (García-Herrera *et al.*, 2005) – but such early *in situ* observations are nevertheless of tremendous value for climate research, including bridging the overlap with paleoclimatic data.

Additionally, steps will be taken to reprocess original ICOADS data to improve completeness and data quality, particularly of legacy data sources. ICOADS will continue to work with providers of high quality DM data, including WOD, PMEL, SAMOS, and the GCC. Other

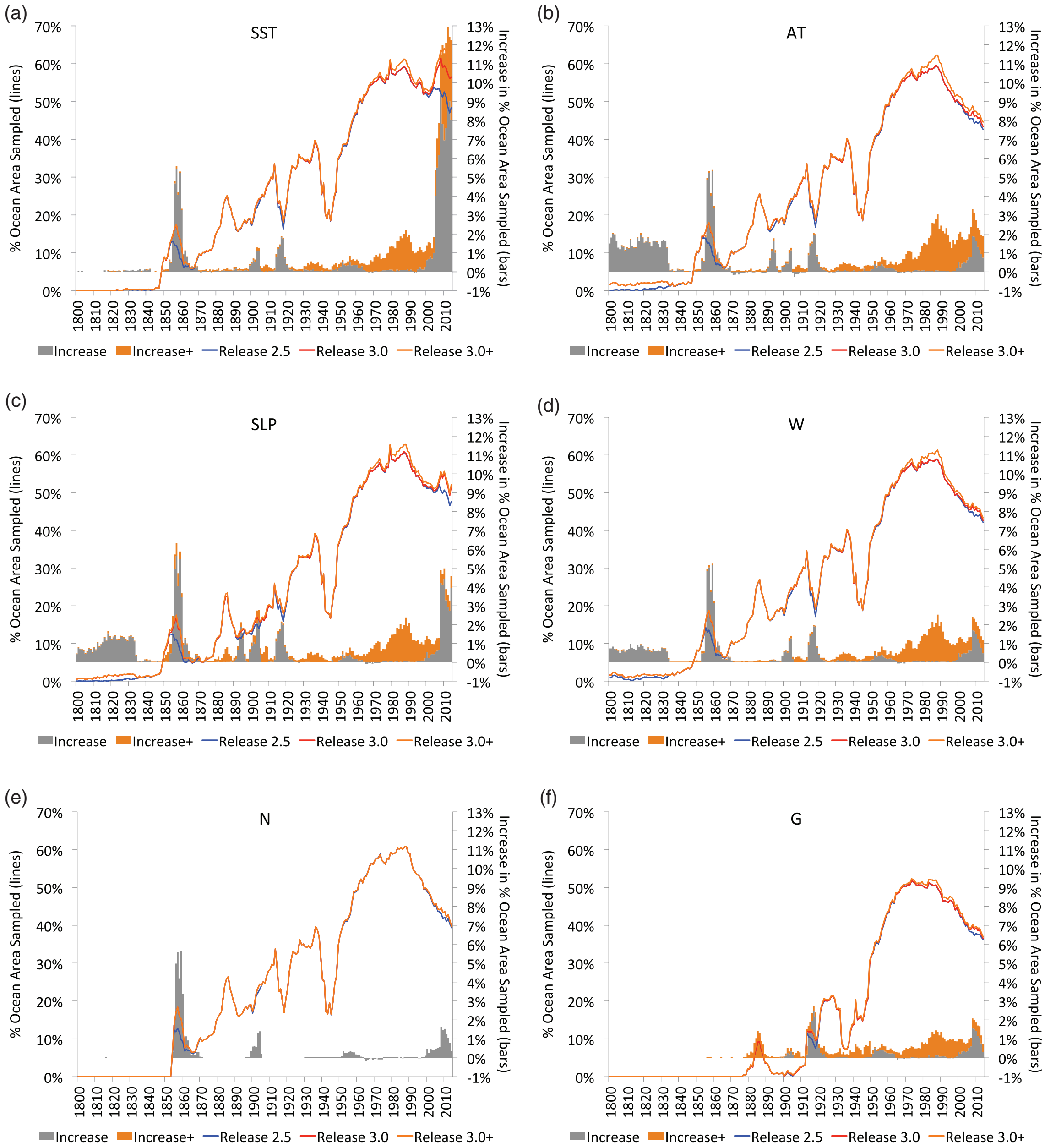

Figure 5. As for Figure 4, except the Figure 4 results for each variable are underlaid with the annual potential percentage of global ocean and coastal area sampled for R3.0 (orange curve, left axis), and with the annual potential percentage increase in global ocean and coastal area sampled (orange bars). In both cases, the potential spatial coverage improvements arise from allowing unfiltered data into ocean/coastal boxes with trimming limits missing, and it should be emphasized, e.g. for early GTS receipts (~1966–1979), that sizable apparent augmentations in coverage in some cases may arise merely from noisy, spatially mislocated data.

new sources of data will be investigated, for example surface meteorological data holdings under the Global Sea Level Observing System (GLOSS) and the Permanent Service for Mean Sea Level (PSMSL) (e.g. Holgate *et al.*, 2013). Inclusion of additional GTS sources will be explored. More data and metadata, and better QC are keys to improving the data set to allow climate research to improve and answer today's most pressing climate questions, to provide improved marine data resources for climate assessments and to permit close scrutiny of recent environmental changes (e.g. Karl *et al.*, 2015).

Platform tracking to ensure consistency of position and time information for observations with IDs and to cluster together observations without IDs likely to have been made on the same platform has been implemented for ICOADS (Carella *et al.*, 2015). Integrating tracking into ICOADS processing is valuable for QC, data set development and reanalysis.

Reprocessing and updating legacy data sources and documentation is another priority. Reprocessing will enable recovery of parameters and metadata fields not transcribed in previous ICOADS releases. A further benefit would be the generation of an improved intermediate (pre-dupelim) file containing duplicates for data sets integrated in ICOADS prior to R2.5. This would allow a more detailed comparison of differences between data sources than currently possible. Reprocessing will enable a reassessment of the dupelim procedure and is expected to improve the quality of the legacy data.

Additional advances in operationalization post R3.0 should benefit climate services in support of societally important monitoring activities and other NRT applications, e.g. through: (1) blending additional international GTS sources, to achieve even more complete coverage, robust to communications and computer failures; (2) more timely integration and utilization of NWP feedback information (e.g. as provided by the Met Office GTS) on data quality; (3) transitioning from monthly to more frequent (e.g. daily) updates (in support of satellite *SST* work, etc.); and (4) expanding capture of GTS and other frequently updated data streams, to include near-surface oceanographic parameters and NRT population of the *Rean-qc* attm. In support of improved *SST* analyses, e.g. the extensive coverage of Argo profiling floats, which have been included in R3.0 only in DM via WOD 2013, appears to be a major factor in enhancements in *SST* coverage since about 2005 as shown in Figure 4.

Additional format improvements and extensions to the IMMA format will consider ways to incorporate and provide higher resolution data as well as more platform and instrument metadata. Other potential format enhancements include: platform tracking information; better preservation of original input values for historical data, such as Beaufort wind force values; linking of original logbook images to individual reports using *UID*s and the more complete capture of meteorological and underway ocean data from automatic systems.

ICOADS will explore the potential for extending the monthly summary statistics to include wave and swell summaries (recommended by JCOMM, 2011). Providing ICOADS in other formats, e.g. netCDF, will expand the user base and possibly enable ICOADS to be served alongside observations in WOD 2013.

To promote further international formalization, there are also plans to initiate ICOADS as a WMO-IOC Centre for Marine-Meteorological and Oceanographic Climate Data (CMOC) under the new Marine Climate Data System (MCDS) (WMO, 2012; International Oceanographic Data and Information Exchange of the IOC (IODE), 2013).

7. Concluding remarks

R3.0 establishes a framework for ICOADS, its partnership and widening user community, to collaborate to further improve the surface marine climatological record. *UID*s, as part of the extended archive data format, allow the association of almost any type of information with an individual report. Further format extensions provide convenient access to a wider range of oceanographic parameters, alongside the meteorological variables and *SST*, expanding the potential ICOADS user-base.

ICOADS R3.0 individual observations and monthly summary products are available to users, from multiple access points at partner organizations, the NCAR, NCEI, and NOAA's Earth System Research Laboratory (ESRL), each with slightly different options designed to serve different user groups. The ICOADS products web page (http://icoads.noaa.gov/products.html, accessed 1 January 2016) links to full information and data distribution websites.

NCAR is currently providing assignment and management of data set digital object identifiers (DOIs) to ICOADS that will promote its official and standard citation in publications and enable the capability to more accurately measure its impact. A new DOI has been minted for the ICOADS R3.0 IMMA data (doi: 10.5065/D6ZS2TR3), and for the monthly summary products (doi: 10.5065/D6V40SFD).

Building on an over 30-year record of innovations in data stewardship and user access services, R3.0 opens up to the general public an even richer historical marine surface record – temporally, spatially, and of broadened scope in data and metadata offered – to a wide range of science, academic and commercial applications.

Acknowledgements

We are grateful to all the people and organizations that have provided new observations, or data improvements, to be included in ICOADS R3.0 (PMEL, WOD, OSD, GOSUD, NDBC, SAMOS, GCC, ICOADS International Partners), digitization projects, such as ACRE and GODAR, including the thousands of volunteers participating in the oldWeather and Weather Detective projects, those running and delivering operational observing programs, and the observers making the observations. We are also grateful for our users and their continuous feedback over the years, which have considerably enhanced ICOADS and continue to improve future releases. We would also like to thank the reviewers and Editors at the International Journal of Climatology for their assistance in proofreading and suggestions that helped to shape this article. SDW, SJL and this study were funded in part by the Climate Observation Division, Climate Program Office, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. SJW and ZJ were supported by the US National Science Foundation through core funding at the National Center for Atmospheric Research. PB was supported by the Joint DECC and Defra Integrated Climate Programme, DECC/Defra (GA01101). EK was supported by NERC through the NOC National Capability Program and DB was supported by NERC through grant NE/J020788/1. SS was also supported by the Climate Observation Division of NOAA's Climate Program Office and by the U.S. National Science Foundation's Oceanographic Instrumentation and Technical Services Program (grant OCE-1447797).

References

Allan R, Ansell T. 2006. A new globally complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2004. *J. Clim.* **19**: 5816–5842, doi: 10.1175/JCLI3937.1.

Allan R, Brohan P, Compo GP, Stone R, Luterbacher J, Brönnimann S. 2011. The international atmospheric circulation reconstructions over the earth (ACRE) initiative. *Bull. Am. Meteorol. Soc.* **92**: 1421–1425, doi: 10.1175/2011BAMS3218.1.

Bentamy A, Grodsky SA, Katsaros K, Mestas-Nunez AM, Blanke B, Desbiolles F. 2013. Improvement in air–sea flux estimates derived from satellite observations. *Int J Remote Sens* **34**: 5243–5526, doi: 10.1080/01431161.2013.787502.

Berry DI, Kent EC. 2009. A new air–sea interaction gridded data set from ICOADS with uncertainty estimates. *Bull. Am. Meteorol. Soc.* **90**: 645–656, doi: 10.1175/2008BAMS2639.1.

Berry DI, Kent EC. 2011. Air–sea fluxes from ICOADS: the construction of a new gridded data set with uncertainty estimates. *Int. J. Climatol.* **31**: 987–1001, doi: 10.1002/joc.2059.

Berry DI, Kent EC, Taylor PK. 2004. An analytical model of heating errors in marine air temperatures from ships. *J. Atmos. Ocean. Technol.* **21**: 1198–1215, doi: 10.1175%2F1520-0426(2004)021%3C1198:AAMOHE%3E2.0.CO;2.

Blunden J, Arndt DS (eds). 2014. State of the climate in 2013. *Bull. Am. Meteorol. Soc.* **95**(7): S1–S257, doi: 10.1175/2014BAMSStateofthe-Climate.1.

Bojinski S, Verstraete M, Peterson TC, Richter C, Simmons A, Zemp M. 2014. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **95**: 1431–1443, doi: 10.1175/BAMS-D-13-00047.1.

Bourlès B, Lumpkin R, McPhaden MJ, Hernandez F, Nobre P, Campos E, Yu L, Planton S, Busalacchi AJ, Moura AD, Servain J, Trotte J. 2008. The PIRATA program: history, accomplishments, and future directions. *Bull. Am. Meteorol. Soc.* **89**(8): 1111–1125, doi: 10.1175/2008BAMS2462.1.

Boyer TP, Antonov JI, Baranova OK, Coleman C, Garcia HE, Grodsky A, Johnson DR, Locarnini RA, Mishonov AV, O'Brien TD, Paver CR, Reagan JR, Seidov D, Smolyar IV, Zweng MM. 2013. World Ocean database 2013. In *NOAA Atlas NESDIS 72*, Levitus S (ed.) and Mishonov A (technical ed.), 209 pp. doi: 10.7289/V5NZ85MT.

Braun DS. 2000. Scientific vision, a passion for observation, and the impetus for change: Germany loans Maury logs to the National Climatic Data Center. *Earth Syst. Monit.* **11**(1): 4–7.

Brohan P, Allan R, Freeman E, Wheeler D, Wilkinson C, Williamson F. 2012. Constraining the temperature history of the past millennium using early instrumental observations. *Clim. Past* **8**: 1551–1563, doi: 10.5194/cp-8-1551-2012.

Carella G, Kent EC, Berry DI. 2015. A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.* , doi: 10.1002/joc.4492.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk M, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**: 1–28, doi: 10.1002/qj.776.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597, doi: 10.1002/qj.828 DOI:doi:%C2%A010.1002/qj.828.

Eastman R, Warren SG, Hahn CJ. 2011. Variations in cloud cover and cloud types over the ocean from surface observations, 1954–2008. *J. Clim.* **24**: 5914–5934, doi: 10.1175/2011JCLI3972.1.

Farrington AJ, Lubker S, Radok U, Woodruff S. 1999. South Atlantic winds and weather during and following the Little Ice Age – a pilot study of English East India Company (EEIC) ship logs. In *Proceedings of the International Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata*, Toledo, Spain, 15–17 September 1997. Diaz HF, Woodruff SD (eds). WMO/TD-No.957, MMROA Report No. 43. World Meteorological Organization: Geneva, Switzerland, 69–72.

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M. 2013. Evaluation of Climate Models. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker T, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK and New York, NY, 741–866, doi: 10.1017/CBO9781107415324.020.

Flint W. 1994. Lightships: Sentinels of the sea-lanes-Part-1. *Mar. Wea. Log*. **38**(4): 10–15. http://www.uscg.mil/history/articles/Lightships.pdf (accessed 3 May 2016).

García-Herrera R, Können GP, Wheeler DA, Prieto MR, Jones PD, Koek FB. 2005. CLIWOC: a climatological database for the World's Oceans 1750–1854. *Clim. Change* **73**: 1–12, doi: 10.1007/s10584-005-6952-6.

GCOS. 2010. Implementation plan for the global observing system for climate in support of the UNFCCC. GCOS-138, World Meteorological Organization, Geneva, Switzerland. http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf (accessed 24 March 2016).

Gulev SK, Grigorieva V. 2006. Variability of the winter wind waves and swell in the North Atlantic and North Pacific as revealed by the voluntary observing ship data. *J. Clim.* **19**: 5667–5685, doi: 10.1175/JCLI3936.1.

Gulev SK, Grigorieva V, Sterl A, Woolf D. 2003. Assessment of the reliability of wave observations from voluntary observing ships: insights from the validation of a global wind wave climatology based on voluntary observing ship data. *J. Geophys. Res. Oceans* **108**(C7): 3236, doi: 10.1029/2002JC001437.

Gulev SK, Jung T, Ruprecht E. 2007. Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part I: uncertainties in climate means. *J. Clim.* **20**: 279–301, doi: 10.1175/JCLI4010.1.

Hahn CJ, Warren SG. 2009. Extended edited cloud reports from ships and land stations over the globe, 1952–1996 (2009 update). Numerical Data Package NDP-026C, Carbon Dioxide Information Analysis Center, Oak Ridge, TN.

Hahn CJ, Warren SG, London J. 1995. The effect of moonlight on observation of cloud cover at night, and application to cloud climatology. *J. Clim.* **8**: 1429–1446, doi: 10.1175/1520-0442(1995)008 < 1429:TEOMOO > 2.0.CO;2.

Hartmann DL, Klein Tank AMG, Rusticucci M, Alexander LV, Brönnimann S, Charabi Y, Dentener FJ, Dlugokencky EJ, Easterling DR, Kaplan A, Soden BJ, Thorne PW, Wild W, Zhai PM. 2013. Observations: Atmosphere and Surface. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker T, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK and New York, NY, 159–254, doi: 10.1017/CBO9781107415324.008.

Hirahara S, Ishii M, Fukuda Y. 2014. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**: 57–75, doi: 10.1175/JCLI-D-12-00837.1.

Holgate SJ, Matthews A, Woodworth PL, Rickards LJ, Tamisiea ME, Bradshaw E, Foden PR, Gordon KM, Jevrejeva S, Pugh J. 2013. New data systems and products at the Permanent Service for Mean Sea Level. *J. Coast. Res.* **29**(3): 493–504, doi: 10.2112/JCOASTRES-D-12-00175.1.

Huang B, Banzon V, Freeman E, Lawrimore J, Liu W, Peterson T, Smith T, Thorne P, Woodruff S, Zhang H. 2015. Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: upgrades and intercomparisons. *J. Clim.* **28**: 931–951, doi: 10.1175/JCLI-D-14-00006.1.

International Oceanographic Data and Information Exchange of the IOC (IODE). 2013. *IOC Committee on International Oceanographic Data and Information Exchange, Twenty-second Session*, Ensenada, Mexico, 11–15 March 2013. IOC/IODE-XXII/3, UNESCO, Paris.

Jackson DL, Wick GA. 2010. Near-surface air temperature retrieval derived from AMSU-A and sea surface temperature observations. *J. Atmos. Ocean. Technol.* **27**: 1769–1776, doi: 10.1175/2010JTECHA1414.1.

Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM). 2011. *Proceedings of the 3rd International Workshop on Advances in the Use of Historical Marine Climate Data (MARCDAT-III)*, Frascati, Italy, 2–6 May 2011. JCOMM Technical Report No. 59. http://www.wmo.int/pages/prog/amp/mmop/jcomm&uscore;reports.html (accessed 24 March 2016).

Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM). 2015. *Proceedings of the Fourth JCOMM Workshop on Advances in Marine Climatology (CLIMAR-4) and of the First ICOADS Value-Added Database (IVAD-1) Workshop*, Asheville, NC, 9–13 June 2015. JCOMM Technical Report No. 79. http://www.jcomm.info/index.php?option=com_ oe&task=viewDocumentRecord&docID=15293 (accessed 24 March 2016).

Josey SA, Kent EC, Taylor PK. 1999. New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology. *J. Clim.* **12**: 2856–2880, doi: 10.1175/1520-0442(1999)012 < 2856:NIITOH > 2.0.CO;2.

Kaplan A, Kushnir Y, Cane MA. 2000. Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *J. Clim.* **13**: 2987–3002, doi: 10.1175/1520-0442(2000)013 < 2987:RSOIOH > 2.0.CO;2.

Karl TR, Arguez A, Huang B, Lawrimore JH, McMahon JR, Menne MJ, Peterson TC, Vose RS, Zhang H. 2015. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**(6242): 1469–1472, doi: 10.1126/science.aaa5632.

Kaspar F, Tinz B, Mächel H, Gates L. 2015. Data rescue of national and international meteorological observations at Deutscher Wetterdienst. *Adv. Sci. Res.* **12**: 57–61, doi: 10.5194/asr-12-57-2015.

Kennedy JJ. 2014. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **52**: 1–32, doi: 10.1002/2013RG000434.

Kennedy JJ, Rayner NA, Smith RO, Saunby M, Parker DE. 2011a. Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. *J. Geophys. Res.* **116**: D14103, doi: 10.1029/2010JD015218.

Kennedy JJ, Rayner NA, Smith RO, Saunby M, Parker DE. 2011b. Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. *J. Geophys. Res.* **116**: D14104, doi: 10.1029/2010JD015220.

Kent EC, Berry DI. 2005. Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *Int. J. Climatol.* **25**(7): 843–856, doi: 10.1002/joc.1167.

Kent EC, Woodruff SD, Berry DI. 2007. Metadata from WMO publication no. 47 and an assessment of voluntary observing ship observation heights in ICOADS. *J. Atmos. Ocean. Technol.* **24**: 214–234, doi: 10.1175/JTECH1949.1.

Kent EC, Rayner NA, Berry DI, Saunby M, Moat BI, Kennedy JJ, Parker DE. 2013. Global analysis of night marine air temperature and its uncertainty since 1880: the HadNMAT2 Data set. *J. Geophys. Res. Atmos.* **118**(3): 1281–1298, doi: 10.1002/jgrd.50152.

Kinzel J, Fennig K, Schröder M, Andersson A, Bumke K, Hollmann R. 2016. Decomposition of random errors inherent to HOAPS-3.2 near-surface humidity estimates using multiple triple collocation analysis. *J. Atmos. Ocean. Technol.*, doi: 10.1175/JTECH-D-15-0122.1 (in press).

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K. 2015. The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorol. Soc. Jpn.* **93**(1): 5–48, doi: 10.2151/jmsj.2015-001.

Krauss W. 1996. On the slippage of deep-drogued drifters. *J. Geophys. Res.* **101**: 1309–1319, doi: 10.1029/95JC02686.

Lindau R. 1995. A new Beaufort equivalent scale. In *Proceedings of International COADS Winds Workshop*, Kiel, Germany, 1 May–2 June 1994. NOAA Environmental Research Laboratories, Boulder, CO and Institut fur Meereskunde/Christian-Albrechts-Universitat, Kiel, 232–252. http://www2.meteo.uni-bonn.de/staff/rlindau/publ/Kiel.Lindau.pdf (accessed 24 March 2016).

Liu W, Huang B, Thorne P, Banzon V, Zhang H, Freeman E, Lawrimore J, Peterson T, Smith T, Woodruff S. 2015. Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part II: parametric and structural uncertainty estimations. *J. Clim.* **28**: 911–930, doi: 10.1175/JCLI-D-14-00007.1.

Martinho F, van der Veer HW, Cabral HN, Pardal MA. 2013. Juvenile nursery colonization patterns for the European flouder (*Platichthys flesus*): a latitudinal approach. *J. Sea Res.* **84**: 61–69, doi: 10.1016/j.seares.2013.07.014.

Maury MF. 1854. Maritime Conference held at Brussels for devising a uniform system of meteorological observations at sea, August and September, 1853. In *Explanations and Sailing Directions to Accompany the Wind and Current Charts*, 6th edn. E. C. and J. Biddle: Philadelphia, PA, 54–96.

McPhaden MJ, Busalacchi AJ, Cheney R, Donguy JR, Gage KS, Halpern D, Ji M, Julian P, Meyers G, Mitchum GT, Niiler PP, Picaut J, Reynolds RW, Smith N, Takeuchi K. 1998. The Tropical Ocean-Global Atmosphere (TOGA) observing system: a decade of progress. *J. Geophys. Res.* **103**: 14169–14240, doi: 10.1029/97JC02906.

McPhaden MJ, Meyers G, Ando K, Masumoto Y, Murty VSN, Ravichandran M, Syamsudin F, Vialard J, Yu L, Yu W. 2009. RAMA: the research moored array for African–Asian–Australian monsoon analysis and prediction. *Bull. Am. Meteorol. Soc.* **90**: 459–480, doi: 10.1175/2008BAMS2608.1.

Minobe S, Maeda A. 2005. A 1° monthly gridded sea-surface temperature data set compiled from ICOADS from 1850 to 2002 and Northern Hemisphere frontal variability. *Int. J. Climatol.* **25**: 881–894, doi: 10.1002/joc.1170.

Mohn H. 1907. Meteorology. No. 4 in Vol. 1 of *Report of the Second Norwegian Arctic expedition in the "Fram" 1898–1902*. http://archive.org/details/reportsecondnor00oslogoog (accessed 3 May 2016).

Poli P, Hersbach H, Berrisford P, Dee D, Simmons A, Laloyaux P. 2015. ERA-20C deterministic. ERA Report Series No. 20, ECMWF, 48 pp. http://www.ecmwf.int/sites/default/files/elibrary/2015/11700-era-20c-deterministic.pdf (accessed 24 March 2016).

Rhein M, Rintoul SR, Aoki S, Campos E, Chambers D, Feely RA, Gulev S, Johnson GC, Josey SA, Kostianoy A, Mauritzen C, Roemmich D, Talley LD, Wang F. 2013. Observations: Ocean. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK and New York, NY, 255–316, doi: 10.1017/CBO9781107415324.010.

Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim GK, Bloom S, Chen J, Collins D, Conaty A, da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Peqion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J. 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**: 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Roberts JB, Clayson CA, Robertson FR, Jackson DL. 2010. Predicting near-surface atmospheric variables from special sensor microwave/imager using neural networks with a first-guess approach. *J. Geophys. Res. Atmos.* **115**: D19113, doi: 10.1029/2009JD013099.

Saha S, Moorthi S, Pan H, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y, Chuang H, Juang HH, Sela J, Iredell M, Treadon R, Kleist D, Delst PV, Keyser D, Derber J, Ek M, Meng J, Wei H, Yang R, Lord S, Van Den Dool H, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M. 2010. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **91**: 1015–1057, doi: 10.1175/2010BAMS3001.2.S1.

da Silva AM, Young CC, Levitus S. 1994. Atlas of surface marine data 1994, volume 1: algorithms and procedures. In *NOAA Atlas NESDIS 6*. US Government Printing Office: Washington, DC, 83 pp.

Slutz RJ, Lubker SJ, Hiscox JD, Woodruff SD, Jenne RL, Joseph RH, Steurer PM, Elms JD. 1985. Comprehensive ocean-atmosphere data set: Release 1. NTIS PB86-105723, NOAA Environmental Research Laboratories, Climate Research Program, Boulder, CO, 268 pp.

Smith SR, Rolph JJ, Briggs K, Bourassa MA. 2009. *Quality-Controlled Underway Oceanographic and Meteorological Data from the Center for Ocean-Atmospheric Predictions Center (COAPS) – Shipboard Automated Meteorological and Oceanographic System (SAMOS)*. National Oceanographic Data Center/NOAA.

Smith SR, Hughes PJ, Bourassa MA. 2011. A comparison of nine monthly air–sea flux products. *Int. J. Climatol.* **31**: 1002–1027, doi: 10.1002/joc.2225.

Smith SR, Freeman E, Lubker SJ, Woodruff SD, Worley SJ, Angel WE, Berry DI, Brohan P, Ji Z, Kent EC. 2016. The International Maritime Meteorological Archive (IMMA) Format. http://icoads.noaa.gov/e-doc/imma/R3.0-imma1.pdf (accessed 31 May 2016).

Thorne PW, Willett KM, Allan RJ, Bojinski S, Christy JR, Fox N, Gilbert S, Jolliffe I, Kennedy JJ, Kent EC, Klein Tank A, Lawrimore J, Parker DE, Rayner N, Simmons A, Song L, Stott PA, Trewin B. 2011. Guiding the creation of a comprehensive surface temperature resource for 21st century climate science. *Bull. Am. Meteorol. Soc.* **92**: ES40–ES47, doi: 10.1175/2011BAMS3124.1.

Tokinaga H, Xie SP. 2011. Wave and anemometer-based sea-surface wind (WASWind) for climate change analysis. *J. Clim.* **24**: 267–285.

Villegas-Hernández H, Lloret J, Muñoz M. 2015. Climate-driven changes in life-history traits of the bastard grunt (Pomadasys incisus) in the north-western Mediterranean. *Mediterranean Marine Science* **16**(1): 21–30, doi: 10.12681/mms.951.

Wallbrink H, Koek F, Brandsma T. 2009. *The US Maury Collection Metadata 1796–1861*. KNMI-225/HISKLIM-11. http://www.knmi.nl/bibliotheek/knmipubmetnummer/knmipub225.pdf (accessed 24 March 2016).

Wilkinson C, Woodruff SD, Brohan P, Claesson S, Freeman E, Koek F, Lubker SJ, Marzin C, Wheeler D. 2011. RECovery of logbooks and international marine data: the RECLAIM project. *Int. J. Climatol.* **31**: 968–979, doi: 10.1002/joc.2102.

WMO. 1955. International list of selected, supplementary and auxiliary ships. WMO-No. 47, World Meteorological Organization, Geneva, Switzerland (Serial publication; recently annual. Editions prior to 1966 were entitled *International List of Selected and Supplementary Ships*).

WMO. 2002. Workshop on advances in marine climatology. *WMO Bull.* **51**(4): 377–380.

WMO. 2010. Guide to meteorological instruments and methods of observation. WMO-No 8, World Meteorological Organization, Geneva, Switzerland.

WMO. 2011. Manual on codes. WMO-No. 306, World Meteorological Organization, Geneva, Switzerland. http://www.wmo.int/pages/prog/www/WMOCodes.html. (accessed 3 May 2016).

WMO. 2012. *Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology*, 4th Session, Yeosu, Republic of Korea, 28–31 May 2012. WMO-No. 1093, World Meteorological Organization, Geneva, Switzerland. http://www.jcomm.info/index.php?option=com_oe&task=viewDocumentRecord&docID=9774. (accessed 3 May 2016).

Wolter K. 1997. Trimming problems and remedies in COADS. *J. Clim.* **10**: 1980–1997, doi: 10.1175/1520-0442(1997)010 < 1980:TPARIC > 2.0.CO;2.

Wolter K, Timlin MS. 2011. El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Int. J. Climatol.* **31**: 1074–1087, doi: 10.1002/joc.2336.

Woodruff SD. 2007. Archival of data other than in IMMT format: the International Maritime Meteorological Archive (IMMA) Format. In *Second Session of the JCOMM Expert Team on Marine Climatology (ETMC)*, Geneva, Switzerland, 26–27 March 2007. JCOMM Meeting Report No. 50, 68–101. http://icoads.noaa.gov/e-doc/imma/ (accessed 24 March 2016).

Woodruff SD, Slutz RJ, Jenne RL, Steurer PM. 1987. A comprehensive ocean-atmosphere data set. *Bull. Am. Meteorol. Soc.* **68**: 1239–1250, doi: 10.1175/1520-0477(1987)068 < 1239:ACOADS > 2.0.CO;2.

Woodruff SD, Lubker SJ, Wolter K, Worley SJ, Elms JD. 1993. Comprehensive ocean–atmosphere data set (COADS) Release 1a: 1980–92. *Earth Syst. Monit.* **4**(1): 1–8.

Woodruff SD, Diaz HF, Elms JD, Worley SJ. 1998. COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth* **23**: 517–526, doi: 10.1016/S0079-1946(98)00064-0.

Woodruff SD, Worley SJ, Arnott JA, Diaz HF, Elms JD, Jackson M., Lubker SJ, Parker DE. 2003. COADS updates and the blend with the UK Met Office Marine Data Bank. In *Advances in the Applications of Marine Climatology – The Dynamic Part of the WMO Guide to the Applications of Marine Meteorolog*y. WMO/TD-No. 1081 (JCOMM Technical Report No. 13), World Meteorological Organization, Geneva, Switzerland, 3–10.

Woodruff SD, Diaz HF, Worley SJ, Reynolds RW, Lubker SJ. 2005. Early ship observational data and ICOADS. *Clim. Change* **73**: 169–194, doi: 10.1007/s10584-005-3456-3.

Woodruff S, Worley S, Lubker S, Ji Z, Freeman E, Berry D, Brohan P, Kent E, Reynolds D, Smith S, Wilkinson C. 2011. ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**: 951–967, doi: 10.1002/joc.2103.

Woods Hole Oceanographic Institute (WHOI). 2002. Historical U.S. lightship and light station data, hydrographic and meteorological observations. http://www.whoi.edu/science/GG/woos/resources/related.htm. (accessed 3 May 2016).

Worley SJ, Woodruff SD, Reynolds RW, Lubker SJ, Lott N. 2005. ICOADS Release 2.1 data and products. *Int. J. Climatol.* **25**: 823–842, doi: 10.1002/joc.1166.