INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **25**: 823–842 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/joc.1166

ICOADS RELEASE 2.1 DATA AND PRODUCTS

STEVEN J. WORLEY,[a,*] SCOTT D. WOODRUFF,[b] RICHARD W. REYNOLDS,[c] SANDRA J. LUBKER[b] and NEAL LOTT[c]
[a] *National Center for Atmospheric Research, Boulder, CO, USA*
[b] *Climate Diagnostics Center, NOAA/OAR, Boulder, CO, USA*
[c] *National Climatic Data Center, NOAA/NESDIS, Asheville, NC, USA*

Received 17 July 2004
Revised 10 September 2004
Accepted 28 September 2004

ABSTRACT

The International Comprehensive Ocean–Atmosphere Data Set (ICOADS), release 2.1 (1784–2002), is the largest available set of *in situ* marine observations. Observations from ships include instrument measurements and visual estimates, and data from moored and drifting buoys are exclusively instrumental. The ICOADS collection is constructed from many diverse data sources, and made inhomogeneous by the changes in observing systems and recording practices used throughout the period of record, which is over two centuries. Nevertheless, it is a key reference data set that documents the long-term environmental state, provides input to a variety of critical climate and other research applications, and serves as a basis for many associated products and analyses.

The observational database is augmented with higher level ICOADS data products. The observed data are synthesized to products by computing statistical summaries, on a monthly basis, for samples within 2° latitude × 2° longitude and 1° × 1° boxes beginning in 1800 and 1960 respectively. For each resolution the summaries are computed using two different data mixtures and quality control criteria. This partially controls and contrasts the effects of changing observing systems and accounts for periods with greater climate variability. The ICOADS observations and products are freely distributed worldwide.

The standard ICOADS release is supplemented in several ways; additional summaries are produced using experimental quality control, additional observations are made available in advance of their formal blending into a release, and metadata that define recent ships' physical characteristics and instruments are available. Copyright © 2005 Royal Meteorological Society.

KEY WORDS: marine data; ship data; buoy data; sea-surface temperature; air temperature; wind; sea-level pressure

1. INTRODUCTION

The project to construct the original US Comprehensive Ocean–Atmosphere Data Set (COADS) was planned in 1981, and continues as a three-way cooperation between the National Oceanic and Atmospheric Administration (NOAA) — its Climate Diagnostics Center (CDC) and National Climatic Data Center (NCDC) — and the National Science Foundation (NSF) National Center for Atmospheric Research (NCAR). These three partners continue to develop and distribute the data and products, which have steadily been expanded and improved. In recognition of significant growing international interest and contributions, a new name, the International COADS (ICOADS), was adopted in 2002 (Diaz *et al*., 2002; Parker *et al*., 2004).

ICOADS data are *in situ* marine meteorological observations, mainly from ships and buoys, and from many different national and international data sources. A wide variety of measured and visually estimated variables are included, including all those necessary to estimate ocean surface fluxes (e.g. Josey *et al*., 1999; Grist and Josey, 2003). This long-term global ocean record has supported research in many other scientific domains; recent examples include assessments of climate change (Folland *et al*., 2001a), global atmospheric reanalyses (Kistler *et al*., 2001; Uppala, 2002), satellite and *in situ* blended analyses (Reynolds *et al*., 2002; Rayner

* Correspondence to: Steven J. Worley, NCAR/SCD, PO Box 3000, Boulder, CO 80307, USA; e-mail: worley@ucar.edu

et al., 2003), long-term historical climate analyses (Kaplan *et al.*, 1998, 2000; Smith and Reynolds, 2003, 2004a,b), ground truth for remotely sensed or pre-instrumental proxy data (Dunbar *et al.*, 1994), changes in coastal geological features (Restrepo *et al.*, 2002), and assessment of global anthropogenic emissions from ships (Corbett *et al.*, 1999; Corbett and Koehler, 2003).

ICOADS is currently at release 2.1 with observations for 1784 to 2002, and with monthly summaries for 2° latitude × 2° longitude boxes covering 1800–2002, and for 1° × 1° boxes covering 1960–2002. Section 2 presents a release history for ICOADS, and Section 3 describes the composition of the currently available data. Section 4 discusses data usage considerations. Section 5 defines the ICOADS observations and products, and discusses associated products and analyses. Section 6 covers ICOADS data access, and Section 7 highlights our major future objectives.

2. RELEASE HISTORY

COADS release 1 was completed in 1985 (Slutz *et al.*, 1985; Woodruff *et al.*, 1987). This made available global observations and monthly statistical summaries for 2° latitude × 2° longitude areas (boxes) for 1854–1979. Many data-handling strategies and processing procedures that were developed for release 1 are still in use today. Table I summarizes the ICOADS releases, data temporal coverage, issuance dates, and other information.

The observational data for release 1 came from a variety of US and international sources, including ship logbook records keyed back to 1854, obtained through bilateral agreements starting in the 1940s, and more recent keyed ship reports, which have been routinely exchanged internationally since World Meteorological Organization (WMO) Resolution 35 in 1963. Marine reports circulated over the WMO Global Telecommunication System (GTS; or its precursors) supplemented the logbook data starting in 1966. In addition to ship data, the historical record was augmented by near-surface sea temperatures extracted from oceanographic profiles back to the late 19th century, and by data for recent decades from environmental (moored and drifting) buoys and other contemporary automated observing systems.

After 1985, release 1 was extended initially by a set of 'interim' products, and later extended and fully replaced by major updates for three segments covering the full period of record: release 1a (Woodruff *et al.*, 1993) initially included 1980–92, and was finally extended to cover 1980–97, through updates made in 1995,

Table I. ICOADS release history. For each release, the total temporal coverage, any updates and extensions, and the composition (for release 2.0 forward, only), in terms of the delayed-mode (DM) and real-time (RT; GTS-only) archives, are listed

Release name	Temporal coverage (issuance date)	Updates and extensions (issuance date)	DM archive	RT archive
Release 1	1854–1979 (1985)	1980–91 (1987–92)[a]		
Release 1a	1980–92 (1993)	1992–93 (1995) 1990–95 (1997) 1980–97 (1999)		
Release 1b	1950–79 (1996)	1970[b] (1999)		
Release 1c	1784–1949 (2001)			
Release 2.0[c]	1784–1997 (2002)		1784–1997	
Release 2.1	1784–2002 (2003)		1784–1997	1998–2002[d]

[a] Following release 1, 'interim' products, constructed using simplified procedures and preliminary input data, were first issued in 1987 (covering 1980–86) and then extended, on an approximately annual basis, to cover 1980–91 finally.
[b] Minor corrections for October–November 1970.
[c] New release nomenclature adopted for the combination of releases 1c, 1b, and 1a (no new data or products).
[d] March–December 1997 observational data were also processed.

1997, and 1999. Release 1b (Woodruff *et al*., 1998) covering 1950–79 and release 1c covering 1784–1949 were completed in 1996 and 2001 respectively. In conjunction with the latter two releases the products were supplemented with new 1° monthly summaries for 1960 and forward, and the 2° summaries were extended back to 1800.

Because releases 1c, 1b, and 1a collectively covered 1784–1997 and had fully replaced release 1, we transitioned to new terminology and adopted 'release 2.0' to refer to ICOADS in 2002. Like release 1, release 2.0 was created from many different data sources, including GTS data, and had been subjected to the full 'quality control' (QC) and other automated data processing steps (see Section 5). To identify this level of data inclusion and processing we defined these observations as the 'delayed-mode' archive (ICOADS.DM, henceforth DM).

In 2003, we extended ICOADS through to 2002 using only GTS data from NOAA's National Centers for Environmental Prediction (NCEP). By limiting the data sources we were able to keep ICOADS more current. However, since the processing is slightly different and does not include all available sources, we defined these observations as the 'real-time' (ICOADS.RT, henceforth RT) archive. At the same time we adopted a new numbering policy to refer to the total collection of observations and products made available to the user community at any one time as a single release number. Thus, release 2.1 identifies the currently available observations for 1784–2002, and monthly summary products for 1800–2002. We will extend and improve both observational archives as often as practicable, aiming toward annual RT extensions and less frequent DM updates. Observational and summary products offered to users at any given time will draw on both archives, including the full period of record from DM supplemented with RT.

3. DATA COMPOSITION

ICOADS release 2.1 has data for over two centuries, 1784–2002, and has been built incrementally from previous releases. The 185 million observations come from many data sources. In the raw input data, there are overlaps (duplication) between some sources, new data-set versions that only partially replace older versions, and cases where some parts of sources are known to be of inferior quality compared with others. This leads to a complicated blending and duplicate elimination process that will not be discussed in detail here. Further information about multiple aspects of the data processing is available at the ICOADS Website (www.cdc.noaa.gov/coads/) and more specifically in its documentation section (www.cdc.noaa.gov/coads/doc.html).

An important key to successful blending and analysis of ICOADS is the ability to track records back to the original sources. Two data fields available on each individual record were designed for tracking: 'deck' (originally referring to punched card decks) and 'source identification' (SID). Table II documents the complete release 2.1 composition in terms of the decks included. Qualitatively, the impact of each deck can be assessed by its period of record and number of records included in ICOADS. Similarly, Table III provides a breakdown of the release 2.1 data mixture by SID, which either provides cross-cutting information within decks, or identifies broader collections of decks. These fields are critical for data mixture assessment and problem analysis by the ICOADS project staff; and they can be useful for data users, in addition to more general categorizations, such as provided by the platform type field (e.g. ship, buoy).

3.1. Delayed-mode archive

The segments of the DM archive are now considered chronologically (see Table I for the release dates), and the highlights of the data additions and modifications, relative to release 1.

3.1.1. 1784–1949 observations. The update for this period (Woodruff *et al*., in press) blended data from the UK Marine Data Bank (MDB) for 1854–1949 (including 0.5 million newly keyed 1935–39 UK merchant ship reports), Russian Marine Meteorological (MARMET) data back to 1888, and about one million recently keyed reports from Japan's Kobe Collection (concentrated in the Pacific) to enrich the data-sparse period around

Table II. ICOADS release 2.1 deck composition. For each deck number, the description, starting and ending years, and number of reports (in thousands) after final blending are listed. Note that some decks exist for longer periods, and have not yet been fully blended into ICOADS (e.g. the UK MDB, decks 201–255, has not been blended for 1950–79). Decks present since release 1 (Slutz *et al.*, 1985) are numbered in normal type, decks added to form ICOADS.DM are in italic, and decks added to form ICOADS.RT are in bold. Further details about individual decks are available on the ICOADS Website

Deck	Description	Start	End	Reports (thousands)
110	US Navy Marine	1945	1951	634
116	US Merchant Marine	1945	1963	6878
117	US Navy Hourlies	1952	1964	11
118	Japanese Ships No. 1 (Kobe Collection data keyed in 1961)	1930	1953	1727
119	Japanese Ships No. 2 (Kobe Collection data keyed in 1961)	1951	1961	905
128	International Marine (US- or foreign-keyed ship data)	1950	1978	15 375
143	Pacific Marine Environmental Laboratory (PMEL) buoys	1976	1977	13
144	PMEL TOGA/TAO buoys[a]	1985	1997	3
145	PMEL (daily) equatorial moorings and island stations[a]	1979	1991	17
150	Pacific (US Responsibility) HSST Netherlands receipts	1939	1961	97
151	Pacific (US Responsibility) HSST German receipts	1862	1960	590
152	Pacific (US Responsibility) HSST UK receipts	1855	1961	52
155	Indian (Netherlands Responsibility) HSST	1861	1960	1115
156	Atlantic (German Responsibility) HSST	1852	1961	6011
184	Great Britain Marine (194 extension)	1953	1961	593
185	USSR Marine IGY	1957	1958	111
186	USSR ice stations	1950	1970	20
187	Japanese whaling fleet	1946	1956	10
188	Norwegian Antarctic whaling factory ships	1932	1939	2
189	Netherlands Marine	1939	1959	232
192	Deutsche Seewarte Marine	1855	1939	5998
193	Netherlands Marine	1800	1938	6276
194	Great Britain Marine	1856	1955	785
195	US Navy ships logs	1941	1946	598
196	Deutsche Seewarte Marine (192 extension)	1949	1954	143
197	Danish (and other) Marine (polar)	1871	1956	23
201–255	UK Marine Data Bank (MDB)[b]	1854	1994	6346
281	US Navy Monthly Aerological Record (MAR)	1926	1945	187
555	US Navy FNMOC Monterey Telecom.	1966	1973	2468
666	Tuna boats	1970	1975	17
667	Inter-American Tropical Tuna Commission (IATTC)	1971	1997	1173
701	US Maury Collection	1784	1863	1346
702	Norwegian Logbook Collection	1867	1889	201
705–707	US Merchant Marine Collection (1912–46)	1910	1946	3503
714	Canadian Marine Environmental Data Service (MEDS) buoys	1978	1997	10 309
731	Russian S.O. Makarov Collection	1804	1891	3
732	Russian Marine Met. Data Set (MARMET) (received at NCAR)	1888	1995	8839
733	Russian AARI North Pole (NP) stations	1937	1991	98
734	Arctic drift stations	1893	1924	12
749	First GARP Global Experiment (FGGE) Level IIb	1978	1979	12
762	Japanese Kobe Collection data (keyed after decks 118–119)	1890	1932	1045
780	Levitus World Ocean Atlas/Database (WOA/WOD)	1874	1996	4479
792	NCEP BUFR GTS: ship data	1998	2002	5139
793	NCEP BUFR GTS: buoy data (FM 13 'SHIP' code)	1998	2002	5532
794	NCEP BUFR GTS: buoy data (FM 18 'BUOY' code)	1998	2002	16 787
795	NCEP BUFR GTS: C-MAN data	1998	2002	2548

Table II. (*Continued*)

Deck	Description	Start	End	Reports (thousands)
849	First GARP Global Experiment (FGGE)	1978	1979	253
850	German FGGE	1978	1979	154
874	Shipboard Environmental (Data) Acquisition System (SEAS)	1995	1997	68
876–882	US National Data Buoy Center (NDBC) data	1972	1979	315
883	NDBC data	1980	1997	12 782
888	US Air Force Global Weather Central (GWC)	1973	1997	6385
889	Autodin (US Dept. of Defense Automated Digital Network)	1972	1995	1039
891	US National Oceanographic Data Center (NODC) surface data	1950	1977	297
892	NCEP ship data	1980	1997	9681
893	NCEP moored buoy data	1986	1997	2230
896	NCEP miscellaneous (OSV, plat, and rig) data	1980	1997	588
897	*Eltanin*	1962	1963	597
898	Japanese	1954	1974	123
899	South African whaling	1900	1955	65
900	Australian	1931	1979	392
901	FOSDIC reconstructions (card images from 16 mm film)	1868	1963	7
902	Great Britain Marine (184 extension)	1957	1961	129
926	International Maritime Meteorological (IMM) data	1954	1997	18 934
927	International Marine (US- or foreign-keyed ship data)[c]	1970	1997	10 881
928	Same as 927 including Ocean Station Vessels (OSV)	1970	1974	4
999	US Air Force Environ. Technical Applications Center (ETAC)	1967	1969	37

[a] Deck 145 contains daily averaged data, and up to the early 1990s TAO deck 144 contains average estimates for 2–8 h depending on the buoy instrument package and power requirements.
[b] Deck 215, within this range is believed to be derived from the same original German punched cards as deck 192.
[c] A mixture of US- and foreign-keyed data exists in deck 927 prior to 1980; starting about 1980, deck 927 is believed to contain only US-keyed ships.

World War I (Manabe, 1999). Also included were data from several recently digitized collections (Diaz and Woodruff, 1999): the US Maury Collection (covering 1784–1863, but concentrated around 1830–60; 1.3 million reports), the Norwegian Logbook Collection (1867–89; 201 thousand reports), the US Merchant Marine 1912–46 Collection (3.5 million reports), Arctic Drift Stations (1893–1938; 16 thousand reports) and the Russian S.O. Makarov Collection (1804–91; 3400 reports). Sea-surface temperature (SST) estimates derived from oceanographic profiles, and some surface meteorological fields, were also added back to 1874 from the *World Ocean Database 1998* (Levitus *et al*., 1998) (WOD98). Many more details are provided in Woodruff *et al*. (in press).

3.1.2. 1950–79 observations. The update for this period (Woodruff *et al*., 1998) pre-dated the MDB blending work. The MDB will be included in this period at a future time. Data additions included Russian MARMET, North Pole manned drifting ice-station data, and oceanographic profile SST estimates from the *World Ocean Atlas 1994* (Levitus and Boyer, 1994), an earlier version of the WOD98. Significant data improvements were made during this processing, including the correction of widespread temperature biases in GTS records (Woodruff *et al*., 1998) and a minor correction for mislocated GTS data in October–November 1970.

3.1.3. 1980–97 observations. Data for this period were originally issued for 1980–92 (Woodruff *et al*., 1993), and then extended to 1995 (Woodruff *et al*., 1998). The latest update (Woodruff *et al*., 2003) involved reprocessing data for 1980–95 and adding data through 1997. One major addition was 17.9 million MDB reports for 1980–94. In the blend, 7% of these data were retained as unique or judged to be of preferable quality compared with other data already in ICOADS. For example, British Navy decks were retained at

Table III. Release 2.1 source ID (SID) composition. For each SID number, the description, starting and ending years, and number of reports (in thousands) after final blending are listed. SIDs present since release 1 (Slutz *et al*., 1985) are numbered in normal type, SIDs added to form ICOADS.DM are in italic, and SIDs added to form ICOADS.RT are in bold. Generally, data within a SID are confined to a single input format, whereas some decks have been received by ICOADS in different formats. Further details about individual SIDs are available on the ICOADS Website

SID	Description	Start	End	Reports (thousands)
1	Atlas	1800	1969	33 986
2	HSST Pacific	1855	1961	715
3	HSST Indian	1861	1960	1115
4	HSST Atlantic	1852	1961	6011
5	Old TDF-11 supplement B	1854	1975	3008
6	Old TDF-11 supplement C	1855	1978	2690
7	Monterey Telecommunications	1966	1969	757
8	Ocean Station Vessels (OSV)	1945	1973	823
9	OSV supplement	1947	1973	59
10	MSQ 486 and 105 omissions	1854	1968	176
11	National Oceanographic Data Center (NODC) surface	1950	1975	221
12	NODC surface supplement	1950	1977	51
13	*Eltanin*	1962	1963	1
14	Japanese	1954	1974	122
15	South African whaling	1900	1955	65
16	Australian	1931	1970	196
17	International Maritime Meteorological (IMM) data	1956	1979	1660
18	′70s decade	1970	1979	12 897
19	IMM ′70s	1978	1979	<1
20	OSV Z (′70s)	1970	1974	3
21	Australian (′70s)	1971	1979	196
22	NCDC: 1980–84 annual receipts	1982	1987	135
23	′70s mislocated data	1973	1979	2
24	Buoy data	1972	1979	192
25–28	NCDC: 1980–85 annual receipts	1962	1985	1561
29	NCDC: NCEP reconversion	1980	1992	8410
30	NCDC: 1980–84 period of record	1965	1984	4200
32–33	NCDC: annual receipts (and duplicates; starting in 1986)	1974	1997	4531
34–45	NCDC: 1986–97 receipts (delayed)	1965	1996	18 394
46–47	International Maritime Met. (IMM) tape archive	1969	1995	7134
48–49	Levitus 1994 World Ocean Atlas (WOA)	1950	1979	2255
50	US National Data Buoy Center (NDBC) data	1980	1997	12 782
51–52	Russian AARI North Pole (NP) stations	1937	1991	98
53	First GARP Global Exp. (FGGE) Level IIb: surface marine data	1978	1979	12
56	Russian S.O. Makarov Collection	1804	1891	3
57	Russian Marine Meteorological Data Set (MARMET)	1888	1993	8839
59	UK IMM corrections	1982	1989	1552
60	French International Maritime Met. (IMM) corrected	1954	1988	160
61	Canadian Marine Environmental Data Service (MEDS) buoys	1978	1997	9583
62	MEDS World Ocean Circulation Experiment (WOCE) buoys	1992	1997	726
66	Pacific Marine Environmental Lab. (PMEL) TOGA/TAO buoys	1985	1992	236
67	PMEL (daily) equatorial moorings and island stations	1979	1991	17
68	Arctic drift stations	1893	1924	12
69	US Maury Collection	1784	1863	1346
70	Inter-American Tropical Tuna Comm. (IATTC) porpoise obs. logs	1979	1997	736
71	IATTC fishing logs	1971	1997	437
72	IMM tape arch. from WMO Global Collecting Centre (GCC)	1982	1997	3870

Table III. (*Continued*)

SID	Description	Start	End	Reports (thousands)
77	NCDC: NCEP reconversion	1994	1997	2801
78	NCDC: US-keyed logbook data reconversion (TD-9972)	1987	1997	307
79	US Air Force Global Weather Central (GWC): DATSAV2 format	1980	1997	1595
84	US Merchant Marine Collection (1912–46): full QC	1910	1944	1929
85	US Merchant Marine Collection (1912–46): partial QC	1910	1946	1246
86	PMEL TOGA/TAO buoys: RAM data	1990	1997	2464
87	PMEL TOGA/TAO buoys: SPOT data	1990	1997	64
88–89[a]	Levitus 1998 World Ocean Database (WOD98)	1874	1996	1924
90–93[a]	UK Marine Data Bank (MDB)	1854	1994	6346
96	Norwegian Logbook Collection	1867	1889	201
97	Japanese Kobe Collection data (IMMT format; 1998 edition)	1890	1932	1045
98	US Merchant Marine Collection (1912–46): full QC (CLICOM)	1914	1944	328
100	NCEP BUFR GTS: operational tanks: converted from message	1998	1999	8930
103	NCEP BUFR GTS: dumped data: converted from BUFR	1999	2002	21 076

[a] Due to the release processing sequence, data from these SIDs have not been included in the 1950–79 period.

relatively high rates compared with other MDB sources. Moored and drifting buoy data were enhanced in this update. GTS receipts after 1990 from Tropical Atmosphere–Ocean (TAO) moorings were replaced by ‘standard archive’ data obtained directly from NOAA’s Pacific Marine Environmental Laboratory (PMEL), which improved diurnal coverage and quality controls; the Pilot Research Moored Array in the Tropical Atlantic (PIRATA) data are also now part of this archive. Similarly, all the updates for this period involved replacing GTS receipts from US coastal moorings and the Coastal-Marine Automated Network (C-MAN) by quality controlled and more complete data from NOAA’s National Data Buoy Center (NDBC). Canada’s Marine Environmental Data Service (MEDS) corrected 1980–85 drifting buoy data for a day assignment error that had impacted some buoys reporting in the last quarter of the day. Also, for 1993–97, processing changes at Service Argos necessitated some modifications in the handling of MEDS QC information to obtain increases in the available drifting buoy data. Other new or improved data sources were included in this update. Also similar to the earlier periods, Russian MARMET data and reports from North Pole manned drifting-ice stations were added, and similar to the earliest period, WOD98 SST estimates and meteorological data were added through 1996.

3.2. Real-time archive

Information about the RT portion of release 2.1 has not previously been published, so here the data composition and processing are described in some detail. During the development of the RT archive (March 1997–2002), GTS data from NCEP and NCDC were compared to ensure good overall quality. The 10 month overlap period, March–December 1997, was used for comparison with DM, but only RT data for 1998–2002 were used for release 2.1.

The GTS data received at NCEP were translated from GTS message strings (i.e. FM 13 SHIP or FM 18 BUOY codes) into the binary universal form for the representation of meteorological data (BUFR; WMO, 1995). NCEP also included one or more of the original input GTS message strings in the BUFR record. Including the strings was extremely beneficial, because it provided access to the original data and allowed accuracy verification for the translated data.

Several representations of the GTS data were used to assess and create the input for release 2.1 RT. First, all relevant fields from the NCEP BUFR and the NCEP attached GTS strings were converted to separate data sets using the new International Maritime Meteorological Archive (IMMA) format (Woodruff, 2004). At NCDC the GTS data stream was converted to IMMA format and the GTS string was attached. We received these data and also converted the NCDC attached string to IMMA format. The various sources were then compared and it was found that:

- The first 31 months (March 1997-September 1999) of the NCEP BUFR translation had some problems and the retranslation of the NCEP GTS string provided better quality and more complete data in this period.
- Checks between NCEP and NCDC GTS strings continue to be helpful to improve the quality of the independent translations at NCEP to BUFR and at NCDC to IMMA. The NCDC GTS data were not used in release 2.1.
- The comparison of overlapping DM and RT data for March–December 1997 (Figure 1) shows that the DM archive has more ship, moored buoy, and C-MAN data than the RT archive. For these platform categories the major differences are due to delayed (keyed) ship receipts, and inclusion of more complete post-processed moored buoy and C-MAN data from PMEL and NDBC.
- In contrast, there are approximately 40% more drifting buoy reports in RT compared with DM. The real-time data stream has near-duplicate buoy reports that MEDS consolidates before they are received by ICOADS and used to create DM.

For the final RT archive we used exclusively data from NCEP. For January 1998 through to September 1999 the retranslation of the attached GTS strings was used, and for October 1999 through to December 2002 the BUFR translation was used. The transition point in 1999 is also coincident with a switch to input data that had been QC processed at NCEP. More details about the processing, remaining problems, and archive analyses are available at the ICOADS Website.

4. DATA USAGE CONSIDERATIONS

Scientific data analyses are affected by the many sources of data used to derive ICOADS. Release 2.1 has most observations in the modern era (Figure 2) and sparse observations notably during the World Wars

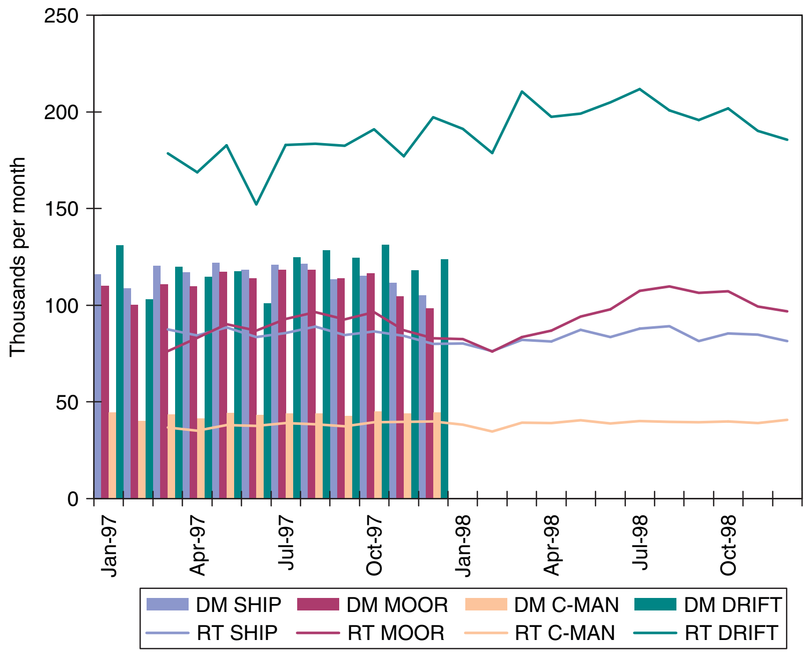

Figure 1. Thousands of reports per month for DM (1997 only; bars) compared with RT (March 1997–December 1998; lines). The data from both archives have been stratified by (approximate) platform type: ships, moored buoys (MOOR), Coastal-Marine Automated Network (C-MAN), and drifting buoys (DRIFT)

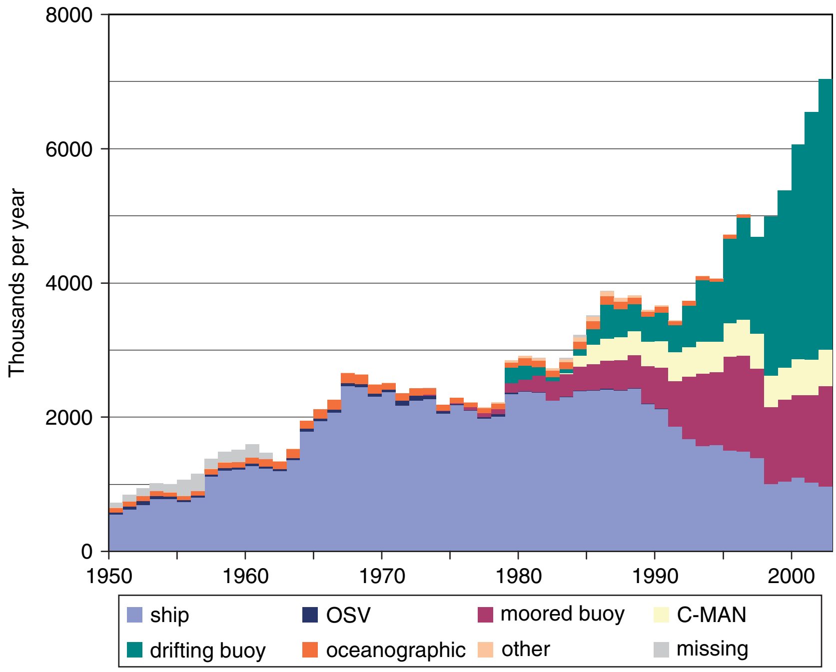

Figure 2. Annual number of reports stratified by platform type in ICOADS release 2.1 for the modern (1950–2002) period. Ships are stratified in three categories: 'ship' reports generally are regular voluntary observing ship observations from merchant and navy vessels, 'OSV' (ocean station vessels) are weather ships usually at stationary observing locations, and 'oceanographic' ships are from the research fleet. Buoys are stratified in two categories, either drifting or moored, 'C-MAN' automated reports are from the US land–ocean boundary zone, a small number of 'other' platforms are grouped together, and reports with unknown platform type are denoted as 'missing'

and in the 19th century (Figure 3). Where possible, each report in ICOADS has been assigned a platform type, and over time that mixture has changed; the earliest periods have observations exclusively from ship platforms, but buoy data later (1979–88) add to increasing ship observations. Starting in about 1989, ship reports decline and buoy data more than compensate, in a numerical sense, for the loss to give a net overall growth (Figures 2 and 3). In the most basic sense this is evidence of the changing observational systems on the global ocean, from earlier periods with shipping routes determining coverage and providing many meteorological observed variables at synoptic periods, to recent periods with buoys (either moored or drifting) and other automated platforms providing fewer observed variables, but often at higher frequency. These changes are convoluted, with instrument and sampling procedure changes, changes in coding of measurements and visual estimates, and variations of instrument package placement on the ocean platforms, which all can impact products and analyses based on the observations. Many researchers have studied these changes, particularly for SSTs and air temperatures, and wind (the following are examples: Barnett, 1984; Wright, 1986, 1988a; Ramage, 1987; Cardone *et al*., 1990; Kent *et al*., 1993; Folland and Parker, 1995; Hahn *et al*., 1995; Kent and Taylor, 1997; Smith and Reynolds, 2002; Rayner *et al*., 2003; Berry *et al*., 2004).

The global ocean has not been uniformly sampled over time. We use SST, a well-sampled variable, to illustrate the changes. The number of SST samples and percentage of ocean (and coastal) area covered, based on 2° boxes, varies dramatically (Figure 4). If we choose at least five samples per month in a 2° area as a reference, SST sampling increases fairly steadily from nearly zero in the 1840s to a peak over 20% around 1916, falling below 10% during the two World Wars, and approaching 30% between them. Growth resumes after World War II, with coverage reaching a general maximum in about 1980 of around 45 to 50%. Referencing Figure 2, the 1980s was when the ship data were at a maximum plateau and buoy data were just beginning. At sampling frequencies of one to ten per

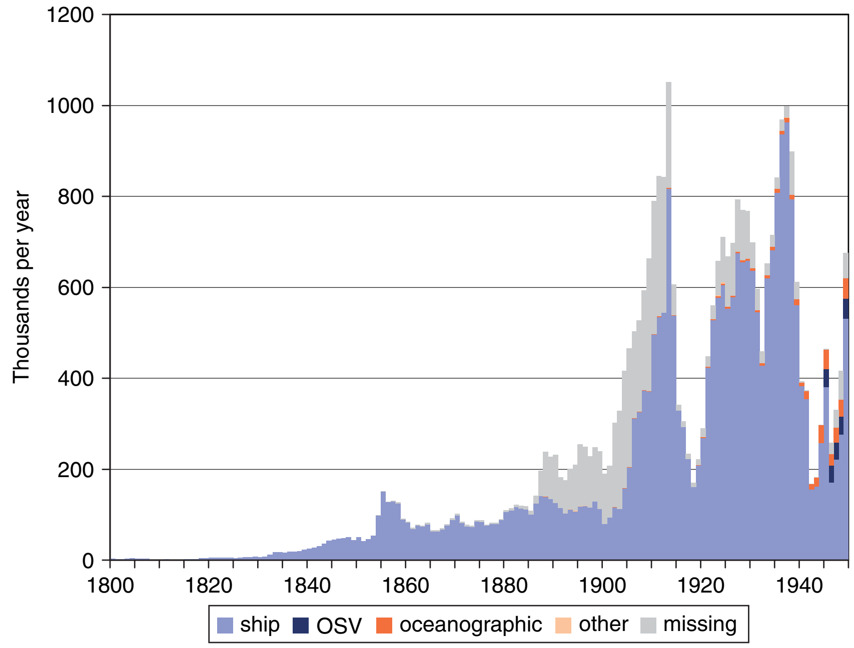

Figure 3. Annual number of reports stratified by platform type in ICOADS release 2.1 for the early (1800–1949) period. Fewer stratification categories (with the same definition as in Figure 2) appear during this period. Reports with 'missing' platform type indicator are likely to be from ships. Note, a different vertical scale range is used compared with that in Figure 2

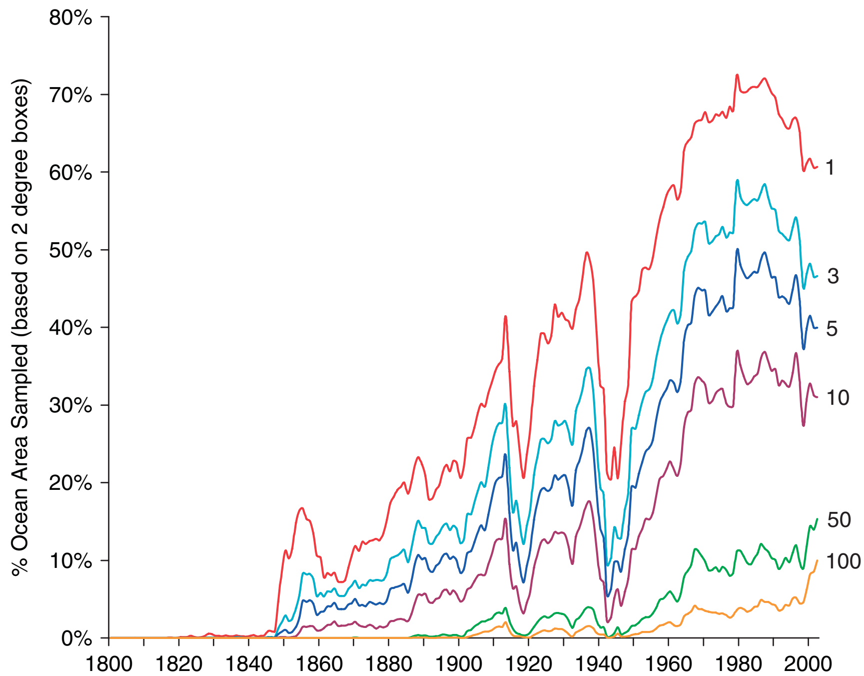

Figure 4. Percentage global ocean and coastal area sampled based on area-weighted 2° boxes (smoothed) for SST, with the number of observations per month (at minimum) in the boxes shown by the numbered curves. Determined from the 'enhanced trimming' (4.5σ) product that includes ship and buoy records. This is an update of Woodruff *et al*. (1987: figure 4)

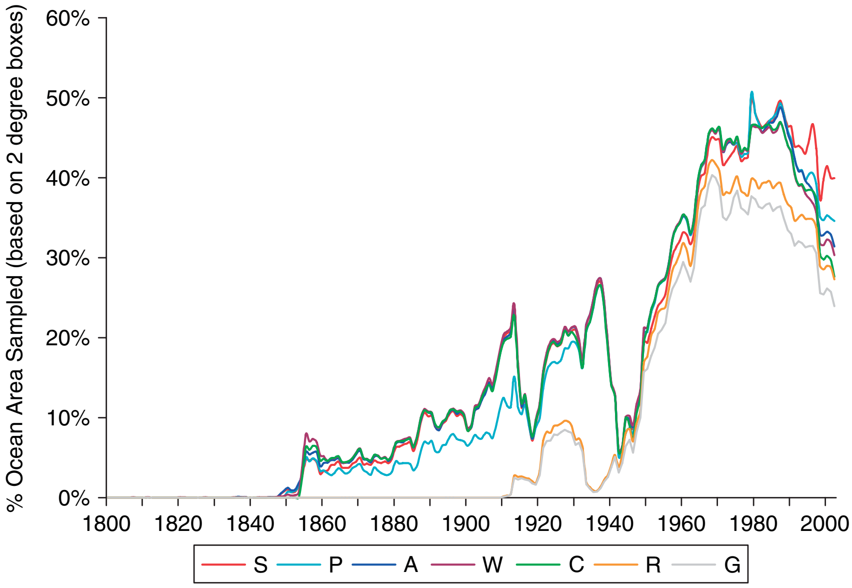

Figure 5. As for Figure 4, except comparing the SST (S) coverage, at five observations per month, with that for sea-level pressure (P), air temperature (A), wind speed (W), total cloudiness (C), and relative humidity (R). Determined from the 'enhanced trimming' (4.5σ) product that includes ship and buoy records. Also plotted is the evaporation parameter (G; see Table IV), which is computed from S, P, A, W, and R, and thus illustrates the extent to which surface fluxes can be computed from the individual observations

month there are general coverage decreases from 1990 onward, in contrast to increases in sampling at 50 and 100 (or more) observations per month. This is probably due to factors such as decreased reports from ships, changing ship routes, and densely spaced observations from moored, or slowly drifting, buoys.

The sampling densities for other selected variables follow similar overall patterns, but differences also exist (Figure 5). Relative humidity data are virtually non-existent prior to about 1910, sea-level pressure is less reported than SST, air temperature, wind speed, and total cloudiness until the 1930s, and SST is the best-sampled variable in the last 10 years (probably due to the impact from drifting buoys). The degree to which ICOADS can support surface fluxes estimates is illustrated with a derived evaporation variable, which requires the availability of several variables at the same observation time (see Figure 5). The coverage of this evaporation parameter closely tracks that of relative humidity, the least prevalent of its constituent variables, but is generally at a lower level, particularly in more recent years.

The distribution of samples within the day has also varied notably over time and illustrates another type of observing-system change. In Figure 6, the UTC hour time for each observation is used to bin the data and compute the percentage of reports in each hour interval over the global. Prior to 1854 (not shown), many of the data were reported (or digitized) only once a day at local noon (Woodruff *et al.*, in press). Between 1854 and roughly 1930, many of the available data were reported six times daily according to local time (Parker *et al.*, 1995), which yields a fairly uniform distribution globally in UTC. An exception is the pattern for 12 UTC, which is due to the US 1912–46 Merchant Marine Collection (Table II, decks 705–707) in which observations were taken largely or exclusively at Greenwich Mean Noon (GMN) into the 1930s. Around 1933–41 there is a noticeable shift, by 3 h, of some reports from the UTC synoptic hours (0, 6, 12, 18 UTC). This is primarily due to Japanese ships in the Kobe Collection (Table II, deck 118) coupled with the otherwise relatively sparse data during that period. After World War II, in contrast, most reporting was on a UTC synoptic schedule. This results in dominance, at nearly 25%, of the observations at the 0, 6, 12, 18 UTC hours. The decrease in six-hourly synoptic data starting about 1980 and corresponding increase in three-hourly or hourly data is the effect of increasing numbers of buoy data reports.

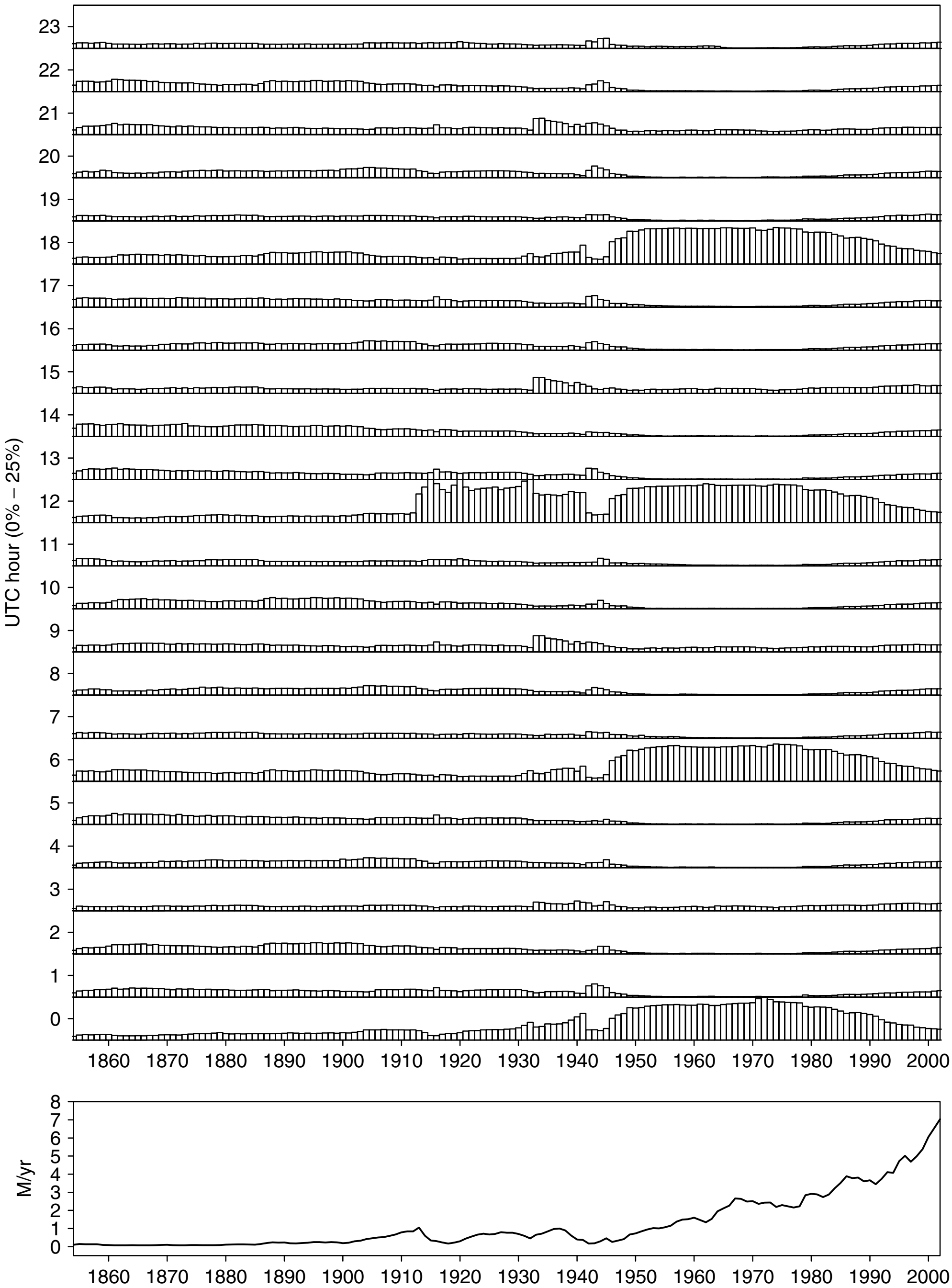

Figure 6. Global percentage of reports per UTC hour, for 1854–2002. Amplitude between horizontal reference lines is equivalent to 25%. The bottom panel is the total number of reports. This is an update of Woodruff *et al*. (1998: figure 4)

5. OBSERVATIONS AND PRODUCTS

5.1. ICOADS release 2.1

The ICOADS observational data formats contain many data and fields, including those for standard meteorological instrument readings (e.g. SST, air temperature, dew-point or wet-bulb temperature, and sea-level pressure), winds (measured or visually estimated), weather, cloudiness, and sea state (wind wave and swell) observations. Methods and accuracy of sampling are preserved in the records where they have been provided or are known. Additional details about all the fields are available at the ICOADS Website.

The observations are condensed into monthly summary products in 2° and 1° latitude by longitude areas (boxes). The products are unanalysed (i.e. data-void areas are not filled) with the 2° and 1° resolution products beginning in 1800 and 1960 respectively. Prior to 1960 (1800) the data were not sufficiently dense to support a global 1° (2°) product. Eight observed and 14 derived variables (Table IV) are summarized using 10 statistics (Table V), forming global year–month time series at each spatial resolution.

Not all observations are used to create the products: the observations are compared with climatological values and outliers are rejected (trimmed) from the computation following the QC procedures designed for release 1 (Slutz *et al*., 1985) and modified to handle newly added data and time periods outside 1854–1979 better (Woodruff *et al*., 1993, 2003). Briefly, the automated QC processing performs some consistency checks, compares commonly used variables against large-scale climatologies, checks for position errors that place observations on land, and adds QC flags to the records. Data editing and platform (ship or buoy) track checks are not performed during ICOADS processing (but are applied by some data providers). Nor are data adjustments made for biases due to sampling methods or instrumentation.

Table IV. Variables in the ICOADS monthly summary products. Eight variables are referred to as 'observed' (related more closely to the original observations), and the remaining 14 are derived quantities

Abbreviation	Variable
'Observed'	
S	SST
A	Air temperature
W	Scalar wind
U	Vector wind eastward component
V	Vector wind northward component
P	Sea-level pressure
C	Total cloudiness
Q	Specific humidity
Derived[a]	
R	Relative humidity
D	$S - A$ = sea–air temperature difference
E	$(S - A)W$
F	$Q_S - Q$ = (sat. Q at S) $- Q$
G	$FW = (Q_S - Q)W$ (evaporation parameter)
X	WU (wind stress)
Y	WV (wind stress parameter)
I	UA (sensible-heat — transport parameter)
J	VA (sensible heat — transport parameter)
K	UQ (latent-heat — transport parameter)
L	VQ (latent heat — parameter)
M	FU
N	FV
B_1	$B = W^3$ (high resolution)
B_2	$B = W^3$ (low resolution)

[a] Note that transfer coefficients have not been applied to the pseudo-fluxes and other derived variables, e.g. UQ is simply the product of individual observations of U and Q.

Table V. Statistics, computed for each variable, in the ICOADS monthly summary products

Abbreviation	Statistic
s_1	1/6 sextile (estimate of $m - 1s$)
s_3	3/6 sextile (the median)
s_5	5/6 sextile (estimate of $m + 1s$)
m	Mean
n	Number of observations
s	Standard deviation[a]
d	Mean day-of-month of observations
h_t	Fraction of observations in daylight
x	Mean longitude of observations
y	Mean latitude of observations

[a] A separate estimate e of the standard deviation can be computed as $e = (s_5 - s_1)/2$.

There are two statistical product lines derived by applying different trimming limits and accepting different platform type mixtures. Secondary QC flags, set during other ICOADS processing or by data providers, are also applied. The standard trimming option is designed to select observations taken from ships and rejects data that exceed 3.5 estimated standard deviations around the smoothed median value for each box and climatological period (1854–1909, 1910–49, and 1950–79; and using the earliest or latest climatological period for trimming data before 1854 or after 1979). The enhanced trimming option is designed to select almost all available observations (e.g. ships and buoys) and rejects data that exceed 4.5 estimated standard deviations around the smoothed median value. The standard product is preferred when a consistent platform type (i.e. ships) is desired, whereas the enhanced product is best when maximum spatial coverage is critical. The degree of trimming can affect the products. For example, the more restricted standard trimming has been shown to remove excessive data in extreme SST events, e.g. 1877–78 and 1982–83 El Niño–southern oscillation occurrences (Wolter, 1997; Smith and Reynolds, 2003).

5.2. Associated products and analyses

ICOADS has served as a foundation for many associated products and analyses developed by the scientific community, with efforts frequently focused on additional QC, the reduction of data biases, and large-area analyses (with error estimation). These products are often specific to a single variable (e.g. SST or wind) or important derived quantities (e.g. heat budgets), and, bearing in mind the limitations and uncertainties of these advanced methods (e.g. Hurrell and Trenberth, 1999; Folland *et al*., 2001b), can be beneficial for many users.

SST, for example, is a critical climate variable that has received extensive analysis development, in comparison with the ICOADS monthly summaries. ICOADS observations are a primary data source in analyses extending back into the 19th century, such as Kaplan *et al*. (1998), Rayner *et al*. (2003), and Smith and Reynolds (2003, 2004a), as well as in analyses requiring satellite data (from around 1981 forward), e.g. Reynolds *et al*. (2002). Uncertainties should decrease closer to near-current dates (e.g. from denser and more accurate sampling) — but note that these products also employ different QC and analysis methods, rely to varying degrees on satellite data, on sea-ice data to constrain polar SST, and on bias adjustments for historical changes in measurement methods.

Similar improvements have been made in derived quantities like ocean–atmosphere fluxes. A number of flux and other climatologies were developed in the 1980s, which advanced and analysed the ICOADS monthly summary statistics (e.g. Oberhuber, 1988; Wright, 1988b). Subsequently, analysed products, such as monthly tropical wind pseudo-stresses (Smith *et al*., 2004), have been created using ICOADS observations. Or bias-adjusted observations have been used in products such as monthly flux analyses for 1945–93 (da

Silva *et al*., 1994), and climatologies constrained with estimated ocean advective fluxes (Josey *et al*., 1999; Grist and Josey, 2003) or utilizing other improvements, including time-dependent adjustments of the Beaufort wind scale (Lindau, 2001).

Some meteorological variables in ICOADS are subjected to limited QC and are not reduced to monthly summary statistics. Two such variables are clouds (with the exception of total cloudiness) and ocean waves. These data must be independently evaluated to produce good quality products. For example, North Atlantic and global wave climatologies (Gulev and Hasse, 1998; Gulev *et al*., 2003), and global distributions of total cloud cover and cloud-type amounts over the ocean have been derived (Warren *et al*., 1988) from ICOADS.

There are many more associated data products that are not mentioned here. The complementary development of associated products is possible because the ICOADS project has always offered the individual observations and simple forms of monthly summary statistics. Increased computing capability has enabled many research projects to begin with the observations and develop the specific analyses they require. We expect and support the continued development of products associated with ICOADS. By this process, more is learned about the characteristics of the data, higher quality products are made available, and the scientific content of ICOADS is available to a broader research community.

6. DATA ACCESS

The ICOADS observations and monthly statistics have unrestricted distribution worldwide and are readily available. Only the data access options provided by the US ICOADS partners are described here. These sources are generally the most up to date and provide supporting sources for documentation, software, and consultation. A single point for data access originates from the project Website. Table VI outlines more specifically the roles of these contributing partners, and the products, available data formats, subsetting and filtering options, and online locations. In general, NCAR provides binary and ASCII formatted data files for observations and monthly summary statistics. The ASCII file data processing enables subsetting by time, space, variable, and a choice of filtering options (standard, enhanced, or self-determined) in the case of observations. CDC provides data in netCDF file format, and through a user interface (Live Access Server) offers subsetting, graphical, and multiple file format outputs for the monthly summary statistics.

In addition, ICOADS is supplemented with observations and monthly summary products (see Table VI), which fall outside release 2.1. When possible, and prior to full ICOADS data processing, new observations are made available and denoted as supplementary (or 'add on') collections. These supplementary observations are largely unique and not yet blended into a formal release, but are provided in standard ICOADS formats. Through this approach we are able to provide access to new data collections. Current offerings include data from the Climatological Database for the World's Oceans (García-Herrera *et al*., in press), which was focused on rescuing early (1750–1854) European ship data (mainly text-based wind estimates), and the latest (2003) edition of the Kobe Collection. Similarly, to provide data to more current dates, a preliminary set of observations and 2° monthly summary statistics are available from a GTS-based NCEP real-time (NRT) dataset. The data processing is simplified and the observational format is a brief ASCII record.

The long-term scientific usage and the inherently complex composition of ICOADS have resulted in a substantial collection of documentation, higher level derived data products, database analyses, and ancillary data. The project Website provides access to these data, metadata, and other resources:

- Format descriptions and software.
- Project publications (most in digital form), and a selected research bibliography.
- Analyses of data processing metrics generated when creating new releases.
- Definition of some outstanding problems.
- Experimental data sets, e.g. global monthly SST (1800–2002) was summarized in 2° boxes using an adaptive QC method where the reference climatology used to trim outlier data varied slowly in space and time according to the large-scale climate signal. This is in contrast to the fixed reference climatologies used to generate the standard and enhanced data products.

Table VI. ICOADS data, metadata, and documentation access

Product	Format[a]	Partner	Subset[b]	Online site
Links to all data				www.cdc.noaa.gov/coads/products.html
Observations:				
Release 2.1	b,a	NCAR	t,s,v	dss.ucar.edu/datasets/ds540.0/
Release 2.1	a	NCDC	t,s	[www.ncdc.noaa.gov][c]
Supplementary[d]	b,a	NCAR	none	dss.ucar.edu/datasets/ds530.0/
NRT[e]	a	CDC	none	www.cdc.noaa.gov/coads/nrt.html
Monthly Statistics:				
Release 2.1	b,a	NCAR	t,s,v	dss.ucar.edu/datasets/ds540.1/
Release 2.1	n,a,o	CDC[f]	t,s,v[g]	www.cdc.noaa.gov/coads/products.html
NRT	n,a,o	CDC[f]	t,s,v[g]	www.cdc.noaa.gov/coads/products.html
Metadata				www.cdc.noaa.gov/coads/metadata/
Documentation				www.cdc.noaa.gov/coads/doc.html
Publications				www.cdc.noaa.gov/coads/publications.html
Workshop recommendations				www.cdc.noaa.gov/coads/climar2/recs.html

[a] Data formats: binary (b), ASCII (a), netCDF (n), others (o) (spreadsheet, etc.).
[b] Subsetting and filtering options: time (t), space (s), and variable (v).
[c] NCDC plans to offer Web-based access to the observations beginning in late 2005.
[d] Supplementary (or 'add-on') data sets are separate collections with a high percentage of unique data that were not included in release 2.1.
[e] NCEP real-time (NRT) data, and 2° statistics, are based on GTS data and updated each month (about 5 days after the end of the data month). The observational records contain only a few primary variables and have received only limited QC, and the 2° monthly summaries (the mean and number of observations for each of 13 variables) are not fully consistent with ICOADS processing. These products provide a preliminary extension of release 2.1 data to near-current dates, i.e. beyond 2002.
[f] CDC stores both the release 2.1 and NRT statistics in standardized netCDF format. These products can be accessed either through CDC Web interfaces, or unified access is provided through a Live Access Server (LAS). The LAS is an interactive interface for selecting data for graphical display or file download.
[g] Each statistic and variable is stored and provided in a separate file, e.g. the mean and standard deviation of SST are two separate netCDF files.

- Links to analysed data sets based on ICOADS and published by other projects, e.g. extended reconstructed SST (Smith and Reynolds, 2003, 2004a) and sea-level pressure beginning in 1854 (Smith and Reynolds, 2004b), and 1° climatologies for many variables from the Joint Institute for the Study of the Atmosphere and Ocean (JISAO).
- Metadata for ocean platforms, e.g. ship metadata (1955–98) from WMO (1955) Publication No. 47.
- Links to related projects and initiatives that complement ICOADS, e.g. the CLIWOC project (www.ucm.es/info/cliwoc/) and the Voluntary Observing Ship Climate (VOSClim) project (http://www.ncdc.noaa.gov/oa/climate/vosclim/vosclim.html).
- Consolidated recommendations from recent ICOADS-related workshops (Diaz *et al*., 2002; Parker *et al*., 2004).

7. FUTURE OBJECTIVES

The major thrusts for ICOADS are adoption of the IMMA format for all observational data, a real-time access system at NCDC, implementation of a regular update cycle to include new data sources, focusing initially on annual RT updates, and improved metadata.

The IMMA format (Woodruff, 2004) is likely to be recommended soon as a standard format to the Joint WMO/Intergovernmental Oceanographic Commission (IOC) Technical Commission for Oceanography and

Marine Meteorology (JCOMM). This ASCII-based format is ideal for marine surface observations spanning many centuries and has been a coordinated development between ICOADS and JCOMM. It has an expanded fixed-length 'core' part that contains more data fields than in the previously available ICOADS formats. By itself, the core will be sufficient for many data users, because it includes most of the regularly used data. The full IMMA format consists of the core plus an arbitrary number of data attachments. The attachments can be general (e.g. ship metadata) or they can be specific to data sources or time periods (e.g. historical or contemporary data). These attachments make the records longer and of variable length, but they serve a critical role for a permanent archive by providing flexibility and retention of original data forms. The GTS data received at NCDC, for example, are translated to the IMMA format and the original GTS message is attached. This means that reprocessing of the original data is always possible, in the event of an incorrect or incomplete translation into the IMMA regular fields. The CLIWOC project translated their database into IMMA format, and ICOADS would benefit from having other data contributors begin using IMMA so that data are translated most accurately (i.e. by technicians most familiar with the data) and they can be incorporated more directly into ICOADS.

NCDC is preparing to serve the ICOADS observations through its Climate Data Online system (cdo.ncdc.noaa.gov) beginning in late 2005. The more recent data will be online initially, followed gradually by the full period of record for release 2.1. The system will allow spatial and temporal subset selections of ICOADS observations and other *in situ* data sets, and a geographical information system (GIS) map service interface will make these data easily available to the GIS user community. Later, in cooperation with the US Navy, Web-accessible data summaries will be offered, e.g. for a specified time, space, and ICOADS variable, frequency distributions and other statistics will be computed on demand.

ICOADS has developed and received data from many national and international data contributors in recent years. Integration of these into ICOADS is planned (Figure 7). However, each new input, and especially the oldest collections, requires careful study, analysis, and assessment before blending into ICOADS. Historical documentation about sampling practices, instrumentation and calibrations, and coding schemes are all critically

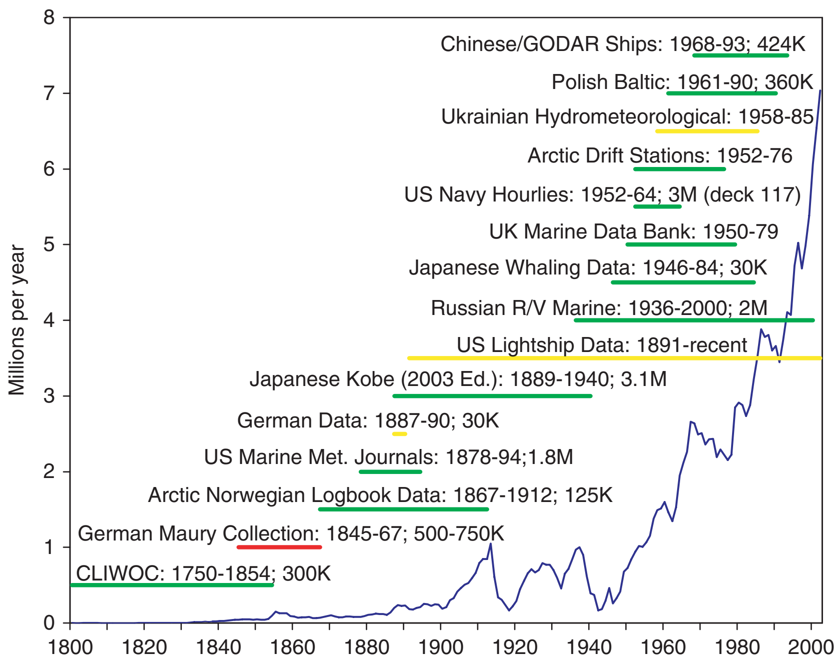

Figure 7. The time periods of candidate datasets to be blended into ICOADS are spanned by horizontal coloured lines: green candidates are fully digitized and may need format translation and QC work, yellow are partially digitized, and red are in the planning stages for digitization. Each data-set name is appended with the date range and approximate number of observations if known. The solid curve is the number of observations in release 2.1

important. This data detective work does not always yield a definitive answer, so occasionally the informed best guess is the action taken. Each new data integration processing is compared with earlier releases, and significant anomalies (e.g. changes in monthly mean values) are investigated in more detail. Sometimes the data anomalies are reasonable and at other times they indicate the need for further detective work, correction, and data reprocessing. The ICOADS project detects and fixes many errors, but it also benefits from data user feedback. Close inspection of the data during scientific investigation can reveal some very subtle and important errors. The creation of new ICOADS releases is robust. All the data are drawn from a set of original files and are not based on the set of records in the most current release. For release 2.1 approximately 100% more data were processed than were passed to the final output. This is largely due to a high percentage of duplicate or near-duplicate records in the various GTS data streams, and the inclusion of other large national archives (e.g. the UK MDB and Russian MARMET) that contain many of the same data.

Increased metadata are important for more accurate future long-term scientific investigations that are based on ICOADS. Some metadata are included in ICOADS observational records. These basic 'indicators' (e.g. SST measurement method) are included whenever they are available, but they often are absent, and occasionally questions exist about the coding, reporting, or past processing of these indicators. More research on these metadata, with associated studies of coding and reporting practices, could improve the metadata quality. Platform descriptions and instrumental metadata are also critically important. The major source for ship metadata is WMO (1955), which was published only in paper form prior to 1973. Digital forms of these metadata for 1973–98 are available in separate annual files and in several different formats (www.cdc.noaa.gov/coads/metadata/), including a consistent reformatting of the 1973–92 metadata by Kent and Oakley (1995) (more recent metadata, through to early 2004, are available from the WMO). Digital images from the 1955–72 editions of the publication are also available, from a recent project under the Climate Data Modernization Program (managed by NCDC). Ideally, all these ship metadata, and an extension to 2002, should be unified into a standard digital format and attached to appropriate IMMA data records as a supplement. A similar effort for buoy metadata should also be undertaken. The data and metadata would then be tightly coupled and would make it easier for bias adjustment studies and development of higher level homogeneous data sets (e.g. ocean surface winds adjusted to 10 m height).

Working with metadata to correct biases and create more homogeneous observations can be as challenging as working with the data itself. Some advancements have been made using historical ship metadata and modelling ship structure characteristics. More homogeneous SST estimates have been created by understanding the sampling methods and biases introduced by changing instrumentation (Folland and Parker, 1995). Air temperatures and wind estimates have been improved by using metadata that define instrument placement and through studies of air flow and diurnal heating on ships (Kent *et al*., 1993; Yelland *et al*., 1998; Berry *et al*., 2004; Thomas *et al*., 2005). Further refinement and understanding of ship data and metadata is one purpose of the VOSClim project. Through this effort, and through correlations with the historical metadata, it may be possible to prepare further homogeneous data products from ICOADS.

8. CONCLUSION

ICOADS began as a US-led effort and has been successively improved over the past two decades. Sustained and increasing international contributions have significantly benefited the project and the worldwide user community. However, there is much work still to be done. ICOADS needs to be enhanced with the newly available and rescued data, throughout the two-century period of record, and kept up to date with modern data. It is a critical data resource for studying climate and climate change, but improvements in the data, metadata, and quality control are necessary to enable an even greater benefit to the scientific community.

ACKNOWLEDGEMENTS

The original COADS project, and the continuing US contribution toward the new international database, ICOADS, is the result of a cooperative project between NOAA — specifically its Office of OAR/CDC, its NESDIS/NCDC, and the CIRES, conducted jointly with the University of Colorado — and NCAR. The

NOAA portion of ICOADS is currently supported by the OGP, and by the ESDIM and CDMP programmes. We wish to thank the two anonymous reviewers, whose comments and suggestions helped improve the paper. We are grateful to the many valued national and international contributors to ICOADS, including Jim Arnott, Henry Diaz, Joe Elms, Florence Fetterer, Ricardo García-Herrera, Guo Fengyi, Mike Jackson, Frits Koek, Günther Können, Linda Mangum, Forrest Miller, Xiao-Wei Quan, Marsel Shaimardanov, Tom Smith, Alex Sterin, Hendrik Wallbrink, Dennis Wheeler, and Erik Wishman. Long-standing assistance has been received from Chris Folland and David Parker at the UK Met Office, Robert Keeley (MEDS), Elizabeth Kent (Southampton Oceanography Centre), Sydney Levitus and the Ocean Climate Laboratory of NOAA's National Oceanographic Data Center (NODC), Teruko Manabe at the WMO and formerly at the Japan Meteorological Agency, and Val Swail at Environment Canada.

REFERENCES

Note: references followed by † are available online at: http://www.cdc.noaa.gov/coads/publications.html or from the underlying selected research bibliography.

Barnett TP. 1984. Long-term trends in surface temperature over the oceans. *Monthly Weather Review* **112**: 303–312.

Berry DI, Kent EC, Taylor PK. 2004. An analytical model of heating errors in marine air temperature from ships. *Journal of Atmospheric and Oceanic Technology* **21**(8): 1198–1215.

Cardone VJ, Greenwood JG, Cane MA. 1990. On trends in historical marine wind data. *Journal of Climate* **3**: 113–127.

Corbett JJ, Koehler H. 2003. Improving the accuracy of ship emissions inventories. *Journal of Geophysical Research* **108**(D20): 4650. DOI: 10.1029/2003JD003751.

Corbett JJ, Fischbeck PS, Pandis SN. 1999. Global nitrogen and sulfur inventories for oceangoing ships. *Journal of Geophysical Research* **104**(D3): 3457–3470.

Diaz H, Folland C, Manabe T, Parker D, Reynolds R, Woodruff S. 2002. Workshop on advances in the use of historical marine climate data. *WMO Bulletin* **51**(4): 377–380.†

Diaz HF, Woodruff SD (eds). 1999. *Proceedings of the International Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata (Toledo, Spain, 15–17 September 1997)*. WMO/TD–No.957, MMROA Report No. 43. World Meteorological Organization: Geneva.†

Dunbar RB, Wellington GM, Colgan MW, Glynn PW. 1994. Eastern Pacific sea surface temperature since 1600 A.D.: the δO^{18} record of climate variability in Galapagos corals. *Paleoceanography* **9**(2): 291–315.

Folland CK, Parker DE. 1995. Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society* **121**: 319–367.

Folland CK, Karl TR, Christy JR, Clarke RA, Gruza GV, Jouzel J, Mann ME, Oerlemans J, Salinger MJ, Wang S-W, *et al*. 2001a. Observed climate variability and change. In *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai K, Maskell K, Johnson CA (eds). Cambridge University Press: Cambridge; 99–181.

Folland CK, Rayner NA, Brown SJ, Smith TM, Shen SSP, Parker DE, Macadam I, Jones PD, Jones RN, Nicholls N, Sexton DMH. 2001b. Global temperature change and its uncertainties since 1861. *Geophysical Research Letters* **28**: 2621–2624.

García-Herrera R, Können GP, Wheeler D, Prieto MR, Jones PD, Koek FB. In press. CLIWOC: a climatological database for the world's oceans 1750–1854. *Climatic Change*.

Grist JP, Josey SA. 2003. Inverse analysis adjustment of the SOC air–sea flux climatology using ocean heat transport constraints. *Journal of Climate* **16**: 3274–3295.

Gulev SK, Hasse L. 1998. North Atlantic wind waves and wind stress fields from voluntary observing ship data. *Journal of Physical Oceanography* **28**(6): 1107–1130. DOI: 10.1175/1520-0485(1998)028.

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 global wind wave climatology from the VOS data. *Journal of Geophysical Research* **108**(C7): 3236. DOI: 10.1029/2002JC001437.

Hahn CJ, Warren SG, London J. 1995. The effect of moonlight on observation of cloud cover at night, and application to cloud climatology. *Journal of Climate* **8**: 1429–1445.

Hurrell JW, Trenberth KE. 1999. Global sea surface temperature analyses: multiple problems and their implications for climate analysis, modeling, and reanalysis. *Bulletin of the American Meteorological Society* **80**: 2661–2678.

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. *Journal of Climate* **12**: 2856–2880.

Kaplan A, Cane MA, Kushnir Y, Clement A, Blumenthal M, Rajagopalan B. 1998. Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research* **103**: 18 567–18 589.

Kaplan A, Kushnir Y, Cane MA. 2000. Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *Journal of Climate* **13**: 2987–3002.

Kent E, Oakley D. 1995. Report on characteristics of WMO47 metadata. Southampton Oceanography Centre, James Rennell Division, internal report.

Kent EC, Taylor PK. 1997. Choice of a Beaufort equivalent scale. *Journal of Atmospheric and Oceanic Technology* **14**: 228–242.

Kent EC, Taylor PK, Truscott BS, Hopkins JS. 1993. The accuracy of voluntary observing ships' meteorological observations — results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology* **10**: 591–608.

Kistler R, Kalnay E, Collins W, Saha S, White G, Woolen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van del Dool H, Jenne R, Fiorino M. 2001. The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society* **82**: 247–267.

Levitus S, Boyer TP. 1994. *World Ocean Atlas 1994, Volume 4: Temperature*. *NOAA Atlas NESDIS* 4. US Government Printing Office: Washington, DC.

Levitus S, Boyer TP, Conkright ME, O'Brien T, Antonov J, Stephens C, Stathoplos L, Johnson D, Gelfeld R. 1998. *World Ocean Database 1998: Volume 1: Introduction*. *NOAA Atlas NESDIS* 18. US Government Printing Office: Washington, DC.

Lindau R. 2001. *Climate Atlas of the Atlantic Ocean: Derived from the Comprehensive Ocean–Atmosphere Data Set (COADS)*. Springer: Berlin.

Manabe T. 1999. The digitized Kobe Collection, phase I: historical surface marine meteorological observations in the archive of the Japan Meteorological Agency. *Bulletin of the American Meteorological Society* **80**: 2703–2715.

Oberhuber JM. 1988. An Atlas Based on the 'COADS' data set: the budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. Report No. 15 (March 1988), Max-Planck-Institute für Meterologie, Hamburg.

Parker D, Kent E, Woodruff S, Dehenauw D, Harrison DE, Manabe T, Meitus M, Swail V, Worley S. 2004. The Second JCOMM Workshop on Advances in Marine Climatology (CLIMAR-II). *WMO Bulletin* **53**(2): 157–159.†

Parker DE, Folland CK, Jackson M. 1995. Marine surface temperature: observed variations and data requirements. *Climatic Change* **31**: 559–600.

Ramage CS. 1987. Secular changes in reported surface wind speeds over the ocean. *Journal of Climate and Applied Meteorology* **26**: 525–528.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research* **108**(D14): 4407. DOI: 10.1029/2002JD002670.

Restrepo JD, Kjerfve B, Correa ID, González J. 2002. Morphodynamics of a high discharge tropical delta, San Juan River, Pacific coast of Columbia. *Marine Geology* **192**: 355–381.

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W. 2002. An improved *in situ* and satellite SST analysis for climate. *Journal of Climate* **15**: 1609–1625.

da Silva AM, Young CC, Levitus S. 1994. *Atlas of Surface Marine Data 1994, Volume 1: Algorithms and Procedures*. NOAA Atlas NESDIS 6. US Government Printing Office: Washington, DC.

Slutz RJ, Lubker SJ, Hiscox JD, Woodruff SD, Jenne RL, Joseph DH, Steurer PM, Elms JD. 1985. Comprehensive Ocean–Atmosphere Data Set; Release 1. NOAA Environmental Research Laboratories: Boulder, CO.†

Smith SR, Servain J, Legler DM, Stricherz JN, Bourassa MA, O'Brien JJ. 2004. *In situ*-based pseudo-wind stress products for the tropical oceans. *Bulletin of the American Meteorological Society* **85**: 979–994.

Smith TM, Reynolds RW. 2002. Bias corrections for historical sea surface temperatures based on marine air temperatures. *Journal of Climate* **15**: 73–87.

Smith TM, Reynolds RW. 2003. Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *Journal of Climate* **16**: 1495–1510.

Smith TM, Reynolds RW. 2004a. Improved extended reconstruction of SST (1854–1997). *Journal of Climate* **17**: 2466–2477.

Smith TM, Reynolds RW. 2004b. Reconstruction of monthly mean oceanic sea level pressure based on COADS and station data. *Journal of Atmospheric and Oceanic Technology* **21**(8): 1272–1282.

Thomas BR, Kent EC, Swail VR. 2005. Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology* **25**: 979–995; this issue.

Uppala S. 2002. ECMWF reanalysis, 1957–2001, ERA-40. In *Workshop on Re-analysis*, 5–9 November 2001. ERA-40 Project Report Series No. 3, European Centre for Medium-Range Weather Forecasts.

Warren SG, Hahn CJ, London J, Chervin RM, Jenne RL. 1988. Global distribution of total cloud cover and cloud type amounts over the ocean. NCAR Technical Note 317+ STR (December 1988), National Center for Atmospheric Research: Boulder, CO.

WMO. 1955. *International List of Selected, Supplementary and Auxiliary Ships*. WMO-No. 47. WMO: Geneva. (Serial publication; recently annual. Editions prior to 1966 were entitled *International List of Selected and Supplementary Ships*.).

WMO. 1995. *Manual on Codes*. WMO-No. 306. WMO: Geneva (including supplements through to no. 3 (VIII.2001)).

Wolter K. 1997. Trimming problems and remedies in COADS. *Journal of Climate* **10**: 1980–1997.

Woodruff SD. 2004. Archival of Data Other than in IMMT format: the International Maritime Meteorological Archive (IMMA) format. Report submitted to the first session of the JCOMM Expert Team on Marine Climatology (ETMC-I) Gdynia, Poland (7–10 July 2004) (update of JCOMM-SGMC-VIII/Doc.17, Asheville, NC, USA (10–14 April 2000)). http://www.cdc.noaa.gov/coads/e-doc/imma/ [accessed 26 June 2004].

Woodruff SD, Slutz RJ, Jenne RL, Steurer PM. 1987. A comprehensive ocean–atmosphere data set. *Bulletin of the American Meteorological Society* **68**: 1239–1250.

Woodruff SD, Lubker SJ, Wolter K, Worley SJ, Elms JD. 1993. Comprehensive Ocean–Atmosphere Data Set (COADS) release 1a: 1980–92. *Earth System Monitor* **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. *Physics and Chemistry of the Earth* **23**: 517–526.†

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 Meteorology*. WMO/TD-No. 1081 (JCOMM Technical Report No. 13), 3–10 (CD-ROM).†

Woodruff SD, Diaz HF, Worley SJ, Reynolds RW, Lubker SJ. In press. Early ship observational data and ICOADS. *Climatic Change*.

Wright PB. 1986. Problems in the use of ship observations for the study of interdecadal climate changes. *Monthly Weather Review* **114**: 1028–1034.

Wright PB. 1988a. On the reality of climatic changes in wind over the Pacific. *Journal of Climatology* **8**: 521–527.

Wright PB. 1988b. An atlas based on the 'COADS' data set: fields of mean wind, cloudiness and humidity at the surface of the global ocean. Report No. 14 (March 1988), Max-Planck-Institute für Meterologie, Hamburg.

Yelland MJ, Moat BI, Taylor PK, Pascal RW, Hutchings J, Cornell VC. 1998. Wind stress measurements from the open ocean corrected for air flow distortion by the ship. *Journal of Physical Oceanography* **28**(7): 1511–1526.

