Updating COADS-Problems and Opportunities

Scott D.Woodruff and Sandra J.Lubker
NOAA/Climate Monitoring and Diagnostics Laboratory
Mei Yuen Liu Cooperative Institutefor Research in Environmental Sciences

Abstract

Release 1 of the Comprehensive Ocean-Atmosphere Data Set (1854-1979)—the most extensive record now available of past climatic conditions at the ocean's surface has been provisionally extended through 1991 by a limited set of COADS "interim" products. The availability of important new data sources, coupled with advances in computer storage and processing technology, present opportunities to enhance the coverage of COADS in time and space, and to restructure COADS products as required to better meet the needs of climate researchers.

Planned updates are described that will incorporate a revised update strategy, with the aim to make available from each update a consistently processed data set, to the maximum extent possible, for the total period of record. Release la is the initial stage of the new strategy. This extension of Release 1 products through 1991 is planned for completion by the end of 1992. An update of the basic observational data (individual marine reports) back to World War II, Release 1b, is planned for availability during 1993. Release 2, a major phase in the new strategy, will involve a complete reprocessing of the period 1854-date, to include new data sources, to utilize improved processing techniques, and to incorporate newly defined user requirements.

Release 2 is planned for completion in the mid 1990s; however, the exact timing is dependent on three major considerations:

- Unresolved data problems. Examples are presented, such as random and systematic mislocation of ship reports.
- Platform and diurnal inhomogeneities. Examples, such as differences between data from ships versus buoys, or between data from daytime versus nighttime, are presented from a recently completed NOAA report.
- The extent and complexity of newly defined user requirements, such as the need for finer spatial or temporal resolution in summary statistics for some variables. As background, results are described from a questionnaire distributed in preparation for this workshop.

1. Introduction

Since completion of Release 1 of the Comprehensive Ocean-Atmosphere Data Set (1854-1979), a set of COADS "interim" products has been produced and updated to extend key portions of the data set past 1979.[1] These products are generally compatible in format and organization with other COADS products, but were constructed using simplified procedures and preliminary input data. The interim products are now widely used, e.g., 1980-87 interim data are included on the World WeatherDisc CD-ROM (WeatherDisc Associates, 1989), and meet many research requirements for COADS data extending past 1979.

However, new or improved data sources are now becoming available that could serve to enrich both the relatively modem COADS record (1950-date) and the earlier historical period (back to 1854, or earlier). Improvements in the coverage and quality of the surface marine archive could benefit major new projects such as a global reanalysis of atmospheric data (Jenne, 1992), whose products will be widely used. In addition, knowledge has grown of some systematic data problems that impact the existing COADS products, e.g., resulting from data transmission or processing errors, or from changes in instrumentation or observing practices. For example, biases between data from different platform types and times-of-day, and resulting from quality controls that were overly restrictive when applied to data from extreme climate events, are now better recognized and will require enhancements or additions to COADS products.

Fortunately, the passage of time since Release 1 of COADS has also brought extensive improvements in computer technology. These include the emergence of inexpensive and powerful hardware storage devices and computing platforms, which should allow updates to be made more frequently and effectively, and enable implementation of widely needed improvements and corrections to COADS products. Further technological advances include hypertext-based systems (Moninger et al., 1992), which may permit more effective collection of and access to electronic "metadata" (information about COADS, including comments and technical documentation). Other systems under development incorporate advances in visualization techniques and network-based data distribution (Mock and Messenger, 1992).

This paper describes a revised update strategy for COADS that has been designed with these scientific considerations and technological advances in mind, with the aim to steadily improve the quality and usability of the data set. Background is given on cooperative work with various countries and institutions to collect the needed data inputs, examples are provided on known data biases and errors, and details are included on newly defined user requirements (Appendix).

2. Revised update strategy

In the future we plan to offer COADS products for the total period of record, currently 1854-1991, to which consistent processing methods have been applied. It should be noted that prior to COADS there have been many changes in processing methods, as well as in instrumentation, observational practices, and so forth (e.g., Woodruff et al., 1987). There are also some differences in the interim processing in comparison with Release 1, such as usage of a simplified procedure for elimination of duplicates of individual marine reports. The revised plan seeks to avoid introducing any new inhomogeneities through variations in COADS processing methods within the available record.

In the first stage of the revised update strategy, a 1980-91 update designed to retain maximum consistency with Release 1 (1854-1979) data products is underway. This update, Release la, is

[1] COADS (Slutz et al., 1985; Woodruff et al., 1987) is the result of a continuing cooperative project between the National Oceanic and Atmospheric Administration (NOAA)-its Environmental Research Laboratories (ERL), National Climatic Data Center (NCDC), and Cooperative Institute for Research in Environmental Sciences (CIRES; joint with the University of Colorado)-and the National Science Foundation's National Center for Atmospheric Research (NCAR). COADS products are available from NCAR or NCDC.

planned for completion during 1992. To meet urgent data requirements for projects such as global reanalysis, a preliminary update of the basic observational data back to about 1947, Release 1b, is planned during 1993 (summary statistics are tentatively planned during 1994). By the mid 1990s we plan to re-process the entire period (1854-date), using new and improved techniques as well as additional data inputs; this will form a new set of products to be designated as Release 2.

In addition to a continuing need for major updates to be done periodically (Ramage, 1986), yearly issuance of the interim products has made apparent a wide interest in having full annual updates to COADS summary products. For practical reasons, updates to the interim products have taken the form of appending a single year of data to the end of the existing product series. However, in the future it may become feasible to annually reprocess the last several years of the archive to include more delayed data, such as digitized ship logbook data from foreign countries (frequently delayed 2- 5 years; see Jenne, 1992, this volume).

3. Update data inputs

This section gives an overall description of data inputs planned for Releases 1a, 1b, and 2 of COADS (Table 1). For additional details refer to other papers in this volume (Elms, 1992; Parker, 1992; Uwai and Komura, 1992; Worley, 1992; and Yudin et al., 1992).

a) Data inputs for Release 1a

In contrast to the limited data inputs (see Table 1) used for the interim products, Release la is planned to include much additional or improved data for 1980-91. Data from NCDC will form the bulk of the input for Release la. Included in the NCDC set are U.S.-keyed logbook data, and International Maritime Meteorological (IMM) exchange logbook data obtained by NCDC under WMO's (1963) Resolution 35 or bilateral agreements. Also included are international ship and buoy data transmitted in near real-time over the Global Telecommunication System (GTS). In the U.S., GTS data are gathered by NOAA's National Meteorological Center (NMC), by U.S. Air Force Global Weather Central (GWC), and by the U.S. Navy Fleet Numerical Oceanography Center (FNOC). After declassification, FNOC supplies "Autodin" data from U.S. Navy vessels (see section 5 of this paper for a discussion of problems in GTS inputs).

Table 1. Input data sets planned for inclusion in the basic observational data products for COADS Releases 1a, 1b, and 2. Summary statistical products may exclude some individual marine reports (e.g., coastal, island, or pier data) due to questions about compatibility with open ocean data.

Description	Availability	1a	1b	2
National Climatic Data Center (NCDC annual file:				
U.S. keyed logbook data	July*†	yes	yes	yes
International Maritime Met. (IMM) exchange (logbook) data	July*†	yes	yes	yes
National Meteorological Center (NMC) data (1972-)	July*†	yes	yes	yes
U.S. Navy Autodin (declassified) data (1980)	July*†	yes	yes	yes
U.S. Air Force Global Weather Central (GWC) data (1973-81)	now†	yes	yes	yes
U.S. Navy Fleet Numerical Oceanography Center (FNOC) data	now	no	no	maybe
National Oceanographic Data Center (NODC) surface data (1900-)	now	maybe	yes	yes
National Data Buoy Center (NDBC) data	July*†	yes	yes	yes
USSR ship data (1888-1989)	now	maybe	yes	yes
French International Exchange (IMM) corrections (1954-88)	now	maybe	yes	yes
UK International Exchange (IMM) corrections (1982-89)	now	maybe	yes	yes
TOGA Maritime Climatology Data Set (MCDS (1985-90)	now	no	no	yes
Voluntary Special Obs. Prog. —N. Atlantic (VSOP-NA) (1988-90)	now	no	no	yes
Pacific Maritime Environmental Laboratory (PMEL) data (1980-)	now	yes	yes	yes
Marine Environmental Data Service (MEDS) buoy data (1980-)	now	yes	yes	yes
MEDS daily seawater files (1914-85)	now	no	no	yes
Polar Science Center (PCS) ice buoy data (1979-)	1994	no	no	yes
UK Met. Office Main Marine Data Bank (MDB) (1854-)	Mar. ‘93	no	no	yes
U.S. World War I and II data (1912-46)	Oct. ‘93	no	no	yes
Inter-American Tropical Tuna Commission data (1972-)	Feb. ‘93	no	maybe	yes
ORSTOM Indian Ocean Subsurface Temp. Data Set (1906-90)	now	no	no	yes
German Seewetteramt Data Archive (1826-1945)	?	no	no	yes
German digitized data (1887-1890)	?	no	no	maybe
Japanese Kobe Imperial Marine Observatory data (pre-933)	?	no	no	maybe
Maury Collection (~1796-1861)	?	no	no	maybe
Norwegian Ship Logbooks (1867-1890)	?	no	no	maybe

* The “annual file” of data received by NCDC during a given year (including any delayed data for previous years) is available by each July of the following year. Generally few IMM data for a given year are received during that year; many IMM data are delayed 2-5 (or more) years (see Jenne, 1992).

† Included in the currently available 1980-91 interim products, as follows: U.S.-keyed, Autodin, and NMC data were included for each year. GWC data were included only for 1980-81, and NDBC data were included only for 1980-84. Starting with 1985 data, updates effectively took the form of appending a single year to the previously available record; thus only the relatively incomplete set of IMM data received during a given data year (see previous footnote) could be included, and delayed data were not merged into prior years (cf., Woodruff and Lubker, 1986)

In addition to the NCDC data, NOAA's National Data Buoy Center (NDBC) has available a relatively complete, quality controlled set of data from U.S. moored buoys and from shore stations operating within the Coastal-Marine Automated Network (C-MAN); that set will be included in Release la. In recognition of the need for adequate quality control of the international drifting buoy archive, Canada's Marine Environmental Data Service (MEDS) has cooperated by preparing a global set of drifting (and moored) buoy data using the basic NMC archive for 1980-85, plus drifting buoy data collected by MEDS for 1986-91 (Worley, 1992). Automated platform data also will be included in Release 1a from NOAA's Pacific Marine Environmental Laboratory Atlas buoys and other equatorial moorings or "flat" island stations (Worley, 1992).

b) Data inputs for Release 1b or Release 2

All of the Release la data inputs also are planned for inclusion in Releases 1b and 2, plus any delayed international exchange (IMM) data that may become available in time to be included. In addition, the Tropical Ocean Global Atmosphere (TOGA) Programme is collecting a file of ship logbook and GTS data for 30°N-30°S in its Marine Climatology Data Set (MCDS), and France and the UK have supplied international exchange data corrected for systematic processing errors (see section 4). Ship logbook data also will be included in Release 2 from WMO's Voluntary Special Observing Programme—North Atlantic (VSOP-NA); detailed documentation is available from this project about instrumentation and observing practices aboard the 45 participating ships (Hopkins and Taylor, 1992). Other inputs planned for Release 2 include Canadian "daily seawater" (e.g., pier SST) measurements obtained from MEDS, and data from Arctic buoys on ice collected by the University of Washington's Polar Science Center (PSC), with added quality control by MEDS.

A very extensive addition to Releases 1b and 2 will comprise near-surface sea temperatures taken by expendable and mechanical bathythermographs or other oceanographic sensors, provided by NOAA's National Oceanographic Data Center (NODC) (Worley, 1992). For Release 2, additional near-surface sea temperatures will be extracted from ORSTOM's (Institut Français de Recherche Scientifique pour le Développement en Coopération) Indian Ocean Subsurface Temperature (IOSBT) Data Set, which includes temperature profile data from Indian Ocean tuna fishing vessels.

Because of the sparsity of ships-of-opportunity and other platforms in the equatorial and eastern tropical Pacific, an important development has been an agreement reached between ERL and the Inter-American Tropical Tuna Commission (IATTC) to obtain meteorological data starting in the early 1970s taken aboard tuna fishing boats and also by separate porpoise observers accompanying the tuna fishermen. In the past, severe confidentiality problems have limited access to these data (Sadler, 1986), and some constraints on the release of individual reports remain in place under the terms negotiated between ERL and IATTC. However, IATTC is providing quality controlled data for the eastern and western Pacific for use in COADS, with pseudo-IDs in place of actual ship names (Worley, 1992).

c) Historical data inputs for Release 2

Several important inputs for Release 2 extend back into the 19th Century. For example, global ship data from the former Soviet Union were obtained through a USSR-U.S. bilateral agreement (Yudin et al., 1992). Release 2 will also include additional data being digitized to fill in gaps during World Wars I and II, although it now appears few original U.S. logbooks still exist for World War II (Elms, 1992). Other possible additions to early years include previously undigitized records from the Japanese Kobe Marine Observatory Collection (Uwai and Komura, 1992), from the Maury Collection (Elms, 1992), and from about 600 Norwegian ship logbooks. Germany is also keying ships' logbooks from the late 1800s, and will provide records from the Seewetteramt Data Archive to replace data included in Release 1 from WMO's Historical Sea Surface Temperature (HSST) Data Project; the HSST format unfortunately excluded elements such as present weather and subsidiary cloud fields.

A major element of Release 2 will involve the planned merge of COADS with the UK Meteorological Office Main Marine Data Bank (MDB) for the period 1854 to date. A preliminary comparison of COADS with the UK Meteorological Office Main Marine Data Bank (MDB) (Woodruff, 1990) was completed as a research contribution in support of a scientific assessment by the WMO/UNEP Intergovernmental Panel on Climate Change (IPCC, 1990). Individual ship logbook reports for six selected 10° boxes in the Atlantic, Pacific, and Indian Oceans for 1854-1979 were compared. Checks between the data sets for duplicates suggest the possibility of significant gains from a merge: 63% duplicate; 28%/9% unique from COADS/MDB (thus -9% more data for COADS). As a result detailed planning has been initiated for a merge of COADS with MDB, to be performed in the U.S. in conjunction with Release 2 (Parker, 1992).

The COADS/MDB merge should also offer the opportunity for correction of significant problems in both data sets, some of which were discovered in the course of the IPCC support study. For example, in "Dutch" (source deck 193) data, which make up 12% of the pre-1970 COADS, sea level pressure was not translated from millimeters to millibars by NCDC, while in MDB it appears the UK Met. Office did translate the pressure but failed to make a correction for gravity. These pressure data are not included in COADS Release 1 monthly summaries, although the uncorrected data are included in the supplemental fields of the corresponding ship reports (i.e., with the field shown as missing in the usual report position).

4. Data problems

A number of other biases or errors have been identified in COADS data (e.g., Woodruff et al., 1987; Wolter et al., 1989; Morrissey, 1990; Pierson, 1990). Research efforts and possible collaborative work that may lead to the correction of systematic observing biases, such as incorporation of a revised Beaufort equivalence scale (Cardone et al., 1990; Isemer and Hasse, 1991), are discussed in sections 5 and 7 and elsewhere in this volume; corrections for many such observational biases probably should reside in "value-added" products separate from the basic COADS archive. Systematic data transmission or processing errors, on the other hand, will be identified and corrected, if feasible, in preparation for Release 2. However, correction of some errors may have to be delayed even beyond Release 2.

For example, random spatial or temporal mislocation of ship data has been estimated to occur at rates ranging as high as 6%-8% in GTS data (Jenne, 1986). Mislocated data may be identified during "track checking," in which continuity checks are applied for platform movement through time and space and for changes in weather elements (selected sources, such as MEDS drifting buoy data, have already been track checked). However, current track checking procedures require an adequate ship call sign, or other platform ID, for each individual marine report, which may not always exist even in recent (e.g., 1980s) data. Ship IDs may be less reliable or completely unavailable in many earlier years, due to the use of poorly documented numbering schemes or non-digitization of ID information.[2]

New data problems have been discovered both in the 1854-1979 period and in the interim data through 199 1; for further information, see Elms (1990). One example is that French logbook data during approximately 1954-1988 and within 90°E-90°W across the dateline (~1/4 million reports) are displaced by ten degrees of longitude; this problem was caused by an ambiguity in the WMO format for international exchange, which lead to incompatible conversions between countries. In addition, UK international exchange data have an error in the indicator for the method of sea surface temperature measurement for 1982-89, similar to apparently unresolvable problems in the indicator for U.S. data in many earlier years (Slutz et al., 1985). Fortunately, replacement data have been received from France and the UK to help correct these problems for Releases 1b and 2 (see section 3).

Other problems relate to code differences between GTS and logbook data. For example, in 1982 WMO added the station/weather indicator (i_x) to the GTS code; this indicator is important for proper interpretation of present and past weather fields in order to distinguish between weather not reported because it was "good" (not significant), versus when it was intentionally not reported. Unfortunately, i_x was omitted from the international exchange (logbook) format until March 1985, and some countries may continue to omit this indicator from logbook data due to documentation problems.[3]

Similarly, the ID keyed from logbook data may not always match the ID in GTS data (e.g., logbook reports from ocean weather ships when "on station"). For example, the UK supplies in its logbook data a ship number that has the advantage of being permanently assigned to each UK ship, whereas any matching GTS reports contain the ship's international call sign. It is important to note that annual metadata about ship characteristics (WMO, 1955-) currently are keyed only to call sign. Existing COADS processing to eliminate duplicate reports gives preference to a logbook

---

[2] For example, HSST reports (16% of the pre-1970 Release 1 data) include no form of ship ID. We plan to prepare information about the number of reports for which an ID is available. Although documentation may be lacking for some early ship-numbering schemes, we expect that in many cases a number was assigned only to one vessel in a given period or area, and thus the number will still be useful for track checking.

[3] The station/weather indicator (i_x) provides an example of the problems that can develop when real- time data codes are changed for operational meteorology, in this case to reduce GTS message traffic, without sufficient consideration of the impact on the continuity of long-term data records. From a climatological perspective it is just as important to know the frequency of no significant as for significant weather. When i_x is missing, the data user generally must assume that there was no significant weather to report, thus biasing the results toward fair weather. Note that i_x was not utilized in construction of, and is not available within, the 1980-91 COADS interim Compressed Marine Reports (CMR).

report when a matching GTS report of "equal" quality is identified. However, changes may be necessary to ensure that all useful information is retained before discarding GTS reports in favor of logbook reports.

5. Source and diurnal differences

Extensive comparisons among different surface marine data sets were undertaken in recognition of possible biases between data taken from ships and other platform types (e.g., Wilkerson and Earle, 1990), and between data taken during daytime and nighttime hours (e.g., Folland et al., 1984). A second goal of the comparisons was to assess the value of FNOC data as a near real-time continuation of COADS.[4] A NOAA special report (Woodruff et al., 1991) gives the results of the comparisons, and incorporates a series of recommendations to help insure archival integrity.

In the first part of that study, GTS data from NMC for December 1986 and June 1988 were divided up by platform type (ships, moored buoys, drifting buoys, etc.). Separate monthly summaries for 2° latitude x 2° longitude boxes were then calculated and compared for each platform type, with a further separation for data taken during daytime and nighttime hours. These comparisons confirmed the existence of significant platform and diurnal differences. Examples of such differences are generally higher wind speeds from ships than from buoys, and higher than expected day/night differences in the air temperature compared to that of the sea surface

Secondly, the subsets of NMC data were compared with similar subsets of FNOC data, in terms of temporal and spatial coverage, and in terms of response to duplicate elimination and to other quality control steps (GWC and Autodin data were also included in a more limited set of comparisons). One important finding was that, due to different cutoffs for data receipt and other factors, these GTS data sets are not identical (e.g., ~10% of the ship reports may be missing if only NMC data are included). As a result, FNOC or other GTS data are under consideration as possible future additions to COADS (Release 2 or later). Significant errors or longstanding biases in current GTS data sources were also clearly identified as a result of this work.[5]

[4] Unclassified data from FNOC have been used at ERL since 1987 in production of a monthly *Surface Marine Bulletin* for climate monitoring. GTS ship and buoy reports obtained by FNOC are processed at ERL using essentially the same software used for production of COADS interim products. Comparable GTS data derived from the basic NMC data set are under consideration as a possible supplement for the FNOC data.

[5] COADS currently contains two FNOC source decks: Monterey Telecom (1966-73) and U.S. Navy declassified "Autodin" data (1972-). Other GTS sources in COADS include NMC data (1980-) and U.S. Air Force Global Weather Central (GWC) data (1973-81). See Woodruff et al. (1991) for more information on data problems in these GTS sources.

Table 2. Products planned for COADS Release 1a (products are numbered to correspond with Table 5 in Woodruff et al., 1987). Note that several infrequently used products from Release 1 are not planned for Release 1a. In addition, the format for Compressed Marine Reports (CMR) is planned for replacement by the revised LMR (variable-length, binary) format or a fixed length version of it. A new format also is under development for Release 2 to replace the previous ASCII format (TD1129) for individual observations.

Product	Distribution*
1. Long Marine Reports (LMR)	NCAR
2. Inventories (INV)	"
15. Monthly Summaries Trimmed Timesort (MST.T)	"
16. Monthly Summaries Trimmed Boxsort (MST.B)	"
18. Monthly Summary Trimmed Groups (MSTG)	"

* NCDC may also distribute some of these products in the future.

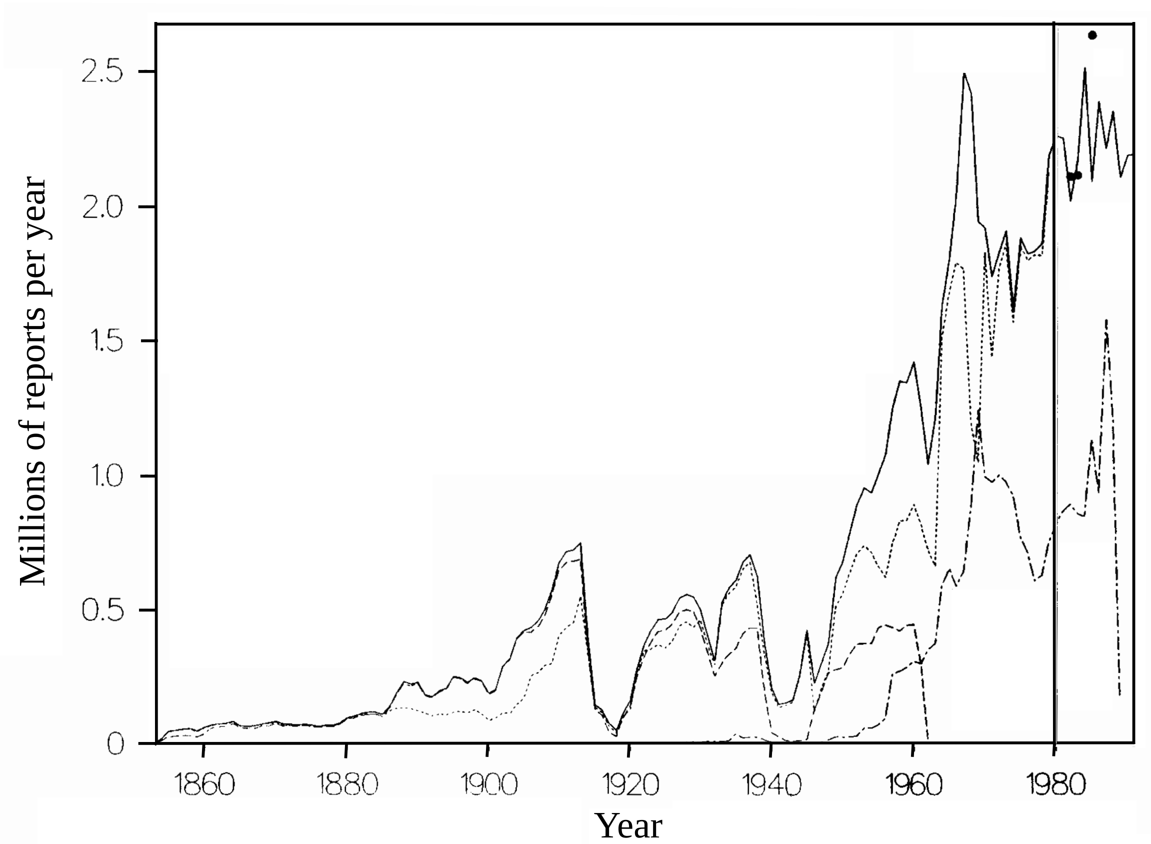

Figure 1. Annual global marine reports after duplicate elimination: for COADS Release 1 (1854-1979) and the COADS interim products (1980-91) (solid curve). Available numbers of reports for preliminary Release 1a data for 1982-83 and 1985 are shown as solid dots. Shown for comparison (see Woodruff et al., 1987) are the Atlas File (dotted, through 1969) continued by the 70s Decade (dotted, 1970-79), and the HSST data set (dashed, through 1961). Also shown is the USSR ship data set through 1989 (dash-dot); see Worley (1992).

6. Releases la and 1b: processing and products

Release la will include a selection of commonly used COADS products (Table 2). Although data products will be consistent in processing with Release 1, various product format improvements are planned. For example, any ship call signs or other platform identifiers will be readily available in a revised packed-binary format for individual marine reports (LMR). In addition, for the convenience of many scientific applications such as global reanalysis, the individual reports will be available in a pure synoptic sort (year, month, day, hour, 10° box).

Work is currently in progress to convert all of the individual marine reports from the input data sets for Release 1a into the revised LMR format, to set quality control flags, and to eliminate or flag duplicate reports. For input to global reanalysis or other projects we now have available relatively complete sets of preliminary data for 1982-83 and 1985, processed through the synoptic sort (Figure 1). The complete 1980-91 observational data set for Release la, and corresponding monthly summaries for 2° boxes, are planned for completion by the end of 1992.

Similarly as for the interim products, we plan to use the 1950-79 trimming limits to flag suspect 1980-91 data in the individual marine reports for Release 1a, and to eliminate such data from the 2° monthly summaries. However, it is clear that using this approach for extreme climate conditions (e.g., tropical Pacific sea surface temperatures during the 1982-83 ENSO event) introduces some unwanted distortion of the statistics (see Wolter, 1992); therefore, an experimental set of statistics with different trimming limits also is planned for availability by late 1993.

To meet urgent data requirements for global reanalysis, Release 1b will extend the basic observational data (products 1 and 2 in Table 2) back to about 1947; this update, including a synoptic order sort, is planned for completion by May 1993. The 1947 starting date has been chosen to avoid the World War II data gap (see Figure 1) and to coincide with the availability of other data types (Jenne, 1992). Depending on the amount of new data available for Release 1b and other considerations, it is likely that we will also update the 2° monthly statistics (products 15, 16, and 18 in Table 2) for the period 1947-date during 1994 as a second phase of Release 1b, which may also incorporate some improvements in the trimming procedure.

7. Release 2: plans and unresolved issues

The general goal for Release 2 is to update COADS for the entire period 1854 (or earlier) to date, to include surface marine data of crucial importance to climate research that are now becoming available. Completion of Releases 1a and 1b will serve as preparatory steps toward Release 2, in providing updates to the basic observational data after World War II. For Release 2, the emphasis is planned on gathering and improving earlier records, since coverage may be very sparse compared to modem coverage (see Figure 1). As part of these efforts, early historical ship records will be included from a variety of international digitization projects or data exchange agreements (section 3), and numerous data corrections or improvements are planned, such as making early ship ID information available whenever possible (section 4).

Another goal for Release 2 is to incorporate changes in technology or widely needed user requirements that have emerged since Release 1 (see Appendix). For example, some users desire summary data for selected variables at resolutions finer temporally than one month, or finer spatially than 2° latitude x 2° longitude boxes (e.g., for 5-day averages and 1° latitude x 1° longitude boxes), at least in regions and times containing sufficiently dense data coverage. Similarly, changes to Release 1 processing are felt to be necessary in the calculation of additional summary statistics that could be used to correct for biases resulting from combining data from different platform types (e.g., ships versus buoys) and times-of-day (see section 5).

However, a practical consideration for proposed separations by platform type and time-of-day, coupled with possible increases in temporal or spatial resolution, is that the resulting files may be rendered virtually unusable in direct form because of too many sparsely populated boxes, and excessive storage volume. Table 3 shows that the total storage volume of statistics after various separations and increases in resolution will tend to exceed the size of the complete set of individual observations. We therefore anticipate that many of the higher resolution studies will use the individual ship reports rather than the statistics.

Table 3. Estimated storage volume for possible new summary statistics for 1980-91 data, in comparison to the corresponding volume of Long Marine Reports (LMR).*

Processing modification	Estimated increase factor	Cumulative volume (gigabytes)	Volume of LMR
(No change)	1	0.401	2.47
5-day resolution	6	2.41	
1° latitude x 1° longitude resolution	4	9.62	
Separation by platform type	~2	19.2	
Separation by daytime/nighttime	2	38.5	

* Initial volume (“no change”) is based on the full MST (binary) format defined for Release 1, which includes 14 statistics for 19 variables, assuming 6,000 (51%) of the world's ocean/coastal 2 boxes contain at least one observation (see Figure 4 in Woodruff et al., 1987). Volume for LMR is estimated to average 95 bytes per report, with 26 million reports estimated during 1980-91. For convenience, volumes are given in gigabytes (109 8-bit bytes). The estimated factors of increase (applied cumulatively) yield volumes at the upper limit, since we have not taken into account that any new boxes without data would actually be suppressed.

Conceivably, one could distribute to some users with specialized requirements, software that would produce statistical summaries from the individual observations, and that would include the flexibility to output summaries at different resolutions or to process subsets of the total observational data set. Providing software tools to enable users to access the individual observations in this way could also help facilitate research into data biases resulting from changes in instrumentation or observing practices, since only at this stage can data or metadata pertaining to the characteristics of individual vessels be effectively utilized (e.g., sea surface temperature measurement method, anemometer height).

Nevertheless, at this stage of planning we expect that day/night statistics for some variables will be produced for Release 2 and possibly back to ~1947 as part of Release 1b, at least for selected variables (sea surface and air temperatures, cloudiness, and wind). We also expect that 1° summaries and submonthly resolution may be available for the last 30-year period, for instance, but perhaps not for all variables. In addition, a “blended” set of statistics will be compiled for general use from the separate or high-resolution files, e.g., where the broadest spatial coverage is desired.

In preparation for Release 2, some fundamental COADS processing changes also are planned. Among these are changes to the duplicate-elimination procedure, because some unique (e.g., hourly buoy) data were lost from Release 1, and in other cases actual duplicates were not eliminated. However, improving the duplicate elimination procedure to identify small numbers of additional duplicates that may be widely separated in space or time (Lander and Morrissey, 1987; Steurer, 1987) will probably be delayed beyond Release 2, because of the difficulty of implementing generalized platform track-checking, particularly for earlier data where ship IDs may be less reliable or unavailable (see section 4).

Two distinct procedures for quality control (QC) were applied for Release 1: a) NCDC’s basic QC for surface marine data, and b) the trimming procedure, which was used to flag suspect individual observations in the Compressed Marine Reports (CMR) and to exclude data from the 2° monthly summaries (a subset of flags resulting from the NCDC procedure was used to select the observations included in CMR; otherwise the two procedures were independent).

We have tentative plans to revise the NCDC procedure by using a digitized version of the U.S. Navy (1981) marine atlas to flag climatologically suspect data (see Elms, 1990). Although preliminary testing points to some localized disagreement with a more recent atlas incorporating corrected sea surface temperatures from the UK data bank (Bottomley et al., 1990), the limits in use at this stage are so broad (based on 4.8s and 5.8s) that we anticipate no negative consequences to using the older atlas. However, for Release 2 we plan to consider the possibility of a closer integration of the trimming procedure with the NCDC-defined QC, which may yield additional improvements in the identification of climatologically suspect data. Also, pre-assigned quality control indicators may be used for Release 2, such as supplied with NMDS buoy data (a lot of effort went into the MEDS person-computer QC, so we should also archive these flags).

Research and implementation of a modified trimming procedure will also continue in order to accommodate extreme climate events, again because limits used for trimming were found to be overly restrictive for such less common climate anomalies (see section 6). A related problem is the absence of empirical limits in areas, such as those traversed by drifting buoys, where historical data are sparse or non-existent. A number of different approaches are under consideration for resolving these issues (Wolter, 1992).

One unresolved issue is to what extent the trimming, and possibly other quality control steps, should be performed separately by platform type or time-of-day, to reflect any similar separations that may be utilized for later calculation of statistics. Some separations, possibly also incorporating a higher resolution in time or space, may yield statistics that are too sparse for the derivation of meaningful trimming limits. Another unresolved question is how improvements in the trimming method, for which research efforts have thus far concentrated on sea surface temperature, can be extended to variables with different time/space variability such as pressure, or to bivariate quantities such as wind.

Even with processing improvements, the climatic significance of many trends in COADS data still cannot be reliably interpreted because of the host of complex inhomogeneities that apply to early years (particularly before 1947), or even in recent years because of poorly understood effects arising from increasing numbers of anemometer (versus Beaufort-estimated) wind observations, etc. Some researchers have already addressed selected problems. For example, Bottomley et al. (1990) and Farmer et al. (1989) have adjusted sea surface temperature data based on assumptions about bucket types and climatological conditions. Cardone et al. (1990) and Isemer and Hasse (1991) have recommended usage of a revised Beaufort wind equivalence scale, and noted the effects of increasing shipboard anemometer heights.

It is our hope that collaborative work, fostered in part by this workshop, will lead to the development of data sets incorporating homogeneity adjustments, and analysed fields incorporating satellite data that could provide the highest resolution and best coverage possible. At the same time, however, it is critical that original, unadjusted data remain permanently available from the basic COADS archive.

Acknowledgments

We are grateful to H. Diaz, J. Elms, R. Jenne, U. Radok, K. Wolter, and S. Worley for helpful comments. Thanks also to those who responded to the questionnaire, which provided useful background for the preparation of this paper. The NOAA portion of COADS update activities has been funded in part by the Climate and Global Change program, and in previous years by the Equatorial Pacific Ocean Climate Studies (EPOCS) program.

Appendix: Questionnaire results

Early in 1992 a questionnaire was distributed to the COADS mailing list[6] soliciting feedback from the user community on requirements for future COADS updates and regarding possible interest in collaborative work. Twenty responses (not necessarily to all questions) were received (one anonymous). Most of the responses were from the U.S., but responses were also received from Canada, China, Germany, Japan, and the UK. Following is a summary of major questions and results from the questionnaire:

• *Are changes needed in the temporal or spatial resolution of the statistical summaries (currently, monthly summaries for 2° latitude x 2° longitude boxes)?* Opinion was split on whether a finer temporal resolution (e.g., 5-15 days) was needed, but the majority favored an increase in spatial resolution to 1° latitude x 1° longitude boxes. In view of the large storage requirements and many sparsely populated boxes that may result from finer time/space resolution, there was no consensus

[6] Additions or changes to the COADS mailing list may be brought to the attention of Scott Woodruff. Mailing address: NOAA/ERL (R/E/CG2); 325 Broadway, Boulder, CO 80303; USA. Phone: 303-497-6747; Fax: 303-497-6750. Omnet: NOAA.CRD; e-mail: sdw@noaacrd.colorado.edu.

on whether increases in resolution should be confined to certain variables (temperatures and wind were most frequently mentioned) or to certain regions and times with “sufficient” data. Others felt that monthly/2° products, perhaps derived from higher resolution statistics, should continue to be made available.

It was noted that a spatial resolution finer than 1° boxes could be misleading in many cases due to insufficient accuracy in reported positions (e.g., from early ship data), but that a weekly time resolution would be useful for compatibility with satellite sea surface temperatures (since 1982). It was also noted that higher resolution climatologies (e.g., 5-day/l°), possibly synthesized from lower resolution data, could be used to improve calculation of anomalies in regions or times with strong gradients. One suggestion was for a varying grid-size depending on data density; another suggestion was that users could generate fields of appropriate resolution from the individual marine reports.

Numbers of observations per time/space box deemed to be “sufficient” ranged from 40 down to 3, and may vary depending on the time/space variability of the phenomenon, region, and season being studied, the actual distribution of observations within the box, etc. For example, one response suggested that at least 7 observations be required per monthly box, which should span more than a 7-day period. This and other topics were felt by some to require further research. In any case, it was noted that the numbers of observations per time/space box, and other statistics, were especially important in data-sparse regions where individual data errors may have greater impact, and that such information should continue to be made available to users.

• *Are additional variables or statistics needed?* Only a few responses indicated a desire for additional quantities (cube of wind speed, height of anemometer, humidex/vapor pressures, cloud type/height summaries compatible with satellite data). It was noted that a flag indicating daytime/nighttime would be useful to accompany the individual observations. A 4-cell frequency count for time-of-day for each time/space box was suggested as a possible addition to the statistics.

• *Should the statistical summaries be separated by platform type and by time-of-day?* Although more study was felt necessary at least for diurnal problems, opinions leaned toward the need for separations both by platform type (e.g., for buoys starting in 1979) and by time-of-day. For time-of-day separation, one response suggested considering the possibility of compatibility with operational analyses or satellite summaries; but others commented that separations should be made according to local sun time (e.g., sun above or below horizon). Another response indicated that time-of-day biases could be investigated more effectively with other data sources. Some felt separations should be made for all variables; otherwise, wind and temperatures were most frequently mentioned. For cloudiness it was noted, e.g., that cirrus is more visible during daytime.

One suggestion was to handle quasi- stationary or relatively high-frequency data (e.g., from ocean weather stations, and moored or drifting buoys) differently than ship data. Another suggestion was that ship data might themselves need to be separated by ship size or other factors. For ships, the desirability of a separation between estimated and measured winds was noted, as well as the need to address anemometer height differences.

If separate sets of statistics were available to users, there was also strong interest in making a “blended” product available. There were a few suggestions for the method of blending, e.g., based on a multiple regression model, or using differences from climatologies. One response indicated a desire for a product restricted only to homogeneous data, for which any necessary adjustments had been made, and which might have to be limited to recent decades and be of lower resolution.

• *Should homogeneity adjustments be attempted and for which variables?* There was some strong interest in collaboration, although others felt adjustments could not now be attempted or should be left to individual users. Sea surface and air temperatures were foremost among the variables for which it was felt corrections should be attempted, followed by wind (e.g., after World War II). Other variables were mentioned (pressure, dewpoint, wave height). It was suggested that adequate buoy data may now exist to help correct ship sea surface temperatures and perhaps other variables. Previous work relevant to adjustments was also noted (e.g., Bottomley et al., 1990; Cardone, 1990). One concern was that both adjusted and unadjusted data must always remain available.

• *What are the needs for metadata (information about COADS data) and systems to access metadata?* Interest was expressed in access to a variety of metadata, such as information about observing practices or instrumentation (e.g., pictures of instruments), platform dimensions and other information including buoy locations and ship routes, inventory products, and COADS documentation (e.g., to help in accessing data products). Metadata concerning errors and biases were felt to be especially desirable. Responses varied on the desirability of an interactive “help” system to access metadata; such a system might need to provide maps and graphics, with equations of less importance. One comment noted the possible need to translate metadata between languages, e.g., from English to Japanese. Another comment was that the metadata files could potentially grow as large as the data themselves.

• *Additional comments*: Interest was expressed in alternative distribution media (e.g., CD-ROM or 8 mm tape), e.g., for storage of individual observations for ocean areas. The need was also suggested for intercomparisons between COADS and other climatic data sets (e.g., existing climatologies, operational or satellite analyses). One response suggested retaining an emphasis on basic processing (e.g., identifying bad data), as opposed to “science” processing.

References

Bottomley, M. C.K. Folland, J. Hsiung, R.E. Newell, and D.E. Parker, 1990: *Global Ocean* Sur*face Temperature Atlas "GOSTA."* MIT Press, Boston, 20 pp., 313 plates.

Cardone, V.J., J.G. Greenwood, and M.A. Cane, 1990: On trends in historical marine wind data. *Journal of Climate*, **3**, 113-127.

Elms, J.D., 1990: Enhancement for the global marine data base. *Second AMS Symposium on Global Change Studies*, American Meteorological Society, Boston, pp. J19-J23.

Elms, J.D., 1992: Status of NCDC keying of historical marine data. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

Farmer, G., Wigley, T.M.L., Jones, P.D., and Salmon, M., 1989: *Documenting and Explaining Recent Global-mean Temperature Changes.* Natural Environmental Research Council Report. Climatic Research Unit, University of East Anglia, Norwich, U.K., 141 pp.

Folland, C.K., Parker, D.E., and Kates, F.E., 1984: Worldwide marine temperature fluctuations 1856-1981. *Nature*, **310**, 670-673.

Hopkins, J.S. and Taylor, P.K., 1992: Voluntary Special Observing Programme-North Atlantic. *WMO Bulletin*, **41**, 76-79.

IPCC (Intergovernmental Panel on Climate Change), 1990: *Climate Change-The IPCC Scientific Assessment.* Houghton, J.T., Jenkins, G.J., and Ephraums, J.J. (eds.), Cambridge University Press, 365 pp.

Isemer, H.-J., and Hasse, L., 1991: The scientific Beaufort equivalent scale: Effects on wind statistics and climatological air-sea flux estimates in the North Atlantic Ocean. *Journal of Climate*, **4**, 819- 836.

Jenne, R.L., 1986: Technical working group report. *In:* Woodruff, S.D., (ed.) *Proceedings of a COADS Workshop, Boulder, Colorado, January 22-24, 1986*, NOAA Environmental Research Laboratories, Boulder, Colorado, pp. 183-189.

Jenne, R.L., 1992: The importance of COADS for Global Reanalysis. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colo. (this volume).

Lander, M.A., and Morrissey, M.L., 1987: Unexpected duplicate ship reports in the Comprehensive Ocean -Atmosphere Data Set (COADS). *Tropical Ocean-Atmosphere Newsletter*, **38**, 13-14.

Mock, D.R., and Messenger, A.R., 1992: Climate Research Data Tools (CRDtools) and COADS. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

Moninger, W.R., Woodruff, S.D., Fozzard, R.L., and Slutz, R.J., 1992: COADS Metalog:

Dynamic COADS documentation. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

Morrissey, M.L., 1990: An evaluation of ship data in the equatorial western Pacific. *Journal of Climate*, **3**, 99-112.

Parker, D.E., 1992: Blending of COADS and UK Meteorological Office marine data sets. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

Pierson, W.J., 1990: Examples of, reasons for, and consequences of the poor quality of wind data from ships for the marine boundary layer: Implications for remote sensing. *Journal of Geophysical Research*, **95**, 13313-13340.

Ramage, C.S., 1986: Scientific working group report. *In:* Woodruff, S.D. (ed.) *Proceedings of a COADS Workshop, Boulder, Colorado, January 22-24, 1986*, NOAA Environmental Research Laboratories, Boulder, Colorado, 72-77.

Sadler, J.C., 1986: Observations from fishing fleets of the Inter-American Tropical Tuna Commission. *In:* Woodruff, S.D. (ed.) *Proceedings of a COADS Workshop, Boulder, Colorado, January 22-24, 1986*, NOAA Environmental Research Laboratories, Boulder, Colorado, 49- 51.

Slutz, R.J., Lubker, S.J., Hiscox, J.D., Woodruff, S.D., Jenne, R.L., Joseph, D.H., Steurer, P.M., and Elms, J.D., 1985: *Comprehensive Ocean-Atmosphere Data Set; Release 1*. NOAA Environmental Research Laboratories, Boulder, Colorado, 268 pp. (NTIS PB86-105721).

Steurer, P.M., 1987: Reply: unexpected duplicate ship reports in the Comprehensive Ocean Atmosphere Data Set (COADS). *Tropical Ocean-Atmosphere Newsletter*, **40**, 12-14.

U.S. Navy, 1981: *Marine Climatic Atlas of the World. Vol. IX: World-wide Means and Standard Deviations*. NAVAIR 50-lC-65, U.S. Government Printing Office, Washington, D.C., 169 pp.

Uwai, T. and Komura, K., 1992: The collection of historical ships' data in Kobe Marine Observatory. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

WeatherDisc Associates, 1989: World WeatherDisc: A comprehensive description of the weather and climate of the planet earth on a single compact disc (CD-ROM), WeatherDisc Associates, Seattle, Wash.

Wilkerson, J.C. and Earle, M.D., 1990: A study of differences between environmental reports by ships in the Voluntary Observing Program and measurements from NOAA buoys. *Journal of Geophysical Research*, **95**, 3373-3385.

Wolter, K., 1992: Sifting out erroneous observations in COADS-the trimming problem. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

Wolter, K., Lubker, S.J., and Woodruff, S.D., 1989: Trimming-a potential error source in COADS. *Tropical Ocean-Atmosphere Newsletter*, **51**, 4-7.

Woodruff, S.D., 1990: Preliminary comparison of COADS (US) and MDB (UK) ship reports. *In:* Parker, D. (ed.) *Observed Climate Variations and Change: Contributions in Support of Section 7 of the Scientific Assessment of Climate Change of Working Group I of the IPCC*, WMO/UNEP publication.

Woodruff, S.D., and Lubker, S.J., 1986: COADS 1980-85 update. *In:* Woodruff, S.D. (ed.) *Proceedings of a COADS Workshop, Boulder, Colorado, January 22-24, 1986*, NOAA Environmental Research Laboratories, Boulder, Colorado, 36-48.

Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A comprehensive ocean atmosphere data set. *Bulletin of the American Meteorological Society*, **68**, 1239-1250.

Woodruff, S.D., Lubker, S.J., Quayle, R.G., Radok, U., and Doggett, E.D., 1991: *Differences Within and Among Surface Marine Datasets.* NOAA Environmental Research Laboratories, Boulder, Colorado, 216 pp.

Worley, S.J., 1992: Status of other new data sets for COADS. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).

WMO (World Meteorological Organization), 1955: *International List of Selected, Supplementary and Auxiliary Ships.* WMO No. 47, Geneva, Switzerland.

WMO (World Meteorological Organization), 1963: *Fourth World Meteorological Congress Abridged Report with Resolutions.* WMO No. 142, Geneva, Switzerland.

Yudin, K.B., Ylyanich, I.G., Popova, V.N., and Krankanovskaya, Ye.M., 1992: Russian Marine Meteorological Data Set. *In:* Diaz, H.F., Wolter, K., and Woodruff, S.D., (eds.) *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992*, NOAA Environmental Research Laboratories, Boulder, Colorado, (this volume).