

Scott D. Woodruff<sup>\*1</sup>, Steven J. Worley<sup>2</sup>, and Richard W. Reynolds<sup>3</sup>

<sup>1</sup>Climate Diagnostics Center, OAR, NOAA, Boulder, CO

<sup>2</sup>National Center for Atmospheric Research, Boulder, CO

<sup>3</sup>National Climatic Data Center, NESDIS, NOAA, Asheville, NC

## 1. INTRODUCTION

This paper reviews the current status of data and information handling for the widely used Comprehensive Ocean-Atmosphere Data Set (COADS). It also describes how new technologies have changed the ways the products and metadata are processed and distributed, since data set development began in 1981 (Woodruff et al., 1987). Ongoing international efforts to improve the flow of operational and historical marine data and metadata—for which the COADS project has served as a critical focal point and broadly recognized example—are also discussed.

Processing and distribution of the complex and heterogeneous data collection that makes up COADS continues to be a challenging task. The bulk of the observations consist of marine meteorological reports from voluntary observing ships, now extending back for over two centuries. These have been supplemented since the late 19th century by data from oceanographic research vessels, and in recent decades by measurements from moored and drifting environmental buoys. The resultant quality controlled observations and gridded monthly summary products have always received unrestricted public distribution.

## 2. DATA SET DEVELOPMENT

In 2001, we reached a major milestone for COADS: the release of higher quality data and products that extend from 1997 back to 1784. Our most recent efforts concentrated on the period prior to 1950. This required major efforts in data archaeology to locate and digitize from US and international archives (Diaz and Woodruff, 1999) millions of historical ship records.

Observational and statistical products based on these and other data (Table 1) were completed in 2001, comprising COADS Release 1c (1784-1949). This represented a key milestone for COADS. Releases 1b (1950-79) and 1a (1980-97), similar updates to more contemporary periods, were finished during 1996-97 (Woodruff et al., 1998). Thus upon completion of Release 1c, the original Release 1 (1854-1979; Slutz et al., 1985) data were fully replaced by a consistent set of data and metadata products covering the extended period 1784-1997, with many additions of new and improved data.

To reach this point, the data are subjected to a series of processing steps. The first stage is used to convert the marine reports into a common format. This format is designed with the flexibility to retain elements unique to data sources or that cannot readily be converted into standardized fields. Then extensive quality control (QC) is applied, and duplicate marine reports (e.g., received from real-time and delayed-mode sources) are eliminated. Finally, higher-level products are created in the form of global monthly summary statistics using 2° latitude x 2° longitude boxes (and for 1°x1° boxes since 1960).

## 3. PROCESSING IMPROVEMENTS

Improvements to the QC are a key area of ongoing work. The present QC uses climatological sextiles based on the original Release 1 data to define upper and lower limits for "good" data. This "trimming" procedure needs to be improved because some erroneous data pass the QC tests when they should fail, while other data fail during periods with large anomalous climate signals when they should pass.

Wolter (1997) suggested a more "adaptive" QC approach, by addressing the scatter of observations around the individual year-month median rather than around a climatological value. We have begun implementing a revised QC procedure along these lines, starting with sea surface temperature (SST), and later to be extended to other primary variables.

For use in the revised QC, we are computing analyses using optimum interpolation (Reynolds and Smith, 1994) to determine the large-scale SST signal. Once this signal is removed, the climatological monthly standard deviation ( $\sigma$ ) can be recomputed. The new  $\sigma$  is smaller, and thus should automatically become more stringent in eliminating "bad" data, because it is not influenced by large-scale signals (Fig. 1).

In addition, to better serve time-critical research, we are developing improved observational and summary products in near-real-time. These observational data will be available in a new International Maritime Meteorological Archive (IMMA) format. Data in this ASCII format will be output from the real-time data system as a prototype, with the long-term goal of international standardization. This synchronization of COADS with the real-time system and eventual coordination with data collection world-wide will augment the available data and metadata, and permit users to easily and consistently extend the formal COADS (now ending at 1997) to current dates.

At present there is no effective, internationally agreed format for exchange of keyed historical data. The format needs flexibility to preserve crucial original

---

\* Corresponding author address: Scott D. Woodruff, NOAA/OAR Climate Diagnostics Center (R/CDC1), 325 Broadway, Boulder, CO 80305; e-mail: [sdw@cdc.noaa.gov](mailto:sdw@cdc.noaa.gov).

data elements and metadata. This will help facilitate analyses of data biases and discontinuities arising from changes in instrumentation and observing practices. Moreover, the format should be expandable, to meet new requirements that are not presently anticipated, but also simple enough that it is practical to implement by World Meteorological Organization (WMO) Member countries.

To meet these requirements, the prototype IMMA format will be presented to the Joint WMO/IOC (Intergovernmental Oceanographic Commission) Technical Commission for Oceanography and Marine Meteorology (JCOMM) in approximately one year (JCOMM, 2001), with a view to eventual submission to the Commission for formal adoption.

#### 4. INFORMATION HANDLING

Technological advances have significantly changed the ways COADS is accessed and the extent to which it is available around the world. All access methods currently available from the developers are linked to at the project Website ([www.cdc.noaa.gov/coads/](http://www.cdc.noaa.gov/coads/)). These include FTP access to all products and observations in the binary archive formats, online request forms to provide ASCII and binary subsets, and data in netCDF file format. The subsetting options for observations not only provide users with temporal and spatial selections, but also choices in how to apply the QC and data filtering flags. Temporal and spatial subsettings are available for all monthly summary products.

Some other data sites have made portions of COADS available through online data access systems. We will do the same for the full COADS products collection in the future. A two-year effort between NOAA's Pacific Marine Environmental Laboratory (PMEL) and the COADS project has been funded to implement a "Live Access Server" (LAS), through which users would have access to a very high-quality browsing interface to the gridded products on the Web. The server will provide subsetting and downloading in a variety of formats with enriched metadata content. This new service along with those already in place will offer the users many options for data access. The coordination of these archive resources and the various access methods will always be described on the project Website. This provides COADS users with a single entry point for data and information, hopefully minimizing any confusion over data availability.

Future studies seeking to make the most accurate environmental assessments will likely require additional metadata concerning the specific platforms from which the observations are taken. As the data sources become more diverse and the period of record becomes longer, it becomes increasingly important to collect and organize the relevant metadata. To support this requirement the project Website has a growing body of metadata, and we plan to expand our efforts in this area. The Website includes some platform metadata (e.g. ship instrumentation details for 1973-98 from WMO, 1955-), documentation about data sources (e.g. the US Maury Collection), data and product

assessments by the COADS team (e.g., of WMO's BUFR format), and a publication list that identifies important studies related to COADS.

#### ACKNOWLEDGEMENTS

We are grateful to Xiao-Wei Quan for Fig. 1.

COADS is the result of a continuing cooperative project between the National Oceanic and Atmospheric Administration (NOAA)—specifically its Office of Oceanic and Atmospheric Research (OAR)/Climate Diagnostics Center (CDC), its National Environmental Satellite, Data and Information Service (NESDIS)/National Climatic Data Center (NCDC), and the Cooperative Institute for Research in Environmental Sciences (CIRES, conducted jointly with the University of Colorado)—and the National Science Foundation's National Center for Atmospheric Research (NCAR). The NOAA portion of COADS is currently supported by the NOAA Climate and Global Change (C&GC) Program and by the NOAA Environmental Services Data and Information Management (ESDIM) Program.

#### REFERENCES

Diaz, H.F., and S.D. Woodruff (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, 114 pp.

JCOMM (Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology), 2001: *First Session, JCOMM-I, Akureyri, Iceland, 19-29 June 2001, Abridged Report with Resolutions*.

Manabe, T., 1999: The Digitized Kobe Collection, Phase I: Historical Surface Marine Meteorological Observations in the Archive of the Japan Meteorological Agency. *Bull. Amer. Met. Soc.*, **80**, 2,703-2,716.

Reynolds, R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses. *J. Climate*, **7**, 929-948.

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steurer, and J.D. Elms, 1985: *Comprehensive Ocean-Atmosphere Data Set; Release 1*, 268 pp. [Available from NOAA Climate Diagnostics Center, Boulder, CO.]

WMO (World Meteorological Organization), 1955-: International List of Selected, Supplementary and Auxiliary Ships, WMO No. 47, Geneva, Switzerland [annual publication].

Wolter, K., 1997: Trimming problems and remedies in COADS. *J. Climate*, **10**, 1,980-1,997.

Woodruff, S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A comprehensive ocean-atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 1,239-1,250.

Woodruff, S.D., H.F. Diaz, J.D. Elms, and S.J. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth*, **23**, 517-527.

Table 1. Primary additions to COADS for the most recently completed update (1784-1949). For each collection, the table includes the number of reports (before elimination of duplicates), period-of-record, and noteworthy comments.

Collection	Rpts	Covering	Comments
UK Main Marine Data Bank	12.1M	1854-1949	Also blended for 1980-97
US Merchant Marine Collection	3.5M	1912-1946	US WW II records mostly discarded about 1974
US Maury Collection	1.3M	1784-1863	Keyed by China; only available data for 1784-1803
Japanese Kobe Collection	1M	1890-1932	Key additions to Pacific in WW I; Manabe (1999)
World Ocean Database 1998	405K	1874-1949	SST estimates from oceanographic profiles
Russian MARMET Data	268K	1888-1949	Also blended through 1997
Norwegian Logbook Collection	201K	1867-1889	About 200K reports remain undigitized
Arctic Drift Ships and Stations	16K	1893-1925	Norwegian expedition ships <i>Fram</i> and <i>Maud</i>
Russian Makarov Collection	3.5K	1804-1891	27 ships including the <i>Vitiaz</i>

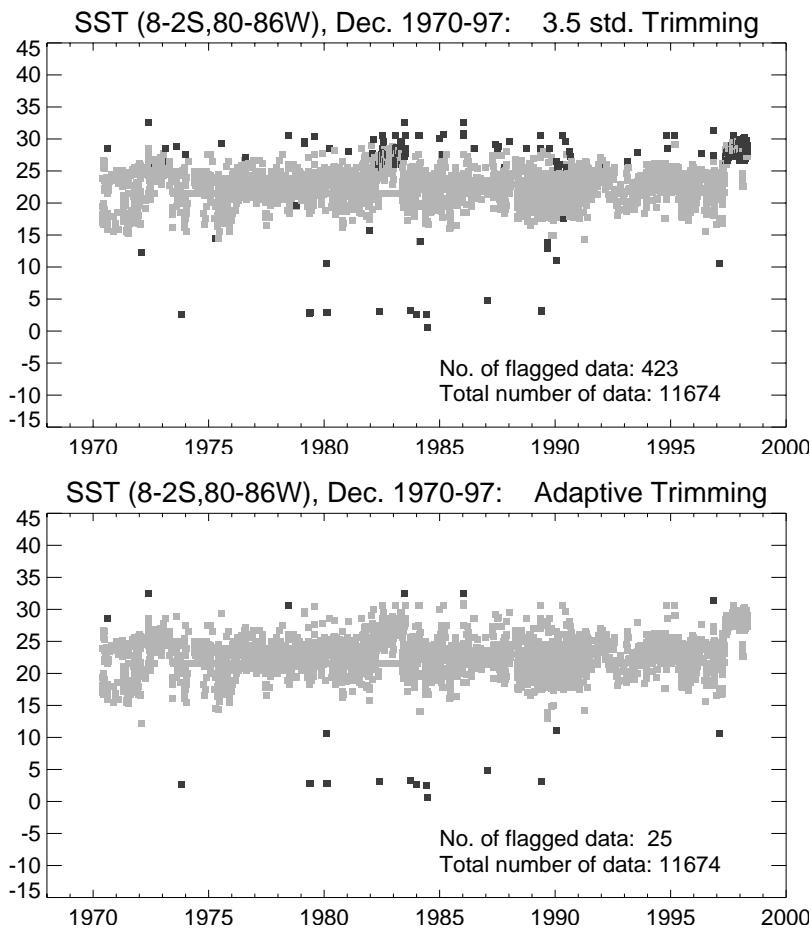


Figure 1. SST observations ( $^{\circ}$ C) for the Decembers of 1970-97 in the indicated region of the eastern tropical Pacific. The observations were accepted (gray) or rejected (black) using: (upper panel) original trimming at  $(3.5\sigma)$ ; (lower panel) new adaptive trimming. Note: points were plotted near the ends of the years listed on the horizontal axis.