COADS UPDATES AND THE BLEND WITH THE UK MET OFFICE MARINE DATA BANK

S.D. Woodruff[1], S.J. Worley[2], J.A. Arnott[3], H.F. Diaz[1],
J.D. Elms[4], M. Jackson[3], S.J. Lubker[1], and D.E. Parker[3]

ABSTRACT

Ongoing data and metadata enhancements to the Comprehensive Ocean-Atmosphere Data Set (COADS) are described, including the blend with the UK Met Office Marine Data Bank (MDB). MDB data available since 1854 are being used, together with other new or improved sources, to enhance data coverage and quality within the presently available period-of-record (1854-1997). In addition, some newly available historical data will be used to extend coverage back to 1784. Data composition and coverage are discussed, and future plans outlined, including improved products to help address data continuity problems arising from observational, instrumental, and processing changes. Improved and expanded metadata also are becoming available as part of an upgrade of the COADS web site (http://www.cdc.noaa.gov/coads), which includes details about how to request data products.

INTRODUCTION

The Comprehensive Ocean-Atmosphere Data Set (COADS) is the most extensive set of surface marine meteorological data presently available for the world ocean, now covering the 1854–1997 period. Surface meteorological observations from ships of opportunity are available for the entire period-of-record. These have been supplemented in more recent years by increasing amounts of data from moored and drifting buoys, oceanographic Research Vessels (R/Vs) and fishing vessels.

Extensive efforts also are underway to enhance the quality and completeness of earlier ship records. These include a blend of COADS with the UK Met Office Marine Data Bank (MDB), and national and international efforts to digitize additional logbook data and metadata (Diaz and Woodruff, 1999). The updated data are providing crucial input for the Intergovernmental Panel on Climate Change (IPCC) Scientific Assessments, and internationally for several centres that compute global atmospheric reanalyses.

Sections 2 and 3 of this paper describe efforts over the last several years toward a complete replacement and update of COADS Release 1 (1854–1979) (Slutz *et al.*, 1985; Woodruff *et al.*, 1987), including improved observational and summary products.

COADS provides a relatively uniform database for a wide variety of scientific investigations, and its products are distributed openly and without restrictions. These characteristics have been critical in developing broad international participation.

Section 4 highlights variations in data composition and coverage of the presently available 1854–1997 data. These variations are compounded by other data inhomogeneities arising, for example, from instrumental, observational and processing changes. Section 5 outlines future plans, including data and metadata improvements targeted to help address these issues.

RECENT AND ONGOING UPDATES

RELEASE 1A (1980–97)

Release 1a data, originally completed for 1980–92 (Woodruff *et al.*, 1993), have been updated and extended several times in response to requirements for reanalysis projects and demands from the user community for updated products. The most recent update, completed in June 1999, involved a complete reprocessing of

(1) NOAA/OAR Climate Diagnostics Center, Boulder, Colorado, USA
(2) National Center for Atmospheric Research, Boulder, Colorado, USA
(3) Hadley Centre, Meteorological Office, Bracknell, UK
(4) NOAA/NCDC, Asheville, North Carolina, USA

the previously available data for the 1980–95 period, plus an extension through to 1997.

A major element of the 1980–97 update was the blend with the MDB for 1980-94. The Met Office implemented and documented a conversion from MDB 'flatfile' formats into the Long Marine Report (LMR) format used for COADS production processing (a fixed-length LMRF format is generally is distributed to users). The Met Office provided 17.9 million MDB reports for 1980-94, and 7 per cent of these data were retained as unique or judged to be of preferable quality compared with data already in COADS (Figure 1), and British Navy decks were retained at relatively high rates compared to other MDB sources (Table 1).

For earlier periods we estimate that gains from the blend will be higher: Woodruff (1990) estimated 63 per cent duplicate and 28 per cent/9 per cent unique from COADS/MDB based on tests against Release 1 data for six sample 10° latitude × 10° longitude boxes.

Some major enhancements were made as part of the 1980-97 update to data from moored and drifting buoys. Global Telecommunication System (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 offer improved diurnal coverage and quality controls; the Pilot Research Moored Array in the Tropical Atlantic (PIRATA) is also now part of this archive. Canada's Marine Environmental Data Service (MEDS) corrected 1980-85 drifting buoy data for a day misassignment problem (+1, only impacting 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 quality control information to obtain increases in the available drifting buoy data.

Other new or improved data sources were included in this update. Sea surface temperature (SST) estimates derived from the uppermost levels of oceanographic profiles, and some surface meteorological fields, were added through 1996 from the World Ocean Database 1998 (WOD98; Levitus *et al.*, 1998). Additional data from Russia's large marine archive of ship data (MARMET) were included in 1995, as well as Russian Arctic and Antarctic Research Institute (AARI) North Pole (NP) Station (manned drifting ice floe) data through 1991, obtained through the University of Washington's Polar Science Center.

RELEASE 1B (1950-79)

This update (completed in November 1996) pre-dated our work on the blend with the MDB (to be fully implemented in a future update of this time period), but provided improvements in data quality and coverage in comparison to previously available Release 1 data for 1950–79. Data additions included Russian MARMET

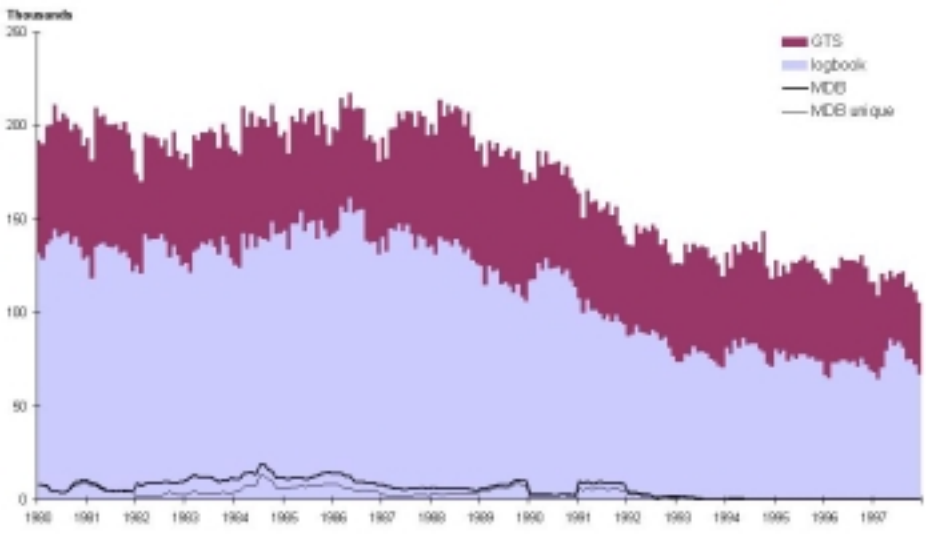

Figure 1—Monthly bars show the numbers of ship reports (from LMRF) output for 1980–97, received via the Global Telecommunication System (GTS) or in delayed mode (generally keyed logbook data). Within the logbook category, curves show the total retention of MDB data (dark), and within that the number of unique MDB reports (light); i.e. "total retention" includes unique MDB reports plus others that were considered preferable duplicates (compared with COADS, plus possibly with other MDB reports).

Table 1—Numbers (K=thousands and M=millions) and percentages of MDB data by deck ("series" number) output (retained, e.g. as unique or "best" duplicate) in the blend with COADS, 1980–97. Overall, 93 per cent of the reports were eliminated from the MDB as duplicates.

Deck	Description	Reports	Output LMR: %
221	"MARIDS" and trawlers	1.55K	6
223	Selected ships	1.52K	<1
224	Ocean Weather Stations	59	6
229	British Navy (HM) ships	10.6K	92
233	Selected ships	48.8K	2
234	Ocean Weather Stations	781	4
239	British Navy (HM) ships	42.6K	34
254	Int. Maritime Met. (IMM)	1.18M	8
255	Undocumented sources	2.78K	8
Total		1.29M	7

and NP data and an earlier oceanographic archive designated as the World Ocean Atlas 1994 (WOA94; Levitus and Boyer, 1994).

Significant data corrections were also made during Release 1b processing, including the correction of widespread temperature biases in GTS records (see Woodruff *et al.*, 1998). In 1999, a minor correction to Release 1b October-November 1970 data also was made to remove a small number of mislocated GTS observations.

RELEASE 1C (1784–1949)

We are nearing completion of the reprocessing of this time period (planned by early 2001). The update will blend MDB data for 1854-1949 (including 0.5M newly keyed 1935-39 UK merchant data), Russian 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 the First World War (Manabe, 1999).

We will also include data from several recently digitized collections (described in more detail in Diaz and Woodruff, 1999 and Elms *et al.*, 1999): the US Maury Collection (covering 1784-1863, but concentrated around 1830–60; 1.3M reports), the Norwegian Logbook Collection (1867-89; 201K), the US Merchant Marine 1912–46 Collection (3.5M), Arctic Drift Stations (1893-1938; 16K) and the Russian S.O. Makarov Collection (1804–91; 3,500 reports).

However, conversions of early logbook data to modern units in the LMR format involve complex and scientifically important translation issues, and resources to implement the conversions are limited. Therefore, some conversions may have to be delayed until a future update, or be scaled back to key data elements, which are considered to be SST, sea level pressure (SLP), air temperature and wind.

PRODUCTS: OBSERVATIONS, STATISTICS, METADATA

Releases 1a and 1b already offer significant improvements in comparison to the original Release 1 products, as discussed in sections 3.1 and 3.2. Completion of Release 1c will make the entire archive available in uniform observational and summary formats. Improved metadata are also being made available as part of these updates, as discussed in section 3.3. The COADS web site (http://www.cdc.noaa.gov/coads) provides links to, or information on how to request, data and metadata products.

INDIVIDUAL MARINE REPORTS (OBSERVATIONS)

To develop Release 1a, the LMR format used for COADS production processing was updated. Also, a fixed-length LMRF format was developed to satisfy the majority of current user requirements for individual observations and to replace the Release 1 Compressed Marine Report (CMR) format. The number of data fields was expanded in LMRF, compared to CMR, and critical new metadata, such as platform type and identification, were added to track the increasing number and diversity of data sources.

The LMRF archive is maintained in a packed-binary format, which offers computational and storage-volume efficiencies and is appropriate for long-term global studies with multiple variable requirements. However, for smaller scale studies (temporally or spatially), a simple ASCII format is more portable between computers and is easier to use. As part of the blend with the MDB we have started to design an abbreviated ASCII format that would include basic data elements and most probably some pre-applied quality controls (i.e. suspect data elements would be eliminated).

We note that a highly abbreviated ASCII format is also available to meet requirements for observational data that are more recent than those available in COADS (i.e. now later than 1997). These near-real-time data are updated on a monthly basis by NOAA's National Centers for Environmental Prediction (NCEP) and can be accessed through the COADS web site.

MONTHLY SUMMARY STATISTICS

The COADS observations are statistically summarized for each year-month and in 2° or 1° latitude × longitude boxes. Table 2 summarizes the temporal and geographical coverage, and box resolution, for the most popular statistical products. Detailed descriptions of these and secondary products are available from the COADS web site.

The Monthly Summary Trimmed Groups (MSTG) product was developed during Release 1 and is available for the full period-of-record. User suggestions, and our experience, resulted in the development of an improved product, the Monthly Summary Groups (MSG), which contains more statistics and variables (see Woodruff *et al.*, 1998) and will fully replace MSTG upon the completion of Release 1c.

Studies of events such as the 1982–83 El Niño revealed that the "trimming" (quality control) processing developed for Release 1 was too conservative and resulted in the distortion or elimination of some large climate signals (Wolter, 1997). Concerns have also been raised about the effects of mixing ship data with data from other platform types such as drifting and moored buoys (e.g. Woodruff *et al.*, 1993).

To help mitigate these problems, and so that researchers can study the effects, two separate sets of statistics were computed for Releases 1a and 1b. The 'standard' statistics are derived from ship data only, using the restricted Release 1 trimming limits. In contrast, the 'enhanced' statistics are derived from ship and other platform types (e.g. drifting and moored buoys), using relaxed trimming limits to better preserve climate anomalies. Similar strategies are under consideration for Release 1c (1784–1949) statistics (discussed in Wolter *et al.*, 1999).

METADATA

Major improvements to the COADS web site were completed in June 1999. These included software, electronic documentation (e-doc) and inventories for currently available products; selected on-line publications (see the references for examples); and annual ship instrumental metadata available in digital form since 1973 gathered in WMO publication No. 47 (1955-). The WMO-No. 47 files from 1973–94 were reprocessed by Elizabeth Kent of the UK Southampton Oceanography Centre (discussed in Kent *et al.*, 1999).

Product	*Period*	*Domain*	*Resolution*
MSTG	1854–1997	global	2° × 2°
MSG	1950–1997	global	2° × 2°
MSG	1960–1997	global	1° × 1°
MSG	1960–1997	equatorial belt†	1° × 1°

Table 2—COADS year-month summary statistics: product abbreviations, present temporal and geographic availability, and latitude × longitude box resolution.

* Products: Monthly Summary Trimmed Groups (MSTG) and Monthly Summary Groups (MSG). Each MSTG (MSG) contains eight (10) statistics for each of four variables, and overall the products comprise 19 (22) observed and derived variables.

† 10.5°N-10.5°S, with gridding offset 0.5° from the global product and a row of 1° boxes straddling the equator.

The metadata available on the web site will continue to grow in the future. We plan to add early UK and US documentation that is not readily available, along with discussions and descriptions of the data problems that we analyze and answers to frequently asked questions.

DATA COMPOSITION AND COVERAGE

Ship data volume has been declining since peaks in the 1980s (Figures 1-2). This trend is presently influenced by delays in receiving keyed ship logbook data (in future, delayed-mode ship data will be provided by electronic means such as on diskettes or through telecommunication services like Inmarsat). However, most of the likely changes will stem from factors such as the decline in global ship traffic, increases in ship size and shifting shipboard priorities for reporting weather observations.

Conversely, there has been substantial growth in the number of drifting and moored buoy reports since around 1979 (Figure 2). The moored TAO array has significantly improved the quantity and quality of reports in the tropics, and drifting buoys provide vital reports, for example, in the far southern latitudes (although only SST and SLP in general).

As shown in Figure 1, the GTS contributed a relatively stable amount, but increasing fraction, to the total ship data mixture in recent decades; in 1997, delayed-mode ship data still composed over half of the mixture. Figure 3 shows leading contributions of the International Maritime Meteorological (IMM) keyed ship logbook data exchanged internationally since WMO Resolution 35 (Cg-IV) was adopted in 1963.

The pre-1950 data mixture (Figure 4) will be significantly enhanced to the extent that we will be able to include newly digitized data as part of Release 1c. Also, as illustrated in Figure 4, Dutch (deck 193) data predominate prior to the 1880s. The recovery and gravity correction of deck 193 SLP data that are available among the LMR supplementary data form an important goal for Release 1c processing to enrich SLP coverage (Figure 5).

FUTURE PLANS

Analyses of the COADS historical record have been affected by many changes since the 19th century as regards instrumentation, observing and reporting practices and processing (e.g. quality controls). Historical metadata to help address these heterogeneities are often nonexistent or incomplete. Many of these issues have been discussed in past COADS workshops (e.g. Diaz and Isemer, 1995) and in the open literature (e.g. Parker *et al.*, 1995).

In more recent decades, data from drifting and moored buoys have expanded data coverage. However, in a trend that will continue as future global ocean observing systems are implemented (Molinari, 1999), new platform types have further complicated the data mixture, and metadata are still frequently inadequate (e.g. information on past buoy arrays, although efforts have been

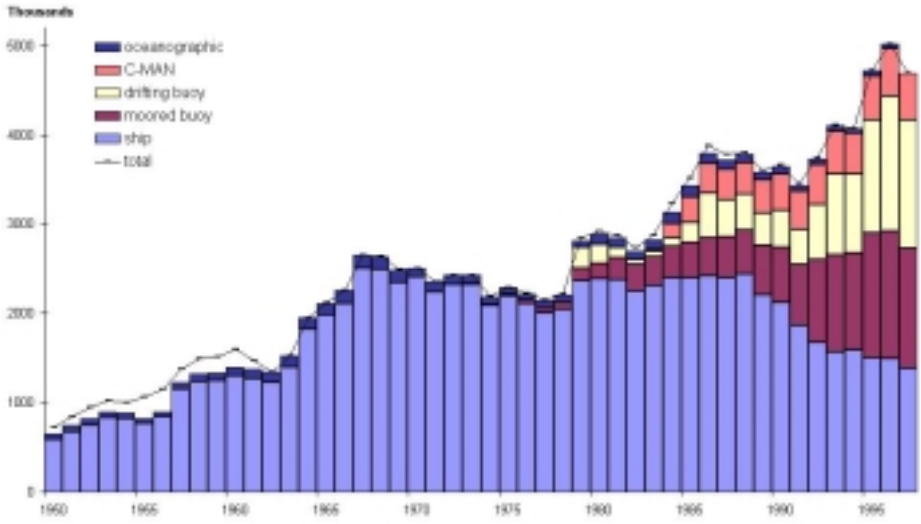

Figure 2—Annual bars show reports (LMRF) output for 1950–97, with the numbers from different platform types plotted in descending order, from bottom to top, of total magnitude: ship, moored and drifting buoy, Coastal-Marine Automated Network (C-MAN) and ocean R/Vs reports. The LMRF total line is sometimes larger due to miscellaneous additional platform types or reports for which platform type was not determined.

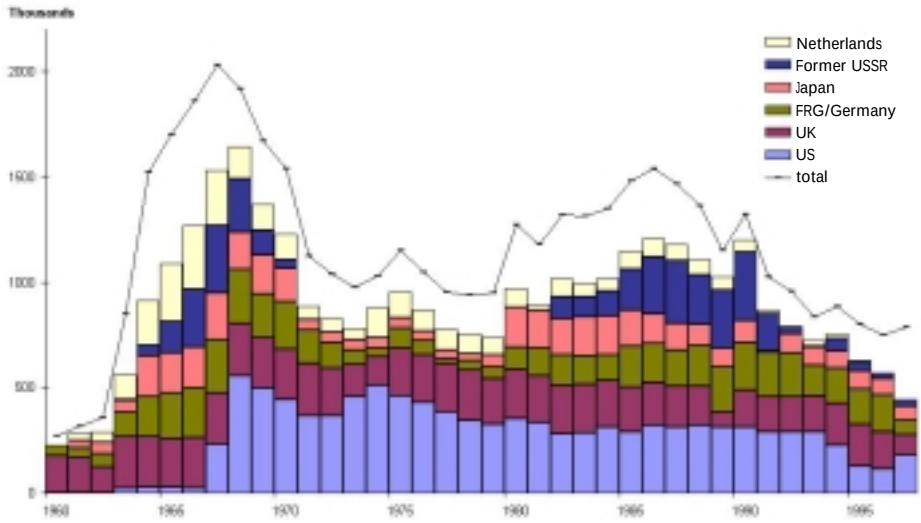

Figure 3—Annual bars show the six largest national contributions of IMM (digitized ship logbook) reports (from LMRF) output for 1960-97, plotted in descending order, from bottom to top, of total magnitude. These, plus IMM contributions from other countries, yield the total line.

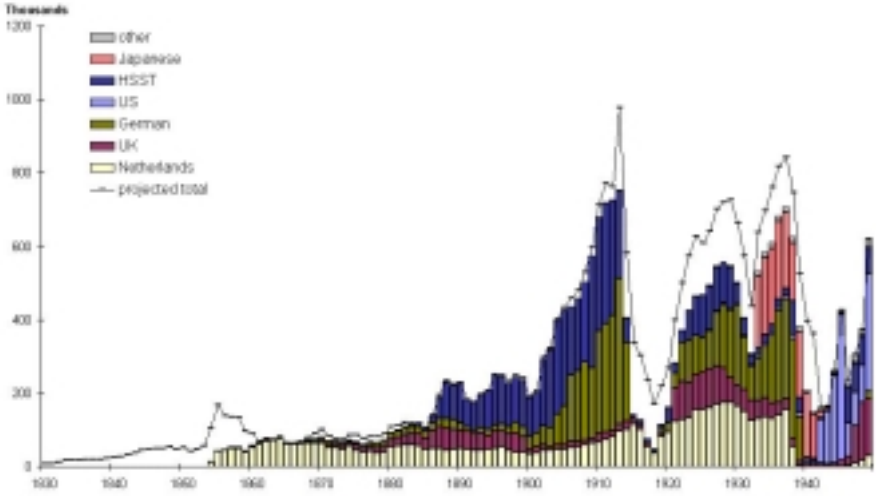

Figure 4—Annual bars show the global source deck make up of COADS Release 1 (CMR) for 1854–1949, roughly grouped according to national categories, plus Historical Sea Surface Temperature (HSST) Project data (after Figure 3 in Woodruff et al., 1998). Note the virtual absence of US data prior to 1941; and the predominance of Japanese data during 1940–41, of US data during 1942—45, and of Dutch data prior to around 1880 (it is undocumented whether the Dutch and some other early decks might actually be international compilations). The projected total line indicates the gains expected dating back to 1830 over the present COADS data from available Kobe, Maury, Norwegian, US Merchant Marine (1912–46) and Arctic Station collections (as discussed in section 2.3).

initiated to archive buoy metadata in accordance with WMO-No. 47, 1955). Even within a category such as ship data, we have the contemporary mixture of merchant and Navy, ocean R/Vs and fishing vessels which ply the oceans for different reasons and in doing so may collect data with different biases (e.g. some fishing vessels seek calm tropical regions resulting in a low wind bias).

So far we have dealt with these problems using the simplest approach. As discussed in section 3.2, separate sets of statistics are computed in an attempt to examine platform heterogeneity questions and to reduce invalid data-exclusions during climatic events with large variability.

When Release 1c and the COADS-MDB blend are completed we will be in a position to process the full archive (1784 through to the late 1990s) and better address these problems. This is an important development phase for COADS. We are hoping to complete what will be known as Release 2 towards the end of 2001.

To be successful, Release 2 will require better user access to metadata and data and improved quality controls. In conjunction with new ASCII format products, temporal and spatial subsetting of the global long-term archive will be provided. As discussed in Woodruff *et al.* (1998), we plan to concentrate metadata improvements and bias adjustments on individual observations. However, calculations of some new $2° \times 2°$ and $1° \times 1°$ statistical products, including night-time air temperatures, will be developed; further separations, such as platform types and source decks, may also be desirable.

ACKNOWLEDGEMENTS

COADS is the result of an ongoing cooperative project between the National Oceanic and Atmospheric Administration (NOAA)—specifically its Office of

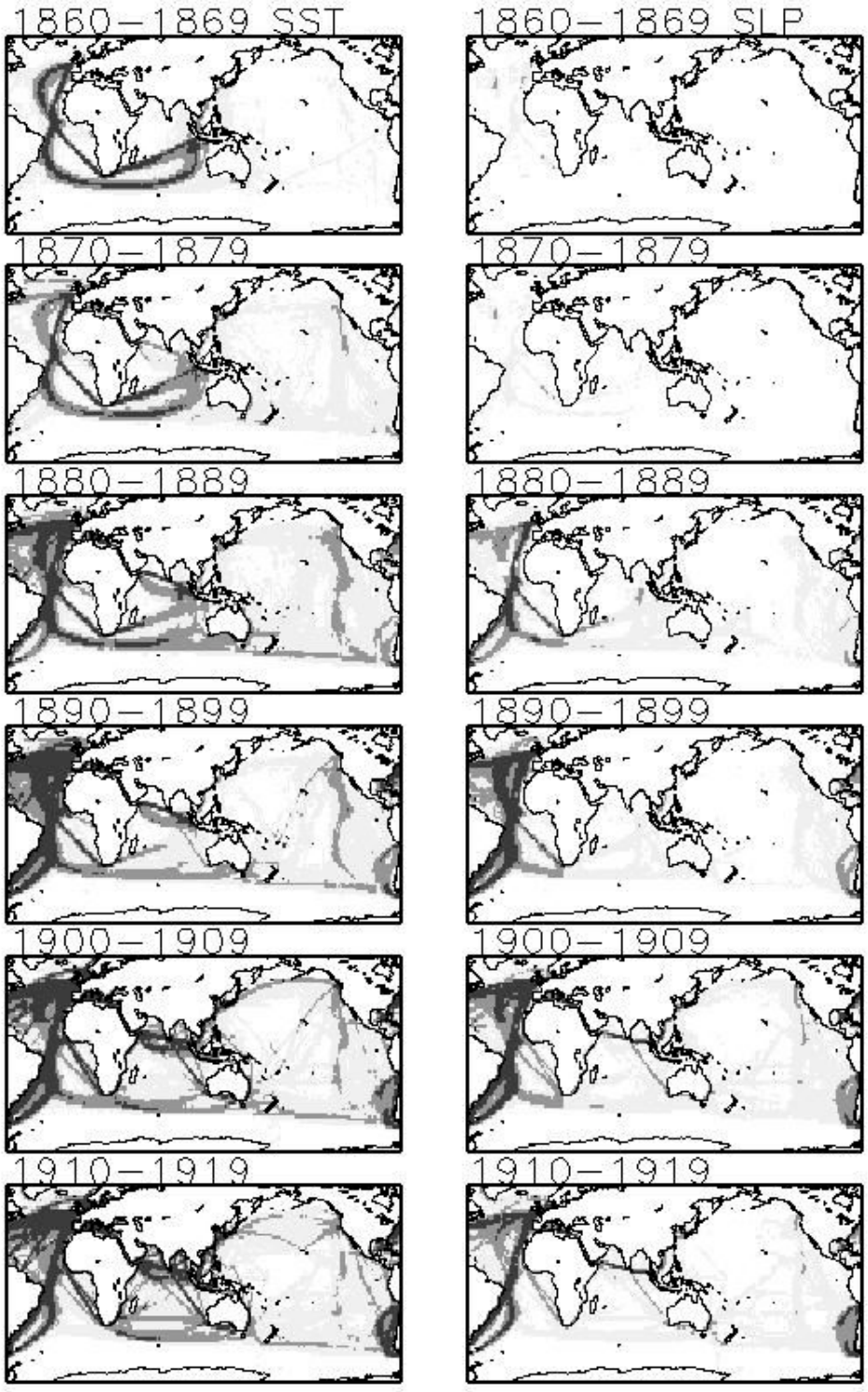

Figure 5—Decadal totals of SST (left) and SLP (right) observations (70°N–78°S; 68°W–68°W) for six early decades of Release 1 data (after Figure 5 in Woodruff et al., 1987). Three increasingly dark shadings show at least 10, 100, or 400 observations in a 2° × 2° box per decade, i.e. respective averages over 120 months of 0.08, 0.83, or 3.33 observations per month (note: all the observations for a decade could fall into as few as one year-month).

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) programme and the NOAA Environmental Services Data and Information Management (ESDIM) programme.

REFERENCES

Diaz, H.F. and H.-J. Isemer (Eds.), 1995: *Proceedings of the International COADS Winds Workshop* (Kiel, Germany, 31 May-2 June 1994). NOAA Environmental Research Laboratories, Boulder, Colo., 301 pp.

Diaz, H.F. and S.D. Woodruff (Eds.), 1999: *Proceedings of International Workshop on Digitization and Preparation of Historical Surface Marine Data and Metadata* (15-17 September 1997, Toledo, Spain). WMO/TD-No.957, MMROA Report No. 43.

Elms, J.D., S.D. Woodruff, D.E. Parker and S.J. Worley, 1999: Newly digitized historical marine data and metadata becoming available for COADS and the UK Met Office Marine Data Bank. *Preprints: WMO Workshop on Advances in Marine Climatology—CLIMAR99* (Vancouver, B.C., Canada, September 8-15, 1999). Environment Canada, Downsview, Ontario, 11-22.

Kent, E.C., P.K. Taylor and S.A. Josey, 1999: Improving global flux climatology: the role of metadata. *Preprints: The Workshop on Advances in Marine Climatology—CLIMAR99* (Vancouver, B.C., Canada, September 8-15, 1999). Environment Canada, Downsview, Ontario, 292-299.

Levitus, S. and T.P. Boyer, 1994: *NOAA Atlas NESDIS 4, World Ocean Atlas 1994, Volume 4: Temperature.* National Oceanographic Data Center, Ocean Climate Laboratory, Washington, DC, 117 pp.

Levitus, S., T.P. Boyer, M.E. Conkright, T. O'Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson and R. Gelfeld, 1998: *NOAA Atlas NESDIS 18, World Ocean Database 1998, Volume 1: Introduction.* National Oceanographic Data Center, Ocean Climate Laboratory, Silver Spring, MD, 346 pp.

Manabe, T., 1999: The digitized Kobe Collection, phase I: historical surface marine meteorological observations in the archive of the Japan Meteorological Agency. *Bull. Amer. Meteor. Soc.*, 80, 2703-2715.

Molinari, R.L., 1999: Lessons learned from operating global ocean observing networks. *Bull. Amer. Meteor. Soc.*, 80, 1413-1419.

Parker, D.E., C.K. Folland and M. Jackson, 1995: Marine surface temperature: observed variations and data requirements. *Climatic Change,* 31, 559-600.

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.* NOAA Environmental Research Laboratories, Climate Research Program, Boulder, CO, 268 pp.

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

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

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

Wolter, K., S.J. Lubker, and S.D. Woodruff, 1999: Quality control in recent and pending COADS releases. *Preprints: WMO Workshop on Advances in Marine Climatology—CLIMAR99* (Vancouver, B.C., Canada, September 8-15, 1999). Environment Canada, Downsview, Ontario, 312-318.

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

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, 1239-1250.

Woodruff, S.D., S.J. Lubker, K. Wolter, S.J. Worley and J.D. Elms, 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980-92. *Earth System Monitor,* 4, No. 1, 1-8 [On-line at: http://www.cdc.noaa.gov/coads/coads1a.html].

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 [On-line at: http://www.cdc.noaa.gov/coads/egs_paper.html].